Neutral Genetic Processes Influence MHC Evolution in Threatened Gopher Tortoises (Gopherus polyphemus)

Jean P. Elbers, Rachel W. Clostio, and Sabrina S. Taylor

Abstract

Levels of adaptive genetic variation influence how species deal with environmental and ecological change, but these levels are frequently inferred using neutral genetic markers. Major histocompatibility complex (MHC) genes play a key role in the adaptive branch of the immune system and have been used extensively to estimate levels of adaptive genetic variation. Parts of the peptide binding region, sites where MHC molecules directly interact with pathogen and self-proteins, were sequenced from a MHC class I (95/441 tortoises) and class II (245/441 tortoises) gene in threatened and nontreated populations of gopher tortoises (Gopherus polyphemus), and adaptive genetic variation at MHC genes was compared to neutral genetic variation derived from 10 microsatellite loci (441 tortoises). Genetic diversity at the MHC class II locus and microsatellites was greater in populations in the nontreated portion of the gopher tortoise's range (MHC class II difference in mean $A$ = 8.11, $A_R$ = 0.79, $H_O$ = 0.51, and $H_E$ = 0.16; microsatellite difference in mean $A$ = 1.05 and $A_R$ = 0.47). Only MHC class II sequences showed evidence of positive selection ($d_N/d_S > 1$, $Z = 1.81$, $P = 0.04$). Historical gene flow as estimated with Migrate-N was greater than recent migration estimated with BayesAss, suggesting that populations were better connected in the past when habitat was less fragmented. MHC genetic differentiation was correlated with microsatellite differentiation ($Mantel r = 0.431$, $P = 0.001$) suggesting neutral genetic processes are influencing MHC evolution, and advantageous MHC alleles could be lost due to genetic drift.

Subject area: Conservation genetics and biodiversity; Population structure and phylogeography

Key words: conservation genetics, genetic drift, gopher tortoise, Gopherus polyphemus, major histocompatibility complex, microsatellites

Adaptive genetic variation may confer evolutionary potential, but is often inferred using neutral genetic markers (van Tienderen et al. 2002). Neutral genetic markers such as microsatellites have been frequently used to inform management plans for species of conservation concern by estimating overall levels of genetic diversity and delineating populations harboring unique neutral genetic variation, but levels of neutral genetic variation do not directly influence adaptive traits and may not accurately reflect localized adaptation as a result of differing selective pressures (Holderereger et al. 2006). For example, neutral genetic variation may indicate little or no population structure, but the same analysis with adaptive genetic variation may identify distinct populations worthy of separate management.
The gopher tortoise (*Gopherus polyphemus*) is a species of conservation concern and a keystone species of longleaf pine forests through its production of burrows and dispersal of seeds (Eisenberg 1983). Gopher tortoises occur across the gulf coast states from southeastern Louisiana to southern Florida, and populations are currently divided into 2 major management units by the United States Fish and Wildlife Service (USFWS): western populations—those west of the Alabama and Tombigbee Rivers in Alabama—and the remaining eastern populations. Western gopher tortoise populations are federally listed as threatened (USFWS 1987) and have lower reproductive success and genetic diversity as estimated with neutral microsatellite markers compared to eastern gopher tortoise populations (Ennen et al. 2010; Clostio et al. 2012).

During the past century, gopher tortoise populations have declined dramatically due to habitat loss, road mortality, food and pet trade, and possibly infectious disease (Auffenberg and Franz 1982; USFWS 2013). Important habitat such as longleaf pine forests have been cleared such that only ca. 3% of their historical cover remains (Allen et al. 2006). In addition to habitat loss, gopher tortoises are also susceptible to an infectious and occasionally fatal upper respiratory tract disease (URT). Gopher tortoises subjected to *Mycoplasma agassizii* and/or *Mycoplasma testudineum* develop URTD clinical signs (Brown et al. 1999, 2004), so Mycoplasmas have been the focus of most URTD studies. URTD susceptibility may have a genetic basis as Mycoplasmas are widespread but clinically ill tortoises occur mainly in eastern gopher tortoise populations in Georgia and Florida (e.g., Wendland et al. 2010; McGuire et al. 2014). As tortoises differ in basic immune response assays, and populations are genetically differentiated based on neutral markers, different alleles for immune response genes may be present among gopher tortoise populations (Seigel et al. 2003; Kahn 2006; Clostio et al. 2012).

Genes such as those of the major histocompatibility complex (MHC) have been used to estimate levels of adaptive genetic variation in a variety of vertebrates (reviewed in Bernatchez and Landry 2003) and are important because they play a key role in the adaptive branch of the immune system. MHC class I and II genes encode MHC I and II molecules, respectively, which are glycoproteins that bind and present self or foreign peptides to T cells (Klein 1986). MHC class I molecules occur on nearly all somatic cells and present intracellular peptides to cytotoxic T cells, while class II molecules occur on antigen presenting cells and present extracellular peptides to helper T cells (Janeway et al. 2001). Both types of MHC molecules are highly polymorphic and bind and present peptides on a highly variable peptide binding region (PBR, Marsh et al. 1999). PBR nucleotides affect the shape and conformation of the PBR for both types of molecules, thus an individual’s MHC gene repertoire directly influences what types of peptides can be bound and what pathogens can be recognized (Fremont et al. 1992). Because the PBR is so important to MHC molecule function, most studies assessing adaptive genetic variation sequence parts of the PBR.

