GENOMIC DIFFERENTIATION OF BIG BLUESTEM (*ANDROPOGON GERARDII*) ALONG THE GREAT PLAINS' ENVIRONMENTAL GRADIENT

by

MIRANDA M. GRAY

B.S., Chatham University, 2009

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Plant Pathology
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2012

Approved by:

Co-Major Professor
Dr. Eduard D. Akhunov

Approved by:
Co-Major Professor
Dr. Loretta C. Johnson

Abstract

Big bluestem (*Andropogon gerardii* Vitman) is an ecologically dominant grass of the North American grasslands with precipitation-dependent productivity. However, climatic predictions for big bluestem's dominant range in the Great Plains include increased periods of drought. The main objectives of this research were to determine the extent of neutral and non-neutral genetic differentiation and diversity among putative big bluestem ecotypes using amplified fragment length polymorphism (AFLP) markers. This is the first study of both neutral and non-neutral genetic diversity of big bluestem which also includes source populations of well-described ecotypes studied in reciprocal common gardens. A total of 378 plants were genotyped from 11 source prairies, originating from one of three ecoregions (Central Kansas, Eastern Kansas, and Illinois). Using two AFLP primer sets, 387 polymorphic markers (error rate 9.18%) were found. Un-rooted neighbor joining tree and principle-component analyses showed continuous genetic differentiation between Kansas and Illinois putative ecotypes, with genetic overlap occurring between Kansas ecotypes. Analysis of molecular variance showed high diversity within-prairie sites (80%) relative to across-prairies (11%), and across- ecoregions (9%) ($p<0.001$). Within-prairie genetic diversity levels were similar among ecoregions (84-92%), with the highest genetic variation maintained in Illinois prairies (92%). Population structure analyses supported $K=6$ genetic clusters across the environmental gradient, with Kansas prairies belonging to three main genetic groups, and Illinois prairies having largely divergent allele frequencies from Kansas prairies. Interestingly, BAYESCAN analysis of the three putative ecotypes identified eight F_{ST} -outlier AFLP loci under potential diversifying selection. Frequency patterns of loci under diversifying selection were further linked to geo-environmental descriptors including precipitation, temperature severity, diurnal temperature variation, prairie location, and

elevation. The observed allele frequency divergence between Kansas and Illinois ecotypes suggests tallgrass restorations should consider possible maladaptation of non-local ecotypes and genetic swamping. However, high within-prairie genetic variation may help individual big bluestem populations withstand climatic variability.

Table of Contents

List of Figures

List of Tables

Acknowledgements

The work behind this thesis would not have been possible without the support of several people. First, I would like to thank all National Parks, landowners, and seed collectors whom graciously donated seed for this study. I would like to express my sincere gratitude to my co-major professors: Dr. Loretta Johnson for the opportunity to conduct this project, support, and professional mentorship and Dr. Eduard Akhunov for mentoring and helpful discussion regarding data analysis. Many thanks are extended to Dr. Paul St. Amand for training in AFLP work, discussions, and kind mentoring. Mary Knapp, KS State Climatologist generously provided all weather data for this project. To Drs. Karen Garrett and Chris Toomajian, I extend gratitude for constructive criticism which improved this thesis. Dr. Guihua Bai and laboratory members are warmly thanked for welcoming me into their lab to perform this work. Thanks to Dr. Alina Akhunova for continuous support. Hanquan Liang, Yanni Lun, Drs. Shichen Wang, Sanjay Chepella, Susan Brown, and Cyrille Saintenac are thanked for technical and/or bioinformatics assistance. Richard Wynia & John Row (USDA Plant Materials Center, Manhattan, KS) provided storage space and Lauren Wheeler helped with leaf collection. This master's thesis was supported through funding from the USDA (Grant # 2008-35100-04545), Ecological Genomics Institute at Kansas State University, a Kansas Native Plant Society Mary Bancroft Award, the Sigma Xi Grants-in-Aid of Research program, and an NSF Graduate Fellows in K-12 Education Program (KSU GK-12 Program, NSF DGE-0841414). Last, but certainly not least, I am extremely grateful for the patience and encouragement from my loving family and fiancé Bryan Bischof. To my Plant Pathology Family (you know who you are): "there's no place like home," so thanks for making Kansas a second home.

Dedication

This work is dedicated to my late grandmother and friend, Eleanor May Ring, who always made herself just a phone call away.

Chapter 1 - Literature Review

The Threatened U.S. Grasslands and Biodiversity

A child said What is the grass? fetching it to me with full hands, / How could I answer the child? I do not know what it is any more than he. / I guess it must be the flag of my disposition, out of hopeful green stuff woven.

-Walt Whitman, *Leaves of Grass*

The grasslands biome once contributed to 42% of the total plant cover world-wide, dispersed across all continents (Bailey 1998, Anderson 2006). Grasslands are important for carbon sequestration (Seastedt and Knapp 1993), biological diversity and ungulate grazing (McNaughhton 1984, Coughenour 1985, Samson and Knopf 1994). The area that once extended south from Canada to the Mexican border, east from the Rocky Mountains to western Indiana and Wisconsin is referred to as the Great Plains of North America (Samson and Knopf 1994). This is a landscape that Walt Whitman (1819-1892) once referred to as the "limitless and lonesome prairie" ("Song of Myself" in *Leaves of Grass* 1855). Still, the expanse of the once seemingly "limitless" tallgrass prairie is dwindling. Today, only about 4% of tallgrass prairie remains (Samson and Knopf 1994), primarily due to conversion to row crop agriculture and anthropogenic fragmentation.

Grasslands harbor substantial biodiversity for flora and fauna. Within the United States, prairies are a high priority for conservation of biodiversity, and perhaps the highest (Samson and Knopf 994). Worldwide, almost half of 234 Centers of Plant Diversity (CPDs) include grassland habitat

(WRI 2000). These CPDs, found in most regions of the world, represent areas with high diversity where conservation practices could protect a large number of grassland species. Furthermore, of 136 terrestrial ecoregions identified as outstanding examples of the world's diverse ecosystems, 35 are grasslands, supporting some of the most important grassland biodiversity in the world today (WRI 2000). Unfortunately, protected grasslands comprise only 7.6 percent of total grassland area (WRI 2000). Thus it becomes imperative to survey the current genetic diversity and possible locally-adapted ecotypes of the existing Central grasslands in the United States. Understanding genetic differentiation and genetic diversity levels of important grassland species will help inform species' conservation and restoration of the endangered grassland ecosystem.

The U.S. grasslands cover a total of 3 million km^2 and is a recently formed (~20,000 years), dynamic landscape with a number of endemic species and fragmented habitat regions (Axelrod 1985). The vegetative structure of the grasslands has been greatly influenced by burning, grazing, and anthropogenic changes, with fires maintaining grassland structure and growth (Axelrod 1985, Anderson 2006). Grasslands are also importantly and largely shaped, both structurally and functionally, by water availability (Fay et al. 2003). Across the U.S. Midwest, a sharp precipitation gradient corresponding to an increase in primary productivity from the shortgrass steppes in the West to the tallgrass steppes of the central United States exists (Transeau 1935, Sala et al. 1988) (**Figure 1.1**).

Variability in rainfall and temperature are major abiotic factors controlling productivity, structure and many ecosystem processes of tallgrass prairie (Axelrod 1985, Knapp et al. 2002). Rainfall is a major factor controlling aboveground net primary productivity (ANPP) worldwide (Sharp et al.

2004) and especially in tallgrass prairies. The amount of rainfall has been found to be a limiting factor in affecting plant community composition and forb ANPP (Knapp et al 2001). Water availability depends on a number of factors, including temperature and frequency of rainfall events (Sala et al. 1988). Thus, changes in both precipitation and temperature are expected to pose a serious threat to the tallgrass prairie ecosystem (Knapp et al. 2002). For the Great Plains region, it is predicted that longer periods of drought will occur in the future with more intense but punctuated rainfall evens, and droughts have been observed to indeed be severe for this region (IPCC 2007, NOAA 2012) Yet, we do not know the extent to which prairies harbor sufficient genetic variation to be able to respond to natural and human caused changes such as climatic shifts.

Andropogon gerardii: an Ecologically-Dominant Prairie Grass

Andropogon gerardii (big bluestem) is a warm-season ecological-dominant of the North American tallgrass prairie. Big bluestem comprises approximately 70% of the biomass of the tallgrass prairie ecosystem (Gale et al. 1990) and is a co-dominant species along with *Sorghastrum nutans* and *Panicum virgatum* (Riley and Vogel 1982). Big bluestem's has a wide distribution spanning the United States and Canada, with its dominant distribution east of the Rockies (**Figure 1.2**). Big bluestem propagates vegetative growth in the form of reproductive and non-reproductive tillers (McKendrick et al. 1975). After a season of dormancy, the bud forms following cold weather, and leads to the elongation of the bud usually around April, signaling the start of the growing season of big bluestem (Owsley 2003). A perennial grass, big bluestem is long-lived through maintenance of rhizomes, and root cores have been observed to

last for approximately 3-5 years. The stages of growth of big bluestem tillers include five main stages, common to perennial forage grasses: 1) germination, 2) vegetative growth, 3) elongation, 4) reproductive growth, and 5) seed ripening (Moore et al. 1991). Pollen is largely wind-pollinated. Less than 5% pollen from the three-pronged spikelet (hairy inflorescence) is wind-has been tracked to travel ~30 meters (Jones and Newell 1946). Seed dispersal is important to the maintenance of populations and also to diversity levels. Ripe seed of big bluestem is positioned approximately 1-3 meters in the air, free-standing. The hairiness of the seed as well as position allow for likely seed attachment to bison fur as well as humans (Keeler 2000). This suggests that seed dispersal is wide-ranging especially in the extant prairie landscape, with bison transporting seed to where furs are shed in the springtime. Big bluestem is an obligate outcrosser with low to non-existent selfing (Normann et al. 2003), with strong self-incompatibility (no seed set when 27 accessions were selfed).

As with many other grass systems (Stebbins 1971), big bluestem consists of a large polyploid genome. The base chromosome number in big bluestem is ten chromosomes (Gould 1956). As an autopolyploid, each chromosome may be duplicated six-nine times with intermediate numbers of chromosomes (between 60-90 chromosomes) possibly the result of enneaploids producing an unbalanced set of gametes (Keeler and Davis 1999, Normann et al. 2003). Ploidy level variation studies found mixed cytotypes occurring in natural habitat (6X-9X) within the species (Keeler and Davis 1999).

Ecological and Genetic Studies

Big bluestem has served as a model species for prairie ecology for nearly a hundred years. The use of big bluestem for research has included studies on climate effects, community structure, physiological responses, and restoration effectiveness (Epstein et al. 1998, Knapp et al. 2001, Silletti and Knapp 2002, Fay et al. 2002, Fay et al. 2003). Most studies of this ecological dominant have historically taken an ecological and eco-physiological approach (Knapp et al. 1993, Gustafson 2004, Tompkins 2010). While these studies have provided a plethora of information regarding water limitations, productivity, physiology and distribution of species across its range, the utility of population genetics combined with a large sub-continental landscape/environmental approach has rarely been undertaken (except see Tompkins 2011, Rouse et al. 2011, and Price et al. 2012).

Previous studies using neutral genetic markers to detect genetic diversity in big bluestem have been performed. Using 37 random amplified polymorphic DNA (RAPDs), Gustafson et al. (1999) found that genetic diversity retained within seven Arkansas remnant prairies dispersed within a similar region was greater than among-prairie genetic diversity, with within-prairie diversity ranging between 82.7-99.3% and comprising 89% of the total present genetic variation. However, this study did not find evidence for prairie genetic structuring along the landscape sampled. The high within-prairie diversity discovered is consistent for outcrossing systems (Hamrick and Godt 1996). Furthermore, when attempting to correlate genetic diversity with size of the study prairie, no significant association was detected. However, the lack of association detected between increasing geographical distance and genetic dissimilarity of individuals is a factor that must be considered when restoring tallgrass prairie as this suggests that the genetic

variation is influenced more highly by selection pressures. Contrasting results were found in a study by Rouse et al. (2011) of natural populations of big bluestem sampled across the U.S. Midwest precipitation gradient; in this study, dissimilarity among individuals increased with increasing geographic distance. The presence of fine-scale genetic structuring in big bluestem has been demonstrated, with an interesting high frequency of certain genotypic clones that may have implications for community diversity (Avolio et al. 2011). Even at small scales (1 m^2) genotypic diversity is noted, with certain genotypes maintain higher frequencies in plots, suggestive of competitive hierarchy among genotypes of big bluestem at finer scales. Genotype frequency has in plants which can propagate through vegetative growth such as big bluestem (through tillers) may signify selection for that genotype in a given environment.

