Final Report – Revision 1

White-tailed Deer in Arkansas: Genetic Connectivity and Chronic Wasting Disease susceptibility

Submitted to: M. Cory Gray and Chris Middaugh Arkansas Game and Fish Commission 771 Jordan Drive Monticello, AR 72655

Prepared by:

Marlis R. Douglas Tyler K. Chafin Bradley T. Martin Zachery D. Zbinden Michael E. Douglas

Arkansas Conservation & Molecular Ecology Laboratory University of Arkansas Biological Sciences 850 W Dickson St Fayetteville, AR 72701

> Respectfully submitted: 02 August 2019

Table of Contents

Executive Summary

Overview:

This study represents a genetic evaluation of 1,720 white-tailed deer sampled across 75 counties in Arkansas from 2016-2019. It was conducted to better understand the potential risk and spread of Chronic Wasting Disease (CWD) in the state (Fig. 1). The genetic patterns discovered in our study reflect the dispersal of deer through the diverse and complex environments of the state (Figs. 2-4), and help make predictions about the potential spread of CWD from the current Management Zone, but also provide a framework for the adaptive management of both deer and CWD in Arkansas.

Objectives:

- 1. **We used a genomic approach (SNP genotyping) to infer population connectivity of white-tailed deer in Arkansas.** Population genetic patterns reflect deer dispersal through complex environments and demonstrate levels of population connectivity within an anthropogenically modified landscape (Amaral et al., 2016).
- 2. **We also assayed levels of polymorphisms at the** *PRNP* **gene. This allowed us to identify gene variants that are associated with reduced (or elevated) susceptibility to CWD.** The distribution and frequency of *PRNP* gene variants can help identify subpopulations that might be less susceptible to CWD. This, in turn, informs risk assessment and the potential for disease spread (Blanchong et al., 2009; Johnson et al., 2006; Kelly et al., 2014).
- 3. **We also examined our genetic data within the context of environmental and habitat variables to identify landscape characteristics that can promote or inhibit dispersal of white-tailed deer in Arkansas.** Understanding landscape characteristics conducive to deer dispersal can help management to focus on those areas where CWD might more rapidly spread into deer herds that are currently CWD-free.

Approach:

We used a novel and cutting-edge genomic approach (termed 'SNP-Genotyping;' Fig. 5) to analyze our deer samples. This method is more effective than traditional methods (used over the last 20+ years in wildlife management) for detecting subtle population structure, and thus herd connectivity, because it allows hundreds of samples to be screened across thousands of genetic markers. Thus, it is both efficient and cost-effective. We also sequenced a section of the *PNRP* gene to identify variants that might be associated with CWD susceptibility. Given these data, we then determined the spatial distribution of these variants across the landscape. A total of 1,720 samples were evaluated in 75 counties in Arkansas, with the majority taken from the CWD Management Zone (Table 1, Fig. 6).

Key Findings:

- **1.** Analysis of genome-wide genetic variation at ~35,000 SNP loci (genotypes) per individual revealed **eight subpopulations** (gene pools) across Arkansas (Figs. 6-11).
- **2.** Interestingly, our analyses uncovered signatures of **ancestry from earlier** out-of-state **translocation events**, as well as **restocking efforts from within the state** (Fig. 11). These analyses unequivocally demonstrate the power of our analyses to detect genetic traces at both historic and contemporary levels.
- **3. Modelling migration rates** (a proxy for **dispersal**) underscored that some areas of the state displayed **reduced connectivity**, suggesting restricted movement of deer between some regions (Fig. 12).
- **4.** Our analyses were able to break down the variability in **migration rates** into spatial and environmental components. This, in turn, revealed the manner by which climate and land-cover influenced deer dispersal across the state, with the strongest signal indicating the **Arkansas River as a major barrier to deer dispersal**.
- **5. One variant of the** *PRNP* **gene** in white-tailed deer from Arkansas appears to be associated with **reduced susceptibility to CWD**; this result is consistent with similar studies from other states (Tables 2-4).
- **6.** We also found that the **distribution and frequencies of** *PRNP* **variants differed across the state**. Two frequent variants associated with higher (haplotype B) and lower (haplotype C) susceptibility showed distinct distributions (Fig. 18-21), However, sampling density was too low in some areas to support robust statistical analyses (Figs. 16+17).
- **7.** We also found **evidence for natural selection of one** *PRNP* **gene variant. Haplotype C** (with the 96S mutation) was significantly elevated among **older age classes** and is **consistent with** an i**ncreased probability of survival in individuals with this variant** (Fig. 21).
- **8.** The signals in our genetic data for white-tailed deer within the focal CWD zone [e.g., selection of haplotype C (above); CWD prevalence rate and distribution] suggests the possibility that the area **was exposed to CWD for at least 20-25 years, possibly longer**. This demonstrates the slow incubation of this disease in wild populations and underscores the necessity for continued monitoring in the state as one aspect of wildlife management.
- **9. Additional sampling/analyses are recommended** in areas of the state that are represented in our study at low sampling density. This would allow us to better understand of the distribution of *PRNP* **gene variants associated with low/high susceptibility**, as well as clarify movements of deer within and among subpopulations. It would also increase our resolution with regard to the manner by which the **landscape filters the movements** of deer and CWD spread.

I. Introduction

Chronic Wasting Disease (CWD)

Chronic wasting disease (CWD) is a fatal neurodegenerative disease found in white-tailed deer (*Odocoileus virginianus*) and other cervids (Family Cervidae; Joly et al., 2003, Williams and Young, 1980). The pathogen is a misfolded protein, called prion (Miller and Williams, 2003), which is naturally produced by the deer, but becomes infectious when not folded normally. Because prions are derived from a naturally produced protein, an organism's immune system does not recognize it as alien and hence cannot develop immunity against the disease (Prusiner, 1998). However, certain variants of the *PRNP* gene that produces the prion protein seem to be associated with a lower susceptibility to CWD (Blanchong et al., 2009; Brandt et al., 2015, 2018; Johnson et al., 2003, 2006; Kelly et al., 2008; Wilson et al., 2009).

Transmission of CWD occurs through these prions, either through direct contact among deer, or indirectly through exposure to prions in the environment (Gough and Maddison, 2010; Williams et al., 2002). In North America, CWD has caused economic losses in states where it emerged, primarily due to revenue loss associated with reduced hunting (Arnot et al., 2009) and management of CWD is essential to protect wildlife, as well as local economies. To gather necessary data on CWD occurrence and potential spread, several states have implemented disease surveillance and management plans (e.g., Manjerovic et al., 2014).

One important aspect to proactively manage the disease spread is movement of deer to determine the risk of CWD spread among areas. Behavior of deer facilitates spreading of CWD, such as male dispersal over long distances (e.g., >100 km; Kelly et al., 2010), or philopatry of females that can exacerbate vertical and horizontal transmission within matriarchal groups (Grear et al., 2010).

CWD in Arkansas

CWD in Arkansas was first confirmed in February 2016 in a sample from 2.5-year female elk (*Cerphus elaphus*) legally harvested in October of 2015 near Pruitt (Newton County). A few days later, testing also confirmed CWD in a 2.5-year female white-tailed deer found dead at Ponca (also Newton County). During 14-24 March 2016, biologists from the Arkansas Game and Fish Commission (AGFC) and other agencies collected 266 white-tailed deer samples from a 'CWD Focus Area' spanning 125,000-acres in Newton County. Test results indicated 62 individuals (23%) were CWD-positive, with prevalence differing slightly between genders (females= 20%; males=32%). Additional sampling subsequently occurred statewide, to include the collection of road-killed deer. A 2.5-year road-killed male in Pope County (45 miles south of the initial CWD Focus Area) also tested CWD-positive, and prompted a question: Did this individual disperse into the area from another location, or did it indicate that CWD was already established outside the CWD Focal Area? Additional state-wide sampling efforts were initiated in conjunction with hunter harvest and bi-annual surveys to establish a state-wide baseline of CWD occurrence.

As of April 2019, a total of 598 white-tailed deer and 19 elk tested positive for CWD in Arkansas. (Fig. 1). Based on occurrences of confirmed CWD-positive animals, AGFC established a CWD Management Zone (MZ) that include counties within a 10-mile radius of the sampling locations. Currently (June 2019), the MZ in Arkansas encompasses 19 counties located in northwestern Arkansas.

Since discovery of CWD in Arkansas, AGFC has made concerted efforts to proactively manage CWD prevalence and potential disease spread. Management actions include relaxed regulations with regards to bag limits and antler size restrictions and focus on additional harvest of the male deer population, one of the best-known management practices to slow disease spread. In addition, restrictions are placed on baiting, feeding, hunting and moving of wildlife within the Management Zone (Fig. 1). Supplemental feeding concentrates animals, increasing the frequency of direct contact between animals and increases the chances of disease transmission. As new CWD+ cases are discovered, the Management Zone is expanded to counties that fall within a 10-mile buffer radius around that location. During the 2018/2019

hunting season, 241 new cases of CWD+ deer and five CWD+ elk were detected. Following this discovery, the CWD Management Zone was expanded to include Baxter, Scott and Stone counties. Arkansas has a compulsory testing requirement for any elk harvested in the state and offers a voluntary program for harvested deer. Testing is facilitated by a system of decentralized drop-off containers, that will be expanded in 2019.

Study Objectives and Approach

The objectives of this study were to assess population connectivity and quantify distribution and frequency of *PRNP* gene variants among deer herds across all 75 counties in Arkansas to (a) determine potential of CWD spread beyond the current Management Zone due to dispersing of infected deer, and (b) evaluate if there were herds in Arkansas that might be more susceptibility or resistant to CWD. We employed a genomic approach (SNP genotyping) to evaluate genetic structure for inference of population connectivity and assessed variation at the *PRNP* gene using sequence analysis. We then compared spatial patterns of genetic diversity and environmental features to identify landscape characteristics that might be either conducive to or inhibit deer dispersal, and hence spread of CWD into other areas.

Population genetic patterns (Figs. 2-4) reflect dispersal of organisms through complex environments and can reveal if population connectivity is indeed present among areas within an anthropogenically-modified landscape (Amaral et al., 2016). Distribution and frequency of *PRNP* gene variants may help identify subpopulations that more or less susceptible to CWD and help inform risk assessment of disease spread. Spatial analysis of genetic and landscape associations can inform pro-active management plans to mitigate future disease spread.

SNP-genotyping is a state-of the art genomic approach for the examination of fine-scale population structure (Fig. 2) and population connectivity (Fig. 3). Admixture analysis (Fig. 4) of SNP-genotypes assays genetic diversity at thousands of genetic markers through a method called RADseq, rather than the 10-20 markers commonly employed in traditional population genetic studies (Fig. 5). This achieves a much higher resolution to detect subtle genetic

structure (Morin et al., 2004), but also allows to distinguish local from dispersed animals (Senn et al., 2013; Tokarska et al., 2009) under the assumption that a resident will significantly associate with a local family, whereas a transient would not.

The **genetic approach developed in this study** and the database established from genotyping SNP polymorphism and *PRNP* gene variation can eventually be used to address a multitude of important questions related to the biology of white-tailed deer in Arkansas and the potential impact of CWD on the state's deer populations. For example, continued SNP and *PRNP* genotyping of white-tailed deer samples could help **monitor deer population dynamics** and their **response to CWD infection** (e.g., selection and adaptation). Comparison of SNP genoptypes of Arkansas deer *versus* those from other states could potentially inform our understanding of the **origin of CWD in Arkansas**. And a more a more in-depth evaluation of *PRNP* **gene variation** across the state could increase our understanding of how polymorphisms (different forms) can lead to different prion forms that moderate **individual susceptibility** to and **progression of CWD infection** and help determine how variants of the prion protein influence the natural selection of the *PRNP* gene. The genetic data this project produced constitute the beginning of what we hope will eventually become a **large database** that can be used to address important CWD research objectives, such as those listed above, with Arkansas taking a leading role in adaptive management of CWD in white-tailed deer and other cervids.

II. Research Objectives

Key objectives of this study were to assess:

(1) Population connectivity of white-tailed deer across the entire state of Arkansas and

(2) Occurrence of *PRNP* **gene variants known to be associated with lower or higher CWD susceptibility and spatial distribution across Arkansas**

and

(3) Association of landscape features with observed patterns of genetic structure to evaluate potential spread of CWD into other areas in Arkansas.

To infer population connectivity, we assayed genetic population structure using a reducedrepresentation genomic approach that allows for simultaneous screening of thousands of variable (polymorphic) genetic markers across the genome of each individual deer (Fig. 5). The genetic variation screened comprises Single-Nucleotide-Polymorphisms (SNPs), the most common mutation across the genome. SNP genotyping has several advantages over microsatellite analysis traditionally used in wildlife management and conservation including **better resolution to detect subtle genetic population structure** - as often present amongst populations in wildlife species with moderate to high dispersal rates (such as white-tailed deer) - **higher efficiency** and **lower costs for screening hundreds of samples** (Peterson et al., 2012; Puckett, 2017), and can be more reliably repeated, and thus allows for comparisons amongst studies (Kelly et al., 2011).

III. Methods

Sampling and Tissue Acquisition

Tissue samples were provided by the Arkansas Game and Fish Commission (AGFC) and consisted primarily of earplugs or tongue pieces. Samples either represented road kills, hunter harvested animals submitted for CWD testing, targeted kills by AGFC in response to report of a suspected CWD-positive deer, or as a by-product of other agency efforts (e.g. health surveys).

A total of 1,720 samples were processed for this study. Of these ~500 were obtained during Phase 1+2 of the project (Douglas et al., 2018) and primarily represented white-tailed deer from the CWD Management Zone (MZ) as it was defined during 2017, with the majority collected in two CWD-Focal Areas (FA) located in Newton (N=111) and Pope (N=56) counties.

For Phase 3, an additional 1,207 white-tailed deer were obtained. These represented samples collected from all 75 counties in Arkansas (Fig. 6; Table 1, Supplement Table S1). In addition, 29 white-tailed deer collected from native populations of white-tailed deer in Wisconsin (kindly provided by Dr. Michael D. Samuel) were included as a reference to examine potential retained genetic variation from past agency-conducted translocations of white-tailed deer (Donaldson et al., 1951). Wisconsin samples consisted of extracted genomic DNA or tissue samples.

All samples were compiled into a database and assigned a unique lab-specific number that was used to track samples through all lab procedures and can be cross-referenced with the AGFC field-tag (Supplement Table S1). This provides a quick oversight of genetic data generated for each individual sample and allows to associate genetic with ecological data provided by AGFC, such as gender, age, or GIS coordinates.

Lab Work – DNA Extraction

Lab procedures to generate the genetic data followed standard, published procedures, with slight modifications to optimize each step of data generation for this study. For details see Appendix 1 through 4.

The first step involved obtaining DNA from each individual tissue sample (Table S1) that then was used as DNA template to generate the different types of genetic data (see below). Several DNA extraction methods were tested and a modified version of the QIAamp Fast DNA Tissue Kit extraction protocol was identified as yielding the best quantity of high-quality DNA A kit (Appendix 1).

DNA quality is an important factor in obtaining reliable, reproducible genetic data. DNA can degrade if exposed to the environment, such as is the case for road kills. Degraded DNA can cause spurious results, which in turn can lead to erroneous conclusions. Thus, special efforts were made to ascertain the quality of the DNA used to generate the genetic data for this study (see Figure A1-1A and A1-1B in Appendix 1).

Lab Work - SNP Data Generation

To evaluate population structure and infer population connectivity among white-tailed deer herds in Arkansas, we screened genetic variation at thousands of genetic markers called SNPs (=Single Nucleotide Polymorphism). SNPs are the most common mutations in the genome and are used in wildlife management to infer genetic population structure to identify conservation units (Funk et al., 2012), or detect hybridization across closely related species, including cervids (Russel et al., 2019).

To generate the SNP data for each white-tailed deer sample, we developed protocols for a genomic approach called double-digest restriction site associated DNA sequencing, or ddRAD (Peterson et al., 2012) using a combination of *in silico* (computational) and *in vitro* (wet lab) methods (Chafin et al., 2018). Because sequencing the entire genome for hundreds of

individuals is both cost-prohibitive and would yield an unmanageable amount of data, the intent of the ddRAD approach is to reduce the genome in a consistent and repeatable manner so the same loci (regions of the genome) can be analyzed for each white-tailed deer. Details on protocol optimization, SNP data generation and bioinformatics processing are provided in Appendix 2.