Parts of the PBR from a MHC class I and class II gene were sequenced from gopher tortoise samples collected across the species’ range. This data could benefit gopher tortoises threatened by infectious disease by identifying populations depauperate in variation, which could influence susceptibility (Elbers and Taylor 2016). MHC variation was also compared to variation at 10 previously genotyped microsatellite loci (Clostio et al. 2012) to determine if neutral genetic processes or selection are influencing MHC evolution. Whether western gopher tortoise populations have reduced genetic variation was also assessed by comparing diversity estimates for MHC and microsatellite loci from western and eastern populations. Because gopher tortoise populations are generally small and isolated, it was predicted that drift would strongly influence MHC and microsatellite variation. Based on previous results showing lower genetic variation in western compared to eastern populations (Ennen et al. 2010), it was also predicted that western gopher tortoise populations would have lower MHC and microsatellite diversity than eastern gopher tortoise populations.

**Methods**

All gopher tortoise samples in this work were used in a previous genetic study where they were genotyped at 10 microsatellite loci (Clostio et al. 2012). Populations west of the confluence of the Tombigbee and Alabama Rivers are western populations and remaining populations are eastern populations (Table 1, Figure 1).

**Primer Design**

Gopher tortoise-specific MHC primers were developed using the MHC class I exon 3 sequence (GenBank accession GQ495891) from *Malaclemys terrapin* (diamond-backed terrapin, McCafferty et al. 2013) and the MHC class II β exon 2 primers MHC-UP and MHC-DP for *Alligator sinensis* (Chinese alligator, Li et al. 2008). The MHC class I exon 3 primers likely amplify a nonclassical locus (see Discussion). After sequencing PCR amplicons following the authors’ recommended thermocycler protocols but using gopher tortoise DNA as a template, the Clontech Universal Genome Walking Kit (catalog no. 638904, Clontech Laboratories, Inc., Mountain View, CA) was used to lengthen the consensus sequences by primer walking via vectorette PCR (Siebert et al. 1995). Final primers were developed for genotyping using lengthened sequences Gopo-UA for the MHC class I locus (GenBank accession KU949600) and Gopo-DAB for the MHC class II locus (GenBank accession KU949601) using the primer design software Oligo 7.0 (Molecular Biology Insights, Inc., Cascade, CO).

**DNA Extraction and Amplification**

A 184-bp MHC class I exon 3 fragment was amplified, which along with exon 2 forms the PBR of class I molecules. The forward primer GopoMhclex3SF (CTTCTCAATGGATGTAAGGCTG) and the reverse primer GopoMhclex3SR (TTTAAAGCCACTCGATGCAGGT) both located within exon 3 were used. A 199-bp MHC class II β exon 2 fragment was also amplified, which along with exon 2 of a class II α gene forms the PBR of class II molecules. The forward primer GopoMhcIIβex2SF (TTACTTCAGGACGGCACCAG) and the reverse primer GopoMhcIIβex2SR (ATCTCTCCGTTGATGGTGAA) both located within exon 2 were used. PCR's were performed in 25 µL reactions containing 2.5 µL of 10X reaction buffer, 1.5 mM MgCl₂, 200 µM dNTPs, 0.2 µM each primer, 100 ng genomic DNA, and 0.625 units of Taq polymerase (New England Biolabs, Ipswich, MA). Thermocycler conditions consisted of initial denaturation of 2 min at 95 °C, followed by 35 cycles of 30 s at 95 °C, 30 s at 58 °C, and 30 s at 72 °C, with a final elongation step of 10 min at 72 °C.
Cloning and Sequencing
Amplicons were purified and sequenced by Beckman Coulter Genomics Single Pass PCR (Danvers, MA). Sequences were edited with Sequencher 5.1 (Genes Codes Corporation, Ann Arbor, MI), and the number of amplified loci were verified via cloning and sequencing 8–12 clones per amplicon using pGEM T-Easy kits (Promega Corporation, Madison, WI).

Data Analysis
For microsatellites, variation was quantified using allelic-based methods with 10 microsatellites from Clostio et al. (2012). Allelic diversity within populations was quantified by measuring the number of alleles per locus ($A$), observed heterozygosity ($H_O$), and expected heterozygosity ($H_E$) with the program Arlequin 3.5.1.2 (Excoffier and Lischer 2010) and allelic richness ($AR$) using HP-RARE v.June6-2006 (Kalinowski 2005).

For MHC loci, the PHASE algorithm (Stephens et al. 2001) implemented in DnaSP 5.10.01 (Librado and Rozas 2009) was used to determine alleles for heterozygotes, and PGDSpider 2.0.8.3 (Lischer and Excoffier 2012) was used to assign allele numbers and convert from FASTA to various file formats.

MHC variation was quantified using allelic- and sequence-based methods. As before, allelic diversity within populations was quantified with number of alleles per locus and observed and expected heterozygosities using Arlequin and allelic richness using HP-RARE. Sequence diversity was estimated using haplotype and nucleotide diversities ($\pi$) with the program DnaSP.

Historical positive selection was tested for by comparing the ratio of nonsynonymous to synonymous substitutions ($d_N/d_S$) over the coding regions of MHC loci for all alleles for each locus using Nei and Gojobori's method (Nei and Gojobori 1986) with the Jukes and Cantor correction implemented in MEGA 5.1 (Tamura et al. 2011).

