Characterization of two MHC II genes (DOB, DRB) in white-tailed deer (Odocoileus virginianus)

Natascha M. D. Ivy-Israel 1*, Carolyn E. Moore 1, Tonia S. Schwartz 2 and Stephen S. Ditchkoff 1

Abstract

Background: The major histocompatibility complex (MHC) is responsible for detecting and addressing foreign pathogens inside the body. While the general structure of MHC genes is relatively well conserved among mammalian species, it is notably different among ruminants due to a chromosomal inversion that splits MHC type II genes into two subregions (IIa, IIb). Recombination rates are reportedly high between these subregions, and a lack of linkage has been documented in domestic ruminants. However, no study has yet examined the degree of linkage between these subregions in a wild ruminant. The white-tailed deer (Odocoileus virginianus), a popular ruminant of the Cervidae family, is habitually plagued by pathogens in its natural environment (e.g. Haemonchus contortus, Elaeophora). Due to the association between MHC haplotypes and disease susceptibility, a deeper understanding of MHC polymorphism and linkage between MHC genes can further aid in this species’ successful management. We sequenced MHC-DRB exon 2 (IIa) and MHC-DOB exon 2 (IIb) on the MiSeq platform from an enclosed white-tailed deer population located in Alabama.

Results: We identified 12 new MHC-DRB alleles, and resampled 7 alleles, which along with other published alleles brings the total number of documented alleles in white-tailed deer to 30 for MHC-DRB exon 2. The first examination of MHC-DOB in white-tailed deer found significantly less polymorphism (11 alleles), as was expected of a non-classical MHC gene. While MHC-DRB was found to be under positive, diversifying selection, MHC-DOB was found to be under purifying selection for white-tailed deer. We found no significant linkage disequilibrium between MHC-DRB and MHC-DOB, suggesting that these loci are unlikely to be closely linked.

Conclusions: Overall, this study identified 12 new MHC-DRB exon 2 alleles and characterized a new, non-classical, MHC II gene (MHC-DOB) for white-tailed deer. We also found a lack of significant linkage between these two loci, which supports previous findings of a chromosomal inversion within the MHC type II gene region in ruminants, and suggests that white-tailed deer may have a recombination hotspot between these MHC regions similar to that found for Bos taurus.

Keywords: Major histocompatibility complex, Linkage disequilibrium, Odocoileus virginianus, Ruminant, Chromosomal inversion, MiSeq

* Correspondence: nataschaisrael@yahoo.com
1School of Forestry and Wildlife Sciences, Auburn University, Auburn, AL 36849, USA
Full list of author information is available at the end of the article

© The Author(s). 2020 Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.
Background

The major histocompatibility complex (MHC) is a well-studied group of genes whose protein products are responsible for recognizing and addressing foreign pathogens present in the body [38, 43, 86]. While type I MHC genes are present on all nucleated cells, type II MHC genes occur only on immune cells, such as dendritic cells and macrophages [40]. These class II MHC molecules consist of two membrane-spanning chains (α and β), both of which are produced by MHC genes [105]. Immune cells, also known as antigen presenting cells (APC), engulf exogenous particles. Once broken down, the peptide fragments are bound to MHC gene products (MHC-DR and MHC-DQ), which present them at the surface of the APC. To achieve this, non-classical MHC genes, such as MHC-DM and MHC-DO, produce accessory proteins used to effectively load antigens onto classical, peptide-binding MHC gene products, which then travel to the APC’s surface to present the antigen to the immune system [40, 54, 72]. If an antigen is recognized as foreign, helper T cells will bind to the presented antigen and release lymphokines, which attract other cells to the area. These helper T cells will also bind to B cell lymphocytes, which will ultimately stimulate the production of antibodies for this particular antigen [40]. The MHC is therefore a crucial component of a vertebrate’s immune system.

The peptide binding region (PBR) determines which of the exogenous peptide fragments the MHC proteins can bind. These regions are often found to be under positive diversifying selection [10, 88], increasing the probability the gene products from different alleles of an MHC gene can bind different antigens. Through this differential antigen presentation, some allele combinations (haplotypes) can confer greater susceptibility or resistance to certain pathogens. While this is strongly documented among humans [2, 22, 32, 34, 66] and domesticated animals [48, 70, 92], it is an emerging field among wildlife species.

While the synteny of MHC genes is relatively well conserved among mammalian species [103], it is notably different among ruminants (Bos taurus, [6]; Ovis aries, [29]). More specifically, while the MHC II genes of most placental mammals are organized as centromere – (~ 20 Mb of non-MHC DNA) – MHC-DM/DO – MHC-DQ/DR – MHC III/I genes [31, 103], the ruminant MHC II genes are organized as centromere – MHC-DM/DO – (~ 20 Mb of non-MHC DNA) – MHC-DQ/DR – MHC III/I genes [16, 82]. The unique organization of MHC II genes found among ruminants is thought to be due to a chromosomal inversion in an ancestral mammal [8] that has split MHC II genes into two subregions: Ila (MHC-DR/DQ) and I Ib (MHC-DM/DO/DY, TAP, among others). A similar MHC II gene configuration was found in the finless Yangtze porpoise (Neophocaena asiaeorientalis asiaeorientalis, [83]), but not in swine (Sus scrofa, [77]), which suggests the inversion occurred after the phylogenetic split between ruminants and Suidae but before ruminants split from Cetacea. The ruminant MHC II subregions are separated by at least 15 cM (centimorgans; [6, 101]), which is markedly greater than what is found in humans (~ 3 cM; [98]) and mice (~ 1.5 cM; Mus musculus, [95]). Significant recombination rates have been observed in the interval between MHC-DR and MHC-DY genes in Bos taurus [68, 69], further suggesting a recombination hotspot between the two subregions in ruminants.

Studies on deer species have found the MHC-DR and MHC-DQ genotypes to be important for disease resistance. Li et al. [50] found that one MHC-DR/DQ haplotype was associated with resistance to purulent disease, a multifactorial disease [51], among forest musk deer (Moschus berezovskii), whereas two other haplotypes increased susceptibility to the disease. Different MHC-DR haplotypes in Iberian red deer (Cervus elaphus hispanicus) influenced susceptibility to a variety of different pathogens [27]. For example, while one haplotype was associated with a decreased occurrence of tuberculosis but increased Elaphostrongylus cervi scores, deer with a different haplotype experienced the opposite trend. Similarly, when using phylogenetic groupings of MHC-DRB exon 2 alleles as haplotypes, Ditchkoff et al. [24] found that white-tailed deer (Odocoileus virginianus) with different haplotypes experienced different levels of abomasal nematodes (including Haemonchus contortus) and ectoparasitism by ticks.

White-tailed deer are a well-studied wild ruminant due to their popularity as a game species [39]. While proper management has significantly improved their numbers, they are continually threatened by pathogens such as Haemonchus contortus [73, 74], Elaeophora [19], and the epizootic hemorrhagic disease virus and blue-tongue virus [28]. These diseases can have significant impacts on deer populations and a deeper understanding of factors that influence susceptibility to these diseases is crucial to managing their populations. While the MHC-DRB region has been previously characterized in white-tailed deer [99, 100], other MHC genes (such as MHC-DOB) and the possibility of the inverted MHC genetic configuration found among other ruminants have not. To date there are 18 documented MHC-DRB exon 2 alleles in white-tailed deer [99, 100]. Newer sequencing technologies (Next Generation Sequencing), however, may reveal greater polymorphism due to their greatly increased mutation detection rate (sensitivity; [14, 15, 41]). Additionally, while homologs of bovine MHC Iib genes have been identified in several species belonging to Cervidae [96], no study has directly examined the possibility of an inverted MHC II configuration in a cervid. While
the white-tailed deer genome has been sequenced [87], the MHC-DRB and MHC-D0B genomic regions are found on different scaffolds that prevent a large-scale understanding of their arrangement.

To fully capture the association between MHC haplotypes and disease susceptibility, we must first have a better understanding of the polymorphism that exists at these genes. We therefore aim to further quantify MHC-DRB exon 2 polymorphism in white-tailed deer and characterize an additional MHC gene (MHC-D0B exon 2). Since these genes are predicted to lie on different MHC II subregions separated by the inversion seen in other ruminants, we also tested for linkage disequilibrium between MHC-DRB and MHC-D0B in white-tailed deer. Since the immunological functions of MHC-DRB and MHC-D0B are vastly different, with MHC-DRB being the classical antigen-presenting protein and MHC-D0B being an accessory loading protein, selection may have influenced these genes differently. We therefore also assessed how these two genes have evolved to better understand selection pressures on MHC polymorphism of white-tailed deer.