Information regarding genetic diversity of local vs. non-local seed is potentially important to restoration efforts. Local remnant populations have been shown to have a higher genetic diversity compared with sites planted with artificial outside seed sources (Selbo and Snow 2005). Introduction of foreign seed source to local populations could increase genetic diversity and the likelihood of species persistence and adaptation in a changing system. However, the use of non-local plant materials in restoration efforts presents a suite of issues (McKay et al. 2005) and thus particular attention should be paid to the range over which a particular plant species is adapted if this plant is to be implemented in restoration and/or conservation efforts. However, some studies on prairie land restoration caution against using outside seed source to increase genetic diversity and adaptive capabilities (Havens 1998, Montalvo and Ellstrand 2001, Hufford and Mazer 2003, Gustafson et al. 2004) as this may result in a maladaptation.

Ecotypes

Ecotypic adaptive genetic variation has been widely studied across plant species, and previous work has uncovered differences in plant biomass, freezing and drought tolerance levels, root growth and phenology across ecotypes (Smith et al. 1946, Clausen and Heisey 1958, McMillan 1969, McKell et al. 1962, Leimu and Fisher 2008, and Dionne et al. 2009). Ecotypic differentiation studies have proven useful to several fields including agriculture, plant breeding, and restoration ecology (Hufford and Mazer 2003, Juenger et al. 2005), as well as to broaden basic understanding of plant population structure and dynamics. Studies involving ecotype differentiation have taken advantage of variation across the environmental extremes of a climatic region like the U.S. central grasslands to address complexities in plant structure and diversity in traits (McMillan 1969, Mintenko et al. 2002, Etterson 2004). However, rarely has a multidisciplinary approach been employed which merges traditional ecological studies with modern molecular techniques to study ecotype neutral and functional variation across such a grand spatial scale (Knight et al. 2006).

Phenotypic variation was noted in big bluestem as early as the 1960s, with seminal studies by McMillan (1965a, 1965b). McMillan sampled six big bluestem ecotypes along a latitudinal gradient in the United States, and planted each ecotype in a common garden in Texas (McMillan 1965a), but also in a light-controlled greenhouse setting (McMillan 1965b). McMillan noted that the native geographic location of the source population (plant ecotype) played a role in the vegetative biomass, with southern-most ecotypes of big bluestem having greater biomass in their native site (Texas), and bluestem native to northern parts of the United States having less and shorter flowering stalks. Gustafson previously also noted the presence of visible phenotypic

differences in big bluestem communities across environmental gradients in the 1970s, indicating the possibility of locally adapted ecotypes. However, no studies have linked phenotypic differences in the species to more in-depth genetic studies of both neutral and non-neutral genome diversity.

The work outlined here seeks to fill a gap between ecological and diversity studies by providing such a genetic analysis of prairie sources that are also seeded in reciprocal gardens to test for adaptation of plant ecotypes. In the reciprocal garden tests, Johnson et al. (in prep) take advantage of historically-observed phenotypic variation across prairies of big bluestem by planting naturally-occurring ecotypes along the U.S. Midwest precipitation gradient in seeded gardens and in single transplants, without competition, along the gradient (**Figure 1.3**). The seeded gardens planted include natural competitors (forbs and grasses native to tallgrass prairie) to simulate natural community response to ecotype plantings. Data from the first two years of field studies suggests that the western-most (Central Kansas) ecotype has native-site advantage, with increased cover in its dry home environment in comparison to the Illinois ecotype from the most mesic end of the precipitation regime (Johnson et al. in prep). Furthermore, in even drier locations at the edge of the species' native range, the western-most occurring big bluestem ecotype still out-competes the Illinois and the Eastern Kansas ecotype, suggesting local adaptation to drier climates. As an extension of these studies of local adaptation and ecotypic differentiation, we utilize a genome scan approach using neutral markers (amplified fragment length polymorphisms, or "AFLPs") but also relate these markers to the environmental cline along which ecotypes of big bluestem occur. This work is presented herein.

Amplified Fragment Length Polymorphisms (AFLP) and Genome Scans

Amplified fragment length polymorphisms (AFLPs), as neutral markers, have great utility for non-model systems as no prior genomic sequence is required (Vos et al. 1995). The major steps of the AFLP reaction include 1) restriction of genomic DNA, 2) ligation or annealing of AFLP adapters, and 3) two steps of fragment amplification (pre-amplification and selective amplification) using different sets of primer pairs (**Figure 1.4**). The advantage of the AFLP method is that after being applied to a non-model species, the number of bands or markers generated per primer pair is large (~150 bands/primer pair) and the bands are highly reproducible (except see genotyping error control, Bonin et al. 2004). Currently, the use of capillary fluorescence has allowed for avoidance of radioisotope staining of gels and greater resolution of AFLP bands (bands with one base pair difference can easily be distinguished using capillary fluorescence although this is not true with the radioisotope method). The main disadvantage of AFLPs as markers is their dominant nature, which does not allow heterozygosity levels to be consistently estimated as heterozygous genotypes cannot distinguished from dominant homozygous individuals (Vos et al. 1995).

In the last few years, there has been an increase in number of studies using "AFLP genome scans" to test for local adaptation of natural populations (for some examples see Balding and Beaumont 2004, Freedman et al. 2010, Keller et al. 2011, and Collin and Fumagalli 2011), with the hope of gleaning more information from traditional AFLP neutral marker studies. Neutral markers have been used heavily in conservation of non-model, ecologically-relevant species in the past (Ronikier 2002, Lucchini 2003). In part, newer AFLP genome scan studies involve studying organisms with wide distributions over environmental gradients where environment

may be correlated to the AFLP markers generated (Balding and Beaumont 2004, Luikart et al. 2003, Oleksyk et al. 2010). The environmental gradients studied in these studies include altitudinal variation, precipitation regimes, and more (Keller et al. 2011, Lee and Mitchell-Olds 2012). Teasing apart environmental factors, geography, demography, and selective processes within or across species is a daunting task that has been made possible in recent years by developments in statistical genetics (for example, Bayesian methods). Such advances and the use of AFLP genome scans have allowed the study of how these different factors (environment, geography, etc.) may associate with observed genetic differentiation and population structuring in natural populations (Gaggiotti et al. 2009). In the hunt for "ecologically relevant adaptive variation" (Karrenberg and Widmer 2008), these developments are key.

References

Anderson, R. 2006. Evolution and origin of the Central Grassland of North America: climate, fire, and mammalian grazers. *Journal of the Torrey Botanical Society* 133 (4):626-647.

Avolio, M.L., C.C. Chang, and M.D. Smith. 2011. Assessing fine-scale genotypic structure of a dominant species in native grasslands. *American Midland Naturalist* 165: 211–224.

Axelrod, D.I. 1985. Rise of the grassland biome, central North America. *Botanical Review* 51: 163-201.

Bailey, R.G. 1998. *Ecoregions: The Ecosystems Geography of the Oceans and Continents.* Springer, New York.

Beaumont, M.A. and D.J. Balding. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* 13: 969-980.

Bonin, A., E. Bellemain, P.B. Eidesen, F. Pompanon, C. Brochmann, and P. Taberlet. 2004. How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* 13: 3261-3273.

Burke, I., W.K. Lauenroth, and W.J. Parton. 1997. Regional and temporal variation in net primary production and nitrogen mineralization in grasslands. *Ecology* 78(5): 1330-1340.

Chan et al. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* 327(5963): 302-305.

Clausen, J., D.D. Keck, and W.M. Heisey. 1948. Experimental studies on the nature of species. III. Environmental responses of climactic races of *Achillea*. Carnegie Institute of Washington, Publication 581.

Collin, H. and L. Fumagalli. 2011. Evidence for morphological and adaptive genetic divergence between lake and stream habitats in European minnows (*Phoxinus phoxinus,* Cyprinidae). *Molecular Ecology* 20: 4490-4502.

Coughenour, M.B. 1985. Graminoid responses to grazing by large herbivores: adaptations, exaptations, and interacting processes. *Annals of Missouri Botanical Garden* 72: 852-863.

Dionne, J., S. Rochefort, D.R. Huff, Y. Desjardins, A. Bertrand, and Y. Castonguay. 2009. Variability for freezing tolerance among 42 ecotypes of green-type annual bluegrass. *Crop Science Society of America* 50: 321-336.

Epstein, H.E., W. K. Lauenroth, I. C. Burke, and D. P. Coffin. 1998. Regional productivities of plant species in the Great Plains of the United States *Plant Ecology* 134:173-195.

Etterson, J. 2004. Evolutionary potential of Chamaecrista fasciculate in relation to climate change II. Genetic architecture of three populations reciprocally planted along an environmental gradient in the Great Plains. *Evolution* 58(7): 1459-1471.

Fay, P.A., J.D. Carlisle, B.T. Danner, M.S. Lett, J.K. McCarron, C. Stewart, A.K. Knapp, J.M. Blair, and S.L. Collins. 2002. Altered rainfall patterns, gas exchange, and growth in grasses and forbs. *International Journal of Plant Science* 163(4): 549-557.

Fay, P.A., J.D. Carlisle, A.K. Knapp, J.M. Blair, and S.L. Collins. 2003. Productivity responses to altered rainfall patterns in a C4 dominated grassland. *Oecologia* 137: 245-251.

Freedman, A.H., H.A. Thomassen, W. Buermann, and T.B. Smith. 2010. Genomic signals of diversification along ecological gradients in a tropical lizard. *Molecular Ecology* 19: 3773-3788.

Gaggiotti, O.E., D. Bekkevold, H.B.H. Jorgensen, M. Foll, G.R. Carvalho, C. Andre, and D.E. Ruzzante. 2009. Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: Atlantic herring as a case study. *Evolution* 63(11): 2939-2951.

Gale, W.J., M.B. Kirkham, E.T. Kanemasu, and C.E. Owensby. 1990. Net carbon exchange in canopies of burned and unburned tallgrass prairie. *Theoretical Applied Climatology* 42: 237-244.

Gould, F.W. 1956. Chromosome counts and cytotaxonomic notes on grasses of the tribe Andropogoneae. *American Journal of Botany* 43: 395-404.

Gustafson, D.J., D.J. Gibson, and D.L. Nickrent. 2004. Competitive relationships of *Andropogon gerardii* (Big Bluestem) from remnant and restored native populations and select cultivated varieties. *Functional Ecology* 18: 451-457.

Havens, K. 1998. The genetics of plant restoration. *Restoration and Management Notes* 16: 68-72.

Hoekstra, H.E., R.J. Hirschmann, R.J. Bundey, P. Insel, and J.P. Crossland. 2006. A single amino acid mutation contributes to adaptive color pattern in beach mice. *Science* 313: 101-104.

Hufford, K.M. and S.J. Mazer. 2003. Plant ecotypes: genetic differentiation in the age of ecological restoration. *Trends in Ecology and Evolution* 18: 147-155

IPCC Report. 2007. Climate change 2007: Physical Science Basis. Approved at 10th session of Working Group I. at the IPCC. IPCC, Paris, France, February 2007.

Johnson et al. in prep. Differential establishment, cover, and gas exchange among ecotypes of a dominant grass big bluestem *Andropogon gerardii* in reciprocal gardens along a Great Plains climatic gradient.

Jones, M.D. and L.C. Newell. 1946. Pollination cycles and pollen dispersal in relation to grass improvement. *Nebraska Agricultural Experimental Station Research Bulletin* 14

Juenger, T., M. Purugganan, and T.F.C. Mackay. 2000. Quantitative trait loci for floral morphology in Arabidopsis thaliana. *Genetics* 156: 1379-1392.

Karrenberg, S. and A. Widmer. 2008. Ecologically relevant genetic variation from a non-*Arabidopsis* perspective. *Current Opinion in Plant Biology* 11: 156-162.