### Table 1. Genetic diversity values from 10 microsatellites for western and eastern gopher tortoise populations

<table>
<thead>
<tr>
<th>Population (abbreviation)</th>
<th>N</th>
<th>State</th>
<th>Lat</th>
<th>Long</th>
<th>$A$</th>
<th>$AR$</th>
<th>$H_O$</th>
<th>$H_E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Western gopher tortoise populations</td>
<td>277</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sandy Hollow WMA (SH)</td>
<td>5</td>
<td>Louisiana</td>
<td>30.80</td>
<td>-90.37</td>
<td>2.83</td>
<td>1.92</td>
<td>0.55</td>
<td>0.57</td>
</tr>
<tr>
<td>Florida Gas Transmission Line (FGP)</td>
<td>36</td>
<td>Louisiana</td>
<td>30.78</td>
<td>-90.00</td>
<td>3.80</td>
<td>2.16</td>
<td>0.37</td>
<td>0.41</td>
</tr>
<tr>
<td>Ben’s Creek WMA (BC)</td>
<td>21</td>
<td>Louisiana</td>
<td>30.83</td>
<td>-89.96</td>
<td>3.50</td>
<td>2.12</td>
<td>0.45</td>
<td>0.51</td>
</tr>
<tr>
<td>Private Site, Blonde (ST)</td>
<td>4</td>
<td>Louisiana</td>
<td>30.65</td>
<td>-90.06</td>
<td>2.37</td>
<td>1.99</td>
<td>0.32</td>
<td>0.34</td>
</tr>
<tr>
<td>Bogue Chitto WMA (BCG)</td>
<td>4</td>
<td>Mississippi</td>
<td>30.67</td>
<td>-89.79</td>
<td>3.17</td>
<td>2.15</td>
<td>0.75</td>
<td>0.67</td>
</tr>
<tr>
<td>Marian County WMA (MC)</td>
<td>23</td>
<td>Mississippi</td>
<td>31.17</td>
<td>-89.70</td>
<td>3.44</td>
<td>2.11</td>
<td>0.40</td>
<td>0.44</td>
</tr>
<tr>
<td>Desoto National Forest (DNF1)</td>
<td>12</td>
<td>Mississippi</td>
<td>30.67</td>
<td>-89.10</td>
<td>3.25</td>
<td>2.23</td>
<td>0.57</td>
<td>0.56</td>
</tr>
<tr>
<td>Private Site, Wiggins (DNF2)</td>
<td>25</td>
<td>Mississippi</td>
<td>30.58</td>
<td>-89.04</td>
<td>4.50</td>
<td>2.36</td>
<td>0.38</td>
<td>0.45</td>
</tr>
<tr>
<td>Desoto National Forest (DNF3)</td>
<td>45</td>
<td>Mississippi</td>
<td>31.00</td>
<td>-89.07</td>
<td>4.33</td>
<td>2.27</td>
<td>0.43</td>
<td>0.49</td>
</tr>
<tr>
<td>Camp Shelby Training Facility (CS)</td>
<td>21</td>
<td>Mississippi</td>
<td>31.14</td>
<td>-89.13</td>
<td>4.00</td>
<td>2.25</td>
<td>0.49</td>
<td>0.55</td>
</tr>
<tr>
<td>Turkey Fork Recreational Area (TF)</td>
<td>12</td>
<td>Mississippi</td>
<td>31.35</td>
<td>-88.72</td>
<td>3.88</td>
<td>2.32</td>
<td>0.55</td>
<td>0.57</td>
</tr>
<tr>
<td>Westervelt Property, Sandhill (SAH)</td>
<td>7</td>
<td>Mississippi</td>
<td>31.32</td>
<td>-88.61</td>
<td>3.43</td>
<td>2.26</td>
<td>0.47</td>
<td>0.63</td>
</tr>
<tr>
<td>Westervelt Property, Leaksville (LK)</td>
<td>14</td>
<td>Mississippi</td>
<td>31.12</td>
<td>-88.45</td>
<td>3.63</td>
<td>2.27</td>
<td>0.49</td>
<td>0.55</td>
</tr>
<tr>
<td>Grand Bay National Wildlife Refuge (GB)</td>
<td>3</td>
<td>Alabama</td>
<td>30.44</td>
<td>-88.38</td>
<td>2.71</td>
<td>2.20</td>
<td>0.57</td>
<td>0.58</td>
</tr>
<tr>
<td>Meeker Creek Tract, Saraland (AL)</td>
<td>45</td>
<td>Alabama</td>
<td>30.90</td>
<td>-88.15</td>
<td>4.70</td>
<td>2.51</td>
<td>0.47</td>
<td>0.50</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.58</td>
<td>2.21</td>
<td>0.48</td>
<td>0.53</td>
</tr>
<tr>
<td><strong>SE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.16</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>Eastern gopher tortoise populations</td>
<td>164</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solon Dixon, Andalusia (SD)</td>
<td>20</td>
<td>Alabama</td>
<td>31.16</td>
<td>-86.70</td>
<td>6.10</td>
<td>3.25</td>
<td>0.66</td>
<td>0.66</td>
</tr>
<tr>
<td>Jones Ecological Research Center (GG)</td>
<td>26</td>
<td>Georgia</td>
<td>31.23</td>
<td>-84.47</td>
<td>4.90</td>
<td>2.70</td>
<td>0.57</td>
<td>0.59</td>
</tr>
<tr>
<td>Wade Tract Preserve, Thomasville (WT)</td>
<td>26</td>
<td>Georgia</td>
<td>30.75</td>
<td>-84.00</td>
<td>5.00</td>
<td>2.91</td>
<td>0.59</td>
<td>0.64</td>
</tr>
<tr>
<td>Private Site, Statesboro (SCI)</td>
<td>24</td>
<td>Georgia</td>
<td>32.45</td>
<td>-81.78</td>
<td>4.20</td>
<td>2.61</td>
<td>0.42</td>
<td>0.56</td>
</tr>
<tr>
<td>Private Site, Darien (SRS)</td>
<td>22</td>
<td>Georgia</td>
<td>31.37</td>
<td>-81.43</td>
<td>4.36</td>
<td>2.48</td>
<td>0.42</td>
<td>0.52</td>
</tr>
<tr>
<td>Tillman Sandhill Reserve, Jasper Co. (SC)</td>
<td>23</td>
<td>South Carolina</td>
<td>32.50</td>
<td>-81.20</td>
<td>4.56</td>
<td>2.30</td>
<td>0.43</td>
<td>0.51</td>
</tr>
<tr>
<td>Private Site, Nassau Co. (FL)</td>
<td>19</td>
<td>Florida</td>
<td>30.59</td>
<td>-81.56</td>
<td>4.70</td>
<td>2.54</td>
<td>0.38</td>
<td>0.50</td>
</tr>
<tr>
<td>Private Site, Lake Co. (Lake)</td>
<td>4</td>
<td>Florida</td>
<td>28.38</td>
<td>-81.70</td>
<td>3.00</td>
<td>2.62</td>
<td>0.47</td>
<td>0.64</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.63</td>
<td>2.68</td>
<td>0.49</td>
<td>0.58</td>
</tr>
<tr>
<td><strong>SE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.31</td>
<td>0.10</td>
<td>0.04</td>
<td>0.02</td>
</tr>
</tbody>
</table>