SNP Analysis – Population Structure

The SNP data were used to determine genetic population structure of white-tailed deer across all 75 counties in Arkansas. This information can be used to infer how 'connected' deer herds are across the landscape through dispersal of individual deer from one herd to another.

Population structure was inferred from SNP data using an assignment test to determine the most likely number of genetic populations (i.e., gene pools) represented by the current set of samples (for details see Appendix 3). To illustrate spatial extent of inferred populations, results were visualized by plotting each individual as a pie-chart, with colors indicating proportion of the individual's genome reflecting a particular gene pool (Fig. 7).

In addition to the full dataset (N=1,143 deer from Arkansas and N=29 additional samples from Wisconsin for reference.), we also replicated these analyses across several subsets: (i) partitioned by sex (Fig. 8); and (ii) across 10 down-sampled replicates (20% of individuals), generated using a random sample weighted inversely by spatial sampling density (Fig. 9). The latter was performed so as to evaluate the impact of uneven sampling on our analyses. Sampling density is particularly important to consider with regards to distribution of genetic subpopulations (e.g., Fig. 10), because of the uncertainty associated with areas of low sampling density.

Another approach to visualize distribution of gene pools used a modeling approach called Empirical Bayesian Kriging (EBK) (Gribov and Krivoruchko, 2012). Instead of plotting gene pools represented by each individual, EBK models the **probability of genetic populations across**

the landscape (i.e., EBK probability surfaces; Fig. 11). These were also used as a first-pass qualitative evaluation of landscape features, such as rivers and roads, to inform predictor variable selection for in-depth spatial analyses (described below).

SNP Analysis – Estimating Migration

Natural populations commonly exist as clusters of genetically closely-related individuals (e.g., family clans) within a given area. Populations in geographic proximity are genetically similar, but not as closely related as individuals within a given cluster (Fig. 2). Because individuals are limited in the distance they can move across the landscape, populations commonly exhibit a pattern of 'isolation by distance' in which genetic relatedness between populations declines as a function of distance (Wright, 1943).

This pattern of genetic similarity in proximity *versus* genetic differences across distances occurs naturally and is referred to as genetic population structure (Fig. 2). It emerges over generations, as related individuals remain in the same herd (e.g., female fawns generally remain with the maternal herd). Some individuals might move away from the maternal herd (i.e., disperse), as is the case with male fawns once they reach maturity (Fig. 3). Because dispersing individuals carry their unique genetic profile with them, they distribute genetic variation across the landscape (i.e., gene flow). If dispersed animals interbreed with resident individuals, the resulting offspring's genetic profile reflects admixed ancestry of both parental gene pools (Fig. 4).

When habitats vary (e.g. elevation gradients, vegetation type), some aspects of the landscape will invariably have an effect on the probability of individuals moving through that space – for example the presence of barriers such as rivers or roads, or variation in the suitability of habitat. As individual dispersal declines, so does gene flow (admixture of genetically different populations). As a consequence, those landscape features affecting individual movement accumulate over time as variation in genetic differences over space. **Variation in genetic connectivity (='gene flow')** can be estimated by examining spatial patterns

of relatedness, under the general assumption that areas in which genetic relatedness decays very quickly have little gene flow, and areas in which genetic relatedness is retained over large distances have high gene flow. **We visualized this as an 'effective migration surface'** using the program EEMS (Petkova et al., 2015), as a means to examine underlying landscape resistance (Fig. 12).

These results are complementary to the assignment tests (above), in that they model different, but related, evolutionary processes. EEMS models migration, while assignment tests model a *product* of varied migration over space, namely population structure. The primary purpose of our subsequent analyses was to understand how features of the environment interact as determinants of migration and population structure.

SNP Analysis – Examining Relative Dispersal by Sex and Age Class

A commonly recognized pattern in mammals displaying strong male-male competition is a sexbiased dispersal, with juveniles or young males often moving the longest distances (Dobson, 1982). To examine the potential for a) sex; b) age; and c) age X sex bias in relative dispersal, **we examined how genetic relatedness decays as a function of distance from each individual**. To do so, we randomly sampled 5,000 SNP loci (to reduce computation time) and calculated pairwise genetic dissimilarity. We then partitioned the data by sex, and by age class, and calculated mean genetic dissimilarity of individuals within a varyiable radius (e.g. 5km, 10km, 50km) (Fig. 13). The primary assumption here is that dispersing individuals (immigrants) will be on average more genetically dissimilar (=less related) to resident individuals than residents are to one another. These calculations were limited to individuals which had neighboring samples within a 5km radius, thus creating an implicit restriction to high-density sampling zones (Fig. 9). We also qualitatively examined sex-bias on population structure.

PRNP/ PRNPPSG – Data Generation

Genomic DNA was also used as a template to evaluate **genetic variability (=polymorphism) of the prion protein** *PRNP* **gene**. Standard procedures for amplifying and sequencing genes were followed, using modified protocol from previous studies (Brandt et al., 2015, Johnson et al., 2006). For details see Appendix 4. To ascertain if the detected polymorphisms were indeed in the functional *PRNP* gene, presence of the non-coding *PRNP* pseudogene (*PRNPPSG)* was evaluated (O'Rourke et al., 2004).

Sequences were manually verified and compared to a reference database of *PRNP* gene sequences from previous studies in other states (Kelly et al., 2008; Brandt et al., 2015; 2018) that are available from public databases (i.e., NCBI GenBank database). All sequences were compiled into a single database for subsequent data analyses.

PRNP Analysis – Haplotype Data

As diploid organisms, **deer have two alleles for each gene.** Because of the technical methodology of DNA sequencing, these two parental alleles are sequenced *together*, meaning that as an initial step in the analyses of the *PRNP* gene variation, it is first required to distinguish the exact sequence(s) for each allele. To do so, we used a statistical approach called 'phasing' (Stephens et al., 2001) that evaluates common associations between mutations along the *PRNP* sequence which may be indicative of them being commonly inherited- that is, they co-occur on the same chromosomal sequence, and were passed as a singular unit (='haplotype') from one parent to an offspring. When data are present from many hundreds of individuals, these associations can be quantified and used to 'phase' (separate) the two parental alleles (haplotypes) present in each individual. This is a necessary step, because downstream analyses are based on two distinct alleles (haplotypes) forming the genotype (diplotype) of each individual. For details on the phasing analyses see Appendix 4; scripts to format inputs and parse results of haplotype phasing are available at *github.com/tkchafin/haploTools*. In order to

mitigate spurious results, we exclusively retained haplotypes that were phased with high confidence (>0.90 posterior probability).

Haplotypes were then categorized (Table 2) according to the nomenclature of Brandt et al. (2015, 2018). To visualize similarity amongst haplotypes, we constructed a haplotype network using the median-joining algorithm employed by POPART (Leigh et al., 2015); a haplotype network reflects numbers of nucleotide substitutions (point mutations) among the different *PRNP* sequences, and can be particularly informative for understanding the genetic differences among haplotypes (Fig. 14). Haplotype frequencies were calculated globally (i.e., state-wide; Fig. 15), by-county (Supplement Table S2), and by CWD status (positive *vs.* negative) both statewide (Table 3) and Newton County only (Table 4), because our more robust sampling density and higher rate of CWD prevalence in Newton County allows for more robust statistical analyses. Scripts for creating these input files can be found at *github.com/tkchafin/scripts*.

PRNP Analysis – Spatial Distribution of PNRP Variants

To interpolate *PRNP* **haplotype distributions across the state** and examine **correlation with the spatial distribution of CWD** (see Fig. 16), we first had to compute frequencies. We did so by grouping samples into 'pseudo-populations' by dividing the state into a minimum number of non-overlapping polygons, each containing 5-10 sampling localities. This resulted in N=211 polygons (Fig. 17). Haplotype frequencies were then computed within each polygon and interpolated using Empirical Bayesian Kriging. We used this approach rather than computing frequencies by-county in order to take advantage of the increased sampling resolution we have in parts of the state. **Spatial distributions of relative frequencies** for the seven most frequent haplotypes in our data set were then plotted (Fig. 18), with additional details provided for haplotype B (Fig. 19) and haplotype C (Fig. 20), implicated with increased and reduced susceptibility to CWD, respectively.

PRNP Analysis – Signatures of Selection on PRNP Variants

In order to detect **potential biases of specific** *PRNP* **haplotypes on the probability of CWD transmission**, we followed Brandt et al. (2015, 2018) in first quantifying effect sizes in the form of an **odds ratio** (**OR**). The odds ratio is a common measure of association between an exposure and an outcome. Here, *for each haplotype*, we consider the probability of contracting CWD (='outcome'), given presence/ absence of the focal haplotype (='exposure'). The resulting ratio provides information on the **relative risk**, where OR=1 means that individuals with the focal haplotype (e.g. "A", "B") are equally represented among CWD-positive and CWD-negative groups, OR>1 means the focal haplotype is over-represented in CWD-positive samples, and OR<1 means the focal haplotype is under-represented in CWD-positive samples. Note that we specifically use terminology referring to their *relative representation*- this is because multiple factors can affect the odds ratio, and thus may not necessarily translate to conclusions regarding the *susceptibility* associated with particular haplotypes. For example, population structure (as above) may create a coincidental bias, when disease prevalence (or detection) is spatially structured. Additionally, an over-representation of a haplotype among CWD-positive individuals could be due to either an increased probability of contraction, or a *survival bias*, wherein individuals having the focal haplotype are more likely to survive with the disease, thus skewing the odds ratio in a random sample. Consequently, it is important to further consider haplotypes that are implicated by the odds ratio (Tables 3 and 4).

To further evaluate **candidate susceptibility variants** (**CSVs**), we first asked whether the relative representation of the haplotype shows an **effect among age classes**. Here, our assumption was that if a haplotype affects the probability of survival to adulthood (e.g., by reducing CWD risk), then it should increase or decrease in relative frequency in older age classes. Do test this, we first computed the relative representation of CSVs among CWD status groups (CWD+ *versus* CWD- individuals) as an odds ratio within each sampled age class (Table 5). We then tested if this relationship was predicted by age using linear regression (Fig. 21).

To further **test if selection is acting on CSVs showing an age bias**, we computed their relative fitness in the form of a selection coefficient (*s*). This is a common population genetic

statistic used to quantify fitness differences among genotypes, where *s*>0 means a fitness advantage, and *s*<0 means a fitness disadvantage. In order to calculate *s*, we first calculated counts of focal and non-focal haplotypes within each age class, and then treated the age class as a *time-series sample*. This was then used to compute *s* using a statistical model that assumed an idealized population (Mathieson and McVean, 2013).

Landscape Genetics - Spatial Analysis

To identify if landscape features influence dispersal of deer, and consequently population connectivity, genetic data were next evaluated within a spatial context and contrasted against sets of environmental variables describing abiotic (e.g., climatic) and biotic (e.g., vegetations) characteristics (Table 6, Appendix 5 and 6). We first generated a 10 km² hexagonal tessellation grid, overlain across the state of Arkansas. This produced 13,378 tiles (Appendix 5), within which estimates of white-tailed deer migration (logM) were calculated based on genetic information (i.e., SNP differences) among individuals in and around each tile (Fig. 12).

Values for 35 environmental factors from a variety of sources (Table 6, Appendix 6) were also determined for each tile. The goal of the spatial analysis was to 1) determine the **amount of autocorrelation of the estimate of white-tailed deer migration** (logM); 2) determine the **amount of variation in migration** that could be **explained by extrinsic/environmental factors**; 3) create a l**ist of important explanatory factors**; and 4) draw **inferences** about the role of **abiotic and biotic processes driving variation in deer dispersal**.

The dataset was first filtered to include tiles with a **migration rate** (logM) generated from a genetic sampling density > 0.01 to overcome heterogeneity of variance among sampling densities (i.e. low densities had high variance). Standard data transformations followed to prepare the data for subsequent analyses, satisfy assumptions, and reduce multiple collinearity (e.g. transforms to fit normality, and Z-score standardization). Then the dataset was reduced to include only factors that significantly related to the migration parameter (logM) via linear regression and t-tests.

Next, we needed to determine how much of the **variation of the migration rate (logM) could be explained purely by space** (i.e. autocorrelation). The spatial coordinates of the hexagonal tiles were used to deconstruct the spatial arrangement into various functions representing different combinations of 'neighborhoods'. These functions result in each tile having a value that in total creates an eigenvector associated with each function. Each eigenvector can be tested for a relationship with migration (logM), and significant relationships are indicative of spatial autocorrelation in the parameter. A set of significant eigenvectors were retained as the spatial dataset for use in variation partitioning.

Finally, partial linear regression was used to partition variation of migration (logM) among spatial and environmental explanatory datasets. This allowed us to determine how much variation of logM could be explained by autocorrelation alone, by environmental factors alone, and by the combination of the two. This step was critical for drawing inferences about the mechanisms driving white-tailed deer dispersal. Additional details on analyses and parameter selection are provided in Appendix 5.

VI. Results

Sampling

A total of 1,720 tissue samples collected in 75 Arkansas counties from 2016-2019 were processed for this study (Fig. 6). Of these, subsets of samples (Supplement Table S1) were subjected to genomic DNA extraction (N=1,720), sequencing of ~800 nucleotides of the *PRNP* gene (N=1,460) and the *PRNPpsg* pseudogene (N=1,459), and generation of SNP data (N=1,208). SNP data were generated for an additional 30 samples from Wisconsin (not shown in Table S1). Not all samples yielded data of sufficient quality, and these were excluded from subsequent analyses. Four samples from Miller county were identified as representing elk and were also excluded from subsequent analyses. Sampling density was not evenly distributed across the state (Fig. 9), which needs to be taken into consideration when interpreting results and drawing inferences.

SNP Data, Population Structure and Landscape Genomics

SNP data were successfully obtained for 1,226 white-tailed deer samples, including those from Wisconsin. Sequencing of the ddRAD libraries yielded an average of 25,584 (s=8,639) independent genomic loci per sample. Retaining only loci overlapping in a minimum of ~50% of samples yielded a total of 35,642 loci. We then removed loci showing signs of having been overmerged (e.g. individuals with >2 alleles), yielding a final filtered dataset of 35,420 loci, from which 2,655,584 SNPs were discovered. Of these, 54,102 were removed for being only found in a single individual. Further bioinformatic processing randomly selected one SNP per genomic locus (to reduce signal redundancy), yielding a final SNP dataset of 35,099 used for analysis of population structure.

Analysis of population structure performed in ADMIXTURE revealed an optimal number of clusters of K=8, which reflect genetically distinct gene pools (*k*=1 through *k*=8). Samples were

then classified according to their probability of membership to the inferred gene pools (=ancestry) and spatially oriented (Fig. 6). The result shows white-tailed deer to consist of eight weakly differentiated genetic subpopulations that seem to be generally geographically defined (Fig. 10 shows a highly oversimplified version of the spatial distribution of the eight gene pools).

The most apparent pattern is a **north-south division defined by the Arkansas River valley** (Fig. 6). South of the Arkansas River there are two primary genetic subpopulations, loosely geographically bounded by I-30 to the north and the Ouachita River to the south (Fig. 10). Of these, the eastern population is assigned to two gene pools (Fig. 6: *k*=1,3). Note that the latter (*k*=3) is also assigned to individuals in the north-central region of the state (primarily in Baxter, Cleburne, Fulton, Independence, Izard, Lawrence, Randolph, Searcy, Sharp, Stone, Van Buren counties), which may reflect either historical translocations, or is an artefact of weak genetic differentiation, rather than non-random, long-distance dispersal of white-tailed deer between northcentral and southwestern Arkansas (i.e., population connectivity). Coincidentally, the southeastern genetic subpopulation is dominated by gene pool *k*=8 (Fig. 6) that also subsumed all Wisconsin samples, possibly suggesting retention of genetic variation from past translocations.