Keeler, K.H. and G.A. Davis. 1999. Comparison of common cytotypes of *Andropogon gerardii* (Andropogoneae Poaceae). *American Journal of Botany* 86: 974-979.

Keeler, K.H. 2000. Influence of past interactions on the prairie today: a hypothesis. *Great Plains Research* 10: 107-126.

Keller, I., A. Taverna, and O. Seehausen. 2011. Evidence of neutral and adaptive genetic divergence between European trout populations sampled along altitudinal gradients. *Molecular Ecology* 20: 1888-1904.

Knapp, A.K., J.K. Koelliker, J.T. Fahnestock, and J.M. Briggs. 1993. Water relations and biomass responses to irrigation across a topographic gradient in tallgrass prairie. *13th North American Prairie Conference* pp. 215-219.

Knapp, A.K., J. M. Briggs, J.M. Blair, and C. L. Turner. 1998. Patterns and controls of aboveground net primary productivity in tallgrass prairie. In: Grassland Dynamics: Long-term Ecological Research in Tallgrass Prairie. Knapp, A.J., Briggs, J.M., Hartnett, D.C., Collins, S.L. (eds). Oxford University Press, New York. Pp: 193-221.

Knapp, A.K. and M.D. Smith. 2001. Variation among biomes in temporal dynamics of aboveground primary production. *Science* 291: 481-484.

Knapp, A. K., P. A. Fay, J. M. Blair, S. L. Collins, M. D. Smith, J. D. Carlisle, C. W. Harper, B. T. Danner BT, M. S. Lett, and J. K. MacCarron. 2002. Rainfall variability, carbon cycling and plant species diversity in a mesic grassland. *Science* 298: 2202-2205.

Knight, C.A., H. Vogel, J. Kroymann, A. Shumate, H. Witsenboer, and T. Mitchell-Olds. 2006. Expression profiling and local adaptation of *Boechera holboellii* populations for water use efficiency across a naturally occurring water stress gradient. *Molecular Ecology* 15: 1229-1237.

Lee C.R. and T. Mitchell-Olds. 2011. Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Molecular Ecology* 20: 4631-4642.

Leimu R. and M. Fischer. 2008. A meta-analysis of local adaptation in plants. *PLoS ONE* 3(12): e4010. doi:10.1371/journal.pone.0004010

Luikart, G., P.R. England, D. Tallmon, S. Jordan, and P. Taberlet. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics* 4: 981-994.

McKay, J. K., C. E. Christian, Harrison S, and K. J. Rice. 2005. "How local is local?" - A review of practical and conceptual issues in the genetics of restoration. *Restoration Ecology* 13: 432-440.

McKendrick, J.D., C.E. Owensby, and R.M. Hyde. 1975. Big bluestem and indiangrass vegetative reproduction and annual reserve carbohydrate and nitrogen cycles. *Agro-Ecosystems* 2: 75-93.

McMillan, C. 1956a. Ecotypic differentiation within four North American prairie grasses. II. Behavioral variation within transplanted community fractions. Ecotypic differentiation within four North American prairie grasses. II. Behavioral variation within transplanted community fractions. *American Journal of Botany* 52: 55-65.

McMillan, C. 1956b. Grassland community fractions from central North America under simulated climates. *American Journal of Botany* 52: 109-116.

McMillan, C. 1959. The role of ecotypic variation in the distribution of the central grassland of North America. *Ecological Monographs* 29: 285-308.

McMillan, C. 1969. Ecotypes and ecosystem function. *American Institute of Biological Sciences* 19(2): 131-134.

McNaughton S.J. Grazing lawns: animals in herds, plant form, and coevolution. *American Naturalist* 124: 863-886.

Mintenko, A.S., S.R. Smith, and D.J. Cattani. 2002. Turfgrass evaluation of native grasses for the Northern Great Plains Region. *Crop Science* 42: 2018-2024.

Montalvo A.M. and N.C. Ellstrand. 2001. Nonlocal transplantation and outbreeding depression in the subshrub *Lotus scoparius* (Fabaceae). *American Journal of Botany* 88: 258-269.

Moore, K.J., L.E. Moser, K.P. Vogel, S.S. Waller, B.E. Johnson, and J.F. Pedersen. 1991. Describing and quantifying growth stages of perennial forage grasses. *Agronomy – Faculty Publications* Paper 507. http://digitalcommons.unl.edu/agronomyfacpub/507

Mueller, U.G. and L.L. Wolfenbarger. 1999. AFLP genotyping and fingerprinting. *Tree* 14(10): 390-394.

National Oceanic & Atmospheric Administration (NOAA). U.S. Department of Commerce. 2012. "U.S. Drought Monitor". Updated July 12 2012.

Norrmann, G. A., C. L. Quarin, and K. H. Keeler. 1997. Evolutionary implications of meiotic chromosome behavior, reproductive biology, and hybridization in 6X and 9X cytotypes of *Andropogon gerardii* (Poaceae). *American Journal of Botany* 84:201-207.

Normann, G.A. and K.H. Keeler. 2003. Cytotypes of *Andropogon gerardii* Vitman (Poaceae): fertility and reproduction of aneuploids. *Botanical Journal of the Linnean Society* 141: 95-103.

Oleksyk, T.K, M.W. Smith, and S.J. O'Brien. 2010. Genome-wide scans for footprints of natural selection. *Philosophical Transactions of the Royal Society* 365: 185-205.

Owsley, C.M. 2000. USDA-NRCS Jimmy Carter Plant Materials Center, Americus, GA. Jun. 200. 16p.

Price, D., P. Salon, and M.D. Casler. 2012. Big bluestem gene pools in the Central and Northeastern United States. *Crop Science* 52: 189-200.

Ronikier M. 2002. The use of AFLP markers in conservation genetics—a case study on *Pulsatilla vernalis* in the Polish lowlands. *Cell Molecular Biology Letters* 7(2B): 677-684.

Rouse, M. N., A.A. Saleh, A. Seck, K.H. Keeler, S.E. Travers, S.H. Hulbert, and K.A. Garrett. 2011. Genomic resistance gene homolog diversity of the dominant tallgrass prairie species across the U.S. Great Plains precipitation gradient. *PLoS ONE* 6(4): e17641.

Sala, O.E., W.J. Parton, L.A. Joyce, and W.K. Lauenroth. 1988. Primary production of the central grassland region of the United States. *Ecology* 69: 40-45.

Samson, F. and F. Knopf. 1994. Prairie Conservation in North-America. *Bioscience* 44: 418-421.

Soil Conservation Service (SCS). 1999. Conservation Reservation Program. U.S. Department of Agriculture, Washington, D.C. *National Bulletin No.* 300-0-8.

Seastedt, T.R. and A.K. Knapp. 1993. Consequences of non-equilibrium resource availability across multiple time scales: the transient maxima hypothesis. *American Naturalist* 141: 621-633.

Selbo, S.M. and A.A. Snow. 2005. Flowering phenology and genetic similarity among local and recently introduced populations of *Andropogon gerardii* in Ohio. *Restoration Ecology* 13: 441-447.

Silletti, A.M. and A.K Knapp. 2002. Assessing the ecological equivalence of the grassland dominants *Andropogon gerardii* and *Sorghastrum nutans*. *Plant Ecology* 163: 15-22.

Smith, D.C., E.L. Nielsen, and H.L. Ahlgren. 1946. Variation in ecotypes of *Poa pratensis*. ***Botanical Gazette* 108: 143.**

Stebbins, G.L. 1971. *Chromosomal Evolution in Higher Plants* pp. 124-154.

Tompkins, R.D., W.C. Stringer, K. Richardson, E.A. Mikhailova, and W.C. Bridges Jr. 2010. Big bluestem (*Andropogon gerardii*; Poaceae) communities in the Carolinas: composition and ecological factors. *Rhodora* 112(952): 378-395.

Tompkins, R.D. 2011. An ecological, genetic, and reproductive study of big bluestem (*Andropogon gerardii*) populations in the Carolinas. Dissertation, the Graduate School of Clemson University, Clemson, SC.

Transeau, E. N. 1935. The prairie peninsula. *Ecology* 16:423-437.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Pelman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 23: 4407–4414.

Whitman, W. *Leaves of Grass*. Philadelphia: David McKay, 1891-92: 29-79. PS 3201 1891 Robarts Library.

World Resources Institute (WRI). 2001. "People and Ecosystems: the fraying web of life". http://www.wri.org/publication/world-resources-2000-2001-people-and-ecosystems-fraying-web-life

Wu, C.A., D.B. Lowry, L.I. Nutter, and J.H. Willis. Natural variation for drought-response traits in the *Mimulus guttatus* species complex. 2010. *Oecologia* 162: 22-23.

Figures and Tables

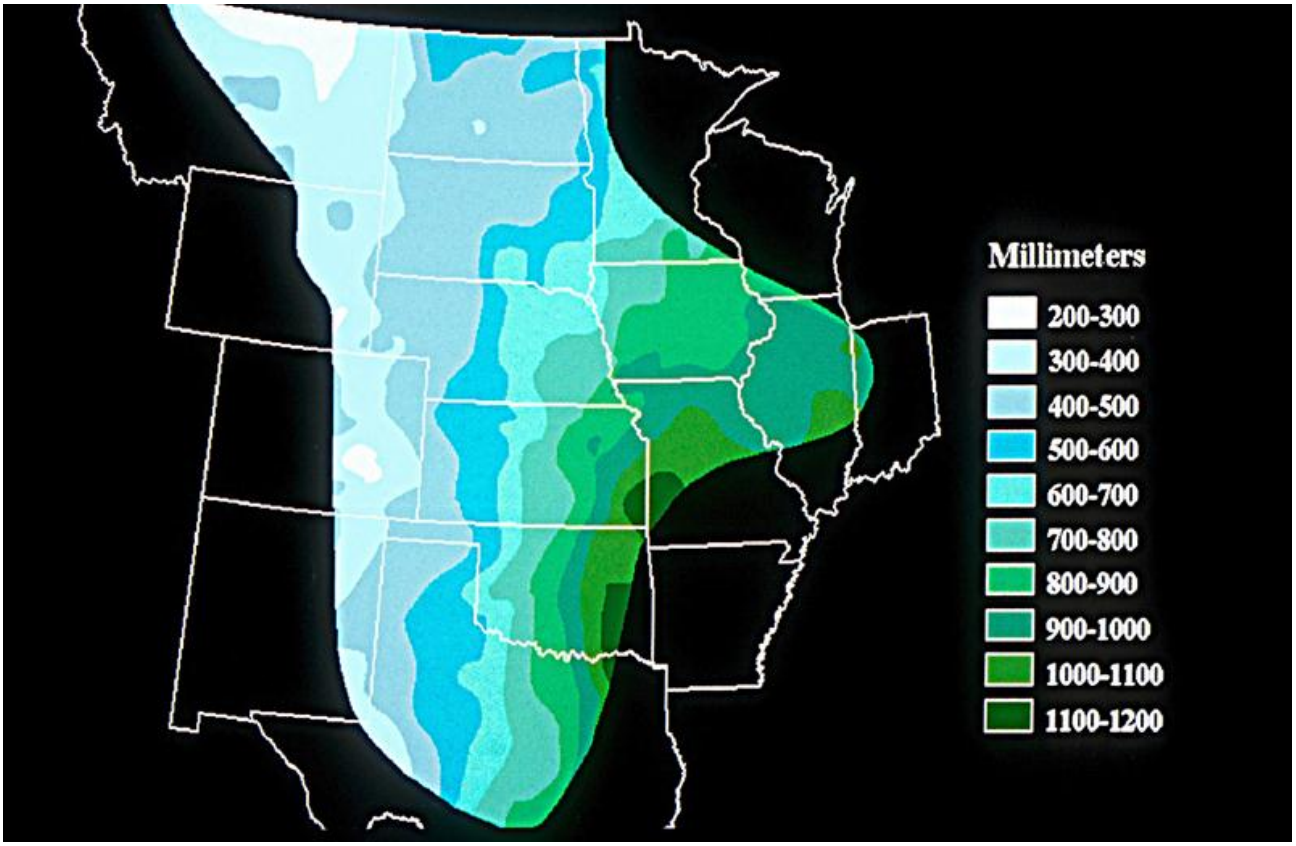

Figure 1.1 Sharp precipitation gradient in the United States U.S. Midwest.