Results
Sequencing
Before filtering, the average number of reads for individuals was 30,549 for MHC-DRB (ranged from 5318 to 84,226 reads with a median of 24,309) and 22,341 for MHC-D0B (ranged from 7029 to 49,892 reads with a median of 21,913). The number of reads recovered per individual after filtering for quality ranged from 4162 to 66,582 reads for MHC-DRB (average = 24,795.5; median = 19,214) and from 2779 to 36,407 reads for MHC-D0B (average = 15,815; median = 15,522). While unfiltered read lengths ranged from 289 to 488 bp for MHC-DRB (average = 306, median = 307) and from 50 to 535 bp for MHC-D0B (average = 397, median = 398), filtered read lengths were 296 to 365 bp long for MHC-DRB (average = 306, median = 307) and 329 to 406 bp long for MHC-D0B (average = 397, median = 398). All raw sequence data is available on the NCBI Sequence Read Archive (SRA accession # PRJNA533917).

No linkage disequilibrium between MHC-DRB and MHC-D0B
We found no significant linkage disequilibrium between MHC-DRB exon 2 and MHC-D0B exon 2 when using only the individuals in the starting population (n = 69; p = 0.95) and when including individuals without pedigree data to this initial population (n = 122; p = 0.37). Since these loci are unlikely to be closely linked, we present the results for each gene separately.

MHC-DRB
Defined 12 new alleles
A total of 19 MHC-DRB exon 2 alleles were found in our white-tailed deer population (n = 373), 12 of which were new alleles (Odvi-DRB*19 – Odvi-DRB*30; Table S2). The seven previously identified alleles found in the population were Odvi-DRB*01, Odvi-DRB*05, Odvi-DRB*06, Odvi-DRB*10, Odvi-DRB*12, Odvi-DRB*14, and Odvi-DRB*16 [99, 100]. The 19 MHC-DRB exon 2 alleles found in our population translated to 18 unique amino acid alleles (82–83 codons), as Odvi-DRB*06 and Odvi-DRB*19 differed by one synonymous substitution (Table S3). Together with the other 12 previously identified MHC-DRB exon 2 alleles [99, 100] not recovered in our population, a total of 30 MHC-DRB exon 2 alleles have been characterized for white-tailed deer. The 12 new MHC-DRB exon 2 alleles have been deposited in Genbank under accession numbers MK952679- MK952690.

Genetic relationship among alleles
The length of MHC-DRB exon 2 alleles was 250 bp for all alleles except for Odvi-DRB*27, which had a 3-bp frameshift deletion. Pairwise nucleotide distances among alleles ranged from 0.40 to 18.40% (mean = 10.99%), and pairwise amino acid distances ranged from 0 to 39.35% (mean = 23.00%). The average number of nucleotide differences (k) was 24.97, and there were 85 polymorphic sites in the MHC-DRB exon 2 sequences.

In assessing the phylogenetic relationships using ModelFinder, the best-fit model according to BIC for the nucleotide MHC-DRB tree was F81 + F + I + G4 (Fig. 1). There were six near-zero internal branches (<0.0040): two in the moose outgroup, two in the roe deer outgroup, one in the reindeer outgroup, and one in the internal branch that separates the reindeer clade from the white-tailed deer sequences. The white-tailed deer/roe deer/reindeer MHC-DRB exon 2 nucleotide sequences separated fairly strongly from the moose outgroup (81% bootstrap support) but not from each other, suggesting the presence of a polytomy. While the roe deer and reindeer sequenced clustered into clear groups, the white-tailed deer sequences did not. However, there were some well-supported clades among these white-tailed deer sequences (up to 100% bootstrap support). When using translated MHC-DRB exon 2 sequences, the best-fit model according to BIC values was PMB + I + G4 (Fig. 2). This tree had seven near-zero internal branches (<0.012) in terms of bootstrap support: four in the moose outgroup and three in the roe deer outgroup. As with the nucleotide tree, the white-tailed deer/roe deer/reindeer sequences separated strongly from the moose outgroup (85% bootstrap support). The roe deer sequences were further separated with 72% bootstrap support, while the white-tailed deer and reindeer sequences...
remained together, though there were some well-supported clades within these remaining sequences (up to 100% bootstrap support).

Population genetic measures and test for selection
When assessing the allele frequencies of the complete dataset (2003–2017), the most common MHC-DRB exon 2 allele was Odvi-DRB*10 (22.4%), followed by Odvi-DRB*20 (14.6%) and Odvi-DRB*14 (12.5%; Fig. 1; Table S2). The most common genotypes were Odvi-DRB*20/Odvi-DRB*10 (7.0%), Odvi-DRB*10/Odvi-DRB*10 (4.8%), and Odvi-DRB*14/Odvi-DRB*10 (4.6%; Table S4). Allele frequencies changed slightly from the founding population (2003–2007) to the more recent population (2016). Frequencies for Odvi-DRB*01 (10.9–6.8%) and Odvi-DRB*14 (16.7–10.2%) decreased over time, whereas Odvi-DRB*10 (18.1–22.5%) and Odvi-DRB*19 (5.8–11.9%) became more frequent in the population. Odvi-DRB*05, a rare allele in our original founding population, was lost from the population over time.

Fig. 1 Maximum likelihood phylogenetic tree for the nucleotide sequences of MHC-DRB exon 2. This tree contains sequences for white-tailed deer (Odvi, Odocoileus virginianus; AF082161-AF082175, AF407169-AF407171, MK952679- MK952690), moose (Alal, Alces alces; X82398, X83278, X83279-X83286; [56]), roe deer (Caca, Capreolus capreolus; KM488213-KM488216, KM488218, KM488220-U90925; [57, 75]), and reindeer (Rata, Rangifer tarandus; AF012716-AF012724; [59]) as these are the closest related species to white-tailed deer with MHC-DRB data. It was rooted with the moose outgroup. Node labels are standard bootstrap support (%). Arrows indicate the presence of near-zero internal branch lengths (< 0.0040), which should be interpreted with caution. Heatmap colors indicate all white-tailed deer MHC-DRB exon 2 alleles and further correspond to Odvi-DRB allele frequencies found in our population, where red is the most common MHC-DRB allele.
While MHC-DRB exon 2 was under Hardy Weinberg equilibrium using the probability-test for both the founding and 2016 populations \((p = 0.88 \text{ and } p = 0.91\text{, respectively})\), score test results for the 2016 population indicated heterozygote excess in the population \((p = 0.05\text{; Table 1})\). Both nucleotide \((\pi)\) and haplotype diversity generally stayed the same over time \((0.09 \text{ to } 0.10 \text{ and } 0.90 \text{ to } 0.88\text{, respectively; Table 1})\). All neutrality test values were positive, and there was an increasing trend from the founding population to the more recent 2016 population (Table 1), suggesting that selection on MHC-DRB is becoming less neutral in our population. Both Fu and Li’s D* and F* test statistics were significant \((p < 0.02)\) for the founding (2.39 and 2.09, respectively) and 2016 (2.56 and 2.45, respectively) populations. These positive results suggest MHC-DRB had an excess of intermediate frequency alleles in our population, which is an indication of balancing selection and/or reduced population size. MHC-DRB showed evidence of positive selection based on the dN/dS ratio of 2.06 (Fig. S1).
Defined 11 alleles

Eleven unique alleles were identified for the extended MHC-DOB nucleotide sequences (Odvi-DOB*01–Odvi-DOB*11) and seven alleles for MHC-DOB exon 2 in our white-tailed deer population (Table S5). Odvi-DOB*01, Odvi-DOB*02, and Odvi-DOB*11 contained the same exon 2 sequence (Odvi-DOB*01-02-11_exon2). Odvi-DOB*03 and Odvi-DOB*10 also contained the same exon 2 (Odvi-DOB*03-05-06-08_exon2), as well as Odvi-DOB*06 and Odvi-DOB*07 (Odvi-DOB*06-07-09_exon2). Odvi-DOB*04, Odvi-DOB*05, Odvi-DOB*08, and Odvi-DOB*09 had unique exon 2 sequences (Odvi-DOB*04_exon2, Odvi-DOB*05_exon2, Odvi-DOB*08_exon2, Odvi-DOB*09_exon2, respectively). The seven MHC-DOB exon 2 alleles translated into three unique amino acid alleles (89 codons). The MHC-DOB alleles have been deposited in Genbank for the extended DOB sequences (accession numbers MK952691-MK952701).