$N$, number of individuals; Lat, latitude; Long, longitude; $A$, alleles per locus; $AR$, allelic richness; $H_O$, observed heterozygosity; $H_E$, expected heterozygosity; SE, standard error.

Figure 1. Map of southeastern United States depicting the location of gopher tortoise populations with abbreviated names (see Table 1 for unabbreviated names) sampled during this project. Populations west of the confluence of the Tombigbee and Alabama Rivers are western populations and remaining populations are eastern populations.

Cloning and Sequencing
Amplicons were purified and sequenced by Beckman Coulter Genomics Single Pass PCR (Danvers, MA). Sequences were edited with Sequencher 5.1 (Genes Codes Corporation, Ann Arbor, MI), and the number of amplified loci were verified via cloning and sequencing 8–12 clones per amplicon using pGEM T-Easy kits (Promega Corporation, Madison, WI).
The HLA-A2 sequence in Bjorkman et al. (1987) (GenBank accession AAA76608.2) and the HLA-DRB1*0101 sequence in Brown et al. (1993) (GenBank accession ABW87990.1) were used to identify putative PBR sites for the MHC class I and class II loci, respectively. The mixed effects model of evolution (MEME, Murrell et al. 2012) implemented in the program Datamonkey (Pond and Frost 2005) was used to test for codons where the proportion of branches with dS > dI was greater than zero. Model selection was performed first, and then MEME was used on MHC class I and class II data. Tajima’s D (Tajima 1989) was computed using DnaSP to assess whether MHC loci sequences deviated from neutral expectations. MHC class I data were not analyzed further because this locus may be a pseudogene or nonclassical MHC locus; sample sizes were small.

Population admixture was estimated using the MHC class II locus and microsatellites following the same procedure as Closiot et al. (2012) except TESS 2.31 (Chen et al. 2007) instead of TESS 2.1 was used, and CLUMPAK web server (Kopelman et al. 2015) was used to visualize population assignments. Briefly, Markov chain Monte Carlo (MCMC) algorithm was run with a burnin of 10 000 sweeps followed by an additional 40 000 sweeps. The spatial interaction parameter was set to 0.5, and the conditional autoregressive model (Durand et al. 2009) was used to assess K ranging from 2 to 23 with 20 replicates for each K-value. The optimum K-value was determined when the deviance information criterion stopped changing. An additional 100 simulations were run at the optimal K-value, and the 20 runs with the lowest deviance information criterion values were retained for use with CLUMPAK using the greedy algorithm.

To better understand gene flow among population clusters identified with TESS, 2 Bayesian analyses were used to estimate historical (Migrate-N 3.6.11; Beerli and Felsenstein 2001; Beerli and Palczewski 2010) and recent patterns of migration (BayesAss 3.0.4; Wilson and Ramalho 2003). Migrate-N estimates migration rates (M = m/μ, where M is the mutation-scaled effective immigration rate, m is the immigration rate, and μ is the mutation rate) and mutation-scaled population sizes (θ = 4Nμμ, where θ is the mutation-scaled population size). For all analyses θ and M were estimated using a stepping stone connection matrix model in a Bayesian framework (see Results and Supplementary Figures S1–S2). 0 and M were estimated with an Fst calculation starting position and long chain MCMC runs recording values every 100 steps. For microsatellite data, a Brownian motion microsatellite model was used and sampled every 4 × 10^9 trees using a uniform prior with a minimum of 0, a maximum of 200, and delta of 10 for 0. For MHC class II data, a DNA sequence model was used and sampled every 4 × 10^9 trees with the default prior.

BayesAss estimates recent patterns of migration by calculating a posterior probability of migrant ancestry for each individual. Several trial runs were run until search parameters produced acceptance rates between 0.2 and 0.6 for migration rate (M), individual migrant ancestry (A), and inbreeding coefficients (F). Multiple runs were performed with optimum search parameters but with different seed values to ensure consistent results. Finally, the log probabilities for each separate run were examined to ensure the MCMC had converged using the program Tracer 1.6 (http://beast.bio.ed.ac.uk/tracer). For all runs for each dataset, 10^9 iterations were performed with 10^6 as burnin, with a sampling interval of 10^6. For MHC class II data, deltaM was 0.1, deltaA was 0.1, and deltaF was 0.7. For microsatellite data, deltaM was 0.1, deltaA was 0.25, and deltaF was 0.4.

To characterize recent migration in population clusters, the highest posterior probability of migrant ancestry was used for each individual to classify individuals as nonmigrants, first generation migrants, or second or greater generation migrants (hereafter second generation migrants).

Isolation by distance was tested for with the MHC II locus by comparing pairwise Fst values calculated using Arlequin with geographic distance calculated with the R (R Core Team 2016) package geosphere (Hijmans 2015). To test if neutral genetic processes were influencing MHC class II locus variation, pairwise Fst values calculated in Arlequin for the MHC class II locus were compared to Fst values from Closiot et al.’s (2012) microsatellite data with Mantel tests implemented by the R package ecodist (Goslee and Urban 2007).