Shading corresponds to mean annual rainfall totals in millimeters (greener shades = more rainfall, with dark green representing 1200 mm rainfall/year and light blue representing 400 mm rainfall/year). *Adapted from Burke et al.* 1997.

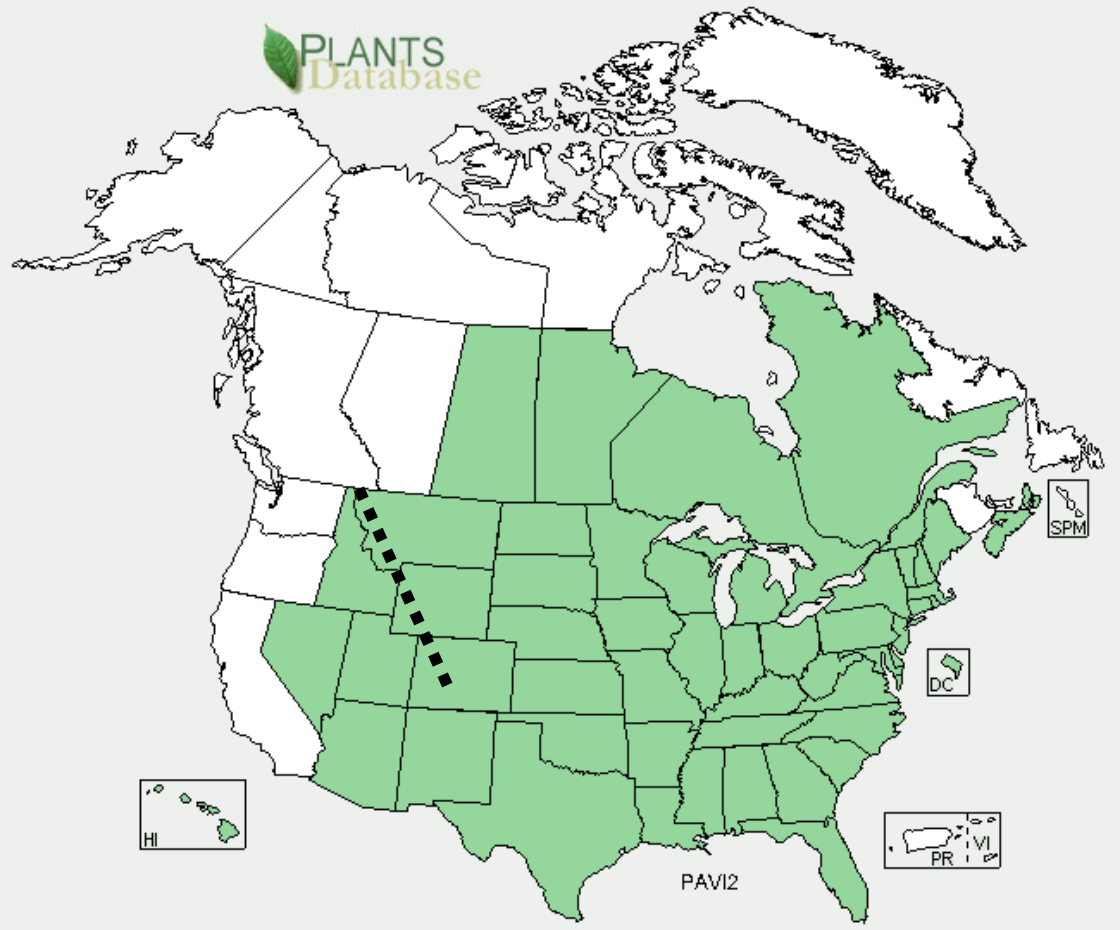

Figure 1.2 Wide distribution of big bluestem (*Andropogon gerardii* Vitman).

Regions of growth include Canada and the United States. Distribution of big bluestem is shaded in green although the current distribution is fragmented and primarily east of the Rocky Mountains (depicted by dotted line). *Adapted from USDA Plants Database.*

Figure 1.3 Ecotypes of big bluestem show morphological variation in common gardens. When grown as singly transplanted plants (above) and seeded gardens (not shown), ecotypes display phenotypic divergence. The Illinois reciprocal garden site is shown here in May 2010, at the start of the growing season. Plants here are grown in non-competitive environments and each prairie source is replicated 10X. Similar growth differences were observed between ecotypes when grown in similar greenhouse conditions with the dwarfed KS ecotypes having more drought-adaptive traits. *Pictured Left:* Illinois ecotype; *Right:* Eastern KS ecotype native to Manhattan, KS. Photo credit: Dr. Sara Baer.

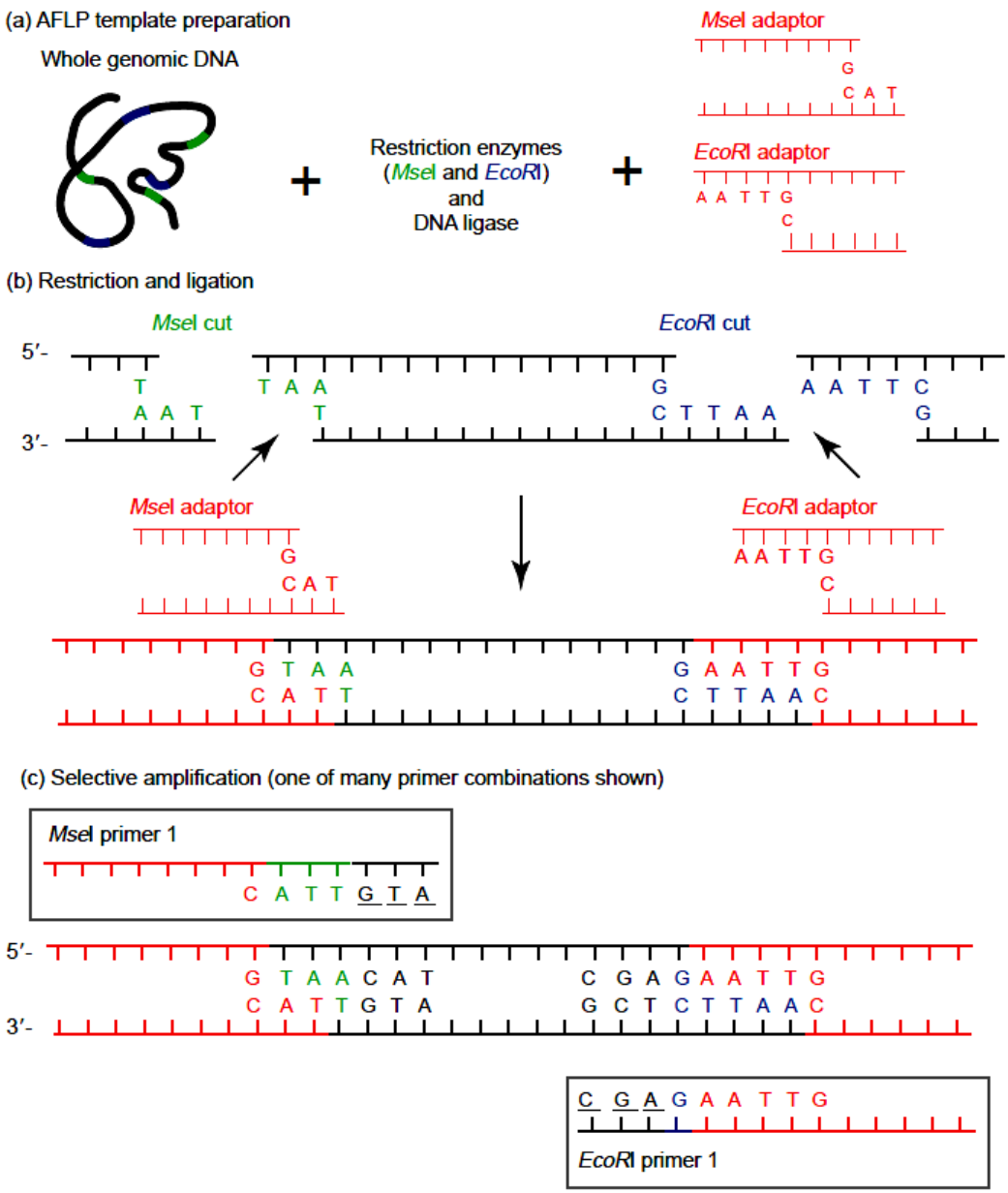

Figure 1.4 Amplified fragment length polymorphism (AFLP) fingerprinting method.

There are three primary steps shown: (a) Whole genome is sheared with restriction enzymes (b) adaptors of known sequence are annealed to restriction sites, and (c) fragments with adaptor sequence attached are PCR-amplified. Credit: *Mueller and Wolfenbarger (1999).*

Chapter 2 - Big Bluestem (*Andropogon gerardii* Vitman) Ecotypes Exhibit Genetic Divergence Along the U.S. Midwest Environmental Gradient

Introduction

Genetic variation makes adaptive evolution possible (Hoekstra 2006, Wu et al. 2010, Chan et al. 2010). Within the last five years, great interest in the ecological relevance of widespread genetic variation in natural populations has been informed by connecting traditional ecological studies to a population genetics and genomics framework (Manel et al. 2003, Luikart et al. 2003, Karrenberg and Widmer 2008, Holdenregger and Wagner 2008, Sork and Waits 2010, Hohenlohe et al. 2010, Lee and Mitchell-Olds 2011). Studying adaptive (non-neutral) genetic variation relative to the existing landscape and environment allow hypotheses to be formed regarding what important environmental and selective pressures exist. Using such landscape genetics approaches, many loci are sampled and F_{ST}-outliers are chosen as candidate loci that may be under selection (e.g., balancing or diversifying selection). A number of studies have attempted to utilize this approach on non-model, ecologically-important systems to answer the questions regarding what environmental variables are most important for species adaptation through natural selection (Gaggiotti et al. 2009, Nielsen et al. 2009, Parisod and Joost 2010, Hancock et al. 2011, Keller et al. 2011, Mattersdorfer et al. 2012). Using these methods, attempts have thus been made to separate the result of natural selection acting on traits with genetic bases from random genetic drift, and complex demography and life histories.

Big bluestem (*Andropogon gerardii* Vitman) is an ecologically-dominant C_4 prairie grass of the U.S. central grasslands. Big bluestem is an obligate outcrosser that is largely wind-pollinated. Its distribution is widespread, including habitat east of the United States Rockies, extending northward into Canada, east to the Carolinas, and south to Texas. However, the primary range of big bluestem in tallgrass prairie ecosystem spans a longitudinal environmental gradient in the U.S. Midwest and overlying much of the Great Plains region. This environmental gradient includes a sharp precipitation cline with a two-fold increase in rainfall moving west-east (58-116 cm rainfall/year). Species' distributed along environmental gradients may exhibit clinal genetic variation, in which local adaptation of the native ecotype may occur and/or species divergence be present, either at large or fine geographic scales (Knight et al. 2006, McKay et al. 2001, Lowry et al. 2009).

Most historic studies of big bluestem have taken an ecological and eco-physiology approach (Sala et al. 1988, Knapp et al. 1993, Gustafson 2004). While these studies have provided a plethora of information regarding water limitations, productivity, and eco-physiology distribution of species across its range, the utility of population genetics combined with a large sub-continental landscape/environmental approach has yet to be taken advantage of in this system. While Tompkins 2011, Rouse et al. 2011, and Price et al. 2012 all address big bluestem neutral genomic differentiation at somewhat larger scales, none of these studies probe both neutral and non-neutral diversity of described big bluestem ecotypes at great depth. Furthermore, such studies have not sampled along environmental gradients into the historic center of the tallgrass prairie ecosystem (Illinois) as we do here. Spatial genetic approaches are instrumental if we are to understand ecotypic differentiation that may potentially help manage

the small number of acres left of tallgrass prairie ecosystem in the United States. As of 1994, only 4% of historical tallgrass prairie remained in the United States (Samson and Knopf 1994), and the USDA Conservation Reserve program consists of 3 million acres of land in Kansas alone to help conserve prairie (SCS 1990). The use of non-local varieties in restoration may have disadvantages, including reduced success and exchange of maladapted genes to local ecotypes through gene flow (Hufford and Mazer 2003, McKay et al. 2005).