Genetic relationships among alleles

In assessing the phylogenetic relationships using ModelFinder, the best-fit model according to BIC for the nucleotide MHC-DOB tree was K2P + I (Fig. 3). There were seven near-zero internal branches (< 0.0028%) in terms of bootstrap support, all of which occurred in the white-tailed deer clades. The red deer and white-tailed deer sequences separated strongly from the cow and sheep sequences (97% bootstrap support). Odvi-DOB*08/Odvi-DOB*09 and Odvi-DOB*06/Odvi-DOB*07 separated from the other white-tailed deer sequences with 67 and 62% bootstrap support, respectively. The translated MHC-DOB tree had Blosum62 as the best-fit model (Fig. 4). There were five near-zero internal branches (< 0.0112%), which all occurred in the white-tailed deer clades. Apart from the fairly strong separation between cow/sheep sequences and the white-tailed deer/red deer sequences (66% bootstrap support), the amino acid sequences for MHC-DOB exon 2 did not retain the same organization as in the MHC-DOB nucleotide tree. The tree clearly demonstrates the three unique MHC-DOB exon 2 translations.

All extended MHC-DOB sequence alleles were 360 bp in length except for Odvi-DOB*01 (359 bp) and Odvi-DOB*02 (361 bp) due to an indel in the intronic region. The pairwise nucleotide distances ranged from 0 to 1.41% (mean = 0.58%). Odvi-DOB*01, Odvi-DOB*02, and Odvi-DOB*11 differed by one indel located in a highly repetitive region of the intron. The greatest amount of dissimilarity was found between Odvi-DOB*07 and Odvi-DOB*09 (4 synonymous substitutions, one nonsynonymous substitution; Table S6). The average number of nucleotide differences (k) was 2.07, and the extended MHC-DOB sequences had seven polymorphic sites. The extended nucleotide MHC-DOB sequences were in Hardy Weinberg Equilibrium for both the founding and 2016 populations (probability test; p = 0.27 and p = 0.55, respectively), and there was no indication of either a heterozygote deficit or excess in the founding population (Table 2). However, there was a heterozygote deficit in the 2016 population (score test; 0.03).

All alleles for the MHC-DOB exon 2 region contained within our extended MHC-DOB sequences had a length of 270 bp. The mean pairwise nucleotide distance for
MHC-DOB exon 2 was 0.78% (0.37–1.50%), while the mean pairwise amino acid distance was 0.86% (0–2.27%). As was seen in the extended MHC-DOB sequences, the greatest amount of dissimilarity was found between Odvi-DOB*0607_exon2 and Odvi-DOB*09_exon2 (3 synonymous substitutions, one nonsynonymous substitution; Table S7). Odvi-DOB*04_exon2 and Odvi-DOB*09_exon2 also differed more than other MHC-DOB exon 2 alleles (2 synonymous substitutions, 2 nonsynonymous substitutions). The MHC-DOB exon 2 sequences contained five polymorphic sites and 2.10 nucleotide differences on average. As with the extended MHC-DOB sequences, the MHC-DOB exon 2 alleles were also in Hardy Weinberg Equilibrium for the founding and 2016 populations (probability test; \( p = 0.57 \) and \( p = 0.33 \), respectively), though there was statistical significance for a heterozygote deficiency in the 2016 population (score test; \( p = 0.01 \); Table 2).

Population genetic measures and test for selection
Odvi-DOB*08 and Odvi-DOB*08_exon2 were the most common alleles for the extended MHC-DOB sequences (28.1%) and MHC-DOB exon 2 (27.8%; Fig. 3; Table S5). The most common genotypes all contained Odvi-DOB*08 and Odvi-DOB*08_exon2 (Table S8, S9). Allele frequencies increased over time for Odvi-DOB*06 (3.5 to 7.1%), Odvi-DOB*09_exon2 (4.2 to 9.7%), and Odvi-DOB*11/Odvi-DOB*010211_exon2 (9.2 to 12.6% and

![Fig. 3 Maximum likelihood phylogenetic tree for the extended nucleotide sequences of MHC-DOB (360 bp). This tree contains sequences for white-tailed deer (Odvi, Odocoileus virginianus; MK952691–MK952701), cow (Bola, Bos taurus; 282493, [112]), sheep (Ovar, Ovis aries; Z49879.1; [107]), and red deer (Ceel, Cervus elaphus; CM008014.1; [7]). It was rooted with the cow outgroup. Node labels are standard bootstrap support (%). Arrows indicate the presence of near-zero internal branch lengths (< 0.0028), which should be interpreted with caution. Heatmap colors indicate all white-tailed deer MHC-DOB alleles and further correspond to Odvi-DOB allele frequencies found in our population, where red is the most common MHC-DOB allele.](image-url)
11.3 to 14.3%, respectively), whereas Odvi-DOB*03 (12.7 to 10.5%), Odvi-DOB*05_exon2 (10.6 to 7.1%), and Odvi-DOB*08/Odvi-DOB*08_exon2 (both from 33.1 to 25.2%) became less frequent in the population.

While nucleotide diversity did not change over time, haplotype diversity increased slightly (Table 2). Neutrality test results indicated an increasing trend from the founding population to the more recent population for both extended MHC-DOB sequences and MHC-DOB exon 2 (Table 2), with Tajima’s D and Fu and Li’s F* approaching statistical significance (0.10 > p > 0.05) for the 2016 population. Tajima’s D increased from 0.87 to 1.31 for the extended MHC-DOB sequences and from 1.38 to 1.77 for MHC-DOB exon 2. Fu’s Fs for MHC-DOB was notably smaller compared to MHC-DRB. The dN/dS ratio for MHC-DOB exon 2 was 0.22, indicating no evidence of positive selection (Fig. S2).

**Discussion**
In this study we aimed to characterize MHC-DRB exon 2 and MHC-DOB exon 2 in our white-tailed deer population and to examine the extent of linkage between these two loci. No significant linkage was found between the second exons of MHC-DRB and MHC-DOB in our white-tailed deer population. While further examination is needed, this finding suggests that white-tailed deer
may have the same chromosomal inversion and recombination hotspot within their MHC type II gene region as found in other ruminants [4, 6, 16, 29, 68, 69]. This MHC II inversion has been documented among cetaceans [83, 84, 110] but not in swine [77], suggesting that this breakpoint may have occurred after the divergence of Cetartiodactyla and Suidae but before the divergence between Cetacea and ruminants (~58 million years ago [111]). A lack of linkage disequilibrium between these two loci allows them to evolve independently from one another. Because they are therefore unlikely to be closely linked, we can contrast their evolutionary patterns as independent loci.

Our white-tailed deer population had 19 MHC-DRB exon 2 alleles, which is comparable to what Van Den Bussche et al. [99] found in a free-ranging white-tailed deer population in Oklahoma (15 MHC-DRB exon 2 alleles, n = 150). Even though white-tailed deer experienced severe, historical population bottlenecks [25, 49], they have managed to retain more MHC-DRB polymorphism than other species that survived similar population reductions, such as moose [56], bison (Bison bison, [58]), fallow deer (Dama dama) and roe deer [59].

The polymorphism found among our founding adult population was less (17 MHC-DRB exon 2 alleles; n = 69) than what was reported in unrelated individuals of red deer (34 MHC-DRB exon 2 alleles, n = 50; [97]), Kankrej cattle (Bos indicus; 24 MHC-DRB exon 2 alleles, n = 50; [9]), and the Indian water buffalo (Bubalus bubalis; 22 MHC-DRB exon 2 alleles, n = 25; [20]). However, white-tailed deer have greater polymorphism than both Kankrej cattle and Indian water buffalo if all 30 MHC-DRB exon 2 alleles found across studies in this species are used for comparison. Red deer also have two MHC-DRB loci whereas white-tailed deer are believed to have only one MHC-DRB locus [99].