Populations north of the Arkansas River are more spatially structured, with six main subdivisions (Fig. 6). The most broadly distributed of these genetic subpopulations (defined by gene pool *k*=5) is found in the Mississippi alluvial plains on the eastern side of the state, extending west across the mainstem of the White River and northward to the confluence of the Black and White Rivers, where it grades into several loosely defined, but distinct genetic subpopulations. The northwestern corner of the state is most genetically heterogeneous, being dominated by four primary endemic gene pools (*k*=2,3,4 and 7). The northern-most genetic subpopulations (dominated by gene pool k=4) primarily encompasses Boone, Carroll, and Marion counties, is bound loosely on the east by the White River, and grades to the west into an area of high admixture in Benton and Washington counties. Immediately south is a narrowly defined genetic subpopulation, defined by gene pool *k*=7 mainly found in Newton and Madison counties with a southward transition loosely defined by the Buffalo River region into two

remaining genetic subpopulations, defined by gene pools *k*=2 and *k*=6, that are bounded on the south by the Arkansas River. All of these genetic subpopulations display **considerable admixture** (multiple colors in each pie chart) and spatially weak transitions, indicating that while there is some reduction in gene flow, **none of the geographical or environmental barriers** separating genetic subpopulations **are 'hard' boundaries** to dispersal of white-tailed deer across Arkansas.

Evidence for Age- and Sex- Biased Dispersal

Restricting population structure analyses to males showed them to generally exhibit less genetic structuring than females (Figs. 8 and 13). Patterns of genetic dissimilarity showed higher distances in age 0 (juvenile) males than females, with a shift towards low dissimilarity in males >5 (Fig. 13A). This suggests that dispersal of male deer in Arkansas has occurred as juveniles (age 0) or yearlings (age 1). By age 5, males have contributed to their local gene pools (i.e., produced offspring with resident females), thereby creating the pattern of lower genetic dissimilarity among neighboring individuals, regardless of distance (Fig. 13C-E). In females, juveniles (age 0) had very low genetic dissimilarity to individuals at low distances (Fig 13B-C), with distances peaking at 1-2 years (Fig. 13B-C).

PRNP Gene Variants and Frequency of Resistant Genotypes in AR

Of the 1,460 samples of white-tailed deer sequenced across the *PRNP* gene, a total of 1,433 yielded useful sequences (>98% success rate). Sequences were trimmed to 720 nucleotides that could be scored unambiguously, with 11 sites showing polymorphism (Table 2). Three sites (nt285, nt299 and nt372) previously reported as polymorphic by Brandt et al. (2015, 2018) were invariable, and one site showed a novel mutation (nt499A/C) that was synonymous (no amino acid change). Variation was also detected at nucleotide position (nt413A/G) known to be associated with a mutation at the *PRNPPSG* pseudogene, a duplicated non-functional copy of the *PRNP* gene; nucleotides at nt413 were all adjusted to reflect the *PRNP* genotype (nt413G/G).

Polymorphism at three sites (nt286, nt367 and nt676) represented non-synonymous substitutions (=NSS) that result in amino acid changes (96S, 122T and 255K, respectively). **The non-synonymous substitution at nt286/A (amino acid 96S) has been associated with reduced susceptibility to CWD in white-tailed deer in other states** (e.g., Brandt et al. 2015, 2018, Johnson et al. 2006, 2011).

The variable sites segregated into **20 distinct haplotypes** (Table 2; Fig. 14), with 16 previously documented in other states (e.g., Kelly et al. 2008, Brandt et al. 2015, 2018) and four novel and potentially unique to Arkansas. Two of the Arkansas haplotypes (AR_2 and AR_3) had the non-synonymous substitution 96S (nt286/A) that has been associated with reduced susceptibility to CWD (i.e., haplotype C of Brandt et al. 2018). Three other haplotypes (I, P and V) also have the 96S amino acid change (Table 2). However, all five haplotypes were at low frequencies (<1% except haplotype I at 1.47%), precluding any statistical associations with CWD susceptibility.

Targeted amplification and sequencing of the *PRNPPSG* pseudogene was successful in 30% of white-tailed deer samples (443 out of 1,459). Comparisons of *PRNP* and *PRNPPSG* haplotypes within these 443 individuals revealed that variation at nucleotide site nt413 appears indeed to be the result of non-targeted amplification of the *PRNPPSG* pseudogene and does not constitute biological variants of the *PRNP* gene.

Haplotype frequencies in our data set (Tables 3 and 4) **differ slightly from those reported in other states** (e.g., Illinois and Wisconsin; Blanchlong, 2009; Brandt et al. 2015, 2018). Four of the most frequent (<10%) haplotypes in Illinois/Wisconsin deer populations were also common among Arkansas samples (Fig. 15; haplotypes A-D). However, the most common haplotype (A) detected in 30% of samples in Illinois/Wisconsin, was only found in 15% of samples in Arkansas, where haplotype B and D were most common (each at \sim 23% of samples). Haplotype C, potentially associated with reduced CWD susceptibility, was detected at the same frequency in Arkansas (15%) as in Illinois/Wisconsin (17%). In addition, two haplotypes (E and G), were also at high frequencies in Arkansas (7% and 11%, respectively), but were found in <5% of samples in the Brandt et al. (2018) study. Out of the 16 rare haplotypes by Brandt et al.

(2018) with frequencies ≤1% (haplotypes K-Z), seven were also present at low frequencies in the Arkansas samples (haplotypes K, L, O, P, R, T, and V). **Haplotype frequencies also differ amongst counties** (Supplement Table S2), although **sample sizes for most counties are too low to statistically quantify a phylogeographic signal**.

With regards to **distribution and frequencies of CSV (Candidate Susceptible Variants) haplotypes between CWD-negative and CWD-positive** white-tailed deer in Arkansas (Fig. 18- 20; Table 3, Supplement Table S2), the present results should be considered **preliminary due to limited numbers of CWD positive samples** in some areas of Arkansas (total N=248, Figs. 16 and 17). The seven most common haplotypes were unevenly distributed across Arkansas (Figs. 18- 20).

Evidence for Natural Selection on PRNP Gene Variants

Two CSVs (haplotypes) of the *PRNP* **gene showed significant deviations in representation among CWD+ and CWD- individuals: Haplotypes C and** B, both also sampled by Brandt et al. (2015, 2018) have been implicated with increased and reduced susceptibility to CWD, respectively. In our data set, haplotype B was significantly over-represented among CWD+ individuals, both when considered state-wide (Odds Ratio *OR*= 2.00; *p*=0.000001; Table 3), and when restricted to Newton County (OR=1.47; p=0.033; Table 4). Haplotype C was significantly under-represented in CWD+ individuals, also both in state-wide (*OR*=0.30; *p*=0.00003; Table 3) and Newton County comparisons (*OR*=0.42; *p*=0.015; Table 4).

To further test for variability in CWD susceptibility as driving selection on Haplotypes B and C, we hypothesized that, if variants do indeed impact CWD susceptibility, their representation (frequencies) should vary among age classes. Here, a haplotype that confers **reduced CWD risk** should increase the probability of reaching older age classes (=**extended life expectancy**), therefore should increase in relative frequency among older individuals. Haplotype C was found to increase in representation with increasing age (Table 5A), both in terms of relative frequency (*p*=0.036; *R^2*=0.635; Fig. 21A) and odds ratio (*p*=0.043; *R^2*=0.601;

Fig. 21A), supporting the hypothesis that it reduced CWD risk. Haplotype B did not show any significant relationship among age classes (Table 5B, Fig. 21B), and its over-representation in the dataset amongst CWD+ deer (Tables 3 and 4) may be an artefact of population structure. The majority of CWD+ deer in our data set were sampled from NW Arkansas (Figs. 16 and 17) and haplotype B could simply be at common haplotype in deer herds in this area, independent of CWD occurrence.

To track how the frequency of Haplotype C changed across years (=age classes), we computed a **selection coefficient (***s***) of 0.1215** for Haplotype C. This corresponds to the strength of selection; where 0.0 means no selection (no impact on survival rate), and 1.0 means total selection (e.g. only individuals having that haplotype would survive). Although it can be difficult to interpret if this selection is direct (=acting on mutations at the *PRNP* gene) or indirect (=acting on mutations elsewhere in the genome that are statistically associated with Haplotype C, or influence *PRNP* gene expression) (Barton and Servedio 2015).

Spatial and Environmental Components of Migration

The 35 environmental factors were distilled into a set of 17 factors that significantly explained white-tailed deer dispersal (Table. 7). Because sampling density was very low in parts of Arkansas, analyses were restricted to areas with sufficient number of samples to conduct statistically meaningful analyses (Fig. 23). The set of 17 meaningful factors included riverine barriers to dispersal (e.g. RIVER_DIST, ECOLYS_LU_Open Water, and SECTION Arkansas River Valley), climatic variables (e.g. BIO8_WETTE), vegetation variables (e.g. West Gulf Coast Mesic Hardwood Forest), and section variables (e.g. Boston Mountains). T**he single most explanatory factor was SECTION_Arkansas River Valley** which when present had a mean logM lower than that of non-Arkansas River valley sections by 0.42. For context, logM ranges from -1 to 1 (Fig. 12). This suggests the **Arkansas River is a significant barrier to white-tailed deer dispersal**. Maps and distributions of the 17 environmental factors are provided in Appendix 6 and results of the spatial analyses are detailed in Appendix 7.

Spatial decomposition via dbMEM extracted 81 meaningful eigenvectors, of which 73 were included in the forward selection procedure, and 65 were retained after correcting for multiple tests. These eigenvectors represent spatial autocorrelation in the variable logM (migration rate). Smaller eigenvectors (e.g. V3) represent broad scale spatial structure and larger eigenvectors (e.g. V81) represent relatively finer scale structure (Fig. 24). Prior to variation partitioning, **only the 17 top-ranked eigenvectors were selected so that the environmental and spatial datasets would have the same number of factors**.

A linear model including both spatial eigenvectors and environmental factors explained 67% of the variance in the variable logM (adj. R^2 = 0.67; p < 0.001). There was a large proportion of spatial autocorrelation in the variable logM (adj. R^2 = 0.64; p < 0.001). However, only 20% of **the variance of logM could be attributed to environmental factors**, the majority of which was **spatially structured environmental variation** (17%). Thus, 47% of the variation of logM was spatially autocorrelated, but could not be explained by the environmental factors presented here (Fig. 25). This suggests that the unexplained autocorrelated structure could be explained by either 1). Unmeasured environmental factors, 2). Biotic factors intrinsic to deer populations (e.g. population size), or 3). A combination of these.

V. Discussion

In this study, we addressed three objectives:

- 1. **We evaluated genetic structure in white-tailed deer collected in Arkansas from 2016- 2019 to infer population connectivity using a genomic (SNP genotyping) approach.** Population genetic patterns reflect dispersal of organisms through complex environments, and hence can reveal if connectivity is indeed present among areas within an anthropogenically modified landscape (Amaral et al., 2016; Kelly et al., 2010).
- 2. **We also assayed the levels of polymorphisms at the** *PRNP* **gene using sequencing analysis and to identify gene variants that might be associated with reduced or increased susceptibility to CWD.** Distribution and frequency of *PRNP* gene variants may allow identifying subpopulations less susceptible to CWD and help inform risk assessment of disease spread (Blanchlong et al., 2009; Johnson et al., 2006; Kelly et al., 2008).
- 3. **We analyzed the genetic structure and population connectivity data within a spatial context to examine environmental and habitat variables that promote or inhibit dispersal of white-tailed deer in Arkansas.** Identifying landscape characteristics that are conducive to dispersal of white-tailed deer can help focus management efforts to areas where CWD might spread more rapidly into currently CWD-free herds (Kelly et al., 2014).

Key Project Findings:

- **1.** Analysis of genome-wide genetic variation using a **novel genomic approach** (i.e., genotyping of ~35,000 SNP loci) revealed **eight subtly differentiated subpopulations** (gene pools) among 1,226 white-tailed deer from across Arkansas (Figs. 6-11).
- **2.** These analyses uncovered signatures of **retained ancestry from** past out-of-state **translocation events**, and of restocking efforts within the state (Fig. 11).
- **3.** Modelling migration rates (a proxy for dispersal) suggested **reduced connectivity in some areas** (Fig. 12).
- **4.** Partitioning variation in **migration rates** into spatial and environmental components revealed roles for climatic and land-cover variables in influencing deer dispersal across the landscape, with the strongest signal being associated with the **Arkansas River as a barrier to dispersal**.
- **5. Variants of the prion-coding** *PRNP* **gene** exist among white-tailed deer in Arkansas, including haplotypes (alleles) identified as being associated in other states with r**educed susceptibility to CWD** (Tables 2-4).
- **6. Distribution and frequencies of** *PRNP* **gene variants differ across the state**; with two frequent variants associated with higher (haplotype B) and lower (haplotype C) **susceptibility showing distinct spatial patterns** (Fig. 18-21). However, sampling density in some areas is too low for robust statistical analyses. (Figs. 16+17).
- **7.** We found **evidence for natural selection on Haplotype C** (96S mutation), with a significant increase in representation among **older age classes** consistent with an i**ncreased survival probability** (Fig. 21). Other *PRNP* variants also share the 96S mutation but were present at too-low frequencies for statistical analysis.
- **8.** Signals in the genetic data (e.g., selection of haplotype C, CWD prevalence rate and distribution) suggests that **the population within the focal CWD zone could have been exposed to the disease for at least 20-25 years, but possibly longer.**
- **9. Additional sampling/analyses** in areas of the state with low sampling density would allow to increase understanding of the distribution of the *PRNP* gene variants associated with low/high susceptibility.

Population Connectivity, Dispersal and Disease Spread

'Distribution' and 'abundance' are not only natural history attributes of species but also the building blocks of ecology (Andrewartha and Birch, 1954). Yet, their functional integrity is being challenged by anthropogenic habitat alterations (Chen et al., 2011), and traits such as **'dispersal'** and '**population connectivity**' have become paramount (Crooks and Sanjayan, 2006) in that they more appropriately mirror wildlife ecology in the Anthropocene (Corlett, 2015).

They are important ecological characteristics that translate to other aspects, such as the **spread of emerging infectious diseases** (EIDs) and are thus key elements that quantify the availability of and contact rates for host species (Tracey et al., 2014). Population connectivity also plays an important role in the disease ecology of vectors (Gog et al., 2002). A vector is an organism that can transmit the causative agent of a disease, as would be white-tailed deer or elk in the case of CWD, with the prions being the disease-causing agent. Studies that evaluate the mechanisms by which disease vectors disperse through the environment, particularly in the context of contemporary **habitat connectivity**, also provide the **ecological context** necessary for a derivation of broad-scale public health considerations (Park, 2012) - or in the case of CWD, wildlife management.

Landscape genetic tools are often employed to characterize population dynamics, population connectivity (gene flow) and movement pathways of disease vectors across heterogeneous landscapes (Biek and Real, 2010), as well as to ascertain the concurrent spread and persistence of infectious agents associated with these movements (Blanchong et al., 2016). For example, gene flow estimates in white-tailed deer corresponded to the rapid expansion of chronic wasting disease (CWD) in northern Illinois/ southern Wisconsin and served to identify those habitats with an elevated risk for infection (Kelly et al., 2014).

Genetic Structure and Population Connectivity

Single Nucleotide Polymorphisms (SNPs) represent a new generation of genomic markers (Ekblom and Galindo, 2010) and have quickly emerged as the primary approach for **genetic studies into wildlife ecology and management** (Fig 5; Morin et al., 2004). Thousands of loci can be efficiently and cost-effectively screened across hundreds of individuals (Peterson et al., 2012; Puckett, 2017), providing sufficient resolution to track processes at ecological time scales – as in disease ecology (Blanchong et al., 2016) – and to parse out the individual components (e.g. local and landscape-level) driving genetic variation.

Our analysis of thousands of SNP loci successfully revealed eight genetic populations among the 1,226 white-tailed deer collected from 75 counties in Arkansas from 2016-2019 (Fig. 7). Genetic populations were spatially structured suggesting that both **broad-scale landscape-level and local-scale, behavioral processes are shaping genetic structure of whitetailed deer in Arkansas.**

At the **landscape scale**, population structure is primarily shaped by dispersal limitations and likely represents a cumulative effect of environmental heterogeneity (e.g. barriers to dispersal). At the **local scale**, social system may serve to naturally 'fragment' metapopulations into social groups (Mussmann et al., 2017), and substructure is often driven by relatedness, including white-tailed deer (Kelly et al., 2011). Hence, kinship in social mammals and its accompanying associations, can be important drivers of disease transmission.