Our study benefits from a co-occurring two-year reciprocal garden experiment conducted across the U.S. Great Plains environmental gradient and extending eastward into Illinois, the historic center of the tallgrass prairie ecosystem. In seeded prairie communities planted across the gradient over two seasons, the Central Kansas ecotype had greater success (based on total plant cover) in its dry home environment than the non-native Eastern KS and Illinois ecotypes. This is indicative of potential local adaptation (Johnson et al. in prep). Ecotypes were also shown to differ in photosynthetic rate, with the Central Kansas ecotype having overall higher photosynthetic rates in a satellite field site located in Colby, Kansas at the most arid end of the reciprocal gardens.

Given an outcrossing system, we expected high within-prairie genetic diversity, with perhaps decreased diversity in more fragmented remnant prairies found in Illinois (Corbett 2004) where genetic drift processes would potentially play a stronger role. We also predicted that the genetic distance between big bluestem individuals would decrease with increasing geographic proximity (isolation by distance) as the majority of big bluestem pollen dispersal occurs over relatively

short distances (~30 meters in the literature though more recent measurements have not been reported to our knowledge).

The specific objectives of this research project were then to 1) probe neutral and non-neutral genetic diversity across the genome of big bluestem putative ecotypes using an AFLP genome scan approach, 2) determine whether ecotypes are genetically differentiated or have similar diversity levels, and 3) discover candidate loci associated with specific geographical and environmental factors which vary across three ecoregions along the U.S. Midwest environmental gradient (Central Kansas, Eastern Kansas, and Illinois). The last objective was included in order to formulate hypotheses regarding what selective pressures may be important in describing highly differentiated outlier loci across the three putative big bluestem ecotypes.

Methods

Seed Collection

Seeds were collected from eleven prairies across the U.S. Midwest environmental gradient including the Great Plains region. Prairies were partitioned into three main ecoregions (Central Kansas [in Hays, Kansas]; Eastern Kansas [in Manhattan, Kansas]; and Illinois [in Carbondale, Illinois]) with varying climates and environmental factors **(Table 2.1)**. All prairies were pristine, i.e., no prior restoration using big bluestem cultivars and unplowed. Prairies were occasionally burned and historically grazed. All of the prairies are currently protected state park land and/or research areas. Prairies ranged in size, and our sampling area covered was thus proportional to

individual prairie size (**Table 2.2**). Several dispersed collection points, at several time points during Fall 2008 were used. Seed from collection points was mixed together and subsampled to attain an unbiased representation of each prairie site.

Leaf Collection and DNA Isolation

Approximately 3.5 g of seeds was physically scarified by hand to increase the species' low germination rate and was densely sown in flats. Seedlings were well-watered and grown in ambient conditions in a greenhouse at temperatures between ~20-25˚C with a 12 hour day length. At two-months, seedlings were singly transplanted into 3 x 4 inch pots with Metro-Mix 510 potting soil and grown until 75-100 mg of young leaves/tillers could be collected per plant. For each prairie, 15-55 leaf samples were collected. Tissue was lyophilized in a freeze-drier (ModulyoD-115, Thermo Savant, Holbrook, New York) for three days and ground to a fine powder with 3.969 mm stainless steel beads (Abbott Ball Company Inc., Hartford, Connecticut) using a Mixer Mill 400 (Retsch Inc., Newton, PA) at 25-30 cycles/sec for ~15 min.

DNA was isolated using a CTAB method (Doyle & Doyle, 1987) and re-suspended in 50-100 ul Tris (10mM) + Triton X-100 (0.003125%) buffer (pH=8.0) overnight. Quality and quantity of the DNA was verified using a spectrophotometer (NanoDrop Technologies, Wilmington, DE), with OD requirements of 260/280=2.0 and 260/230 $\geq$ 1.80 for genotyping. Several randomly selected samples per 96-well plate were checked for degradation on a 0.8% agarose gel as well to verify that the extraction method used was consistent in producing non-degraded, high quality DNA for downstream reactions.

AFLP Genotyping

For all reactions, ~300 ng starting DNA was used; fine adjustments to DNA starting amounts across samples was not made as this was found to not be critical to fingerprint reproducibility nor band intensity values (data not shown). Our overall AFLP protocol was adapted extensively from Vos et al. (1995) and followed some aspects of Rouse et al. (2011) specific to big bluestem. The DNA restriction digestion and adaptor ligation steps of the AFLP protocol were combined and comprised of: ~300 ng genomic DNA (~25 ng/uL), 5 units of EcoRI HF (New England Biolabs, 0.25 uL) and 5 units of MseI (New England Biolabs, 0.5 uL), 100 units of T4 DNA ligase (New England Biolabs, 0.25 uL), 2 uL of 10X ligase buffer (New England Biolabs), 1.0 uL of each adaptor pair (5 pm/uL of EcoRI adaptors; 50 pm/uL of MseI adaptors, Integrated DNA Technologies), and 12 ul ddH_2O for a total reaction volume of 30 uL. The restriction-ligation mixture was incubated at room temperature overnight to ensure complete digestion. Restricted-ligated DNAs were diluted 10X in ddH_2O.

Pre-amplification reactions used primers complementary to the DNA restriction site and adaptor pair and also with an additional one-base pair overhang (EcoRI=5'-AGACTGCGTACCAATTC-A-3'and MseI=5'-GATGAGTCCTGAGTAA-C-5'). Individual pre-amplification PCRs consisted of a final volume of 40 uL per reaction well and included: 10 uL diluted restricted-ligated DNA template, 1.2 uL of each primer (10 uM), 6 uL 5X PCR buffer (Promega), 3 uL $MgCl_2$ (Promega, 25 mM), 0.64 uL dNTPs (5 mM), and 0.75 units of Go Taq Flexi DNA polymerase (Promega, 0.15 uL). PCR steps were as follows: 20˚C, 5 sec; ramp from 20˚C to

70˚C (0.2˚C/sec); 70˚C, 2 min.; 94˚C, 1 min.; 94˚C, 1 min..; 30 cycles of 94˚C, 30 sec; 56˚C, 1 min.; 72˚C, 1 min.; followed by 72˚C, 10 min.; 15˚C, 5 min. Pre-amplified template was diluted 20X with ddH_2O.

A selective PCR was performed using two primer sets with three additional bases (primer set 1: 5'GATGAGTCCTGAGTAA-CTG-3' + 5'HEX-AGACTGCGTACCAATTC-ACC-3'; Primer Set 2: 5'GATGAGTCCTGAGTAACGC-3' + 5'6FAM-AGACTGCGTACCAATTC-AAA-3'). We chose these two selective primer pairs after examining the quality of profiles resulting from eight different primer combinations (data not shown). Each selective PCR had a 20.5 uL final volume and consisted of: 1.5 uL diluted pre-amplified template, 1.62 uL M-side primer (10 uM, M-CTG or M-CGC), 1.62 uL fluorescently labeled E-side primer (10 uM, 5'-6HEX or 5'-6FAM), 4 uL 5X PCR buffer (Promega), 2 uL $MgCl_2$ (Promega, 25 mM), 0.8 uL dNTPs (5 mM), and 1 unit Go Taq Flexi DNA polymerase (Promega, 0.2 uL). The touchdown PCR cycle implemented was as follows: 95˚C for 2 min.; 13 cycles of 65˚C for 30 sec (-0.7C per cycle), 72˚C for 90 sec, and 94˚C for 30 sec; 23 cycles of 56˚C for 30 sec, 72˚C for 90 sec, and 94˚C for 30 sec; 72˚C for 10 min. and 15˚C for 5 min. To optimize the efficiency (overall band intensity) of primer set 2 (M-CGC + 6-FAM), a slight alteration in the PCR profile was made which included touchdown PCR starting at60˚C rather than 65˚C. The selective PCR was diluted 10X in ddH_2O.

A well-mixed solution of 9.5 uL formamide + 0.5 uL GeneScan-500 LIZ internal size standard (Applied Biosystems) was distributed throughout the well plate and 1.5 uL diluted selective template added to each well. Samples were loaded on an ABI Prism 3730 DNA Analyzer

(Applied Biosystems, Foster City, CA) at the DNA and Genotyping Facility, Kansas State University. A 50 cm capillary was used with an electrokinetic injection voltage of 1 kV applied for 10 sec. This lower injection voltage and shorter injection time was found to improve the resolution of AFLP bands of similar molecular weights. This method also improved the repeatability of longer fragments observed in profiles and prevented over-saturation of peak intensities.

Marker Scoring and Error Rate Estimation

Non-normalized profiles were scored using GeneMarker software version 1.97 (SoftGenetics LLC, State College, PA). AFLP panels were auto-created with a 1.0 base pair total width; afterward, bins were manually checked and adjusted to retain smoothly shaped peaks. Irreproducible peaks and extremely wide or irregularly shaped peaks were discarded. A 100 relative fluorescent unit (RFU) minimum peak height was used for peak scoring as this was reliably above the noise of negative controls (PCR reagents and water) included in the study (data not shown).

To verify the consistency of the AFLP technique, a set of four reference DNAs were included in each successive AFLP reaction and genotyping run. In addition, 2-3 independent restriction-ligations and PCRs were performed on one DNA sample per prairie and genotyped. The reference DNAs and replicate samples comprised 4% of total samples genotyped. The AFLP technical error rate estimation was calculated by the dividing total number of mismatched bands by total number of AFLP bands produced overall in the fingerprint (Bonin et al. 2004).

AFLP Data Analysis

Marker statistics, diversity analyses, and correlations between genetic distances and geographic distance between prairie sites were calculated in GENALEX version 6.41 (Peakall & Smouse 2006). We performed an analysis of molecular variance (AMOVA), pooling the data two ways: 1) by prairie site, with the starting null hypothesis that the eleven prairie sites could be considered together as one large, randomly mating population, and 2) at a larger scale based on Kansas as one region and Illinois as the second region, adjusting the null hypothesis so that each of the regions were considered as separate, randomly mating populations. The AMOVA consisted of 999 random permutations to test these two hypotheses. For isolation by distance tests, the null hypothesis of no isolation by distance across prairies was tested using a Mantel test comparing the Euclidean genetic distance and geographic distance matrices. Relatedness among all individuals was depicted using an un-rooted neighbor joining tree where pairwise genetic distance among individuals was calculating using the Dice coefficient of dissimilarity (Dice 1945). The Dice coefficient is analogous to the Nei and Li coefficient (Nei and Li 1979) and Sorensen coefficient (Sorensen 1948). The Dice coefficient of dissimilarity between two individuals (i_1 and i_2) is given by:

$$d(i_1, i_2) = 1 - D(i_1, i_2) = (b + c) / (2a + b + c),$$

where a=band presence in i_1 and i_2, b= band presence in i_1 only, c= band presence in i_2 only. Note that bands absent in both i_1 and i_2 do not affect the resulting distance matrix.

Spatial Population Structure using Bayesian Clustering

To determine the possibility of population structure within and between prairies and ecotypes, the AFLP marker data set was analyzed using STRUCTURE version 2.3.3 with 20,000 burn-in and 500,000 MCMC steps (Pritchard 2000). The possibility of admixture was allowed and the correlated allele frequency setting was selected. The correlated allele frequency setting is appropriate for finding differentiation between ecoregions such as in this study, where prairies belonging to an ecoregion are within a 50-mile radius of one another. Thus, our default settings were chosen to detect fine population structure. . The value of *K* was determined by taking into consideration both the point at which the mean estimated log likelihood did not increase any further and appeared to plateau and where delta K sharply increased (Evanno et al. 2005). STRUCTURE HARVESTER was used for the calculation of delta K following the Evanno method (Earl et al. 2012).