While white-tailed deer already were known to have a reasonably high MHC-DRB polymorphism (18 alleles; [99, 100]), we identified 12 new MHC-DRB exon 2 alleles in our Alabama population. The original 18 MHC-DRB exon 2 alleles were identified using 7 populations from Oklahoma, Iowa, Tennessee, and New York. Alabama deer have an interesting ancestry due to restocking efforts between the 1940’s and 1960’s. Six white-tailed deer subspecies were used for reintroduction [3, 52], thereby creating an admixed white-tailed deer population in Alabama. Our study population was located in Tallapoosa county, which received deer from Georgia, Arkansas, and other counties in Alabama (Clarke, Marengo, Sumter; [53]). Additionally, reintroductions using Clarke county deer occurred after Michigan deer were introduced to Clarke county. Admixture may therefore be contributing to the large number of new MHC-DRB exon 2 alleles identified in this study. Another potential explanation for identifying additional alleles may be our use of newer sequencing technology since the white-tailed deer MHC-DRB was last characterized via SSCP by Van Den Bussche et al. [99, 100]. MiSeq has greater sensitivity relative to SSCP [14, 15, 41], which may have enabled us to detect more nucleotide polymorphisms in the MHC-DRB exon 2 sequences.

### Table 2

| Diversity                  | Founding population | 2016 population | Founding population | 2016 population |
|----------------------------|---------------------|-----------------|---------------------|-----------------|
| Nucleotide diversity (π)  | 0.005               | 0.005           | 0.006               | 0.006           |
| Haplotype diversity       | 0.83                | 0.86            | 0.80                | 0.83            |
| Neutrality                |                     |                 |                     |                 |
| Tajima’s D                | 0.87                | 1.31            | 1.38                | 1.77^           |
| Fu and Li’s F*            | 1.26                | 1.42            | 1.34                | 1.48^           |
| Fu and Li’s D*            | 1.17                | 1.13            | 1.01                | 0.97            |
| Fu’s Fs                   | −0.22               | 0.64            | 0.60                | 1.33            |
| HWE                       |                     |                 |                     |                 |
| Probability Test          | 0.27                | 0.55            | 0.57                | 0.33            |
| Heterozygote Excess       | 0.77                | 0.97            | 0.44                | 0.99            |
| Heterozygote Deficit      | 0.24                | 0.03*           | 0.57                | 0.01*           |

* p < 0.05; ^ 0.10 > p > 0.05
While the outgroup sequences in our MHC-DRB trees clustered together as monophyletic groupings, as is expected for species with reduced MHC allelic diversity due to severe population bottlenecks [56, 59], white-tailed deer MHC-DRB alleles represented paraphyletic groups. This suggests that MHC-DRB polymorphism is most likely greater in white-tailed deer than moose, roe deer and reindeer. Additionally, the lack of separation into geographically structured phylogenetic clades suggests that there is no, or very little, relationship between genetic and geographic distance for white-tailed deer, similar to findings reported by Van Den Bussche et al. [100] and Leberg et al. [49]. Reintroductions may have weakened these geographic associations between MHC-DRB exon 2 alleles. When comparing our MHC-DRB nucleotide tree to the neighbor-joining tree constructed by Van Den Bussche et al. [99, 100], we find little similarity. This is most likely due to differences in tree construction, outgroups, and increased number of Odvi-DRB alleles included in the analyses. However, no quantification of support was provided for their trees, making a direct comparison more challenging as the level of confidence in their tree splits is unknown.

The MHC-DOB tree used sequences from less closely related species (cow, sheep, red deer) due to a current lack of MHC-DOB sequence data for Cervidae, which may have influenced tree inference. The extremely low MHC-DOB polymorphism found within our white-tailed deer population relative to MHC-DRB polymorphism may contribute to the lack of separation seen between the white-tailed deer MHC-DOB exon 2 alleles. Lastly, Ceel-DOB did not strongly separate from Odvi-DOB, which is similar to what Van Den Bussche et al. [99, 100] found for MHC-DRB.

When classifying individuals as homozygotes or heterozygotes for MHC-DRB, we found that the frequency of the second most common sequence was less than 35% but greater than 10% relative to the first allele (63: 35–90:10) for 13.95% of our population. A large proportion of these individuals had either Odvi-DRB*14 or Odvi-DRB*20 as their second allele (25.0 and 38.3%, respectively), which was further validated using pedigree and Sanger data. This could be due to allelic dropout because of a mutation in the priming region of these alleles, thereby reducing their amplification efficiency relative to the first allele. Allelic dropout is fairly common in studies that use PCR and has previously been documented in MHC genes [94]. Positive diversifying selection could also be contributing to MHC-DRB alleles having substitutions in the primer regions.

Varying levels of MHC-DRB diversity have been found among ungulates. For example, the genetic distance between MHC-DRB alleles can range from large (cattle, bison, red deer) to negligible (moose; [59]). Within our white-tailed deer dataset, haplotype diversity (Hd) was similar between the second exons of MHC-DRB and MHC-DOB, while nucleotide diversity (n) was greater for MHC-DRB exon 2 than for MHC-DOB exon 2. Nucleotide diversity equal to 0.1 or greater is considered high [81], suggesting that the white-tailed deer MHC-DRB exon 2 alleles differed considerably from one another.

In evaluating potential selection on MHC-DRB exon 2, which encodes the antigen-binding site of DR molecules [30], at the species level, we found strong evidence of positive, diversifying selection (e.g. dN/dS ratio). Similar results were reported by De et al. [21] for white-tailed deer. Diversifying selection occurs when the amino acid diversity in a gene increases within a species over time [17, 108]. As MHC-DRB produces peptide-binding proteins, increasing this gene’s diversity may enable it to present a greater array of pathogenic antigens to the immune system [1, 13, 40, 93]. In evaluating potential selection within our population, the neutrality tests indicate that there is an excess of intermediate frequency MHC-DRB exon 2 alleles in our population, suggesting that the MHC-DRB alleles are being maintained in the population via balancing selection that is becoming more pronounced in our population over time as there was an increasing trend in all neutrality test values from the 2007 starting population until the 2016 adult population. Additionally, we see an excess of heterozygotes in the 2016 population, and this retention of MHC-DRB alleles may be driven by a heterozygote advantage that exists in our population, especially since MHC genes are codominant. But, since our population sizes for the HWE exact tests were relatively small, the heterozygote excess found in our 2016 population may also be a product of drift. Another explanation for the increasing Tajima’s D and Fu and Li’s F* and D* values found for MHC-DRB may be that while balancing selection acted on the population initially (before population enclosure), a reduced population size is now adding to the generation of intermediate frequency MHC-DRB exon 2 alleles in our population. While our selection analyses are consistent with MHC-DRB exon 2 being under balancing selection in our population, it is important to acknowledge the limitations of our study in both population size and the limited number of loci studied, thereby making it impossible to discount the power of drift in producing these patterns. Demographic processes such as population size decline affect the whole genome, whereas selection typically only affects the target loci and closely linked regions. Therefore, to fully assess the underlying reasons for the excess intermediate frequency MHC-DRB exon 2 alleles, future work would have to take a broader genome approach for our population.

In contrast, MHC-DOB exon 2 appears to have been under purifying selection at the species level (0 > dN/dS).
dS > 1; [11]) thereby eliminating harmful nonsynonymous substitutions from MHC-DOB in white-tailed deer. The intensity of purifying selection depends on how tolerant a genomic region is towards mutations, or how functionally constrained it is. DNA regions in which a mutation is likely to affect their gene product’s function tend to be more functionally constrained and have lower substitution rates. MHC-DOB exon 2, which encodes the extracellular domain of the MHC-DOB protein [5, 62], may therefore be highly constrained for the white-tailed deer species. While haplotype diversity at the nucleotide level was high in our population, most amino acid sequences for MHC-DOB exon 2 had little to no differences among them (mean pairwise difference was < 1%), and we only found 3 unique amino acid sequences. Neutrality tests for MHC-DOB were not significant, suggesting that MHC-DOB seems to be evolving neutrally within our white-tailed deer population. Although the pattern of the selection analyses is consistent with selection against heterozygous individuals (e.g. heterozygote deficiency in the 2016 population), it is also consistent with drift due to the small effective population size and potential inbreeding.