Our analyses suggest both demographic and landscape-level processes are indeed impacting dispersal in white-tailed deer in Arkansas. Weakly differentiated genetic clusters (Fig. 7) seem to be generally geographically defined, and reflect associations with major landscape features, such as the Arkansas River (Figs. 10+11), and broad-scale habitat differences (Table 7, Fig. 25). Elevation is another important factor modulating dispersal of white-tailed deer, although a more comprehensive assessment across the entire state is warranted to statistically evaluate the role of these landscape aspects to dispersal of whitetailed deer in Arkansas; some areas of the state were represented by too few samples (Fig. 9) to conduct statistically robust analyses for spatial association of genetic patterns and landscape

features (Fig. 23, Appendix 7). We also found evidence for age- and sex- biases in dispersal rates (Figs. 8+13), and a strong signal of management interventions (stocking and translocation) on patterns of genetic relatedness (discussed below).

Effects of Historic Restocking and Translocations on Genetic Structure

Following many years of over-hunting and land-use impacts, the Arkansas deer population reached a low ebb of less than 500 individuals (Holder, 1951). Flooding of the Mississippi River in 1927 covered the eastern half of the state and contributed to the decline of deer in this area. Private hunters and AGFC wardens salvaged a small number of individuals, which were used to subsequently re-seed the Delta region. These were distributed in isolated refugia, with known hold-out populations on Caney and Muddy Creeks in the Ouachita Mountains, the Sylamore District of the Ozark Mountains, and the lower bottom lands near the confluence of the Arkansas and Mississippi Rivers. Other isolated groups were known in Pope County, with lesser numbers also found in Woodruff and Cross counties, and a handful of other small remnant populations.

As a mitigation strategy and to bolster population sizes, AGFC implemented a **refuge system**, with restocking of deer playing a major role. Between 1942-1951 a total of 2,343 deer were **translocated across 56 counties** (Donaldson et al., 1951). These supplementations used live-trapped native deer from areas with large deer herds and were intended to reduce crowding in source populations and to boost deer populations in release areas. The majority of these deer were sourced from three areas, with numerous smaller-scale (and poorly documented) translocations by managers and private hunters.

To this end, in 1941 AGFC established the **Howard County Deer Farm** (=720 acres; expanded to 1,305 in 1944) in the southwest corner of the state, within the existing Howard County Refuge (~33,000 acres) provisioned under federal aid to provide a convenient source of individuals for re-stocking (Holder, 1951). Source of the initial Refuge stocks are unknown. Although not all destination records for translocations sourced from the Howard County Farm

are available, it is documented that individuals from the 1945-1946 trapping season were released in Nevada and Little River counties (Holder, 1946), both also located in southwestern Arkansas. The complete extent of stocking from Howard County is unknown.

A second management action was trapping operations initiated in 1942, with the largest source of individuals being the **Sylamore District of the Ozark Mountains** (Hunter, 1952). From 1942-1946, over 900 individuals were trapped in the Sylamore District (Stone and Baxter Counties). Records here are a little more thorough, with known stockings of Sylamore individuals in Baxter, Calhoun, Cleburne, Cleveland, Cross, Hempstead, Howard, Jefferson, Lee, Lincoln, Little River, Logan, Nevada, Newton, Ouachita, Sebastian, Stone, Van Buren, Washington, and White Counties (Holder, 1946). Thus, documented translocations from the Sylamore District in the North-Central Arkansas Ozarks indicate deer have been moved to all corners of the state.

The third major source of individuals were purchased by the Commission from the **Sandhill Game Farm in Babcock,** Wisconsin. The Commission released 107 in 1942, 96 in 1943, and 101 in 1944. We are not aware of any records regarding the destinations and distribution of releases from these purchases. **Several small-scale releases** are also recorded (see Holder, 1951). Arkansas and **Louisiana** deer (unknown origin) were released in Ashley, Franklin, and Washington Counties between 1915-1926, with Ashley County later (~1931) also receiving six individuals from **North Carolina** (and two additional individuals going to Chicot County). An additional 28 deer were imported from the Pisgah National Forest in North Carolina, which were released in Bradley, Drew, Ouachita, and Saline Counties. In 1932-33, 15-18 deer from **Texas** were released in Dallas, Ouachita, and Saline Counties. Additional trapping was also done in the White River National Wildlife Refuge (Hunter, 1952).

Such past translocations often leave some traces of distinct (non-local) genetic diversity in the recipient populations and are likely detected with the high resolution afforded by the SNP approach used in this study, albeit which translocations were indeed successful is highly unpredictable (see Douglas et al., 1999, Douglas and Brunner, 2002). This likely explains some of the anomalous patterns of ancestry seen in our analyses (Figs. 7+11). For example,

common ancestry (=shared gene pools) found in Southwest and Northeast Arkansas (*k*=1 in Fig. 7) likely reflects translocations from the large refuges in the Sylamore DIstrict and from the Howard Game Farm. There is also a signature of Wisconsin ancestry (*k*=8 in Fig. 7), most broadly distributed in the southeastern Gulf Coast Plains region, with smaller pockets in Northeast and Northwest Arkansas. Although detailed records of the destination of the Wisconsin Sandhill Game Farm deer are not known, our approach suggests that stocking was likely heavily concentrated in these areas.

All three Arkansas gene pools that lack spatial cohesion can thus be anecdotally connected to the three major stock sources from the 1940's restoration efforts.

Subpopulation *k*=1 has the highest representation in the Howard County region, while *k*=3 is broadly distributed east of the Sylamore area, and *k*=8 likely reflects Wisconsin ancestry (Fig. 11). One genetic subpopulation (*k*=5) is broadly distributed in the Delta region and primarily found in the eastern half of the state north of the Arkansas River and likely reflects reintroduction of salvaged deer after the 1927 Mississippi River flood. The remaining four genetic population clusters (*k*=2, 4,6, and 7) are 'Ozark endemics'. Although many of the counties in this region were considered to have severely depleted (if not extirpated) populations at the 1930 ebb (Holder, 1951), it is possible that some of this structure reflects genetic variation pre-dating the population crash. At least one such persistent population is documented in Pope County, and which also did not receive any known releases from 1942-1951 (Holder, 1951). **Because severe population bottlenecks (such as this one) can promote large changes in allele frequencies, it can also inflate population structure** (e.g. Douglas et al., 2006). **This in itself may explain the exceptionally high frequency of** *PRNP* **Haplotype B in that region** (Fig. 19; discussed below).

Sex- and Age- Biased Dispersal in Arkansas

The high resolution of SNP data also allowed to examine **age- and sex-related dispersal patterns among white-tailed deer in Arkansas** (Fig. 13). Older males (>5 years) were notably

less genetically dissimilar than younger males, suggesting A) that five-year and above males are contributing substantially more to their local gene pools; and B) that they typically do not move long distances after establishing reproductive success. Local genetic distances peaked for juvenile and yearling males (ages 0 and 1), suggesting that male dispersal likely has already occurred by that age (Fig. 13A). Females, on the other hand, showed notably lower genetic distances as juveniles (Fig. 13B), with local (<10km) genetic distances peaking in the 1-2 year age class (Fig. 13C), suggesting that females move relatively further on average at that age. This is seemingly consistent with the hypothesis of density-dependence driving dispersal in adult females (Lutz et al., 2015), with natal dispersal as a result of mate competition or inbreeding avoidance being dominant in males (Perrin and Mazalov, 2000; Long et al., 2008). The former may explain why 1-2 year females show slightly higher distances at local scales (Fig. 13C), but not long-range scales (Fig. 13D-E).

PRNP Gene Variants

Polymorphism in the *PRNP* **gene detected across 1,433 white-tailed deer collected from the 75 counties in Arkansas is comparable to variation reported from other states** (Table 2-4). Out of the **20** *PRNP* **gene variants** (haplotypes) detected in this study, 16 were previously reported from other states (Blanchlong et al., 2009; Brandt et al. 2015, 2018; Johnson et al., 2006), and generally represent variants found at higher frequencies in these deer populations. **Four novel** *PRNP* **gene variants** were detected, that to date are unique to Arkansas. Two of these could potentially represent biologically relevant variation at the functional *PRNP* gene (CSV: Candidate Susceptibility Variants), because of the 96S mutation (Table 2) that has been associated with lower susceptibility to CWD (Brandt et al., 2018). However, all four haplotypes unique to Arkansas, as well as six others previously reported, were found at frequencies <1%, preventing statistically robust analyses for association with CWD susceptibility (Table 3+4, Fig. 15). While the **mechanism underlying susceptibility remains unknown**, the mounting evidence that some prion protein (*PrP*) variants are statistically associated with lower number of CWD+ animals suggests that *PrP* amino acid composition - determined by the underlying *PNRP* gene
nucleotide variation - might play a role in disease progression (Johnson et al., 2011, Brandt et al., 2018).

PRNP Selection and Time of Chronic Wasting Disease Exposure in Arkansas

Several aspects can be indicative of **time of exposure to a disease**: (1) **selection** for diseaseresistant phenotypes, (2) **prevalence rate** (with higher rates indicating longer time since initial exposure, and (3) **spatial distribution** of disease (initial disease occurrence usually being localized, and extended exposure being more wide-spread).

As in Robinson et al. (2012), we found a signature of selection on the 96S mutation (haplotype C). Given our estimated selection coefficient (*s*=0.1215), **we can yield some insight as to the amount of time the population has been exposed to chronic wasting disease**. In terms of years, in order to reach the level of haplotype C frequency discrepancy in Newton County (the largest CWD prevalence in our sampling) compared to the state-wide average, it would have taken a minimum of 9.78 generations, or 12.78 generations when computing frequencies using only individuals old enough to have reproduced (>2 years old), assuming an idealized population (Crow and Kimura, 1970). Assuming a 2-year minimum generation time (Demarais et al., 2000), which corresponds roughly to the age of sexual maturation for female white-tailed deer, this would correspond to 19.56, or 25.56 years, respectively. However, the real generation time is likely longer, and there are numerous caveats associated with this rudimentary calculation (e.g. the impacts of skewed sex ratios and overlapping generations is ignored). Despite this, **we can posit that the population within the focal CWD zone has likely been exposed to the disease for some time**. A better understanding of chronic wasting disease dynamics and epizootiology (factors controlling CWD presence) in white-tailed deer will likely provide a better estimate of time required to reach the observed prevalence in Arkansas.

Prevalence rate can also be indicative of how long a population has been exposed to an infectious disease. For example, Miller et al. (2000) simulated prevalence rates reaching 1% in a 15-20 year span, and 15% in 37 to 50 years, using plausible transmission rates. However, Miller

36

and Connor (2005) noted variation in prevalence growth rates among sex and age classes, and additionally noted that it is likely non-linear (e.g. growth rate depends on current prevalence). Numerous studies have also found spatial variation in prevalence rate, but association of disease occurrence with landscape features remains unclear.

Almberg et al. (2011) found that prevalence growth rates depend heavily on prion halflife in the environment. Taken together, the above suggest that prevalence growth rate estimates, and hence calendar time predictions for the epidemic, are heavily dependent on the exact model parameterizations. As such, we suggest caution in attempting to apply rate estimates from different contexts to Arkansas. With this in mind, apparent prevalence in the Arkansas CWD Management Zone peaks at 22-30% (C. Middaugh and A. J. Riggs, personal communication), thus it is plausible that CWD exposure in white-tailed deer in Arkansas represents a longer-standing epidemic. The spatial clustering of the disease distribution in Arkansas agrees with a simple spatial epidemic in which prevalence should spread spatially through time (as in Osnas et al., 2009). Given the above, we can posit that endemic CWD in Arkansas could date back potentially as far as the early recovery efforts, e.g. the establishment of Arkansas' refuge system in 1927, or the translocation and restocking efforts starting in 1942 (see below and Holder, 1951 for more information). **However, we once again emphasize that more information regarding population and disease dynamics would be required to make robust estimates.**

Partitioning Spatial Components of Population Structure

Spatial structuring of **biological parameters** (such as genetic population structure or dispersal) can result from extrinsic, **abiotic** determinants which are spatially structured (e.g. rainfall, elevation), or from intrinsic, **biotic** properties (e.g. growth, population size, differential mortality, and competition) (Dray et al., 2006). Our analyses partitioned variation in estimated migration rates (as a proxy for dispersal that cannot be directly measured) into four individual fractions of explained variance (Fig. D1): [a] **non-environmentally explained spatial variance**;

37

[b] **spatially structured environmental variance**; [c] **non-spatial environmental variance**; and [d] **unexplained variance** (Legendre, 1993). The relative sizes of these fractions of explained variance allow us to infer which mechanisms (e.g., extrinsic/abiotic *versus* intrinsic/biotic) are most likely responsible for the observed structure. Fractions [b+c] represent the total variation explained by the measured abiotic parameters, while fraction [a] represents spatial structure not influenced by environmental variation. Therefore, if [a] > [b+c] then **intrinsic factors, such as population size and differential mortality,** are most important; whereas, if [b+c] > [a] then **extrinsic factors, such as rainfall and elevation**, are most important for the spatial structure of the data. **Understanding these relationships is key for building models that predict the spread of CWD.**

Figure D1: Partitioned components of spatial variation in biological parameters, such as estimated migration rates (logM). [a]= non-environmentally explained spatial variance, [b] spatially structured environmental variance, [c] non-spatial environmental variance, and [d] unexplained variance.

The purpose of developing models of these parameters was to determine how variation in each (migration rate and population structure) is explained by spatial (i.e. autocorrelation) and environmental factors (e.g. elevation or climate.). These analyses provide several pieces of information including: 1) the **degree of spatial autocorrelation** in the genetic structuring of deer populations, 2) the degree of **environmental influence** in the genetic structuring of deer populations, 3) a list of **powerful explanatory factors**, and 4)

inferences about the role of **biotic processes**, internal to populations, versus **abiotic processes**, external to them, in driving overall patterns of population genetic structure. **Although our model explained 67% of the variance in estimated migration rate (logM) across Arkansas, only 17% was attributable to environment- with the remainder relating either to intrinsic biotic or unmeasured abiotic factors.**

Of those environmental variables which did explain variance in migration rate (logM), four were climatic (mean temperature during wettest quarter, mean temperature during driest quarter, precipitation of wettest quarter, and precipitation of wettest quarter). Climatic gradients have been shown to influence forage quality and composition relevant to white-tailed deer (Lesage et al., 2000), with subsequent impacts on deer habitat usage and selection (Shipley and Spalinger, 1995) that could culminate in variable population densities (Roseberry and Woolf, 1998). Deer have also shown inverse density-dependence on rates of population growth (Messier, 1991) and on dispersal (Lutz et al., 2015), thus implicating a potential role for habitat quality on net movement, and thereby gene flow. Six of the remaining variables were categorical assignments to National Vegetation Classification types (Short leaf pine; Gulf Coast mesic hardwood, Crowley's Ridge/ Sand forest, Open water, Large river floodplain forest), again implicating a role for habitat composition on gene flow. The remaining variables related to physical barriers (distance to highway or river), or location within either mountainous regions (Boston and Ouachita mountains) or large river valleys (Mississippi and Arkansas rivers), with the strongest impacts being attributable to the Arkansas River. This result echoes previous studies (e.g., Blanchong et al., 2006) that have found an influence of large rivers in generating landscape genetic patterns and **suggests that rates of CWD expansion to new areas will be slowed by large riverine barriers**.

However, a **large component of spatial variation in the migration rate (logM) was unexplained** (=47%; Fig. 25). We here **hypothesize that local demographic histories likely drive much of this variation**. Two components of the specific history of white-tailed deer in Arkansas could drive variance in migration rate (logM) estimates: 1) the **major population bottleneck**

39

peaking around 1930, followed by subsequent population expansion; and 2) **artificial 'gene flow' facilitated by translocation efforts** (both discussed in more detail above).

Population fluctuations (i.e., bottlenecks followed by rapid population growth) are a well-known mechanism of inflating population genetic structuring and can be commonly observed across deep time scales. For example, population structure in natural populations across much of North America echo genetic effects of Pleistocene ice ages, whereby populations contracted, as individuals retreated into relatively fragmented refugia (Hewitt, 1996, 2004). This isolation has several effects on genetic variation across the landscape, but perhaps foremost are 1) the loss of genetic variation *within* populations, due to the sharp decline in population size; and 2) the increase in genetic variation *among* populations, due to stochastic loss of alleles within each isolated refugium (Nei et al., 1975; Tajima, 1989). The result of population growth after the bottleneck and recolonization of new as ranges re-expand is that populations may come into contact again but are much more genetically divergent than before the demographic change. In terms of variants of an individual gene (alleles), many variants will have been lost, and those that remain will have changed drastically in relative frequency among populations (Glenn et al., 2001). The same 'contraction-expansion' cycle is likely reflected in the elevated population structure seen in some areas for white-tailed deer in Arkansas. This may suggest that some natural populations persevered through the early 1900s bottleneck.