Detection of F_{ST}-Outlier Loci

Outlier loci were detected using BAYESCAN version 2.01 (Foll and Gaggiotti 2008) which handles dominant marker data. Data were analyzed in two ways: 1) by prairie site and 2) by ecoregion. Run parameters included 20 pilot runs of length 5,000, 50,000 data burn-in, a thinning interval of 10, and a sample size of 5,000. The prior odds for the neutral model was set to 10 and the inbreeding coefficient (F_{IS} prior) was allowed to vary between 0.0 and 1.0, where 1.0 represents complete inbreeding within the population. Data sets were run with all 387

polymorphic loci, and again with 325 marker loci (excluding highly monomorphic loci present in >90% individuals and minor alleles at <2% frequency across all individuals). The two models which are compared in BAYESCAN are a neutral model (M1) and a model including selection (M2). To choose between the two models using data set N, the Bayes factor (BF) or scale of evidence used for the model with selection (M2) is given by (Foll and Gaggiotti 2008)

$$BF= (P(N|M2) / P(N|M1).$$

BAYESCAN also considers that each locus may be under selection. In this case, the prior odds for the neutral model is set as P(M1/M2) and the posterior odds (PO) is used rather than the Bayes factor and is given by

$$PO = (P(M2|N)/P(M1|N) = BF*P(M2)/P(M1).$$

In the case of this work, the criterion for substantial evidence of selection and rejection of the null hypothesis was a result of $\log_{10}$ (BF) > 0.5 which is equivalent to log (PO) > 0.5. This criterion was used as it is represents a strength of at least substantial evidence for selection using Jeffrey's interpretation. In addition, log (PO) between 1.0 and 1.5 signifies strong evidence for selection, between 1.5 and 2.0 indicates very strong evidence for selection, and log(PO) equivalent to 2.0 or increasing toward infinity is decisive evidence for selection.

Relating Marker Allele Frequency to the Environmental Gradient

In addition to looking for candidate loci under possible selection, we performed a spatial analysis using multiple univariate logistic regression to find significant associations of prairie-specific environmental predictors with the AFLP markers (Joost et al. 2007, 2008). For all weather data, we used the National Oceanic and Atmospheric Administration database and historical records dating back to 1961. Information for each of the eleven prairies was gathered, or when necessary, weather data was taken from a nearby weather station. Eight geo-environmental predictors describing the environmental gradient and prairie location were entered in total. All prairie locations were entered using GPS coordinates. In addition to longitude and latitude coordinates, prairie elevation and the following weather data were also entered: mean annual precipitation in 2011, annual precipitation in 2011, mean annual diurnal temperature variation during the growing season since 1961 (calculated by averaging the daily maximum temperature in ˚C – minimum temperature in ˚C) and a temperature severity index (fraction of days > 35˚C since 1961) (**Table 2.3**). The mean annual precipitation in 2011 is useful as it is a recent measure of climate in the U.S. Midwest.

For the spatial analysis of environment-marker associations, the program MATSAM was used (Joost et al. 2007, 2008). The MATSAM program was run using a component in Matlab® which utilizes a generalized linear model (MacCullagh & Nelder 1989), where the number of models tested equals the product of the total number of markers and the total number of environmental parameters per collection site. For our purposes, significant correlations were defined as models with both McFadden and Efron pseudo R^2>0.3 as well as significant Wald and G-tests.

Results

AFLP Genotyping Results and Error Rate Estimate

In total, 387 polymorphic AFLP loci were identified (mean= 194 bands per primer set, σ=47). Approximately 330 bands were amplified per plant. Most markers were present at ≥25% frequency, with 8% of the data set represented by minor alleles (frequency <2%) (**Figure 2.1**). The overall error rate (number of mismatched bands per replicate sample/total number of bands per replicate profile) was 9.18% and thus within the error range typically reported for AFLP studies (Bonin et al. 2007, Holland et al. 2008, Arrigo et al. 2009, Rouse et al. 2011, Avolio et al. 2011, Price et al. 2012).

Genetic Differentiation and Population Structuring of Phenotypically-Distinct Ecotypes

Un-rooted neighbor-joining tree analyses demonstrated genotypic differentiation of the big bluestem ecotypes, with the greatest genetic similarity observed between the Central Kansas and Eastern Kansas ecotypes (**Figure 2.2**). The Illinois ecotype was split into several unique branches or clusters, largely separated from Kansas ecotypes. A number of tree branches included individuals from several prairie sites, indicating within-prairie genetic differentiation. Nei's pairwise genetic distance measurement showed genetic distances between prairies to be between 0.01-0.08, indicating mild genetic differentiation of prairies (**Table 2.4**). In principle coordinate analysis plots, a similar trend was found in the genetic relationships across individuals within prairies and ecotypes, with the two main clusters of data formed by Illinois ecotype and

Kansas (Central and Eastern) ecotypes; ecotypes were discernible in just one axis (38%), with two axes representing 61% of the present variation (axis 1=38%; axis 2=23%), with the two main clusters of data formed by Illinois ecotype and Kansas ecotypes (**Figure 2.3**). Moderate overlap between Illinois and Kansas ecotypes was observed.

In addition to genetic variation between the Kansas and Illinois of ecotypes, evidence for population structuring based on similar marker allele frequencies was found. The results from STRUCTURE supported the presence of *K=6* clusters through both the number of *K* clusters with highest mean likelihood over replicates and the calculation of the steepest increase in delta *K* (Evanno et al. 2005) (**Figure 2.4, 2.5** respectively). The model converged to this result during both short and long chain lengths (MCMC=10,000 steps and MCMC=500,000 steps, with a data burn-in of 10,000-20,000 each time). STRUCTURE results were consistent with PCA analysis. We observed allele frequencies to be highly similar within Kansas, with three main genetic groups present (**Figure 2.6**) and allele frequencies to be highly similar within prairie (**Figure 2.7**), with some admixed individuals noted.

It is likely that along the longitudinal environmental gradient of the U.S. Midwest, population subdivision or fragmentation of prairie populations exists, especially in more topographically disturbed prairie in Illinois. When considering an isolation-by-distance model (Wright 1942), the total pairwise Euclidean genetic distance between the 3 ecoregions (with each prairie site included in the analyses) increased significantly with increasing geographical distance. However, there was a very weak correlation between genetic distance and geographical distance (**Figure 2.8**, R^2=0.17, regression line fit: $p<0.001$).

Ecoregion-specific AFLP Markers

Several AFLP markers were found to be major alleles (frequency >2%) as well as ecoregion-specific (**Table 2.5**). Four major alleles were private to Illinois and six to Manhattan, Kansas in Eastern Kansas. No private alleles were found segregating only in the Hays, Kansas (Central Kansas) ecoregion at major frequencies. We did find evidence for markers shared between Central Kansas/Eastern Kansas ecoregions and Eastern Kansas/Illinois ecoregions, as might be expected within the same species.

High Within-prairie Genetic Diversity Exists

When considering the 11 prairies as individual sampling units, the analysis of molecular variance partitioned the most variation within-prairie (80%) vs. across-prairies (11%) (**Figure 2.9**, **Table 2.6**, $p<0.001$). The remaining total variation (9%) was partitioned between Kansas vs. Illinois as regions. We were interested in the comparison of Kansas to Illinois on a regional scale as most of the genetic differentiation as well as population structure present in our study occurred between these two regions. When the 3-4 prairie sites within each of the three geographically distinct ecoregions (Central KS, Eastern KS, and Illinois) were considered as one sampling unit, within-ecoregion variation was still considerable, ranging from 84%-92% in each ecoregion ($p<0.001$, data not shown). Thus, despite fragmentation, prairies sampled in Illinois still retain high amounts of genetic variation (92% of total variation).

F_{ST}-outlier Analysis Indicates Presence of Diversifying Selection

In addition to markers under neutral divergence in the genome, we found eleven F_{ST}-outlier loci, highly differentiated in comparison with an overall species F_{ST} of 0.1 determined in BAYESCAN and AFLP-SURV version 1.0 (Foll and Gaggiotti 2008, Vekemans 2002) (**Figure 2.10**). The 11 outlier loci were identified when entering the total set of 387 AFLP markers according to ecoregion/ecotype. When the data were entered in a similar way, but after discarding markers at the highest and lowest frequencies (present in >90% and <2% of individuals) to remove uninformative data from the data set (and thus to avoid the posterior odds being equivalent to prior odds), eight F_{ST} -outliers were observed. Each of these eleven outlier-loci and eight outlier-loci in both data sets were deemed "high outliers" (alpha greater than zero), and thus highly differentiated among ecoregion/ecotype (locus-specific F_{ST}=0.3-0.5, **Figure 2.11**), indicating possible diversifying selection acting on these particular genomic regions. In pairwise-comparisons between ecoregion / ecotype (data not shown), Eastern KS vs. Illinois ecotype had 5 highly differentiated markers (marker 221 [275]*, 255 [292], 219 [250], 200 [228], 203 [232]), Central Kansas vs. Illinois ecotype had 1 highly differentiated marker (marker 255 [292]), and Eastern Kansas vs. Central Kansas ecotype had 1 highly differentiated marker (marker 298 [not found in BAYESCAN when comparing all 3 ecoregions, marker 336]). Three outlier loci recovered using the reduced 325 AFLP marker data set were also recovered in the initial 387 marker data set ([marker 228, 256, and 292]). The other five outlier loci were newly found only after the data set was reduced. **[ ] = previous marker ID number when set of 387 loci were considered.*

Marker Frequencies Correlate with Environmental Predictors along Gradient

The prairie local environments and geographic coordinates within each ecoregion were used in spatial analysis methods. In total, 7 out of 8 environmental variables showed significant correlations with marker frequency patterns, with cumulative seasonal precipitation (from April-August during the growing season of big bluestem) in 2011 having no significant correlation with any AFLP markers (**Table 2.7**). We found 55 total significant models of marker frequency shifts related to the environmental variables (**Figure 2.12**). The most highly significant environmental factors in terms of marker correlations included prairie longitude, mean annual precipitation, temperature severity, and mean annual diurnal temperature variation. In totality, 14 marker loci response variables were associated with environmental predictors.

All eight marker loci (from the reduced 325 marker loci data set) for which we found evidence of diversifying selection in BAYESCAN were also detected in spatial analysis. Thus, 57% of loci detected in spatial analysis appear to be under potential diversifying selection and to have frequencies correlated with environmental variables. We found six total non-outlier marker loci whose frequency patterns correspond to six of the eight geo-environmental variables tested (prairie longitude, 2011 precipitation, temperature severity, diurnal temperature variation, prairie elevation, and prairie latitude), that thus are not associated with selection processes uncovered in BAYESCAN (**Figure 2.13**).

Discussion

Kansas and Illinois Phenotypic-divergent Ecotypes are Genetically Differentiated Along the Large-Scale Environmental Gradient

The use of a variety of metrics for analyzing binary AFLP data is useful for understanding the strength of the genotyping result (Kosman and Leonard 2005). For our study, we used the Dice coefficient to build the neighbor-joining tree, which attaches greater weight to shared AFLP bands (rather than shared 0s or marker absence), which is appropriate for a dominant marker where a null allele may be due to absence of the allele, as well as other factors such as point or insertion mutation in the endonuclease restriction site (Kosman and Leonard 2005). The un-rooted neighbor joining tree calculated using Dice coefficient of similarity gave support to the genetic differentiation of Kansas ecotypes from the Illinois ecotype in PCA analysis calculated using Euclidean distance data, further strengthening the argument for differentiation of Kansas ecotypes from the Illinois ecotype.

Interestingly, even though the prairie populations from Central Kansas and Illinois are ~1000 km apart geographically, our isolation by distance results indicate that geographic distance only explains a small part of the observed differentiation of the ecotypes (R^2=0.17, $p<0.001$). Central and Eastern Kansas ecotype source prairies are roughly 257 km apart from one another; however, Manhattan and Illinois sources are separated by 740 km, pointing toward geographic isolation as a possible mechanism of divergence. In addition to isolation by distance over this sub-continental scale, genetic differentiation of Illinois and Kansas ecotypes may be influenced population founder effects, genetic drift following population subdivision, and selection for

favorable loci (Boileau 1992, Ramsted 2003, Yeung 2010). A high correlation (R^2=0.47) between mean annual rainfall estimates (data not shown) and PCA axis 1 eigenvalue after transformation suggests that environmental factors may play a role in ecotype differentiation.