Conclusions
Overall, this study identified 12 new MHC-DRB exon 2 alleles and characterized a new, non-classical, MHC II gene (MHC-DOB) for white-tailed deer. We also found a lack of significant linkage between these two loci, which suggests there may be a chromosomal inversion in the MHC II region of white-tailed deer. However, more research is required to confirm this. If a chromosomal inversion is indeed found, recombination rates between the two MHC II subregions should be examined, as rates have been found to differ between individuals [68]. Lastly, more MHC polymorphism may be found when using next generation sequencing for white-tailed deer populations from other parts of the Americas. Improving our understanding of MHC II gene structure and polymorphism in white-tailed deer will enable us to further examine how these unique, highly polymorphic genes influence morphology, reproductive success, and overall population dynamics in white-tailed deer.

Methods
Study area
This study took place at the Auburn Captive Facility (ACF) located north of Camp Hill, Alabama. The facility was part of the Piedmont Agricultural Experiment Station, which is owned by Auburn University. Deer sampled during the study were enclosed in a 174-ha facility surrounded by a 2.6-m fence, which was constructed in October 2007. The deer present within the ACF during the study included the original deer that inhabited this area during the fence formation in 2007 and their subsequent offspring. The population size of the adult, founding population was 71, while the effective population size was 64.8, calculated as Ne = 4NmNf / (Nm + Nf), where Nm is the number of breeding males and Nf is the number of breeding females in the population [106]. Subsequent to fencing the area, deer were neither introduced nor hunted within the ACF. Population size varied annually between 100 and 120 deer. Deer had access to supplemental feed in the form of food plots, corn feeders, and ad libitum protein feeders. A creek and its tributaries were present on the property, which provided a reliable water source year-round.

Animal handling
White-tailed deer aged 6 months and older were captured over 10 trapping seasons (October – July each year) from 2007 to 2017 via chemical immobilization. We administered a tranquilizer mixture into the deer’s hindquarter muscle with the use of cartridge fired dart guns (Pneu-Dart model 193) and 0.22 caliber blanks. After adding 4 cc of xylazine (100 mg/ml; Lloyd Laboratories, Shenandoah, IA) to a 5 mL vial of Telazol® (100 mg/ml; Fort Dodge Animal Health, Fort Dodge, IA; [60]), we loaded 2 cc of this tranquilizer mixture into a telemetry dart (2.0 cc, type C, Pneu-Dart Inc., Williamspor, PA), which contained a radio transmitter (Advanced Telemetry Systems, Inc., Isanti, MN) that allowed us to locate the sedated deer via radio telemetry [44]. Tolazine (1.5 ml/45.36 kg) was injected in the shoulder and hindquarter muscle to reverse the sedation once data collection was complete [61]. The animals used in this study were part of a long-term study on reproductive ecology of white-tailed deer [64] and were maintained as part of the study population until they died of natural causes. These methods were approved by the Auburn University Institutional Animal Care and Use Committee (2008–1417, 2008–1421, 2010–1758, 2011–1971, 2013–2372, 2014–2521, 2016–2964, and 2016–2985) and in compliance with the American Society of Mammalogists’ guidelines [91]. Deer received a unique 3-digit identification number at initial capture, which was displayed on ear tags and also freeze branded on the front shoulder and hind quarter of some individuals. Sex and age (tooth wear and replacement aging technique; [89]) were recorded and a 1-cm² notch of tissue was removed from their ear for genetic analysis. This tissue sample was then stored in a – 80°C freezer until DNA analysis could be performed in the laboratory.

Sequencing
We used amplicon-seq to characterize MHC-DRB and MHC-DOB for 381 individuals. A modified version of
**Symbiodinium** DNA isolation methodology proposed by Coffroth et al. [18] was used to extract DNA from tissue samples collected 2007–2013. DNeasy blood and tissue kits (Qiagen, Inc.) were used to extract DNA from tissue samples collected 2014–2017 and from previous samples with low DNA yield. DNA for 381 individuals were shipped to RTL Genomics (Research and Testing Laboratory, Lubbock, TX, United States) for amplicon sequencing of 307 bp of the second exon of **MHC-DRB** (primers LA31 and LA32; [56, 90]) that targeted the peptide binding domain, and 398 bp of the second exon of **MHC-DOB** (Forward: 5′-AAAGCCCCTCCTCTCATAATCC-3′; Reverse: 5′-CCACCAAGGAGACCCACAC-3′). Primers for **MHC-DOB** were constructed using available genomic data for white-tailed deer (Ovir.te.1; NW_018336049.1; [87]).

At RTL Genomics, samples were amplified via a two-step process. First, the forward primer was constructed by combining the Illumina i5 sequencing primer (5′-GTCTCGCATGATGTTATAAGAGACAG-3′) with the specific forward primer for each **MHC-DRB** and **MHC-DOB**, and the reverse primer was created using the Illumina i7 sequencing primer (5′-GTCTGCAGTGTTATAAGAGACAG-3′) and the specific reverse primer for each **MHC-DRB** and **MHC-DOB**. Amplifications were performed in 25 μL reactions using Qiagen HotStar Taq master mix (Qiagen Inc., Valencia, CA), 1 μL of each 5 μM primer, and 1 μL of template on ABI Veriti thermocyclers (Applied Biosystems, Carlsbad, CA). Thermal profiles used for **MHC-DRB** and **MHC-DOB** were identical: 1) 95°C for 5 min; 2) 33 cycles of 95°C for 1 min; 3) 50°C for 30 s; 4) 72°C for 1 min; 5) final extension of 72°C for 10 min; 6) 4°C hold. Products from this first round of amplification were then added to a second PCR based on qualitatively determined concentrations. A second PCR was performed to index the sequence by individual. Primers for the second PCR used primers that were based on the Illumina Nextera PCR primers (Forward: 5′-AAATGATACGGCGACACCCAGAGATCTACAC[5′index]TGCTCGACCCAGAAGTC-3′; Reverse: 5′-CAACGACAGAAGGCCGATACCTGAGAT[5′index]GTCTCGTGAGGGCTGGC-3′) using the following thermal profile: 1) 95°C for 5 min; 2) 10 cycles of 94°C for 30 s; 3) 54°C for 40 s; 4) 72°C for 1 min; 5) final extension of 72°C for 10 min; 6) 4°C hold. All amplification products were visualized with eGels (Life Technologies, Grand Island, NY). The DNA extraction from one individual failed to amplify and was therefore removed from the dataset prior to sequencing. PCR products indexed by individual were pooled in equimolar. Each pool was size-selected in two rounds using SPRIselect Reagent (Beckman Coulter, Indianapolis, IN) in a 0.75 ratio for both rounds. These size-selected pools were then quantified using the Qubit 4 Fluorometer (Life Technologies). Successfully amplified size-selected pools were loaded on an Illumina MiSeq (Illumina, Inc., San Diego, CA) 2 × 300 flow cell at 10 pM. The amplicons were then sequenced for 300 bp length paired-end reads on the Illumina MiSeq platform for a targeted minimum of 10 k reads per individual. Reads were demultiplexed based on the index and sorted to individual files.

To ensure the quality of the sequencing data, raw reads were trimmed using Trimmomatic (v0.35; LEADING 20, TRAILING 20, SLIDINGWINDOW: 6:20, MINLEN: 20; [12]). The quality of these trimmed reads was verified using FastQC (v0.11.8). Trimmed, paired-end reads were then merged using PEAR [109]. These merged reads were further filtered by removing reads with corrupt primers, merged reads shorter than 290 base pairs, and reads whose sequence only occurred once within an individual. Once primers were removed from these filtered merged reads, the amplicon size for **MHC-DRB** and **MHC-DOB** were 250 bp and 360 bp, respectively. While the **MHC-DRB** amplicon only captured exon 2, the **MHC-DOB** amplicon captured exon 2 (270 bp) plus noncoding regions around it. Given this extra data for **MHC-DOB**, we analyzed both the extended **MHC-DOB** sequences and **MHC-DOB** exon 2 for further analyses.