The second process - facilitated gene flow (i.e., translocations) - is also likely causing anomalous patterns in the migration rate logM estimates. Because the procedure we used to calculate migration rates (i.e., EEMS) models gene flow as occurring step-wise among demes (Petkova et al., 2015), long-range translocations may conflate estimates.

Management Implications

Since the discovery of CWD in Arkansas, AGFC has made concerted efforts to proactively gauge its prevalence and adaptively manage its spread. In this sense, two components must be quantified: **Movement of the** disease **among populations** (e.g., its penetration across the landscape), and **its** prevalence within populations (e.g. disease growth). As such, hunting regulations and culling strategies should be two-pronged: **Reduce dispersal of deer from CWD areas** and **decrease population densities**. Below we outline conclusions derived from our genetic analyses and relate how these can be translated within the context of the above two factors. This, in turn, would allow the development of a much broader adaptive management strategy for CWD in white-tailed deer in Arkansas.

Estimated migration patterns indicate that **landscape-level features** (e.g. Arkansas River) have some **impacts on deer movements**. However abiotic factors represent a minority of the variance in migration rate.

2. *Some PRNP gene* **variants** (haplotypes) associate with intrinsic **susceptibility to CWD. One variant (**haplotype C) persists in older deer and thus shows a strong signature of natural selection. Its presence also impacts the probability of individuals persisting in CWD-prevalent regions.

Genetic Pattern Management Implication

While landscape features have an impact on deer movement, and **thus landscape-level CWD transmission**, they are **not hard boundaries** that will prevent spread of CWD. As such, they offer limited guidance with regards to CWD surveillance or selection of target areas for CWD management. Population dynamics (density) drive disease dynamics at the local scale (see below).

Selection on haplotype C suggests variation in the susceptibility of deer to CWD. Thus, analysis of **haplotype data** in deer **might support a density reduction strategy** wherein individuals with 'low genetic susceptibility' to CWD might be retained, whereas those with 'high genetic susceptibility' might be eliminated through intensive culling. However, the **efficacy of differing culling strategies** should also be modeled (Wasserberg et al.,, 2009; Wild et al., 2011). See Potapov et al. (2012) and Uehlinger et al. (2016) for utility of culling in cervids.

- 3. *PRNP* **haplotypes** are **structured spatially**, with state-wide variability in frequencies apparent. This in turn suggests state-wide variability in herdimmunity (Brandt et al. 2018).
- 4. White-tailed deer in Arkansas show a pattern of **age- and sex-biased dispersal**, with greater dispersal by juvenile and sub-adult males (Year 0-1). A slight signal indicates dispersal by young breedingage females (Year 1-2). However, juvenile females move very little, and the most apparent signal of successful mating is found in older (>= Year 5) males.

5. **Patterns of genetic relatedness across the state** show that genetic ancestry has been retained from historic stockings, with signals reflecting both within-state and out-of state. Again, these data demonstrate how powerful the genomic analyses are to detect genetic patterns in Arkansas deer.

The **rate of CWD spread at the landscape level can be modulated to some degree (but not eliminate) by controlling the frequency of particular haplotypes**. However, the **longterm consequences** of this approach are uncertain and **require additional work.** Agent-based modelling may provide a means to examine epidemiological and evolutionary host-disease dynamics.

Sex-biased culling can impact disease spread and prevalence in two ways: 1) Removing **high-tendency dispersers** can reduce disease spread between populations; and 2) Culling of r**eproductive individuals** can reduce rate of growth (or density) within populations. Young males have the highest probability of spreading the disease across space and time. Although females typically do not show natal dispersal, some studies (e.g. Lutz et al., 2015) suggest **density-dependent dispersal** due to competition for fawning habitat, or avoidance of males. This indicates overly high densities as drivers for female migration. Excessive numbers of females can promote large population densities (Holder 1951), thus **female culling should also be considered**.

Efforts should be harnessed among states to **define translocation histories** of deer so as to track the impacts of historic management actions with regards to the current 'mosaic' distribution of CWD nationwide. Exact translocation sources and destinations would be extremely useful with regards to CWD surveillance.

VI. Acknowledgements

We thank Cory Gray, Chris Middaugh, Jenn Ballard and A.J. Riggs from the Arkansas Game and Fish Commission for encouraging and coordinating this study. Their guidance during all phases of the project and insights into the biology of white-tailed deer and emergence of Chronic Wasting Disease in Arkansas are invaluable. Insights about deer biology and management were also provided by Don White from the University of Arkansas-Monticello. Samples of whitetailed deer ear tissues analyzed for this study were kindly collected and provided by staff of the Arkansas Game and Fish Commission. Major contributions to laboratory work and data analyses were made by members of the Arkansas Conservation and Molecular Ecology Laboratory (aCaMEL) at the University of Arkansas – Fayetteville. This project was funded by award RG 2273 and AR-W-F18AF0002 from the Arkansas Game and Fish Commission and contributions from generous endowments by the University of Arkansas to Michael E Douglas (21st Century Chair in Global Change Biology) and Marlis R Douglas (Bruker Professorship in Life Sciences).

VII. References Cited

- **Alexander, D. H., and J. Novembre.** 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19:1655–1664.
- **Almberg, E. S., P. C. Cross, Johnson, C. J., Heisey, D. M., and B. J. Richards.** 2011. Modeling routes of chronic wasting disease transmission: Environmental prion persistence promotes deer population decline and extinction. *PLoS ONE* 6(5):e19896.
- **Amaral, K. E., K. E. M. Palace, K. M. O'Brien, L. E. Fenderson, and A. I. Kovach.** 2016. Anthropogenic habitats facilitate dispersal of an early successional obligate: Implications for restoration of an endangered ecosystem*. PLoS One* 11(3):e0148842.
- **Andrewartha, H. G., and L. C. Birch** 1954. *The distribution and abundance of animals*. University of Chicago Press, Illinois.
- **Arnot C., E. Laate, J. Unterschultz, and W. Adamowicz.** 2009. Chronic wasting disease (CWD): potential economic impact on cervid farming in Alberta. *Journal of Toxicology and Environmental Health* 72:1014–1017.
- **Barton, N. H., and M. R. Servedio.** 2015. The interpretation of selection coefficients. *Evolution* 69(5):1101–1112.
- **Biek, B., and L. A. Real.** 2010. The landscape genetics of infectious disease emergence and spread. *Molecular Ecology* 19:3515–3531.
- **Blanchong, J. A., D. M. Heisey, K. T. Scribner, S. V. Libants, C. Johnson, J. M. Aiken, J. A. Langenberg, and M. D. Samuel**. 2009. Genetic susceptibility to chronic wasting disease in free-ranging whitetailed deer: Complement component C1q and Prnp polymorphisms. *Infection, Genetics and Evolution* 9:1329–1335. doi: 10.1016/j.meegid.2009.08.010
- **Blanchong, J. A., S. J. Robinson, M. D. Samuel, and J. T. Foster**. 2016. Application of genetics and genomics to wildlife epidemiology. *Journal of Wildlife Management* 80:593–608.
- **Borcard, D. and P. Legendre**. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbor matrices. *Ecological Modeling* 153:51–68.
- **Brandt, A. L., A. C. Kelly, M. L. Green, P. Shelton, J. Novakofski, and N. E. Mateus-Pinilla**. 2015. Prion protein gene sequence and chronic wasting disease susceptibility in white-tailed deer (*Odocoileus virginianus*). *Prion* 9:449–462.
- **Brandt, A. L., M. L. Green, Y. Ishida, A. L. Roca, J. Novakofski, and N. E. Mateus-Pinilla**. 2018. Influence of the geographic distribution of prion protein gene sequence variation on patterns of chronic wasting disease spread in white-tailed deer (*Odocoileus virginianus*). *Prion* 12:204–215.
- **Chafin, T. K., B. T. Martin, S. M. Mussmann, M. R. Douglas, and M. E. Douglas**. 2018. FRAGMATIC: *in silico* locus prediction and its utility in optimizing ddRADseq projects. *Conservation Genetics Resources* 10:325–328.
- **Chen, I., J. K. Hill, R. Ohlemueller, D. B. Roy, and C. D. Thomas.** 2011. Rapid range shifts of species associated with high levels of climate warming. *Science* 333:1024-1026.
- **Corlett, R. T.** 2015. The anthropocene concept in ecology and conservation. *Trends in Ecology & Evolution* 30:36–41.
- **Crooks, K. R., and M. Sanjayan (eds)** 2006. *Connectivity conservation*. Cambridge University Press, Cambridge, UK.
- **Crow, J. F., and M. Kimura.** 1970. *An introduction to population genetics theory*. Harper and Row, New York.
- **DaCosta, J. M., and M. D. Sorenson**. 2016. ddRAD-seq phylogenetics based on nucleotide, indel, and presence–absence polymorphisms: Analyses of two avian genera with contrasting histories. *Molecular Phylogenetics and Evolution* 94:122–135.
- **Demarais, S., K. V. Miller, and H. A. Jacobson**. 2000. White-tailed deer. pp. 601-628. *In*: *Ecology and management of large mammals in North America* (S. Demarais and P. R. Krausman, eds.). Prentice Hall, Upper Saddle River, New Jersey.
- **Dobson, F. S.** 1982. Competition for mates and predominant juvenile male dispersal in mammals. *Animal Behaviour* 30(4): 1183–1192.
- **Donaldson, D., C. Hunter, and T. H. Holder.** 1951. *Arkansas' deer herd. Federal Aid Publication Projects 17-D and 20-R*. Arkansas Game and Fish Commission, Little Rock, AR. pp 72.
- **Dorman, C. F., J. Elith, S. Bacher, C. Buchman, G. Carl, G. Carre, and S. Lautenbach.** 2012. Collinearity: A review of methods to deal with it in a simulation study evaluation their performance. *Ecography* 36:27–46.
- **Douglas, M. E., M. R. Douglas, G. W. Schuett, and L. W. Porras**. 2006. Evolution of rattlesnakes (Viperidae; *Crotalus*) in the warm deserts of western North America shaped by Neogene vicariance and Quaternary climate change. *Molecular Ecology* 15:3353–3374.
- **Douglas, M.R., P. C. Brunner, and L. Bernatchez.** 1999. Do assemblages of *Coregonus* (Teleostei, Salmoniformes) in the Central Alpine region of Europe represent species flocks? *Molecular Ecology 8*:589–603.
- **Douglas, M. R., and P. C. Brunner.** 2002. Biodiversity of Central Alpine *Coregonus* (Salmoniformes): Impact of one-hundred years management. *Ecological Applications 12(1):*154–172.
- **Douglas, M. R., M. E. Douglas, D. White, and T. K. Chafin**. 2018. Family group assignment and genotyping of white-tailed deer in Arkansas using Single Nucleotide Polymorphism (SNP) DNA markers extracted from ear tissue samples. *Final Report submitted to the Arkansas Game and Fish Commission,* 30-June-2018, pp. 45.
- **Dray, S., P. Legendre, and P. R. Peres-Neto.** 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbor matrices (PCNM). *Ecological Modelling* 196:483–493.
- **Eaton, D. A. R.** 2014. PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* 30:1844–1849.
- **Edgar, R. C.** 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.
- **Ekblom, R., and J. Galindo.** 2010. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* 107:1–15.
- **Funk, W. C., J. K. McKay, P. A. Hohenlohe, and F. W. Allendorf**. 2012. Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution* 27:489–496.
- **Glenn, T. C., W. Stephan, and M. J. Braun.** 2001. Effects of a population bottleneck on Whooping Crane mitochondrial DNA variation. *Conservation Biology* 13(5):1097–1107.
- **Gog, J., R. Woodroffe, and J. Swinton.** 2002. Disease in endangered metapopulations: The importance of alternative hosts. *Proceedings of the Royal Society of London, Series B* 269:671–676.
- **Gough, K. C., and B. C. Maddison.** 2010. Prion transmission prion excretion and occurrence in the environment. *Prion* 4:275–282.
- **Grear, D. A., M. D. Samuel, K. T. Scribner, B. V. Weckworth, and J. A. Langenberg.** 2010. Influence of genetic relatedness and spatial proximity on chronic wasting disease infection among female white-tailed deer. *Journal of Applied Ecology* 47:532–540.
- **Gribov, A., and K. Krivoruchko**. 2012. New Flexible Non-parametric data transformation for trans-Gaussian Kriging. *Geostatistics Oslo* 2012:51–65.
- **Hewitt, G. M.** 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 58:247–76.
- **Hewitt, G. M.** 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London, Series B* 359:183–195.
- **Holder, T. H.** 1946. Final report of state-wide game restoration project. *Report to Arkansas Game and Fish Commission*. 29p.
- **Holder, T. H.** 1951. A survey of Arkansas game. *Arkansas Game and Fish Commission Federal Aid Publication Project II-R*. 57-79.
- **Hunter, C.** 1952. Trapping and restocking white-tailed deer in Arkansas (Federal Aid Project 17-D). *Paper presented at the Southeastern Wildlife Society Meeting, Savannah, Georgia*.
- **Johnson, C., J. Johnson, M. Clayton, D. McKenzie, and J. Aiken**. 2003. Prion protein gene heterogeneity in free-ranging white-tailed deer within the chronic wasting disease affected region of Wisconsin. *Journal of Wildlife Diseases* 39:576–581.
- **Johnson, C., J. Johnson, J. P. Vanderloo, D. Keane, J. M. Aiken, and D. McKenzie.** 2006. Prion protein polymorphisms in white-tailed deer influence susceptibility to chronic wasting disease. *Journal of General Virololgy* 87:2109–2114.
- **Johnson, C. J., A. Herbst, C. Duque-Velasquez, J. P. Vanderloo, P. Bochsler, R. Chappell, and D. McKenzie**. 2011. Prion protein polymorphisms affect chronic wasting disease progression. *PLoS ONE* 6 doi: 10.1371/journal.pone.0017450
- **Joly D.O., C. A. Ribic, J. A. Langenberg, K. Beheler, C. A. Batha, B. J. Dhuey, R. E. Rolley, G. Bartelt, T. R. Van Deelen, and M. D. Samuel.** 2003. Chronic wasting disease in free-ranging Wisconsin whitetailed deer. *Emerging Infectious Diseases* 9:599–601.
- **Kelly, A. C., N. E. Mateus-Pinilla, J. Diffendorfer, E. Jewell, M. O. Ruiz, J. Killefer, P. Shelton, T. Beissel, and J. Novakofski**. 2008. Prion sequence polymorphisms and chronic wasting disease resistance in Illinois white-tailed deer (*Odocoileus virginianus*). *Prion* 2:28–36.
- **Kelly, A. C., N. E. Mateus-Pinilla, M. R. Douglas, M. E. Douglas, W. Brown M. O. Ruiz, J. Killefer, P. Shelton, T. Beissel, and J. Novakofski.** 2010. Utilizing disease surveillance to examine gene flow and dispersal in white-tailed deer. *Journal of Applied Ecology* 47:1189–1198.
- **Kelly, A. C., N. E. Mateus-Pinilla, M. R. Douglas, M. E. Douglas, and J. Novakofski.** 2011. Microsatellites behaving badly: empirical evaluation of genotyping errors and subsequent impacts on population studies. *Genetics and Molecular Research* 10 (4):2534–2553.
- **Kelly, A. C., N. E. Mateus-Pinilla, W. Brown, M. O. Ruiz, M. R. Douglas, M. E. Douglas, P. Shelton, T. Beissel, and J. Novakofski**. 2014. Genetic assessment of environmental features that influence deer dispersal: Implications for prion-infected populations. *Population Ecology* 56:327–340.
- **Kopelman, N. M., J. Mayzel, M. Jakobsson, N. A. Rosenberg, and I. Mayrose**. 2015. Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* 15:1179–1191.
- **Legendre, P., D. Borcard, and P. R. Peres-Neto.** 2005. Analyzing beta diversity: Partitioning the spatial variation of community composition data. *Ecological Monographs* 75:435–450.
- **Legendre, P.** 1993. Spatial Autocorrelation: Trouble or New Paradigm? *Ecology* 74(6):1659–1673.
- **Leigh, J. W., B. David, and N. Shinichi**. 2015. popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* 6:1110–1116.
- **Lesage, L., M. Crete, J. Huot, and J. Ouellet**. 2000. Quality of plant species utilized by northern whitetailed deer in summer along a climatic gradient. *Ecoscience* 4:439–451.
- **Li, H., J. Ruan, and R. Durbin**. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research* 18:1851–1858.
- **Long, E. S., D. R. Diefenbach, C. S. Rosenberry, and B. D. Wallingford.** 2008. Multiple proximate and ultimate causes of natal dispersal in white-tailed deer. *Behavioral Ecology* 19(6):1235–1242.
- **Lutz, C. L., D. R. Diefenbach, and C. S. Rosenberry.** 2015. Population density influences dispersal in female white-tailed deer. *Journal of Mammalogy*. 96(3):494–501.
- **Manjerovic, M. B., M. L. Green, N. Mateus-Pinilla, and J. Novakofski.** 2014. The importance of localized culling in stabilizing chronic wasting disease prevalence in white-tailed deer populations. *Preventative Veterinary Medicine* 113(1):139–145.
- **Mathieson, I., and G. McVean**. 2013. Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. *Genetics* 193:973–984.
- **Messier, F.** 1991. The significance of limiting and regulating factors on the demography of moose and white-tailed deer. *Journal of Animal Ecology* 60:377–393.
- **Miller, M. W., E. S. Williams, C. W. McCarty, T. R. Spraker, T. J. Kreeger, C. T. Larsen, and E. T. Thorne.** 2000. Epizootiology of chronic wasting disease in free-ranging cervids in Colorado and Wyoming. *Journal of Wildlife Diseases* 36(4):676–690.
- **Miller, M. W., and M. M. Connor.** 2005. Epidemiology of chronic wasting disease in free-ranging mule deer: spatial, temporal, and demographic influences on observed prevalence patterns. *Journal of Wildlife Diseases* 41(2):275–290.
- **Miller, M. W., and E. S. Williams.** 2003. Prion disease: horizontal prion transmission in mule deer. *Nature* 425:35–36.
- **Morin, P. A., G. Luikart, R. K. Wayne, and the SNP Workshop Group***.* 2004. SNPs in ecology, evolution, and conservation. *Trends in Ecology & Evolution* 19:208–216.
- **Murakami, D. and D. A. Griffith**. 2019 Eigenvector spatial filtering for large data sets: fixed and random effects approaches. *Geographical Analysis* 51(1):23–49.
- **Mussmann, S. M., M. R. Douglas, W. J. B. Anthonysamy, M. A. Davis, S. A. Simpson, W. Louis, and M. E. Douglas.** 2017. Genetic rescue, the greater prairie chicken and the problem of conservation reliance in the Anthropocene. *Royal Society Open Science* 4:160736.
- **Nei, M., T. Maruyama, and R. Chakraborty.** 1975. The bottleneck effect and genetic variability in populations. *Evolution* 29(1):1–10.
- **Oksanen, J.** 2015. Multivariate analysis of ecological communities in R: vegan tutorial. University of Oulu, Finland. Available at: http://cc.oulu.fi/~jarioksa/opetus/metodi/vegantutor/
- **O'Rourke, K. I., T. R. Spraker, L. K. Hamburg, T. E. Besser, K. A. Brayton, and D. P. Knowles**. 2004. Polymorphisms in the prion precursor functional gene but not the pseudogene are associated with susceptibility to chronic wasting disease in white-tailed deer. *Journal of General Virology* 85:1339–1346.
- **Osnas, E. E., D. M. Heisey, R. E. Rolley, and M. D. Samuel.** 2009. Spatial and temporal patterns of chronic wasting disease: fine-scale mapping of a wildlife epidemic in Wisconsin. *Ecological Applications* 19(5):1311–1322.
- **Park, A. W.** 2012. Infectious disease in animal metapopulations: The importance of environmental transmission. *Ecology and Evolution* 2:1398–1407.
- **Perrin, N., and V. Mazalov.** 2000. Local competition, inbreeding, and the evolution of sex-biased dispersal. *American Naturalist* 155:116–127.
- **Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra**. 2012. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PloS One* 7:e37135.
- **Petkova, D., J. Novembre, and M. Stephens**. 2015. Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics* 48:94–100.
- **Potapov A., E. Merrill, and M. A. Lewis** 2012. Wildlife disease elimination and density dependence. *Proceedings of the Royal Society, Series B* 279(1741):3139–3145.
- **Prusiner, S. B.** 1998. Prions. *Proceedings of the National Academy of Sciences* 95:13363–13383.
- **Pucket, E. E.** 2017. Variability in total project and per sample genotyping costs under varying study designs including with microsatellites or SNPs to answer conservation genetic questions. *Conservation Genetics Resources* 9:289–304.
- **Robinson S. J., M. D. Samuel, C. J. Johnson, M. Adams, and D. I. McKenzie.** 2012. Emerging prion disease drives host selection in a wildlife population. *Ecological Applications* 22(3):1050–59.
- **Rognes, T., T. Flouri, B. Nichols, C. Quince, and F. Mahé**. 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2409v1.
- **Roseberry, J. L., and A. Woolf.** 1998. Habitat-populations density relationships for white-tailed deer in Illinois. *Wildlife Society Bulletin*. 26(2):252–258.
- **Rosenberg, N. A.** 2004. DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Resources* 4:137–138.
- **Russell, T., C. Cullingham, A. Kommadath, P. Stothard, A. Herbst, and D. Coltman.** 2019. Development of a novel mule deer genomic assembly and species-diagnostic SNP panel for assessing introgression in mule deer, white-tailed deer, and their interspecific hybrids. *G3 Genes|Genomes|Genetics*, 9: g3.200838.2018.
- **Shipley, L. A., and D. E. Spalinger.** 1995. Influence of size and density of browse patches on intake rates and foraging decisions of young moose and white-tailed deer. *Oecologia* 104(1):112–121.
- **Senn, H., R. Ogden, T. Cezard, K. Gharbi, Z. Iqbal, E. Johnson, N. Kamps-Hughes, F. Tosell, and R. McEwing.** 2013. Reference-free SNP discovery for the Eurasian beaver from restriction siteassociated DNA paired-end data. *Molecular Ecology* 22:3141–3150.
- **Stephens, M., N. J. Smith, and P. Donnelly**. 2001. A new statistical method for haplotype reconstruction from population data. *The American Journal of Human Genetics* 68:978–989.
- **Tajima, F.** 1989. The effect of change in population size on DNA polymorphism. *Genetics* 123(3):597– 601.
- **Tokarska, M., T. Marshall, R. Kowalczyk, R. Kowalczyk, J. M. Wojcik, C. Pertoldi, T. N. Kristensen, V. Loeschcke, V. R. Gregersen, and C. Bendixen.** 2009. Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: the case of European bison. *Heredity* 103:326–332.
- **Tracey, J. A., S. N. Bevins, S. Vanderwoude, and K. R. Crooks.** 2014. An agent-based movement model to assess the impact of landscape fragmentation on disease transmission. *Ecosphere* 5:119.
- **Uehlinger, F. D., A. C. Johnson, T. K. Bollinger, and C. L. Waldner.** 2016. Systematic review of management strategies to control chronic wasting disease in wild deer populations in North America. *BMC Veterinary Research* 12:173
- **Vandam, R., E. Kaptijn, and B. Vanschoenwinkel.** 2013. Disentangling the spatio-environmental drivers of human settlement: an eigenvector based variation eecomposition. *PLoS ONE* 8(7):e67726.
- **Wasserberg G., E. E. Osnas, R. E. Rolley, and M. D. Samuel** 2009. Host culling as an adaptive management tool for chronic wasting disease in white-tailed deer: a modeling study**.** *Journal of Applied Ecology* 46(2):457–466.
- **Wild, M. A., N. T. Hobbs, M. S. Graham, and M. W. Miller** 2011. The role of predation in disease control: a comparison of selective and nonselective removal on prion disease dynamics in deer. *Journal of Wildlife Diseases* 47(1):78–93.
- **Williams E. S., and S. Young**. 1980. Chronic wasting disease of captive mule deer spongiform encephalopathy. *Journal of Wildlife Diseases* 16:89–98.
- **Williams E. S., M. W. Miller, T. J. Kreeger, R. H. Kahn, and E. T. Thorne.** 2002. Chronic wasting disease of deer and elk: a review with recommendations for management. *Journal of Wildlife Management* 66:551–563.
- **Wilson G. A., S. M. Nakada, T. K. Bollinger, M. J. Pybus, E. H. Merrill, and D. W. Coltman.** 2009. Polymorphisms at the *PRNP* gene influence susceptibility to chronic wasting disease in two species of deer (*Odocoileus* spp.) in western Canada. *Journal of Toxicology and Environmental Health, Part A* 72:1025–1029.
- **Wright, S.** 1943. Isolation by distance. *Genetics* 28:114–138.