Results

Cloning results showed that each MHC primer pair amplified a single locus (i.e., never more than 2 alleles were present per locus per individual). MHC class I and MHC class II sequences were obtained from 95/441 and 245/441 tortoises, respectively, therefore, diversity may have been underestimated, a possible consequence of mutations in the priming regions.

Polymorphism at Microsatellite Loci

A total of 441 tortoises from 23 populations were genotyped at 10 microsatellite loci (Table 1, Figure 1). The number of alleles per population was greater in eastern than western populations (Wilcoxon rank sum test, W = 17.5, P < 0.01). The same pattern was present for allelic richness (W = 4, P < 0.01). Observed (W = 57, P = 0.87) and expected heterozygosities (W = 39.5, P = 0.196) were equivalent between western and eastern populations.

Polymorphism and Selection at MHC Loci

The MHC class I exon 3 locus exhibited low levels of variation. There were 6 different alleles (GenBank accession numbers KU949614–KU949619) detected from 95 tortoises representing 18 populations, and there were no heterozygous individuals (Supplementary Table S2). Out of 184 sites, 24 were variable, and haplotype and nucleotide diversity were 0.39 and 0.03, respectively. Tajima’s D for all sequences was 0.60 and was not significant at the 0.05 level. dS/dI for all codons across all 6 alleles was not greater than unity according to a Z-test (Z = −0.49, P = 1.00), and the same was true when using only putative PBR sites (Z = −0.49, P = 1.00). MEME found no sites with evidence of episodic diversifying selection at the 0.1 or 0.05 significance levels. Tortoises in western and eastern populations shared alleles Gopo-UAA*1 and Gopo-UAA*3. Only tortoises in western populations possessed alleles Gopo-UAA*2 and Gopo-UAA*4, and only tortoises in eastern populations possessed alleles Gopo-UAA*5 and Gopo-UAA*6. Western and eastern populations had an equivalent number of alleles per locus (W = 20, P = 0.31), allelic richness (W = 21, P = 0.38), observed (W = 28, P = NA), and expected (W = 4, P = 0.8) heterozygosity, and nucleotide diversity (W = 19.5, P = 0.28). Tajima’s D was significant for only the BG population (D = 2.26) in southern Mississippi.

The MHC class II β exon 2 locus was much more variable than the class I locus with 113 different alleles (GenBank accession numbers KU949620–KU949732) detected in 245 tortoises representing 23 populations (Table 2). Out of 199 total sites, 115 were variable, with a haplotype diversity of 0.89 and a nucleotide diversity of 0.13.
Tajima’s $D$ for all sequences was 0.25 and was not significant at the 0.05 level. The ratio of nonsynonymous to synonymous substitutions for all codons across all 113 alleles was not significantly greater than unity according to a $Z$-test ($Z = -0.53, P = 1.00$), but using putative PBR sites it was significantly greater than one ($Z = 1.81, P = 0.04$). MEME identified codons 10, 16, 38, 52, 55, 63, and 64 with $d_N > d_S (P < 0.05)$, but only codon 38 was significant after correcting for multiple tests with the false discovery rate ($Q < 0.05$). Codons 10, 16, 38, and 52 but not 55, 53, and 64 were putative PBR sites. Tortoises in western and eastern populations shared 9 alleles (Supplementary Table S1). Western tortoise populations possessed 31 unique alleles, and eastern tortoise populations possessed 73 unique alleles (Supplementary Table S1). There were more alleles per locus ($W = 22, P = 0.02$) and higher allelic richness ($W = 20, P = 0.01$) and observed ($W = 12, P < 0.01$) and expected ($W = 20, P = 0.02$) heterozygosities in eastern than western populations. Nucleotide diversity was equivalent in western and eastern populations ($W = 87.5, P = 0.08$). Tajima’s $D$ was significant only in the SH (southeastern Louisiana), GB (southern Mississippi), AL (southern Alabama), SRS (eastern Georgia), SC (South Carolina), and FL (northeastern Florida) populations.

**Population Admixture**

Population structure estimated with the MHC class II locus produced $K = 3$, dividing the populations into 3 clusters (Figure 2A).