**Defining new alleles**

Alleles were defined based on nucleotide sequence variation and amino acid sequence variation. All new **MHC-DRB** and **MHC-DOB** alleles followed the nomenclature proposed by Klein et al. [45] and Van Den Bussche et al. [99, 100]. To define alleles and genotypes within an individual, the merged paired-end sequences were organized from the most frequent to least frequent sequence. Individuals were characterized as homozygotes when the frequency of the second most common sequence was less than 10% relative to the first sequence (90:10; median ratio = 99:1). Individuals were classified as heterozygotes if the ratio of the first to second most frequent sequence was less than 90:10 (median ratio = 58:42). For **MHC-DRB**, 60 individuals out of the 380 individuals sampled (15.8%) did not classify as a homozygote or heterozygote using these guidelines, thereby we used pedigree and Sanger sequencing data to confirm their genotypes.

When creating the pedigree for our population, parentage was assigned with 95% confidence [64]. We compared pedigree data with allele assignments made for the clear homozygotes and heterozygotes and found that 95.2% of the parental assignments were consistent with the **MHC-DOB** allele assignments and 96.1% of the parenteral assignments were consistent with the **MHC-DOB** allele assignments. These results reflect the 95% reliability threshold used when creating our pedigree. We therefore did not discard the allele assignments that
were inconsistent with the pedigree data. We also further explored the other less frequent alleles and found that they were chimeric alleles. We removed any individuals who did not have Sanger or pedigree data for confirmation \((n = 5)\), which reduced the sample size for MHC-DRB to 375 individuals. For MHC-DOB, all 380 individuals sampled were clear homozygotes or heterozygotes (extended sequence and exon 2).

New MHC-DRB alleles whose frequencies were less than 0.67% \([5/(2\times375)]\) in the population were not considered true alleles \([67]\) if we could not further validate them using pedigree and Sanger analysis. While 5 MHC-DRB alleles were at less than 0.67% in our population, we were able to validate 3 of these 5 alleles (DRB*28, DRB*29, DRB*30). The other two alleles only occurred once in the population, and these individuals were removed from our DRB dataset for further analysis (Table S1), reducing our sample size further to 373 individuals for MHC-DRB. The minimum allele frequency threshold for MHC-DOB was 0.66% \([5/(2\times380)]\). While all MHC-DOB exon 2 allele frequencies were greater than this minimum, one allele from our extended MHC-DOB sequence data was not. However, we were able to validate this allele using Sanger and pedigree data so it was not removed from the dataset.

The pedigree for our population was created by Newbolt et al. \([64]\). Sanger sequencing \([85]\) was performed for MHC-DRB exon 2 via high throughput sequencing (htSEQ) for individuals born prior to 2008 to augment our MiSeq sequencing data for MHC-DRB.

To investigate whether all characterized alleles belonged to a single MHC-DRB or MHC-DOB locus, we mapped MHC-DRB exon 2 and MHC-DOB exon 2 alleles to the available draft assembly of the white-tailed deer genome (Ovir.te1; \([87]\)). MHC-DRB alleles mapped to two unplaced scaffolds (NW_018337343.1 and NW_018338651.1), and MHC-DOB alleles mapped to one unplaced scaffold (NW_018336049.1). This suggests that there may have been a gene duplication event for MHC-DRB in white-tailed deer, at least for the exon 2 portion of the gene. However, patterns seen in our data for MHC-DRB are more consistent with the presence of one MHC-DRB locus. MHC-DRB sequences that were slightly less frequent than the assigned MHC-DRB alleles for heterozygotes were chimeric sequences of the two most common MHC-DRB alleles for that individual, a commonly known PCR artifact \([36]\). It is unlikely that all of the individuals in our dataset were completely homozygous at both loci and heterozygous for the same MHC-DRB alleles at both loci. Since MHC genes are notoriously difficult to assemble, especially highly variable regions such as MHC-DRB, it is possible that in the draft assembly the MHC-DRB haplotypes were interpreted as different genes and therefore forced onto different scaffolds when assembling this first version of the white-tailed deer genome \([23, 55, 102]\).

**Test for linkage between MHC-DRB and MHC-DOB**

Since MHC-DRB and MHC-DOB may lie on different MHC II subregions in white-tailed deer, as they are separated by an inversion in other ruminants \([8, 16, 82]\), we examined the degree of linkage between MHC-DRB and MHC-DOB among unrelated individuals in our white-tailed deer population using GenePop (Option 2, sub-option 1). Since the second exons of MHC-DRB and MHC-DOB are placed on different scaffolds that were not included in the final assembly of the white-tailed deer genome (MHC-DRB, NW_018337343.1; MHC-DOB, NW_018336049.1; \([87]\)), we were unable to use the available genomic data for addressing this linkage question.

Therefore, to attempt to address this question here, we first assessed linkage disequilibrium (LD) using individuals that are least likely related to one another (individuals born before the fence was constructed in 2007 and that were not offspring from this early group of deer according to pedigree data; \(n = 69\)). We then included individuals without assigned parents in our pedigree data (these individuals could be related to others in the population, but we are not 95% confident that they are; \(n = 122\)) to further assess the possibility of LD between MHC-DRB exon 2 and MHC-DOB exon 2. This method in GenePop \([76, 78]\) tests the null hypothesis that the loci are independent of one another, or in other words, not closely linked \([104]\). We used the default settings where dememorization = 1000, batches = 100, and iterations per batch = 1000.

**Genetic relationships among alleles**

We estimated phylogenetic distances as the number of nucleotide/amino acid differences among alleles and created gene trees for each gene to understand the relationships between MHC-DRB and MHC-DOB. Nucleotide and protein alignments of MHC-DRB and MHC-DOB were created via Geneious (v11.1.5). These alignments were then used to construct phylogenetic trees for the nucleotide and amino acid sequences of both MHC-DRB and MHC-DOB using IQ-TREE (v1.6.9; \([65]\)). IQ-TREE employed maximum likelihood (ML) for tree inference and automatically determined the best-fit model (\(-m\) TEST) via ModelFinder \([42]\) using a standard bootstrap of 1000 replicates (\(-b\) 1000). ModelFinder identifies the best-fitting model of sequence evolution that ultimately produced our data. Phylogenetic support for tree splits was determined via bootstrap values, where a split with \(\geq 95\%\) support is considered statistically significant \([26]\). Outgroups included in the MHC-DRB trees were moose (Alces alces, \([56]\)), roe deer (Capreolus capreolus, \([57, 75]\)), and reindeer (Rangifer tarandus, \([59]\)) as these are
the closest related species to white-tailed deer with MHC-DRB data [33, 71]. More distantly related species could result in a long branch, which may dominate the likelihood and therefore interfere with the tree inference. However, due to a lack of MHC-DOB data among artiodactyls, outgroups for the MHC-DOB trees were more distantly related species including cow (Bos taurus, [112]), sheep (Ovis aries, [107]), and red deer (Cervus elaphus, [7]). The nucleotide tree for MHC-DOB used the extended nucleotide sequences (360 bp), while the amino acid tree used translated MHC-DOB exon 2 sequences. The final ML trees were rooted using FigTree (v1.4.4). The MHC-DOB trees were rooted with the moose outgroup while the MHC-DOB trees were rooted with the cow outgroup. A heatmap was added to the trees via FigTree annotation that corresponded to the Odvi-DRB and Odvi-DOB allele frequencies present in our population.

Population genetic measures and test for selection
Allele and genotype frequencies were generated for each gene via GenePop (v4.2; Option 5, sub-option 1; [76, 78]) to examine the distribution of MHC-DRB and MHC-DOB alleles in our white-tailed deer population. To assess if these allele frequencies were changing over time, Hardy Weinberg Exact Tests were performed on nucleotide data from the adult founding [2003–2007; n = 69 (MHC-DRB), n = 71 (MHC-DOB extended sequence), n = 71 (MHC-DOB exon 2)] and adult 2016 datasets [n = 118 (MHC-DRB), n = 119 (MHC-DOB extended sequence), n = 119 (MHC-DOB exon 2)] with GenePop (Option 1) using the probability-test (sub-option 3; [35, 37]). We also checked for deviations from Hardy Weinberg Equilibrium using score tests [79] to evaluate the presence of heterozygote excess (sub-option 2) or deficiency (sub-option 1) in our founding and 2016 populations. Score tests are more powerful tests than the probability test [79], which will further aid in the detection of deviations from HWE. However, since our population sizes are relatively small, we cannot rule out the effect of drift on these tests. All tests employed default settings (dememorization number = 1000, number of batches = 100, number of iterations per batch = 1000).