VIII. Tables

Table 1: Numbers of white-tailed deer genotyped across ~35,000 SNP loci listed. Samples were collected 75 counties in Arkansas from 2016-2019. Code indicates standard 2-letter abbreviation for each county.

Table 2: Overview of **20** *PRNP* **gene haplotypes** detected across 1,433 white-tailed deer collected from 75 counties in Arkansas (2016-2019). Data are based 720 nucleotides and phased haplotypes were derived from genotype data using PHASE2 (only variable nucleotides shown). Haplotype names indicated by letter (A-V) are as in Brandt et al. (2015); haplotypes denoted AR_# are unique to Arkansas and were not previously reported. Mutations differing from haplotype A are shaded, with green indicating synonymous (no amino acid change) and blue non-synonymous substitutions (NSS; amino acid change in protein). Also listed are amino acid position and nucleotide site as referenced in Brandt et al. (2015).

Table 3: *PRNP* **haplotype frequencies and odds ratio for association with CWD status across Arkansas** (entire state). Haplotypes were derived from unphased sequences of 720 nucleotides of the *PRNP* gene. Haplotype indicated by letters were also reported by Brandt et al. (2015, 2018), whereas AR_# indicates a haplotype unique to Arkansas. Listed are total numbers (N) and relative frequency f(%), and values for deer that either tested CWD-negative (-), CWD-positive (+) or were not tested (?). Odds Ratio (OR) reflects relative representation of a haplotype in CWD+ deer, with OR>1 indicating over-representation and OR<1 under-representation; SE= standard error, CI= 95% confidence interval, Z(OR)= Z-score and p(OR)= probability. **Values in bold are significant**.

Total 2,866 2,420 248 198

Table 4: *PRNP* **haplotype frequencies and odds ratio for association with CWD status for Newton County** only. Haplotypes were derived from unphased sequences of 720 nucleotides of the *PRNP* gene. Haplotype indicated by letters were also reported by Brandt et al. (2015, 2018), whereas AR_# indicates a haplotype unique to Arkansas. Listed are total numbers (N) and relative frequency f(%), and values for deer that either tested CWD-negative (-), CWD-positive (+) or were not tested (?). Odds Ratio (OR) reflects relative representation of a haplotype in CWD+ deer, with OR>1 indicating over-representation and OR<1 underrepresentation; SE= standard error, CI= 95% confidence interval, Z(OR)= Z-score and p(OR)= probability. **Values in bold are significant**.

Table 5: *PRNP* **haplotype frequencies and odds ratio for six age classes** of white-tailed deer collected in Newton County from 2016-2019. Relative frequencies and odds ratio are shown for (A) **haplotype C** (associated with reduced CWD susceptibility) and (B) **haplotype B** (associated with increased CWD susceptibility. Relative frequency of CWD-negative [f(-)] and CWD-positive [f(+)] samples were calculated by dividing numbers CWD-negatives w/C or B, respectively, divided by negatives w/ other haplotypes.

(A) Haplotype C

(B) Haplotype B

Table 6: Descriptions of **35 candidate environmental and climatic predictor variables** used in the spatial analysis.

Table 7. The **17 extrinsic variables that passed through Variation Inflation Factor (VIF) analysis**to remove multiple-collinearity, and data reduction procedure. Only those variables with a significant relationship with migration rate (logM) were retained for further analyses. There are two types of variables presented below: continuous and categorical. For continuous variables, linear regression (LR) against migration rate (logM) was performed and the standard results are provided. For categorical variables student's t-test was performed to test for difference in mean migration rate (DMR) between the presence (1) and absence (0) of the given variable.

IX. Figures

4/23/2010 Time: 3:01:56 PM

Figure 1. Recorded cases of 598 white-tailed deer (circles) and 19 elk (diamonds) that tested positive for Chronic Wasting Disease (CWD+) in Arkansas as of 24-April-2019. The area enclosed in dark green represents the 16 counties included the 2018 CWD Management Zone. Based on new CWD+ cases in Searcy and Scott counties during the 2018/2019 hunting season, the MZ was explanded to include Scott, Baxter and Stone counties. (source: https://www.AGFC.com)

Figure 2. Genetic population structure conceptual graph - colors indicate genetic diversity, with similar colors reflecting genetic similarity. A deer herd comprises of closely related individuals that are genetically similar. This is in part due to female fawns remaining with the mother's herd. Over time, a pattern of genetic similarity and differences across the landscape will emerge that is referred to as genetic population structure. Geographically distant herds will be genetically more distinct (grey *versus* blue deer) than herds in geographic proximity (red and brown deer).

Figure 3. Genetic connectivity conceptual graph – colors indicate genetic diversity, with different colors reflecting genetic differences; arrows indicate dispersal of a deer from one herd to another. In contrast to female fawns, male fawns disperse from the mother's herd and eventually join other herds that are genetically distinct. Such dispersed individuals will show slight genetic differences from local animals. By assessing genetic diversity within and among herds, the rate at which such dispersal occurs can be estimated. This is referred to as population connectivity. Dispersal of a deer is influenced by landscape features, such as rivers and roads being potential barriers; dispersal is more likely to occur over areas that consist of habitat suitable for deer.

Figure 4. Genetic admixture conceptual graph – colors indicate genetic diversity, with homogenous colors reflecting ancestry in a single genetic group (i.e., gene pool), and bi-colored deer reflecting admixed ancestry. If dispersed and local animals produce offspring, the progeny will show genetic signals (ancestery) of both parental gene pools (i.e., admixed ancestry). By estimating the extent of admixture, population connectivity and migration rates can be inferred.

Figure 5. Genomic approach to study genetic structure in white-tailed deer - applicability of RADseq (Restriction-site Associated DNA sequencing). Genetic differences accumulate over hundreds or thousands of generations; large differences reflect evolutionary processes that occurred over long temporal and broad spatial scales (e.g., speciation), whereas small differences reflect more recent, geographically localized processes (e.g., population structure). Genetic markers vary in their capacity to resolve these patterns, with gene sequences generally informative about systematic relationships (Systematics) and fragment analysis (i.e., microsatellite loci) used to infer population structure. RADseq data encapsulate genetic variation at across broad spatial and temporal scales (large and small genetic differences) and thus are suitable to address a variety of research questions, including resolution of taxonomic relationships (Systematics/Phylogeny), geographic patterns of evolutionary lineages (phylogeography), and detailed genetic population structure.

Figure 6: Sampling locations for 1,659 white-tailed deer across Arkansas used in this study. The red shaded area indicates the 16 counties included in the 2018 Chronic Wasting Disease (CWD) Management Zone (MZ), and the points represent collection localities for the individual samples. For details on sample size per county see Table 1. Detailed collection information for each sample provided in Supplements – Table S1.

Figure 8: Population ancestry assignments by sex for female (N=414) and male (N=604) white-tailed deer collected in Arkansas from 2016-2019. Assignment was inferred using the program ADMIXTURE and is based on 33,225 loci (females) and 33,886 loci (males), respectively (Note: number of loci are dependent on samples included in a particular analysis). One variable site (SNP= Single Nucleotide Polymorphism) was retained per locus. Colors represent four genetic subpopulations (K=4) as identified in this analysis. Each individual sample is represented as a pie chart, with colors proportional to the probability of ancestry assigned to one of the four subpopulations (*k*1 through *k*4).

100

25 50

0

200

Miles

100

150

50

25

150

200

Miles

Figure 9: Sample density (=individuals per unit area) for 1,143 white-tailed deer included in genetic population structure analyses. Deer were collected by AGFC in Arkansas in 2016-2019 and individual samples are shown as black dots. Densities are reported as averaged values within a tessellated grid of 10 km² hexagonal tiles, with red reflecting low densities, and green indicating high density.

Figure 10: Oversimplified distribution of eight genetic subpopulations identified across 1,143 whitetailed deer collected in Arkansas from 2016-2019. Loci and sample sizes are identical to Figure 7, but data are presented as dominant ADMIXTURE assignment for each hexagonal tile (N=13,378), defined as the genetic population *k* with the largest mean ancestry assignment within each tile.