The lack of strong isolation by distance along large environmental scales may also be a factor of the sampling scheme. Ehrich and Stensoth (2003) found that with increasing distance of transects sampled, genetic differentiation among *Lemmus sibiricus* decreased. The lack of correlation between geographic and genetic distance in big bluestem natural populations is also in line with the Price et al. (2012) study that covered large geographic and environmental diversity. Avolio et al. (2011), in an examination of fine spatial population structuring of big bluestem genotypes, found that at ecologically-relevant scales (neighborhood scale of 1 m^2), high amounts of genotypic diversity and genetic structuring are obvious, with higher frequencies of certain genotypes. The highest frequency genotype is typically of selective advantage (Stuefer et al. 2009) and at fine population scales competitive advantage may play a role in propagation of certain genotypes. This may explain that sampling at finer scales and along transects may recover stronger isolation by distance mechanism of population structuring.

The observed genetic differentiation between Kansas and Illinois ecotypes is consistent with the phenotypic divergence between ecotypes; when planted in reciprocal seeded gardens along with natural competitors across the environmental gradient, the Central Kansas ecotype was always observed to significantly outcompete in its native environment as well as outside the species' range in a drier environment (Johnson et al. in prep). Allele frequency divergence of the Illinois

ecotype source prairies may suggest maladaptation may occur when the Illinois ecotype is faced with drier environments than its local environment.

Population Structure Exists Despite High Within-prairie Diversity

The amount of genetic variation found within designated natural populations, versus across populations, is informative to population processes and spatial genetic differentiation. When genetic variation was partitioned within-prairies, across-prairies, and across-ecoregions, the highest genetic variation (80%, $p<0.001$) was observed within-prairie. This was confirmed by previous studies of big bluestem genetic diversity levels in remnant prairies in Wisconsin and the Northeast U.S. and Midwest as well as Arkansas remnant prairie [Gustafson et al. 1999 (89% within-prairie diversity), Price et al. 2012 (86% within-prairie diversity), respectively]. Despite differences in marker types used in the Gustafson et al. study (RAPD; 1999) and our study (AFLP), we were able to detect consistently high levels of within-population genetic diversity. The Price et al. (2012) study using AFLP markers was also consistent with levels of within-population diversity found in our study, with 86% of the diversity they found sequestered within-prairies. Price et al. (2012) also showed that natural big bluestem prairie sources from Wisconsin and the Northeast United States form three distinct genetic groups; these genetic groups also had some genetic overlap present between prairies. When they partitioned the total amount of variation into among-prairie variation within each of the three ecoregions, however, it was determined that greater genetic variation is retained among-prairies from certain ecoregions. In contrast, we did not detect differences in the level of genetic variation among prairies partitioned into each of the three ecoregions we include in this study. This suggests that within each

ecoregion a number of diverse individuals exist which maintain the high within-prairie genetic diversity.

High within-prairie genetic diversity was expected for several reasons. In the AFLP data set, several highly diverse individuals were detected, as highlighted by STRUCTURE results (admixed individuals). Furthermore, big bluestem is known to be highly self-incompatible, with viable seed production following selfing events either low or completely absent (Normann et al. 1997, Owsley 2003, Tompkins et al. 2011). Obligate outcrossing across organisms has also traditionally resulted in increased genetic variation (Bomblies et al. 2010, Price et al. 2012, Gustafson et al. 1999). Furthermore, the complex polyploid genome of big bluestem allows for higher genetic variability to be present. However, Rouse et al. (2011) found that AFLP dissimilarity did not increase with known increases in ploidy level. Furthermore, we have tested ploidy level within the Manhattan, Kansas ecotype source prairies as well as in field sites and found that one cytotype predominates (6X) (Johnson and Gaffney, unpublished data). Big bluestem is an autopolyploid, however, so it is predicted that increased ploidy level would result in an increase in chromosome number but not necessarily in number of allelic variants. The opposite may be expected for allopolyploids unlike big bluestem. Thus, we suggest that the high genetic diversity can be more attributed to the outcrossing nature of big bluestem rather than genome size.

Interestingly, population genetic structure exists across big bluestem ecotypes despite high within-prairie diversity. A number of individuals are admixed, sharing similar genetic identities with samples across regions, but the majority of individuals from Illinois or Kansas share similar

allele frequencies. The separation of clusters using distance-based methods into two visible clusters (cluster 1 = Central and Eastern Kansas; cluster 2 = Illinois) further support Bayesian methods and serve to strengthen these results. When structure is visualized at the prairie-level, considerable similarity is found within-prairies; this decrease in admixed individuals within-prairies despite large amount of genetic variability found within-prairies can be caused by biased sampling of genetic clones (genets) of big bluestem. Average clonal size in big bluestem can be up to 3 m^2 (Jurik and Kliebenstein 1999), and potentially a factor even in dispersed sampling and pooled seed collections like ours. However, it may also be a factor of decreased pollen dispersal outside of prairie sites or landscape fragmentation.

Evidence for Diversifying Selection Linked to Environmental Predictor Variables

The eight loci under potential diversifying selection were all correlated with environmental predictor variables. The identification of highly differentiated loci possibly linked to the environment is critical to understanding important environmental variables and to understanding ecotypic differentiation. In this study, we sought to determine how selection, rather than just neutral processes, plays a role in shaping ecotypic differentiation within big bluestem by making use of the widespread occurrence of big bluestem along sharp environmental clines in the grassland ecosystem. In contrast to markers detected under diversifying selection, we did not detect any markers in BAYESCAN that are under balancing or purifying selection, suggesting that these regions of the genome characterized by the AFLP are in regions diverging across ecotypes. Another possibility is that with the low overall genetic differentiation of prairies, we did not have the power to detect markers under potential balancing selection. The presence of diversifying selection would suggest that the ecotypes are possibly influenced by geo-

environmental pressures along the environmental gradient that have led Kansas ecotypes to be genetically differentiated from the Illinois ecotype, despite evidence for gene flow still being maintained between these ecoregions (based on STRUCTURE analyses, see Laurent 2003 for another such example). The conservative nature of Bonferroni correction allows for only strong associations between marker frequency and environmental variables to be detected and thus these results can be considered robust (Foll and Gaggiotti 2008).

Based on spatial analysis methods, no markers were correlated with cumulative seasonal rainfall amounts in 2011 while 10 markers were significantly correlated with 2011 annual precipitation. We expected that mean annual precipitation would play a greater role in divergence of the big bluestem ecotypes given drought-related phenotypic differences within the ecotypes (Johnson et al. in prep) and our finding that PCA eigenvalue 1 was significantly correlated with mean annual rainfall amounts across the three ecoregions (R^2=0.48, data not shown). However, temperature variation also was largely associated with AFLP markers. Temperature variables also may exert selective pressures on populations of big bluestem as sites studied included western regions with more frequent heat stress. Temperature variables (long-term mean annual diurnal variation and fraction of days over 35˚C) were associated with a total of 17 marker loci. Interestingly, two markers (marker 269 and marker 301 in the 387 loci data set) were associated with temperature but not precipitation variables. This suggests that our marker density is high enough to have covered regions of the genome that may be influenced by distinct environmental pressures. All outlier loci identified to be under potential diversifying selection across ecotypes were also correlated with at least one geo-environmental factor, suggesting we have successfully probed both neutral and non-neutral diversity across the big bluestem genome.

Genetic Structuring Implications for Restoring the Dwindling Tallgrass Prairie Ecosystem

The loss of genetic diversity, habitat fragmentation, environmental and demographic changes, and inbreeding depression all contribute to the risk of species' loss (Frankham 2003, Rouse et al. 2011). The small percentage of tallgrass prairie ecosystem left in the U.S. Midwest occurs along the longitudinal precipitation gradient studied here. The genetic information we have derived along this gradient are thus relevant to conservation and restoration practices in the U.S. Midwest occurring along this environmental gradient. While we cannot make broad recommendations at this point in the study, one point of concern for future restoration of tallgrass prairie lies in evidence for marker allele frequency divergence across Illinois and Kansas ecoregions. Allele frequency divergence between Illinois ecotype and Kansas ecotypes has the potential to result in genetic swamping or loss of local adaptation if mixed plantings of these ecotypes were planted along the environmental gradient. Field studies complemented with genetic analyses on these particular prairie gene pools are necessary to make any further conclusions or land management recommendations, however.

Concluding Remarks

We have shown that there is indeed genetic differentiation of the three big bluestem ecotypes (in particular Central and Eastern Kansas ecotypes from Illinois ecotypes) which are dispersed naturally across the United States Midwest environmental cline. The potential to utilize native ecotypes for restoration practices is useful given the high genetic diversity discovered in each of the 11 prairies populations/sites in this study; however, increased genetic similarity among a

large number of individuals from the same prairie indicates that planting mixed populations stands may provide the greatest buffer against future climate change due to maintaining biodiversity. Another benefit of planting mixed ecotypes is to maintain diversity over landscapes which may include small-scale habitat variation or pressures. One might expect planting mixed seed from many populations of big bluestem would thus potentially avoid genetic maladaptation to given regional climates; however the genetic divergence of various gene pools from Kansas vs. Illinois shown here may result in potential genetic swamping. The divergence of allele frequencies across Kansas into Illinois suggests that there is reduced gene flow between these sites, despite weak detection of isolation by distance along the longitudinal environmental gradient. The lack of isolation by geographical distance yet allele frequency divergence can be explained by other isolating factors (reduced population size, founder effects, and habitat-specific environmental pressures) playing a larger role in the divergence of Kansas and Illinois ecotypes or a sampling bias. Interestingly, this weak relationship between geographic distance and genetic distance may suggest environmental adaptation is more significant than geographic isolating factors among Kansas and Illinois ecotypes (as found in Lee and Mitchell-Olds 2011 in *Boechera stricta* natural populations).

A weakness in past studies of genetic diversity in big bluestem and other studies is that genetic markers are studied in isolation without any tests for local adaptation of ecotypes across environmental gradients (Moncado et al. 2005). A great strength of this research is that the prairie sources investigated here are also included in a long-term reciprocal garden test for ecotypic adaptation. Thus, better conclusions can be made over years of the study to inform land management and tallgrass prairie restoration. We would encourage more studies of sequence-

based genetic markers (RAD tags) co-performed with reciprocal garden or transplant tests for making recommendations for restoration.

References

Arrigo, N., J. W. Tuszynski, D. Ehrich, T. Gerdes, and N. Alvarez. 2009. Evaluating the impact of scoring parameters on the structure of intra-specific genetic variation using RawGeno, an R package for automating AFLP scoring. *BMC Bioinformatics* 10:1–14.

Avolio, M.L., C.C. Chang, and M.D. Smith. 2011. Assessing fine-scale genotypic structure of a dominant species in native grasslands. *American Midland Naturalist* 165: 211–224.

Boileau, M.G., P. Hebert, and S.S. Schwartz. 1992. Non-equilibrium gene frequency divergence: persistent founder effects in natural populations. *Journal of Evolutionary Biology* 5: 25-39.

Bomblies, K., L. Yant, R.A. Laitinen, S. Kim, J.D. Hollister, N. Warthmann, J. Fitz, and D. Weigel. 2010. Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of *Arabidopsis thaliana*. *PLoS Genetics* 6(3): e1000890. doi:10.1371/journal.pgen.1000890

Bonin, A., E. Bellemain, P.B. Eidesen, F. Pompanon, C. Brochmann, and P. Taberlet. 2004. How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* 13: 3261-3273.

Bonin, A., D. Ehrich, and S. Manel. 2007. Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Molecular Ecology* 16(18): 3737-58.

Cremieux, L., A. Bischoff, H. Muller-Scharer, and T. Steinger. 2009. Gene flow from foreign provenances into local plant populations: fitness consequences and implications for biodiversity restoration. *American Journal of Botany* 97(1): 94-100.

Dice L.R. 1945. Measures of the amount of ecologic association between species. *Ecology* 26: 297–302.

Doyle J.J., J.L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19: 11-15.

Earl, D.A. and B.M. vonHoldt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4(2): 359-361.

Ehrich, D. and N.C. Stenseth. 2001. Genetic structure of Siberian lemmings (*Lemmus sibiricus*) in a continuous habitat: large patches rather than isolation by distance. *Heredity* 86: 716-730.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611-2620.

Foll, M. and O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* 180: 977–993.

Frankham, R. 2003. Genetics and conservation biology. *C.R. Biologies* 326: S22-S29.