Nucleotide (\( \pi \)) and haplotype (Hd) diversity [63] were calculated for both MHC-DRB exon 2 and MHC-DOB (extended sequence and exon 2) using DnaSP v5.10.00 [80] to assess the genetic diversity present within our population. We also performed several neutrality tests (Tajima’s D, Fu’s Fs, Fu and Li’s F*), Fu and Li’s D*) in DnaSP to examine possible evidence of selection or drift at these loci in our white-tailed deer population. These tests were done using the founding, unrelated population (as described above) and the adults present in the population in 2016 (as described above). Pairwise sequence divergence for nucleotide and amino acid sequences were calculated using MEGA X (v10.0.5; [47]), using maximum composite likelihood for nucleotide distances and a Poisson correction for amino acid distances.

Lastly, to test for evidence of positive selection, SNAP (Synonymous Non-synonymous Analysis Program; v2.1.1; [46]) was used to calculate synonymous and non-synonymous substitution rates and the ratio dN/dS for MHC-DRB exon 2 and MHC-DOB exon 2.

**Supplementary information**
Supplementary information accompanies this paper at https://doi.org/10.1186/s12863-020-00889-5.

**Additional file 1: Table S1.** MHC-DRB exon 2 alleles whose frequencies did not meet the minimum allele frequency (0.67%). No pedigree or sanger data was available to validate these sequences, and they only occurred once in our white-tailed deer population. Table S2. MHC-DOB exon 2 alleles found in our white-tailed deer population and their frequencies. (\(^{a}\) indicates that these alleles translated into the same amino acid sequence). Table S3. Number of nucleotide (below diagonal) and amino acid (above diagonal) differences between MHC-DRB exon 2 alleles for white-tailed deer. Table S4. Genotype frequencies (%) for MHC-DRB exon 2 alleles in our white-tailed deer population. Table S5. MHC-DOB alleles for both the extended sequence (360 bp) and exon 2 (270 bp) found in our white-tailed deer population and their frequencies. (\(^{b}\) and \(^{c}\) indicates that these alleles translated into the same amino acid sequence). Table S6. Number of nucleotide (below diagonal) and amino acid (above diagonal) differences between MHC-DOB exon 2 alleles. Table S7. Number of nucleotide (below diagonal) and amino acid (above diagonal) differences between MHC-DOB exon 2 alleles for white-tailed deer. Table S8. Genotype frequencies (%) for the extended MHC-DOB sequence (360 bp) alleles for white-tailed deer. Table S9. Genotype frequencies (%) for MHC-DOB exon 2 alleles in our white-tailed deer population. Figure S1. Cumulative mean codon-by-codon ratio of synonymous to nonsynonymous substitutions (dS/dN) for MHC-DRB exon 2. Nonsynonymous substitutions are significantly more common than synonymous substitutions for MHC-DOB exon 2 (table S7). Figure S2. Cumulative mean codon-by-codon ratio of synonymous to non-synonymous substitutions (dS/dN) for MHC-DOB exon 2. Synonymous substitutions are overall more common than nonsynonymous substitutions.

**Abbreviation**
MHC: Major histocompatibility complex

**Acknowledgments**
We would like to thank the entire team of volunteers who assisted with data collection, especially C. Newbolt who also provided us with the pedigree data, and V. Jackson for helping with the maintenance of the deer facility. We received valuable advice on the phylogenetic analyses from J. Oaks and R. Klabacka. We would also like to thank Code Blue Scents, Moultrie Feeders, EBSCO Industries and Record Rack for their support.

**Authors’ contributions**
SD provided access to the deer facility as well as data collected since the facility’s installment in 2007. All authors were involved in designing the study. NII and CM aided in capturing deer and collecting morphometric data and blood and tissue samples. CM and NII extracted DNA from tissue samples, which CM used for Sanger sequencing and NII sent out for MiSeq sequencing. Raw reads were processed by NII and TS. These processed sequences were then used by NII and TS for defining new alleles, testing for
linkage between loci, creating phylogenetic trees, and generating population genetic measures. All drafted the manuscript, and all authors read and approved the final manuscript.

Funding
Funding was provided by Code Blue Scents, Moultrie Feeders, EBSCO Industries and Record Rack.

Availability of data and materials
All raw sequence data is available on the NCBI Sequence Read Archive (SRA accession # PRXNA339317). The 12 new MHC-DRB exon 2 alleles and the 11 extended MHC-DQB sequences have been deposited in Genbank under accession numbers MK952679–MK952690 and MK952691–MK952701, respectively.

Ethics approval and consent to participate
All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. All procedures performed in studies involving animals were in accordance with the ethical standards of the institution or practice at which the studies were conducted (Auburn University Institutional Animal Care and Use Committee; 2008–2011, 2010–2017, 2011–1971, 2013–2372, 2014–2521, 2016–2964, and 2016–2985).

Consent for publication
Not Applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
1School of Forestry and Wildlife Sciences, Auburn University, Auburn, AL 36849, USA. 2Department of Biological Sciences, Auburn University, Auburn, AL 36849, USA.

Received: 18 March 2020 Accepted: 16 July 2020
Published online: 29 July 2020

References
1. Acvedo-Whitehouse K, Gulland FDM, Boxen L. MHC class II DRB diversity predicts antigen recognition and is associated with disease severity in California sea lions naturally infected with Leptospira interrogans. Infect Genet Evol. 2018;57:158–65.
2. Allen M, Kalantari M, Ylitalo N, Pettersson B, Hagmar B, Scheibenpflug L, et al. HLA-DQB1 genetic diversity of the bovine lymphocyte antigen DRB3.2 locus in Bovine lymphocyte antigen DRB3.2. Immunogenetics. 1991;33:235–41.
3. Allen RH. History and results of deer restocking in Alabama. Montgomery, Bulletin No. 6: Alabama Department of Conservation; 1965.
4. Amills M, Ramya V, Norimine J, Lewin HA. The major histocompatibility complex of ruminants. Rev Sci Tech Off Int Epiz. 1998;17:108–20.
5. Andersson L, Gustafsson K, Jonsson AK, Rask L. Concerted evolution in a
6. Andersson L, Gustafsson K, Jonsson AK, Rask L. Concerted evolution in a
7. Bernatchez L, Landry C. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? J Evol Biol. 2003;16:363–77.
8. Band M, Larson JH, Womack JE, Lewin HA. A radiation hybrid map of BTA23: linkage between loci, creating phylogenetic trees, and generating population genetic measures. All drafted the manuscript, and all authors read and approved the final manuscript.

9. Biswas S, Akemy JM. Genomic insights into positive selection. Trends Genet. 2006;22:437–46.
10. Bernatchez L, Landry C. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? J Evol Biol. 2003;16:363–77.