Assignment

Figure 11: Probabilistic distribution of eight genetic subpopulations identified in 1,143 white-tailed deer collected in Arkansas from 2016-2019. For each subpopulation (*k*=1 through *k*=8) ADMIXTURE assignment probability was interpolated across the landscape (partitioned into 13,378 hexagonal tiles), where P(*k*)=1 corresponds to a 100% probability of a tile having ancestry of subpopulation *k*, and P(*k*)=0 corresponds to a 0% probability of subpopulation *k* ancestry. Loci and sample sizes are identical to Figure 7, but ancestry probabilities were interpolated from point data using Empirical Bayesian Kriging.

Figure 12: Visualization of population connectivity among white-tailed deer collected in Arkansas from 2016-2019. Population connectivity was estimated as effective migration rates (left) and intra-population diversity (log₁₀ scale; right), calculated using the program EEMS. Analysis was based on 1,143 samples (black dots) evaluated for genetic diversity across 34,214 loci, with one variable site (SNP= Single Nucleotide Polymorphism) retained per locus. Rates were estimated across 13,378 hexagonal tiles and are plotted by colored bin, with bin divisions calculated as natural breaks using the Jenks algorithm in ARCMAP.

Figure 13: Spatial patterns of genetic dissimilarity among deer partitioned by age and sex. Genetic distances between individuals and their neighbors was calculated from 5,000 randomly sampled variable sites (SNP= Single Nucleotide Polymorphism), using different physical distances (x-axis). Results are shown for age classes of males (A) and females (B), with histograms shown for distance slices at 10km (C), 20km (D), and 40km (E).

Figure 14: Haplotype network showing relationship of *PRNP* **gene variants** detected across 1,433 whitetailed deer collected from 75 counties in Arkansas (2016-2019). Data are based on sequence analysis of 720 nucleotides. Circles represent 20 haplotypes with size reflecting frequency of occurrence in entire data set (Table 3) and tick marks number of mutations (nucleotide substitutions) distinguishing one from another (Table 2). Color codes reflect relative frequency among CWD-positive (red) and CWD-negative (green) animals. Letters correspond to haplotype names in Brandt et al. (2015), with haplotypes unique to Arkansas indicated with numbers (AR_#).

Figure 15: **Frequency distribution of 2,866** *PRNP* **haplotypes** detected in white-tailed deer collected in Arkansas 2016-2019. Haplotypes were determined by phasing individual genotypes derived from sequencing 1,433 deer across 720 bp of the *PRNP* gene. Letters (A through V) refer to haplotypes identified by Brandt et al. (2015), whereas numbers (1-4) are haplotypes unique to Arkansas, not previously detected in other states. Frequencies are plotted for all 1,433 samples (=statewide) and a subset of 314 samples from Newton County (N=628 chromosomes). Color codes reflect frequency among CWD-positive (CWD+) and CWD-negative (CWD-) samples; unknown indicates sample was not tested for CWD.

Figure 16: Spatial distribution of 1,433 white-tailed deer samples included in the *PRNP* gene polymorphism analyses. Green dots represent samples that tested negative for CWD (CWD-), whereas red dots represent CWD-positive (CWD+) samples collected in Arkansas 2016-2019. NOTE: Proportion of CWD- *vs* CWD+ do not necessarily reflect absolute CWD prevalence.

Figure 17: Spatial distribution of 211 non-overlapping polygons, each encapsulating 5-10 sampling locations, derived from 1,433 white-tailed deer samples (shown as green that were used to compute *PRNP* haplotype frequencies for interpolation.

Figure 18: **Spatial distribution of relative frequency of seven common** *PRNP* **gene haplotypes**. Data are based on 720 nucleotides sequenced across 1,433 white-tailed deer collected in Arkansas from 2016- 2019. Frequencies were first calculated within arbitrarily defined non-overlapping groupings of 5-10 samples (Figure 17), then interpolated using empirical Bayesian kriging. Frequency is depicted by color, with blue reflecting low occurrence (0-5%) whereas red indicating 46-50% of haplotypes were of this type.

Figure 19: **Spatial distribution of** *PRNP* **haplotype B**, presented as relative frequency. Data are based on 720 nucleotides sequenced across 1,433 white-tailed deer collected in Arkansas from 2016-2019. Frequencies were first calculated within arbitrarily defined non-overlapping groupings of 5-10 samples (Figure 17), then interpolated using empirical Bayesian kriging. Frequency is depicted by color, with blue reflecting low occurrence (0-5%) whereas red indicating 46-50% of haplotypes were of this type.

Figure 20: Spatial distribution of *PRNP* haplotype C, presented as relative frequency. Data are based on 720 nucleotides sequenced across 1,433 white-tailed deer collected in Arkansas from 2016-2019. Frequencies were first calculated within arbitrarily defined non-overlapping groupings of 5-10 samples (Figure 17), then interpolated using empirical Bayesian kriging. Frequency is depicted by color, with blue reflecting low occurrence (0-5%) whereas red indicating 46-50% of haplotypes were of this type.

Figure 21: Relative frequency and odds ratio for two candidate susceptibility variants (CSV) of *PRNP* gene haplotypes detected in white-tailed deer age cohorts (<1 year to 5+ years) sampled in Arkansas from 2016-2019. Haplotype_C (top panel) has been associated with reduced susceptibility to CWD, whereas Haplotype B (lower panel) has been associated with higher susceptibility (Brandt et al. 2018). Data are based on phased haplotypes derived from 720 nucleotides of the *PRNP* gene sequenced across 1,433 deer.

X. Appendices

Appendix 1: DNA Extraction

DNA Extraction and Quantification

Genomic DNA was extracted from all tissues following the QIAamp Fast DNA Tissue Kit protocol (QIAGEN© Corporation, Maryland, USA). To maximize DNA yield from the samples, several DNA extraction methods were tested and a modified version of the QIAamp Fast DNA Tissue Kit extraction protocol was identified as yielding the best quantity of high-quality DNA. Concentration of DNA from each sample was quantified with a Qubit 2.0 Fluorometer (Invitrogen, Inc.) following the standard manufacturer's protocol.

To ascertain the presence of high-quality genomic DNA (i.e., molecular weight >10kb), a 5μl aliquot of the DNA extract was separated on a 2% agarose gel and visualized using GelGreen on a bluelight transluminator (Gel Doc™ EZ Imager; Bio-Rad). Large DNA fragments migrate more slowly than small fragments in a gel, and high-quality DNA forms a distinct band (Fig. A1-1A), whereas degraded or fragmented DNA forms of a 'smear' of small fragments (Fig. A1-1B.

Figure A1-1: Visualization of DNA fragments (white band or smudge) on an agarose gel (dark) based on four samples (each lane represents a sample). (A) Highquality, non-fragmented genomic DNA visible as a distinct bright band in the upper section of the gel. (B) Degraded, fragmented genomic DNA visible as a smear across the gel. Genomic DNA digested with restriction enzymes for ddRAD sequencing would look similar to samples in B.

Appendix 2: SNP Methods

SNP Data Generation – ddRAD Library Preparation

To assay SNP variation across thousands of loci for each white-tailed deer sample, we developed protocols for a genomic approach called double-digest restriction site associated DNA sequencing, or ddRAD (Peterson et al., 2012) using a combination of *in silico* and *in vitro* methods (Chafin et al., 2018). RADseq methods (Fig. A2-1) use restriction enzymes to perform a targeted fragmentation of DNA (i.e., cut the genome into smaller pieces), followed by a size selection to reduce the genome down to some specified number of fragments (i.e., subsample of pieces). Because the fragments selected from this process, and ultimately the fraction of the genome which will be sequenced, depends on 1) where the genome is cleaved (e.g. cut by the restriction enzymes) and 2) which fragments are chosen for sequencing (e.g. by the 'bounds' of the size selection), it is necessary to tailor both of these parameters to a specific target organism.

ddRADseq = Double-digest Restrictionsite Associated sequencing

- 1. Targeted fragmentation of genome with restriction enzymes
- 2. Selection of specific-sized fragments from each individual
- 3. Next-Generation-Sequencing (Illumina) of selected fragments

Figure A2-1: Schematic of ddRADseq approach. Genomes of different individuals are represented in colors. The first step involves targeted fragmentation of genomic DNA with restriction enzymes resulting in similar fragments across individuals. The next step represents size-selection of fragments from each individual to subsample the genome in a consistent and comparable manner. Selected fragments are then barcoded for individual identification and prepared for massive-parallel Next-Generation Sequencing (Illumina Sequencing). The digital output consists of millions of short sequences (reads) that are then bioinformatically processed to identify genetically variable loci.

To select enzymes, we performed simulated digests on multiple reference genomes for cervids and bovids, using the software FRAGMATIC (Chafin et al., 2018). Simulated digests were performed for enzyme combinations *Pst*I/*Msp*I, *Pst*I/*Bsa*HI, *Pst*I/*Nar*I, *Nsi*I/*Msp*I, *Sbf*I/*Msp*I, and *Nsi*I/*Hpy*CHIV, using reference genomes for *Capreolus capreolus*, *Capra hircus*, *Bos taurus*, *Bison bison*, and *Odocoileus virginianus*.

From these results, we selected candidate enzyme pairings for *in vitro* digestion of an exploratory set of white-tailed deer samples (N=8), followed by fragment analysis on the Agilent Tapestation (Agilent, Santa Clara, CA). This allowed quantification of DNA fragments occupying ranges within the distribution of fragments resulting from the restriction digest. We averaged estimates across samples and estimated the distinct number of loci within each region using the following equation, where Γ is the expected haploid genome size, \bar{y} is the mean fragment length for the size range in question (e.g. 250-350 base pairs), \hat{c} is the proportion of DNA mass contained within the fragment length range, and \hat{p} is the expected proportion of sequence-able fragments, as estimated using the *in silico* approach described above.

$$
n\ loci = \frac{2\Gamma * \hat{c} * \hat{p}}{\bar{y}}
$$

This was calculated for each sample for a variety of potential size selection ranges, for each candidate restriction enzyme. Enzyme pairings and size selection parameters were then selected to optimize a) sequencing coverage across individuals, b) number of individuals which could be multiplexed per lane of sequencing (e.g. cost efficiency), and c) number of sampled loci.

We then performed restriction digests using the selected enzymes (*Nsi*I and *Msp*I) as the first step in a modified ddRAD protocol, using 1µg of template DNA per sample. These digests were purified using Ampure XP magnetic bead-based separation (Beckman Coulter, Inc., France). Purified digests were then ligated with individually barcoded adaptor sequences using T4 DNA ligase (New England Biolabs, Inc.) and manufacturer-supplied buffer with a 60min incubation at 37°C, followed by heat-inactivation of enzymes at 65°C. Adapter oligonucleotide fragments were annealed prior to ligation, and diluted to create a 5-10X excess concentration relative to the template DNA.

Individual uniquely-barcoded samples were then pooled, Ampure XP purified, and size-selected using a PippinPrep for automated gel extraction (Sage Science, Inc.). This is advantageous over manual

gel-based methods in that it significantly increases replicability across libraries and diminishes 'small fragment carryover' (DaCosta and Sorenson, 2016).

Adaptors were then extended to include indices for a dual-indexing strategy (Peterson et al., 2012) as well as sequences for anchoring to the Illumina flowcell (Illumina, Inc.) via PCR with the following conditions: initial denaturation at 98°C; 10 cycles of 98°C for 15s, 62°C for 30s, and 72°C for 30s; followed by a terminating extension period at 72°C for 7min. Samples were then submitted for sequencing (N=96 per lane) on the Illumina HiSeq 4000 (single-end; 1x100bp) at the University of Oregon Genomics and Cell Characterization Facility (Eugene, OR).

SNP Data Generation – ddRAD Loci Assembly and Filtering

The digital output of Illumina sequencing consists of millions of short sequences (=reads) that need to be bioinformatically processed to assemble the SNP loci used for analyses. The Illumina data of the ddRAD library was processed and filtered in the assembly pipeline PYRAD (Eaton, 2014). We first demultiplexed reads into per-individual sets, allowing zero barcode mismatches, and deleting any reads with >4 bases falling below a quality threshold of 99% accuracy. Reads were then clustered within individuals to find those representing the same genomic regions (=loci) using a distance threshold of 15% via the VSEARCH algorithm (Rognes et al., 2016), allowing a maximum of 3 indels, and deleting any loci with coverage greater than max(500, mean+2sd)X or less than 20X. PYRAD then performs global alignment within each locus (Edgar, 2004) and distinguishes biological signal (i.e., true SNP variation) from sequencing errors using a maximum likelihood estimation of genotypes (Li et al., 2008). Finally, loci of all samples were then clustered among individuals to identify homology, with any locus present in fewer than 50% of the individuals being removed.

85

Appendix 3: SNP Analysis

SNP Analysis - Population Structure

Population structure was inferred from SNP data using the program ADMIXTURE (Alexander et al., 2009) with the admixturePipeline (github.com/smussmann82/admixturePipeline). In addition to the full dataset (N=1,183), we also replicated these analyses across several subsets: (i) partitioned by sex; and (ii) across ten down-sampled replicates (20% of individuals), generated using a random sample weighted inversely by spatial sampling density. The latter was performed so as to evaluate the impact of uneven sampling on our ADMIXTURE results.

Results were parsed in CLUMPAK (Kopelman et al., 2015). Model selection (for value of *K*, i.e., number of populations) was performed by cross-validation, and results for the best models were recreated using DISTRUCT (github.com/smussmann82/distruct-rerun) (Rosenberg, 2004). To visualize the spatial extent of inferred populations, samples were plotted as pie charts representing probability of assignment to gene pools in ARCMAP (Environmental Systems Research Institute, Inc.). Assignment probabilities were also interpolated using Empirical Bayesian Kriging (EBK) (Gribov and Krivoruchko, 2012). EBK probability surfaces were used as a first-pass qualitative evaluation of landscape features, such as rivers and roads, to inform predictor variable selection for in-depth spatial analyses (described below).

SNP Analysis – Estimating migration

Natural populations commonly exhibit a pattern of 'isolation by distance' in which genetic relatedness declines as a function of distance (Wright, 1943). When habitats vary (e.g. elevation gradients, vegetation type), some aspects of the landscape will invariably have an effect on the probability of individuals moving through that space – for example the presence of barriers such as rivers or roads, or variation in the suitability of habitat. As individual dispersal declines, so does gene flow. As a consequence, those landscape features effecting individual movement accumulate over time as variation in genetic differences over space. Variation in genetic connectivity (='gene flow') can be

86

estimated by examining spatial patterns of relatedness, under the general assumption that areas in which genetic relatedness decays very quickly have little gene flow, and areas in which genetic relatedness is retained over large distances have high gene flow. We visualized this as an 'effective migration surface' using the program EEMS (Petkova et al., 2015), as a means to examine underlying landscape resistance.

These results are complementary to the ADMIXTURE results (above), in that they model different, but related, evolutionary processes. EEMS models migration, while ADMIXTURE models a *product* of varied migration over space, namely population structure. The primary purpose of our subsequent analyses was to understand how features of the environment interact as determinants of migration and population structure.

Appendix 4: *PRNP* **Methods**

PRNP/ PRNPPSG Amplification and Sequencing

Genomic DNA was used as template to amplify a coding section of the *PRNP* gene following a modified protocol from previous studies (Brandt et al., 2015, Johnson et al., 2003). For the functional *PRNP* gene, the forward primer (CWD-13) straddles Intron 2 and Exon 3, with the reverse primer (CWD-LA) located 850bp downstream (Johnson et al., 2003). To ascertain if the detected polymorphisms were indeed in the functional *PRNP* gene, presence of the non-coding *PRNP* pseudogene (*PRNPPSG)* was evaluated by using pseudogene primers 223 and 224 from O'Rourke et al. (2004).

Amplifications for both the functional *PRNP* gene and the *PRNP^{PSG}* pseudogene were performed in 20μl reactions consisting of 10μl Qiagen HotStart Master Mix (1unit HotStartTaq DNA Polymerase, PCR Buffer with 3mM MgCl₂, and 400μM of each dNTP), 8μM each of the forward and reverse primer, 7.4μl RNase-free water, and 1μl of template DNA (~50-100ng). Thermocyling protocols consisted of an initial denaturation step of 15min at 95°C, followed by 10 cycles of 45s denaturation at 95°C, 45s annealing at 57°C, and 75s extension at 72°C, 25 cycles of 30s denaturation at 95°C, 30s annealing at 55°C and 60s extension at 72°C, completed with a final extension step of 5min at 72°C.