<table>
<thead>
<tr>
<th>Population abbreviation</th>
<th>$N$</th>
<th>$A$</th>
<th>$A_r$</th>
<th>$H_o$</th>
<th>$H_e$</th>
<th>$\pi$</th>
<th>Tajima’s $D$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Western gopher tortoise populations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SH</td>
<td>2</td>
<td>4</td>
<td>2.00</td>
<td>0</td>
<td>0.67</td>
<td>0.13</td>
<td>2.31*</td>
</tr>
<tr>
<td>FGP</td>
<td>13</td>
<td>9</td>
<td>2.71</td>
<td>0.31</td>
<td>0.71</td>
<td>0.12</td>
<td>1.48</td>
</tr>
<tr>
<td>BC</td>
<td>9</td>
<td>8</td>
<td>2.99</td>
<td>0.44</td>
<td>0.80</td>
<td>0.13</td>
<td>0.78</td>
</tr>
<tr>
<td>ST</td>
<td>2</td>
<td>3</td>
<td>3.00</td>
<td>0.50</td>
<td>0.83</td>
<td>0.15</td>
<td>0.51</td>
</tr>
<tr>
<td>BG</td>
<td>3</td>
<td>5</td>
<td>3.60</td>
<td>0.67</td>
<td>0.93</td>
<td>0.15</td>
<td>0.55</td>
</tr>
<tr>
<td>MCF</td>
<td>13</td>
<td>8</td>
<td>2.07</td>
<td>0.31</td>
<td>0.47</td>
<td>0.09</td>
<td>-0.77</td>
</tr>
<tr>
<td>DNF1</td>
<td>8</td>
<td>7</td>
<td>2.49</td>
<td>0.38</td>
<td>0.63</td>
<td>0.11</td>
<td>0.94</td>
</tr>
<tr>
<td>DNF2</td>
<td>7</td>
<td>1</td>
<td>1.00</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>DNF3</td>
<td>11</td>
<td>10</td>
<td>2.97</td>
<td>0.36</td>
<td>0.78</td>
<td>0.12</td>
<td>1.91</td>
</tr>
<tr>
<td>CS</td>
<td>6</td>
<td>3</td>
<td>2.08</td>
<td>0.17</td>
<td>0.53</td>
<td>0.10</td>
<td>0.74</td>
</tr>
<tr>
<td>TF</td>
<td>4</td>
<td>4</td>
<td>2.77</td>
<td>0.25</td>
<td>0.75</td>
<td>0.07</td>
<td>0.59</td>
</tr>
<tr>
<td>SAH</td>
<td>5</td>
<td>7</td>
<td>3.50</td>
<td>0.80</td>
<td>0.91</td>
<td>0.12</td>
<td>1.03</td>
</tr>
<tr>
<td>LK</td>
<td>8</td>
<td>7</td>
<td>2.49</td>
<td>0.38</td>
<td>0.63</td>
<td>0.11</td>
<td>0.35</td>
</tr>
<tr>
<td>GB</td>
<td>2</td>
<td>2</td>
<td>2.00</td>
<td>0</td>
<td>0.67</td>
<td>0.12</td>
<td>2.30*</td>
</tr>
<tr>
<td>AL</td>
<td>34</td>
<td>16</td>
<td>3.16</td>
<td>0.59</td>
<td>0.83</td>
<td>0.12</td>
<td>2.43*</td>
</tr>
<tr>
<td>Mean</td>
<td>6.27</td>
<td>2.59</td>
<td>0.34</td>
<td>0.72</td>
<td>0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE</td>
<td>0.98</td>
<td>0.18</td>
<td>0.06</td>
<td>0.04</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Eastern gopher tortoise populations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>9</td>
<td>6</td>
<td>2.54</td>
<td>0.22</td>
<td>0.67</td>
<td>0.12</td>
<td>1.21</td>
</tr>
<tr>
<td>GG</td>
<td>18</td>
<td>25</td>
<td>3.72</td>
<td>0.72</td>
<td>0.95</td>
<td>0.08</td>
<td>0.41</td>
</tr>
<tr>
<td>WT</td>
<td>16</td>
<td>27</td>
<td>3.94</td>
<td>1.00</td>
<td>0.99</td>
<td>0.11</td>
<td>0.93</td>
</tr>
<tr>
<td>SCI</td>
<td>22</td>
<td>18</td>
<td>3.31</td>
<td>1.00</td>
<td>0.87</td>
<td>0.09</td>
<td>1.09</td>
</tr>
<tr>
<td>SRS</td>
<td>18</td>
<td>14</td>
<td>3.42</td>
<td>0.89</td>
<td>0.89</td>
<td>0.08</td>
<td>2.43*</td>
</tr>
<tr>
<td>SC</td>
<td>17</td>
<td>11</td>
<td>3.18</td>
<td>1.00</td>
<td>0.84</td>
<td>0.08</td>
<td>2.04*</td>
</tr>
<tr>
<td>FL</td>
<td>15</td>
<td>8</td>
<td>2.93</td>
<td>1.00</td>
<td>0.79</td>
<td>0.08</td>
<td>3.35*</td>
</tr>
<tr>
<td>Lake</td>
<td>3</td>
<td>6</td>
<td>4.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.12</td>
<td>0.40</td>
</tr>
<tr>
<td>Mean</td>
<td>14.38</td>
<td>3.38</td>
<td>0.85</td>
<td>0.88</td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE</td>
<td>2.92</td>
<td>0.18</td>
<td>0.10</td>
<td>0.04</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$N$, number of individuals; $A$, alleles per locus; $A_r$, allelic richness; $H_o$, observed heterozygosity; $H_e$, expected heterozygosity; $\pi$, nucleotide diversity; SE, standard error.

*Significant, $P < 0.05$.

Figure 2. Population clusters for: (A) 245 individuals sampled from 23 sites using MHC class II exon 2 sequence data and (B) 441 individuals sampled from 23 sites using 10 microsatellite loci. Clusters in (A) and (B) are displayed geographically in Supplementary Figures S1 and S2, respectively.

The first cluster included all western populations and SD (i.e., central Alabama). The second cluster was GG and WT (i.e., western Georgia populations), and the third cluster of populations was from SCI to Lake (i.e., eastern Georgia, South Carolina, and Florida populations). Population structure estimated with the 10 microsatellites produced $K = 5$, splitting the populations into 5 clusters (Figure 2B). The first cluster included all of the western populations. The second cluster included SD (i.e., central Alabama). The third cluster included GG and WT (i.e., western Georgia). The fourth cluster included SCI,
SRS, and SC (i.e., eastern Georgia and South Carolina), and the final cluster included FL and Lake (i.e., Florida) populations.