11. Biswas S, Akemy JM. Genomic insights into positive selection. Trends Genet. 2006;22:437–46.
12. Bolger AM, Lohse M, Usadel B. Trimmmomatic: a flexible trimmer for illumina sequence data. Bioinformatics. 2014;30:2114–20.
13. Borghans JAM, De Boer RJ. Diversity in the immune system. In: Segel LA, Cohen JR, editors. Design principles for immune system & other distributed autonomous systems. UK: Oxford University Press; 2000. p. 161–83.
14. Chennagiri N, White EJ, Frieden A, Lopez E, Lieber DS, Nikiforov A, Ross T, Rask L, Batorsky R, Hansen S, Lip V, Luquette LJ, Macaulay M, Margulies D, Milos PM, Muzney DM, Soreq E, Gibbs RA, Weinstock GM, Womack JE, Skow LC. Comparative analysis of the bovine MHC class IIb sequence identifies inversion breakpoints and three unexplored genes. Anim Genet. 2005;37:121–9.
15. Childers CP, Newkirk HL, Honeycutt DA, Ramachan N, Muzney DM, Soreq E, Gibbs RA, Weinstock GM, Womack JE, Skow LC. Comparative analysis of the bovine MHC class IIb sequence identifies inversion breakpoints and three unexplored genes. Anim Genet. 2005;37:121–9.
16. Chennagiri N, White EJ, Frieden A, Lopez E, Lieber DS, Nikiforov A, Ross T, Rask L, Batorsky R, Hansen S, Lip V, Luquette LJ, Macaulay M, Margulies D, Milos PM, Muzney DM, Soreq E, Gibbs RA, Weinstock GM, Womack JE, Skow LC. Comparative analysis of the bovine MHC class IIb sequence identifies inversion breakpoints and three unexplored genes. Anim Genet. 2005;37:121–9.
17. Cinconicori F, Marcantili P, Arthofer W, Schlick-Steven BC, Steiner FM. Positive diversifying selection is a pervasive adaptive force throughout the Drosophila radiation. Mol Phylogenet Evol. 2017;112:230–43.
18. Coffroth MA, Lasker HR, Diamond ME, Bruenn JA, Bernstein H. Immunogenetics. 1991;33:235–41.
19. de Vries RRP, Mehta NK, Vaidya MD, Gupta MD, Khan PM, van Rood JJ. HLA-linked control of susceptibility to tuberculous leprosy and association with HLA genotypes. Lancet. 1988;2:294–304.
20. Ditthey A, Cox C, Iqbal Z, Nelson MR, McVean G. Improved genome inference in the MHC, using a population reference graph. Nat Genet. 2015;47:682–8.
21. Dutschkoff SS, Hofer SR, Lochlailer RL, Masters RE, Van Den Bussche RA. MHC-DRB evolution provides insight into parasite resistance in white-tailed deer. Southwest Nat. 2005;50:57–64.
22. Ellisworth DL, Honeycutt RL, Silvy NJ, Bickham JW, Klimstra WD. White-tailed deer restoration to the southeastern United States: evaluating mitochondrial DNA and allozyme variation. J Wildl Manag. 1994;58:686–97.
23. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 1985;39:783–91.
24. Fernandez-de-Mera IG, Vicente J, Naranjo V, Fierro Y, Garde JJ, de la Fuente J, Cortazar C. Impact of major histocompatibility complex class II polymorphisms on lycan red deer parasite and wildlife. Infect Genet Evol. 2009;9:1232–9.
25. Fitch AL, Karstad LH. Studies on the pathogenesis of experimental epizootic hemorrhagic disease of white-tailed deer. Can J Com Med. 1971;35:224–6.
26. Gao J, Liu K, Liu H, Blair HT, Li G, Chen C, Tan P, Ma RZ. A complete DNA sequence map of the ovine major histocompatibility complex. BMC Genomics. 2010;11:466.
27. Gaur LK, Nepom GT. Ancestral major histocompatibility complex DRB genes beget conserved patterns of localized polymorphisms. Proc Natl Acad Sci USA. 1996;93:5380–3.
28. Genome Reference Consortium. Genome Reference Consortium Human Build 38 patch release 13 (GRCh38p13), 02/28/2019. BioProject: PRJNA31257, 2019.
29. Godot V, Hanaga S, Beunon I, Tiberghien P, Sarciron E, Gottstein B, Vuitton DA. Resistance/susceptibility to Echinococcus multilocularis infection and cystic profile in humans. Il influence of the HLA B8, DR3, DQ2 haplotype. Clin Exp Immunol. 2001;121:491–8.
30. Guo SW, Thompson EA. Performing the Exact Test of Hardy-Weinberg equilibrium for Multiple Alleles. Biometrics. 1992;48(2):361–72.
31. Haas BJ, Gevers D, Earl AM, Fiegendau M, Ward DH, Gannoukos G, Ciulla D, Tabbaa D, Highander SK, Soregren E, Meite B, DeSantis TZ, Petrosino JF,
87. Seabury CM, Bhattacharai EK, Taylor JF, Viswanathan GG, Cooper SM, Davis DS, Dowd SE, Lockwood ML, Seabury PM. Genome-wide polymorphism and comparative analyses in the white-tailed deer (Odocoileus virginianus): a model for conservation genomics. PLoS One. 2011;6:e15811.

88. Seddon JM, Ellegren H. MHC class II genes in European wolves: a comparison with dogs. Immunogenetics. 2002;54:490–500.

89. Severinghaus CW. Tooth development and wear as criteria of age in white-tailed deer. J Wildl Manag. 1949;13:195–216.

90. Sigurdardottir S, Borsch C, Gustaffson K, Andersson L. Cloning and sequence analysis of 14 DRB alleles of the bovine major histocompatibility complex by using the polymerase chain reaction. Anim Genet. 1991;22:199–209.

91. Sikes RS, Gannon WL. Guidelines of the American Society of Mammalogists for the use of wild mammals in research. J Mammal. 2011;92:235–53.

92. Singh PK, Singh SV, Singh MK, Saxena VK, Horin P, Singh AV, Sohal JS. Effect of genetic variation in the MHC Class II DRB region on resistance and susceptibility to Johne’s disease in endangered Indian Jamunapari goats. Int J Immunogenet. 2012;39:314–20.

93. Sommer S. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool. 2005;2:16.

94. Sommer S, Courtiol A, Mazzoni C. MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. BMC Genomics. 2013;14:542.

95. Steinmetz M, Stephan D, Fischer-Lindahl K. Gene organization and recombination hotspots in the murine major histocompatibility complex. Cell. 1986;44:895–904.

96. Stone RT, Muggli-Cockett NE. BoLa-DIB: species distribution, linkage with MHC, and Northern analysis. Anim Genet. 1993;24:41–5.

97. Swarbrick PA, Schwaiger FW, Epplen JT, Buchan GS, Griffin JFT, Crawford AM. Cloning and sequencing of expressed DRB genes of the red deer (Cervus elaphus) Mhc. Immunogenetics. 1995;42:1–9.

98. Termijtelen A, Meera Khan P, Shaw S, van Rood JJ. Mapping SB in relation to HLA and GLO1 using cells from first-cousin marriage offspring. Immunogenetics. 1983;18:503–12.

99. Van Den Bussche RA. Hoofer SR, Lochmiller RL. Characterization of Mhc-DRB allelic diversity in white-tailed deer (Odocoileus virginianus) provides insight into Mhc-DRB allelic evolution within Cervidae. Immunogenetics. 1999;49:429–37.

100. Van Den Bussche RA, Ross TG, Hoofer SR. Genetic variation at a major histocompatibility locus within and among populations of white-tailed deer (Odocoileus virginianus). J Mammal. 2002;83:31–9.

101. Van Eijk MJT, Beever JE, Da Y, Stewart JA, Nicholas DE, Green CA, Lewin HA. Genetic mapping of BoLa-A, CYP21, DRB3, DYA and PRL on BTA23. Mamm Genome. 1995;6:151–2.

102. Vlijuma A, Mikko S, Hahn D, Skow L, Andenon G, Bergström TF. Genomic structure of the horse major histocompatibility complexity class II region resolved using PacBio long-read sequencing technology. Sci Rep. 2017;7:45518.

103. Wan QH, Zeng CJ, Ni XW, Pan HJ, Fang SG. Giant panda genomic data provide insight into the birth-and-death process of mammalian major histocompatibility complexity class II genes. PLoS One. 2009;4:e4147.

104. Weir BS. Genetic Data Analysis II. Sunderland: Sinauer; 1996. p. 126–455.

105. Wieczorek M, Abualrous ET, Sticht J, Alvaro-Benito M, Stolzenberg S, Noe F, Freund C. Major histocompatibility complex (MHC) class I and MHC class II proteins: conformational plasticity in antigen presentation. Front Immunol. 2017;8:292.

106. Wright S. Size of population and breeding structure in relation to evolution. Science. 1938;87:430–1.

107. Wright H, Redmond J, Ballingall KT. The sheep orthologue of the HLA-DOB gene. Immunogenetics. 1996;43:76–9.

108. Yang Z, Nielsen R, Goldman N, Pedersen AMK. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics. 2000;155:431–49.

109. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics. 2014;30:614–20.

110. Zhang Z, Sun X, Chen M, Li L, Ren W, Xu S, Yang G. Genomic organization and phylogeny of MHC class I loci in cetaceans. J Hered. 2019;esz005. https://doi.org/10.1093/jhered/esz005. Accessed 11 Mar 2020.

111. Zhou X, Xu S, Yang Y, Zhou K, Yang G. Phylogenomic analyses and improved resolution of Cetartiodactyla. Mol Phylogenet Evol. 2011;61:255–64.