Samples where the *PRNP* and *PRNPPSG* amplified were sequenced across both to identify the true polymorphism in the functional *PRNP* gene*.* Amplicons were enzymatically purified, sequenced using BigDye v. 3.1 (Applied Biosystem Inc., Forest City CA) dye-terminator chemistry and resolved on an ABI 3730XL GeneAnalyzer at the University of Illinois Keck Center for Functional and Comparative Genomics. Sequences were manually edited using SEQUENCHER (v 5.4, Gene Codes, Ann Arbor MI) and aligned against a reference database of *PRNP* gene sequences obtained from the NCBI GenBank database (Accession # AF156185.1, AY360o089.1, AY3600091.1).

PRNP Analysis – Haplotype Data

Following alignment, sequences were phased to haplotypes (paired nuclear alleles) using the program PHASE2 (Stephens et al., 2001), which reconstructs haplotypes using a probabilistic model of linkage disequilibrium. Only haplotypes assigned with >90% posterior probability (N=1,433) were retained. Scripts to format inputs and parse results of haplotype phasing are available at *github.com/tkchafin/fasta2phase2*. Haplotypes were then categorized according to the nomenclature of Brandt et al. (2015), and haplotype frequencies were calculated globally, by-county, and by CWD status (positive *vs.* negative). To visualize similarity amongst haplotype, we constructed a haplotype network using the median-joining algorithm employed by POPART (Leigh et al., 2015); a haplotype network reflects numbers of nucleotide substitutions (point mutations) among the different *PRNP* sequences. Scripts for creating these input files can be found at *github.com/tkchafin/scripts*.

Appendix 5 – Spatial Analyses Methods

Spatial Analysis Parameters

All of the following analyses were computed in R version 3.6.0, and all of the code are provided as online supplementary material (https://github.com/zdzbinden). An outline of the spatial analyses is provided in Fig 22. The initial data for the spatial analysis was composed of 35 environmental factors (Table 6, Figs. A6-1 through A6-16) plus the variable of interest, log M (Fig. 12), which were all associated with the 13,378 hexagonal grid nodes. Exploratory analyses revealed higher variance in log M at lower genetic sampling densities, so a threshold of sampling density (>0.01) was found in order to remove heteroscedasticity in the variable. This reduced the number of spatial nodes from 13,378 to 2,665.

Figure A5-1: Schematic of hexagonal network with reference cells in space before (A) and after (B) grid averaging. Lines show tessellation grid and dots represent spatial distribution of genetic samples in a spatial framework.

The analyses consisted of two parts: (1) transforming the environmental data to fit assumptions of the subsequent analyses and reducing the set of candidate environmental variables into a more meaningful and explanatory set of environmental factors, and (2) estimating spatial autocorrelation of the data.

The **data transformation and standardization** included: (a) categorical variables were transformed into 'dummy' binary variables, (b) all variables were checked for deviations from normality and, if non-normal, were corrected using standard transformations, and (c) data were standardized using Z-score scaling.

Reduction of the 35 environmental variables (Table 6) **to an explanatory set of 17 factors** (Table 7) was accomplished using a variety of methods. First, *Variation Inflation Factors* (*VIF*) were calculated for each variable as a means to determine **how correlated each variable** was with others in the set. Variables with a *VIF* > 10 were removed from the set (Dorman et al., 2012). Next, the remaining variables were each tested individually for a **relationship/effect on the variable of interest**, the migration rate (logM). Continuous variables were tested using linear regression and categorical, binary variables were tested for differences in means between the presence and absence of a factor using Student's t-test.

The second part of the analysis involved estimating the **spatial component or autocorrelation of the data**. This was done using *distance-based Moran's Eigenvector Mapping* (dbMEM) (Dray et al., 2006), which is also referred to as eigenvector spatial filtering (ESF) (Murakami and Griffith, 2019) and principal coordinates of neighbor matrices (PCNM) (Borcard and Legendre, 2002). Briefly, this analysis involves decomposing the spatial position of samples (i.e. coordinates) into n-1 eigenvectors by generating more and more complex eigenfunctions that explain 'neighborhoods' of sites based on, in this case, an exponential kernel. The set of eigenvectors is then used in a forward selection procedure to determine **which eigenvectors explain autocorrelation in the parameter of interest**, in this case the migration rate (logM). The forward selection calculates *VIF* and *Akaike Information Criterion* (*AIC*) to select the best set of eigenvectors for explaining/representing the spatial component of the data. The final spatial model lists factors by rank decrease in AIC (i.e. degree the variable made the model better) so that the top number of spatial eigenvectors equal to the number of environmental factors could be

91

selected. This was done so that the environmental and spatial datasets would have equal number of variables for the variation partitioning procedure.

Figure A5-2: Illustration of **Moran's Eigenvector Maps** (MEM). Given the overall region sampled (green shape in background) the Moran's Eigenvector Maps (n-1) decompose the spatial structure of the data into eigenfunctions that range from relatively broad scale structure (MEM=1) to relatively fine scale structure (MEM=1500). Figure after Fig S1 in Vandam and Vanschoenwinkel (2013).

The final step of the analysis was **partitioning the variation of the migration rate** (logM) **among the two datasets**: **environmental** (17 factors) and **spatial** (17 eigenvectors). This process is done using partial multiple linear regression to determine the proportion of variation of logM explained by each set separately (Legendre et al., 2005). Significance of the model and the partitions were tested via permutation procedure with 999 randomizations (Oksanen et al., 2016).

Appendix 6 – Spatial Data – Environmental Maps

Figure A6-1: **Habitat suitability**, reported as the proportion of each tile classified as 'predicted habitat' by the USGS Gap Analysis projections for white-tailed deer.

Figure A6-2: **Slope gradient** (rise as a percentage of run; left) and **elevation** (meters; right) averaged within a 10 km² hexagonal grid. Slope gradients were derived from the Gridded Soil Survey Geographic (gSSURGO) database, and elevation from a five-meter resolution digital elevation model (DEM) published by the Arkansas GIS Office.

Figure A6-3: Available **soil water storage** (0-25cm depth) measured in centimeters, taken from the from the Gridded Soil Survey Geographic (gSSURGO) database maintained by the U.S. Department of Agriculture Natural Resources Conservation Service (USDA NRCS).

Figure A6-4: **Dominant National Vegetation Classification (NVC) Formation Class** of each hexagonal tile. NVC Formation Class represents broad combinations of general dominant growth forms and reflects broad-scale variation in climate and substrate conditions.

Figure A6-5: **Dominant National Vegetation Classification (NVC) Formation Subclass** of each hexagonal tile. NVC Formation Subclass represents combinations of general dominant and diagnostic growth forms and reflects macroclimatic variation primarily due to latitude and continental position.

Figure A6-6: Dominant National Vegetation Classification (NVC) Formation type of each hexagonal tile. NVC Formation type represents a finer scale categorization of the NVC Formation Subclass.

Figure A6-7: **Dominant National Vegetation Classification (NVC) Division** classification of each hexagonal tile. NVC Division represents dominant growth forms and a broad set of diagnostic plant species representing broad-scale biogeographic differences. N. A. = North America; E. = Eastern.

Figure A6-8: **Dominant National Vegetation Classification (NVC) Macrogroup** classification of each hexagonal tile. NVC Macrogroup represents sub-continental to regional differences in mesoclimate, substrate, and disturbance regimes and is a finer-scale version of the NVC Division class.

Figure A6-9: **Dominant National Vegetation Classification (NVC) Group** classification of each hexagonal tile. NVC ecological Group represents finer-scale regional divisions within NVC Macrogroup categories.

Figure A6-10: **Distance of tile to nearest interstate** (left) or **river larger than order 5** (right), measured in kilometers.

Figure A6-11: **BioClim temperature variables 1 through 4**: (A) Mean annual range; (B) Mean diurnal range, defined as the mean of monthly ranges; (C) Isothermality, or the annual range divided by the diurnal range; and (D) Temperature seasonality (standard deviation * 100). Results shown are the mean values for each hexagonal cell, after resampling BioClim rasters.

Figure A6-12: **BioClim temperature variables 5 through 7**: (A) Maximum temperature during warmest month; (B) Minimum temperature during coldest month; and (C) Temperature annual range. Results shown are the mean values for each hexagonal cell, after resampling BioClim rasters.

Figure A6-13: BioClim temperature variables 8 through 11: (A) Mean temperature of wettest quarter (=Q); (B) Mean temperature of driest quarter; (C) Mean temperature of warmest quarter; and (D) Mean temperature of coldest quarter. Results shown are the mean values for each hexagonal cell, after resampling BioClim rasters.

Figure A6-14: BioClim precipitation variables 12 through 15: (A) Mean annual precipitation; (B) Mean precipitation during wettest month; (C) Mean precipitation during driest month; and (D) Precipitation seasonality (coefficient of variation). Results shown are the mean values for each hexagonal cell, after resampling BioClim rasters.

Figure A6-15: BioClim precipitationvariables 16 through 19: (A) Precipitation during wettest quarter; (B) Precipitation during driest quarter; (C) Precipitation during warmest quarter; and (D) Precipitation during coldest quarter. Results shown are the mean values for each hexagonal cell, after resampling BioClim rasters.

Figure A6-16: Bailey's ecoregion classifications at hierarchical levels of: (A) Division; (B) Province; and (C) Section.

Appendix 7 – Spatial Analysis: 17 Relevant Variables

Spatial Analysis Results (In-depth)

The 35 environmental factors were distilled into a set of 17 factors that significantly explained whitetailed deer dispersal (Table 7). This set of meaningful factors included riverine barriers to dispersal (e.g. RIVER_DIST, ECOLYS_LU_Open Water, and SECTION Arkansas River Valley), climatic variables (e.g. BIO8_WETTE), vegetation variables (e.g. West Gulf Coast Mesic Hardwood Forest), and section variables (e.g. Boston Mountains). The single most explanatory factor was SECTION_Arkansas River Valley which when present had a mean logM lower than that of non-Arkansas River valley sections by 0.42. For context logM ranges from -1 to 1 (Fig. 12). This suggests the Arkansas River is a significant barrier to white-tailed deer dispersal. Maps and distributions of the 17 environmental factors and associated results of spatial analyses are provided in Figs. A1-A17).

Spatial decomposition via dbMEM extracted 81 meaningful eigenvectors, of which 73 were included in the forward selection procedure, and 65 were retained after correcting for multiple tests. These eigenvectors represent spatial autocorrelation in the variable logM. Smaller eigenvectors (e.g. V3) represent broad scale spatial structure and larger eigenvectors (e.g. V81) represent relatively finer scale structure (Fig. 24). Prior to variation partitioning, only the 17 top-ranked eigenvectors were selected so that the environmental and spatial datasets would have the same number of factors.

A linear model including both spatial eigenvectors and environmental factors explained 67% of the variance in the variable logM (adj. R^2 = 0.67; p < 0.001). There was a large proportion of spatial autocorrelation in the variable logM (adj.R² = 0.64; p < 0.001). However, only 20% of the variance of logM could be attributed to environmental factors, the majority of which was spatially structured environmental variation (17%). Thus, 47% of the variation of logM was spatially autocorrelated, but could not be explained by the environmental factors presented here (Fig. 25). This suggests that the unexplained autocorrelated structure could be explained by either 1) unmeasured environmental factors, 2) biotic factors intrinsic to deer populations (e.g. population size), or 3) a combination of these.

110

Distance to Interstate Highway

Figure A7-1: Distribution of the **variable ROAD_DIST** across the state. This variable represents the distance to the nearest interstate highway. The relationship between ROAD_DIST and logM is illustrated in the top right corner.

Distance to nearest river stream order > 5

Figure A7-2: Distribution of the **variable RIVER_DIST** across the state. This variable represents the distance to the nearest river (stream order > 5). The relationship between RIVER_DIST and logM is illustrated in the top right corner.

Figure A7-3: Distribution of the **variable HAB** across the state. This variable represents the. The proportion of an area classified as habitable by white-tailed deer. The relationship between HAB and logM is illustrated in the top right corner.

Mean Temperature of Wettest Quarter

Figure A7-4: Distribution of the **variable BIO_WETTEMP** across the state. This variable represents the mean temperature during the wettest quarter of the year. The relationship between BIO8_WETTEMP and logM is illustrated in the top right corner.

Mean Temperature of Driest Quarter

Figure A7-5: Distribution of the **variable BIO9_DRYTEMP** across the state. This variable represents the mean temperature during the driest quarter of the year. The relationship between BIO9_DRYTEMP and logM is illustrated in the top right corner.

Precipitation of Wettest Quarter

Figure A7-6: Distribution of the **variable BIO16WQPR** across the state. This variable represents the precipitation of the wettest quarter of the year. The relationship between BIO16WQPR and logM is illustrated in the top right corner.

Precipitation of Warmest Quarter

Figure A7-7: Distribution of the **variable BIO18WAQP** across the state. This variable represents the precipitation of the warmest quarter of the year. The relationship between BIO18WAQP and logM is illustrated in the top right corner.

U.S. National Vegetation Classification Ozark/Ouachita Short Leaf Pine Bluestem Woodland

Figure A7-8: Distribution of **Ozark/Ouachita Short Leaf Pine Bluestem Woodlands** across the state. The distributions of logM when this variable is present and absent are illustrated in the top right corner.

U.S. National Vegetation Classification West Gulf Coastal Plain Mesic Harwood Forest

Figure A7-9: Distribution of **West Gulf Coastal Plain Mesic Hardwood Forest** across the state. The distributions of logM when this variable is present and absent are illustrated in the top right corner.

U.S. National Vegetation Classification Managed Tree Plantations

Figure A7-10: Distribution of **Managed Tree Plantations** across the state. The distributions of logM when this variable is present and absent are illustrated in the top right corner.

U.S. National Vegetation Classification Open Freshwater

Figure A7-11: Distribution of Open Freshwater (majority of 10 km² area) across the state. The distributions of logM when this variable is present and absent are illustrated in the top right corner.

Figure A7-12: Distribution of **West Gulf Coastal Plain Large River Floodplain Forest** across the state. The distributions of logM when this variable is present and absent are illustrated in the top right corner.

U.S. National Vegetation Classification Crowley's Ridge/Sand Forest

Figure A7-13: Distribution of **Crowley's Ridge/Sand Forest** across the state. The distributions of logM when this variable is present and absent are illustrated in the top right corner.

Bailey's Ecoregion Classification

Figure A7-14: Distribution of the **Boston Mountains** across the state. The distributions of logM when this variable is present and absent are illustrated in the top right corner.

Bailey's Ecoregion Classification Mississippi Alluvial Plain

Figure A7-15: Distribution of **the Mississippi Alluvial Basin** across the state. The distributions of logM when this variable is present and absent are illustrated in the top right corner.

Figure A7-16: Distribution of the **Arkansas River Valley** across the state. The distributions of logM when this variable is present and absent are illustrated in the top right corner.

Figure A7-17: Distribution of the **Ouachita Mountains** across the state. The distributions of logM when this variable is present and absent are illustrated in the top right corner.

Supplemental Material

Table S1: Overview of 1,720 white-tailed deer tissue samples collected in 75 Arkansas counties from 2016- 2019. Samples are listed alpha-numerically by DNA code (=(DNA), and corresponding collection information is provided by referencing AGFC field/tissue number (=field.ID) and county. Listed is also the DNA extraction protocol number (=DNAex). Data columns indicate if samples was extracted (=Dx), and type of data generated: PRNP = sequences for *PRNP* gene, PSG = sequences for *PRNP^{PSG}* pseudogene; SNP = SNP/ddRAD data. '+' = data generation successful; '-' = data generation attempted, but unsuccessful. Samples without field.ID either lack a reference number or are cross-referenced with UTM coordinates in the original AGFC database.

Table S2: Frequency of 20 *PRNP* **haplotypes detected in 1,433 white-tailed deer collected in 75 counties** in Arkansas from 2016-2019. Phased haplotypes were derived from sequence analysis of 720 nucleotides of the *PRNP* gene. Letters indicate haplotypes previously detected in other states, whereas numbers (1- 4) indicate haplotypes unique to Arkansas. Variable sites of haplotypes are listed in Table 2. Samples that tested positive for CWD (+) are listed separately for those counties where CWD was detected.