**Migration Analyses**

For the MHC class II locus, there were more migrants moving from western Georgia to threatened western and central Alabama populations historically (Supplementary Figure S1). Western populations for the class II locus also had a smaller mutation-scaled population size than western Georgia populations, and western Georgia populations had a higher mutation-scaled population size than eastern Georgia, South Carolina, and Florida populations. Historical migration was almost symmetric between western Georgia and eastern Georgia, South Carolina, and Florida populations. Using the 10 microsatellite loci, historical migration rates were asymmetric between each of the 5 clusters except between central Alabama and western Georgia (Supplementary Figure S2). Unlike data from MHC class II locus, mutation-scaled population size was greatest for western populations followed by western Georgia, eastern Georgia and South Carolina, central Alabama, and Florida populations, respectively. It is not appropriate to compare scaled mutation population sizes and migration rates observed between Supplementary Figures S1 and S2 as clusters were composed of different populations for each analysis, and each analysis had different mutation rates for the loci.

Recent migration trends estimated with BayesAss suggest tortoise population clusters are composed of mostly nonmigrant individuals. The MHC class II locus had western and western Georgia populations containing a small proportion of migrants (i.e., <8%, Supplementary Figure S3), and the microsatellite loci had central Alabama, eastern Georgia and South Carolina, and Florida populations with approximately 9%, 2%, and 15% of migrants (Supplementary Figure S3), respectively.

**Population Differentiation and Mantel Tests**

Pairwise $F_{st}$ values for the class II locus ranged from −0.14 to 0.66 with a mean of 0.11 (Figure 3). The 10 microsatellites were less variable in pairwise $F_{st}$ values, which ranged from −0.05 to 0.61 with a mean of 0.27. There was an effect of isolation by distance for the MHC class II locus ($r_M = 0.432$, $P = 0.001$, Figure 3A), and the 10 microsatellites ($r_M = 0.943$, $P = 0.001$, Figure 3B). Pairwise $F_{st}$ values were significantly correlated between the MHC class II locus and microsatellites ($r_M = 0.431$, $P = 0.001$, Figure 4).

**Discussion**

It is rare for MHC studies of reptiles (outside of crocodilian taxa) to sequence both MHC class I and II loci and survey polymorphism from many individuals and populations (Elbers and Taylor 2016). Here, 2 MHC loci were sequenced from gopher tortoises across their range and MHC variation was compared to variation at 10 previously genotyped microsatellite loci. It was found that: 1) eastern populations were typically more diverse than western populations; 2) there are private MHC alleles in both eastern and western populations, which may or may not affect clinical expression of URTD; 3) estimates of Tajima’s D and $d_3/d_4$ indicated that MHC class II sequences deviated from neutrality and historical positive selection influenced MHC class II PBR sites; 4) despite evidence for selection in the PBR of the MHC class II locus, correlations of genetic diversity between microsatellites and each MHC locus were significant, suggesting that MHC loci are strongly influenced by genetic drift, which could result in loss of beneficial MHC alleles; 5) MHC-estimated population structure was consistent with previous microsatellite and mitochondrial DNA inferences; and 6) tortoise populations appear to have been better connected in the past given that Migrate-N
estimated high historical migration rates while recent migration rates estimated in BayesAss were lower. Each of these points is discussed in greater detail below.

There was greater MHC class II and microsatellite diversity in eastern compared to western gopher tortoise populations. Genetic diversity estimated with microsatellites showed that eastern gopher tortoise populations had a greater number of alleles per locus and allelic richness than western gopher tortoise populations, but heterozygosity was similar between eastern and western populations. The MHC class I locus painted a different picture as both eastern and western populations had a greater number of alleles per locus, allelic richness, heterozygosity, and nucleotide diversity. The lack of diversity indicates that this locus may be a nonclassical MHC class I locus (Miller et al. 2007; Glaberman et al. 2008). The MHC class II locus was highly polymorphic and showed that eastern gopher tortoise populations had a greater number of alleles, allelic richness, and heterozygosity than western gopher tortoise populations, but western gopher tortoises still had appreciable diversity, which may result in at least some individuals being resilient to diseases such as URDT. Data from both the MHC class II locus and the microsatellites are in keeping with previous results suggesting western gopher tortoise populations have lower genetic variation than their eastern counterparts (Ennen et al. 2010).

The clinical expression of URDT in eastern but not western gopher tortoise populations is a mystery. It is not known whether certain MHC alleles or possession of multiple different alleles influences URDT susceptibility or severity in gopher tortoises. On the one hand, there are certainly genetic differences in immune gene variation as tortoises in western populations had 2 MHC class I and 31 MHC class II alleles that were not observed in eastern gopher tortoise populations, and eastern gopher tortoises had 2 MHC class I and 73 MHC class II alleles not observed in western gopher tortoises. However, tortoises with URDT have not been sequenced at these MHC loci, so allele associations with URDT susceptibility or resistance are not clear.

Positive selection was not detected acting on the MHC class I locus, which further suggests that it is a nonclassical locus (Hughes and Nei 1989a). Primers were designed from the MHC class I exon 3 sequence from M. terrapin, which McCafferty et al. (2013) thought might be part of a nonclassical MHC class I gene. It is not certain whether the MHC class I locus functions in antigen presentation like a typical classical MHC class I gene, and caution should be exercised regarding significance of population genetic inferences from this locus.

Unlike the class I locus, the MHC class II locus deviated from neutrality and had evidence of positive selection acting on the PBR sites in 3 western and 3 eastern gopher tortoise populations. It is not clear if the positive Tajima’s D values in the southeastern Alabama population and 3 eastern populations are due to balancing selection, population structure, or bottlenecks (Biswas and Akey 2006). Bottlenecks may be partly responsible as Clostio et al. (2012) found these populations had significantly low M ratios based on the method of Garza and Williamson (2001), which can detect multi-generation bottlenecks. The ratio of nonsynonymous to synonymous substitutions was greater than one whether or not PBR sites were defined, which suggests historical positive selection is acting on codons of the class II locus (Hughes and Nei 1989a), a result observed in other reptiles (Badenhorst 2008).

Population structure estimated with MHC class II locus was congruent with groupings identified previously with microsatellite and mitochondrial DNA haplotype clusters (Clostio et al. 2012). Differences may have been observed between MHC and microsatellite-inferred structure because mutation rates may be faster at microsatellite loci, producing finer scale structure as compared to mitochondrial or MHC data (Li et al. 2002).

Gopher tortoise pairwise $F_{ST}$ values for MHC II locus and microsatellites were high in relation to other turtle and tortoise species (e.g., Kuo and Janzen 2004; Escalona et al. 2008; Howeth et al. 2008; Hagerty and Tracy 2010). For example, pairwise $F_{ST}$ values only ranged between 0.003 and 0.132 for the desert tortoise (G. agassizii) populations studied by Hagerty and Tracy (2010). Even the high pairwise $F_{ST}$ value of 0.325 obtained from population comparisons of the yellow-spotted river turtle (Podocnemis unifilis) in South America (Escalona et al. 2008) are low compared to the values obtained from the gopher tortoise (Figure 3). This suggests that some gopher tortoise populations are highly differentiated compared to populations of similar species.

Pairwise $F_{ST}$ values were correlated between the MHC class II locus and microsatellites. This suggests that neutral genetic processes are influencing MHC locus variation. This does not mean that selection is absent as signatures of historical positive selection (Garrigan and Hedrick 2003) were detected acting on codons of the MHC class II locus, rather it is possible that the generally small sizes of gopher tortoise populations allows neutral genetic processes to have a greater influence than if populations were larger, where selection would overpower or mask the effects of genetic drift (Whitlock 2000). If the selection coefficient ($s$) is less than or equal to $(1/2(N_e))$, a locus under selection will behave like a neutrally evolving locus (Wright 1931). For example, an MHC allele with a selection coefficient of 1% would be effectively neutral in a population with an $N_e$ less than or equal to 50 (Frankham et al. 2010). The gopher tortoise populations surveyed here have much larger effective populations than 50 individuals (data not shown), so selection coefficients may be smaller than 1% for many of the alleles observed in this study. Further, although the MHC class II locus may have experienced selection in the past, it may be acting more like a neutral locus in current generations. Genetic drift has been shown to outweigh selection at MHC loci for populations of tuatara (Miller et al. 2008), Peary caribou (Taylor et al. 2012), New Zealand robins (Miller and Lambert 2004), and northern elephant seals (Weber et al. 2004), results that are consistent with this study.

There was an effect of isolation by distance for the MHC and microsatellite loci, which suggests tortoise dispersal limits genetic connectivity as distance among population increases. Gopher tortoises are not particularly vagile and have small home ranges and dispersal distances. For example, gopher tortoise mean home ranges in one Florida population were less than 1 ha, and maximum dispersal distances were less than 1 km (Diemer 1992). This suggests that continued habitat fragmentation will have negative effects on population genetic connectivity for both neutral and adaptive markers.

Migrate-N results indicate that migration was historically more common among population clusters, but BayesAss and isolation by distance results suggest that migration has lessened more recently. Loss of longleaf pine forest (Allen et al. 2006), appears to have reduced genetic connectivity and contributed to loss of genetic diversity in gopher tortoises and other longleaf pine-associated-species (e.g., Stangel et al. 1992; Pauly et al. 2007; Koopman and Carstens 2010; Richter et al. 2014). Although small populations of longleaf pine associated species may have reduced genetic diversity, such populations may be of conservation value because they harbor unique alleles and serve as stepping stones connecting larger populations across the landscape (Stangel et al. 1992).
There are several conservation implications for the findings presented in this article. First, western gopher tortoises have lower genetic variation than their eastern counterparts, which may lower reproductive success in western tortoises, especially in the southeastern Mississippi populations studied by Epplin and Heise (2003). Second, gopher tortoise populations appear to have moderate to high diversity at the MHC class II locus, which is advantageous for the adaptive potential of populations, especially if new pathogens reach gopher tortoise populations (Sommer 2005). Third, neutral genetic processes are influencing MHC evolution in gopher tortoises, which may be due to the generally small sizes of gopher tortoise populations such that selective pressures are low enough that genetic drift is able to appreciably influence genetic differentiation (Willi et al. 2006). Because of small population sizes, it is possible that beneficial alleles may be lost in future gopher tortoise generations due to genetic drift (Fisher 1930). Fourth, these results support Clostio et al.’s (2012) designation of 2 evolutionary significant units for gopher tortoises based on mitochondrial and microsatellite data here also supported by MHC class II data: (1) tortoises west of the Alabama and Tombigbee Rivers, and (2) tortoises east of these rivers. Tortoises in western Georgia probably warrant separate management from those in eastern Georgia due to uniqueness of microsatellite and MHC variation. Finally, gopher tortoise populations appear to have been better connected in the past when habitat was less fragmented, and restoring population connectivity is important for minimizing loss of current levels of genetic variation.

Supplementary Material
Supplementary data are available at Journal of Heredity online.

Funding
This material is based upon work that is supported by the National Institute of Food and Agriculture, US Department of Agriculture, McIntire Stennis project LAB04066 and LAB94169. Funding for J.P.E. was provided by the Lucius Gilbert Foundation.

Acknowledgments
The authors would like to thank LSU High Performance Computing resources, which were used to perform Migrate-N analyses.

Data Availability
Diplodid genotypes for MHC class I and II (DNA sequences and alleles) and microsatellites (alleles) are available from the Dryad repository (doi: 10.5061/dryad.7ck13).

References
Badenhorst L. 2008. The characterization of MHC class II genes of the Nile crocodile (Crocodylus niloticus): an investigation of mechanisms that shape genetic diversity in natural populations. MS Thesis. [Rondebosch (South Africa)]: University of Cape Town.