Gaggiotti, O.E., D. Bekkevold, H.B.H. Jorgensen, M. Foll, G.R. Carvalho, C. Andre, and D.E. Ruzzante. 2009. Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: Atlantic herring as a case study. *Evolution* 63(11): 2939-2951.

Gustafson, D.J., D.J. Gibson, D.L. Nickrent. 1999. Random amplified polymorphic DNA variation among remnant big bluestem (*Andropogon gerardii* Vitman) populations from Arkansas' Grant Prairie. *Molecular Ecology* 8: 1693-1701.

Gustafson, D.J., D.J. Gibson, and D.L. Nickrent. 2004. Competitive relationships of *Andropogon gerardii* (Big Bluestem) from remnant and restored native populations and select cultivated varieties. *Functional Ecology* 18: 451-457.

Hancock, A.M., D.B. Witonsky, G. Alkorta-Aranburu, C.M. Beall, A. Gebremedhin, et al. 2011. Adaptations to climate-mediated selective pressures in humans. *PLoS Genetics* 7(4): e1001375. doi:10.1371/journal.pgen.1001375

Hohenlohe, P.A. S. Bassham, P.D. Etter, N. Stiffler, E.A. Johnson, W.A. Cresko. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* 6(2): e1000862. doi:10.1371/journal.pgen.1000862

Holland, B.R., A.C. Clarke, and H.M. Meudt. 2008. Optimizing automated AFLP scoring parameters to improve phylogenetic resolution. *Systematic Biology* 57(3): 347-366.

Holderegger R. and H.H. Wagner. 2008. Landscape genetics. *Bioscience* 58(3): 199-207.

Hufford, K.M. and S.J. Mazer. 2003. Plant ecotypes: genetic differentiation in the age of ecological restoration. *Trends in Ecology and Evolution* 18:147-155

IPCC Report. 2007. Climate change 2007: Physical Science Basis. Approved at 10th session of Working Group I. at the IPCC. IPCC, Paris, France, February 2007.

Johnson et al. in prep. Differential establishment, cover, and gas exchange among ecotypes of a dominant grass big bluestem *Andropogon gerardii* in reciprocal gardens along a Great Plains climatic gradient.

Jones, M.D. and L.C. Newell. 1946. Pollination cycles and pollen dispersal in relation to grass improvement. *Nebraska Agricultural Experimental Station Research Bulletin* 148

Joost, S., A. Bonin, M.W. Bruford, L. Despres, C. Conord, G. Erhardt, P. Taberlet. 2007. A spatial genetic analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology* 16(18): 3955-3969.

Joost, S., M. Kalbermatten, A. Bonin. 2008. Spatial analysis method (SAM): a software tool for combining molecular and environmental data to identify candidate loci for selection. *Molecular Ecology Research* 8: 957-960.

Keeler, K. H.. and G. A. Davis. 1999. Comparison Of Common Cytotypes of *Andropogon gerardii* (Andropogoneae, Poaceae). *American Journal of Botany* 86(7):974 –979.

Keller, I., A. Taverna, and O. Seehausen. 2011. Evidence of neutral and adaptive genetic divergence between European trout populations sampled along altitudinal gradients. *Molecular Ecology* 20: 1888-1904.

Knapp, A.K., J.K. Koelliker, J.T. Fahnestock, and J.M. Briggs. 1993. Water relations and biomass responses to irrigation across a topographic gradient in tallgrass prairie. *13th North American Prairie Conference* pp. 215-219.

Knight, C.A., H. Vogel, J. Kroymann, A. Shumate, H. Witsenboer, and T. Mitchell-Olds. 2006. Expression profiling and local adaptation of *Boechera holboellii* populations for water use efficiency across a naturally occurring water stress gradient. *Molecular Ecology* 15: 1229-1237.

Kosman, E. and K.J. Leonard. 2005. Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species. *Molecular Ecology* 14: 415-424.

Laurent, R., M. Legault, and L. Bernatchez. 2003. Divergent selection maintains adaptive differentiation despite high gene flow between sympatric rainbow smelt ecotypes (*Osmerus mordax* Mitchill). *Molecular Ecology* 12: 315-330.

Lee C.R. and T. Mitchell-Olds. 2011. Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Molecular Ecology* 20: 4631-4642.

Lowry, D.B., M.C. Hall, D.E. Salt, and J.H. Willis. 2009. Genetic and physiological basis of adaptive salt tolerance divergence between coastal and inland *Mimulus guttatus. New Phytologist* 183: 776-788.

Luikart, G., P.R. England, D. Tallmon, S. Jordan, and P. Taberlet. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics* 4: 981-994.

Manel, S., M.K. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution* 18(4): 189-197.

Mattersdorfer, K. S. Koblmuller, and K.M. SEFC. 2012. AFLP genome scans suggest divergent selection on colour patterning in allopatric colour morphs of a cichlid fish. *Molecular Ecology* 19: 3806-3823.

McCullagh, P. and J.A. Nelder. 1989. Generalized linear models. Chapman and Hall: London.

McKay, J.K., J.G. Bishop, J. Lin, J.H. Richards, A. Sala, and T. Mitchell-Olds. 2001. Local adaptation across a climatic gradient despite small effective population size in the rare sapphire rockcress. *Proceedings of the Royal Society of London, Biological Sciences* 268: 1715-1721.

McKay, J. K., C. E. Christian, Harrison S, and K. J. Rice. 2005. "How local is local?" - A review of practical and conceptual issues in the genetics of restoration. *Restoration Ecology* 13: 432-440.

Moncada, K.M., N.J. Ehlke, G.J. Muehlbauer, C.C. Sheaffer, D.L. Wyse, and L.R. DeHaan. *Crop Science* 47: 2379-2389.

National Oceanic & Atmospheric Administration (NOAA). U.S. Department of Commerce. 2012. "U.S. Drought Monitor". Updated July 12 2012.

Nei , M. and W.H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences* 76: 5269-5273.

Nielsen, E., J. Hemmer-Hansen, N.A. Poulsen, V. Loeschcke, T. Moen, T. Johansen, C. Mittelholzer, G. Taranger, R. Ogden, and G.R. Carvalho. Genomics signatures of local directional selection in a high gene flor marine organism; the Atlantic cod (*Gadus morhua*). 2009. *BMC Evolutionary Biology* 9: 276

Norrmann, G. A., C. L. Quarin, and K. H. Keeler. 1997. Evolutionary implications of meiotic chromosome behavior, reproductive biology, and hybridization in 6X and 9X cytotypes of *Andropogon gerardii* (Poaceae). *American Journal of Botany* 84:201-207.

Owsley, C.M. 2000. USDA-NRCS Jimmy Carter Plant Materials Center, Americus, GA. Jun. 200. 16p.

Parisod, C. and S. Joost. 2010. Divergent selection in trailing- versus leading-edge populations of *Biscutella laevigata*. *Annals of Botany* 105(4): 655-660.

Peakall, R. and P.E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288-295.

Price, D., P. Salon, and M.D. Casler. 2012. Big bluestem gene pools in the Central and Northeastern United States. *Crop Science* 52: 189-200.

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

Ramstad, K.M. C.A. Woody, G.K. Sage, and F.W. Allendorf. 2004. Founding events influence genetic population structure of sockeye salmon (*Oncorhynchus nerka*) in Lake Clark, Alaska. *Molecular Ecology* 13: 277-290.

Rosenberg N.A. 2004. Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes* 4: 137-138.

Rouse, M. N., A.A. Saleh, A. Seck, K.H. Keeler, S.E. Travers, S.H. Hulbert, and K.A. Garrett. 2011. Genomic resistance gene homolog diversity of the dominant tallgrass prairie species across the U.S. Great Plains precipitation gradient. *PLoS ONE* 6(4): e17641.

Samson, F. and F. Knopf. 1994. Prairie Conservation in North-America. *Bioscience* 44: 418-421.

Soil Conservation Service (SCS). 1999. Conservation Reservation Program. U.S. Department of Agriculture, Washington, D.C. *National Bulletin No.* 300-0-8.

Sorensen, T. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskabs Biologiske Skrifter* 5: 1-34.

Sork, V.L., F.W. Davis, R. Westfall, A. Flint, M. Ikegami, H. Wang, and D. Grivet. 2010. Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Nee) in the face of climate change. *Molecular Ecology* 19: 3806-3823.

Stearns, S.C. 1976. Life history tactics: a review of the ideas. *Quarterly Review of Biology* 51: 3-47.

Stuefer, J.F., N.P.R. Anten, H. de Kroon, J.L. Peters, A. Smit-Tiekstra, P.J. Vermeulen, and H.J. During. 2009. Genotypic selection shapes patterns of within-species diversity in experimental plant populations. *Journal of Ecology* 97: 1020-1027.

Tompkins, R.D., W.C. Stringer, and W.C. Bridges Jr. 2011. An outcrossing reciprocity study between remnant big bluestem (*Andropogon gerardii*) populations in the Carolinas. *Ecological Restoration* 29: 339-345.

Vekemans, X. 2002. AFLP-SURV version 1.0. Distributed by the author. Laboratoire de Génétique et Ecologie Végétale, Université Libre de Bruxelles, Belgium.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Pelman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 23: 4407–4414.

Wright, S. 1942. Isolation by distance. *Genetics* 28: 114-138.

Yeung, C.K.L. P. Tsai, R.T. Chesser, R. Lin, C. Yao, X. Tian, and S. Li. 2010. Testing founder effect speciation: divergence population genetics of the spoonbills , C.K.L. P. Tsai, R.T. Chesser, R. Lin, C. Yao, X. Tian, and S. Li. 2010. Testing founder effect speciation: divergence population genetics of the spoonbills *Platalea regia* and *P. minor* (Threskiornithidae, Aves). *Molecular Biology and Evolution* 28(1): 473-482.

Figures and Tables

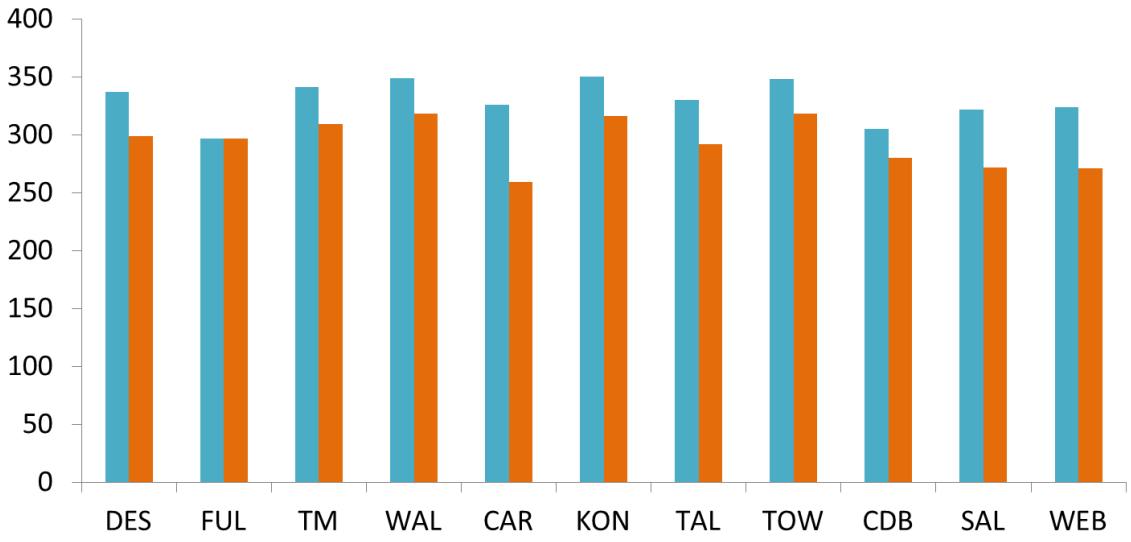

Figure 2.1 AFLP band results across natural prairie sources.

Prairie site is listed on the x-axis and total number of bands on the y-axis. A similar number of total bands (blue) were recovered across two primers for each of the eleven source populations. The majority of total AFLP bands generated per prairie site were at frequencies ≥ 5% (orange). DES, FUL, TM, WAL=Illinois ecotype; CAR, KON, TAL, TOW= Eastern Kansas ecotype; CDB, SAL, WEB = Central Kansas ecotype

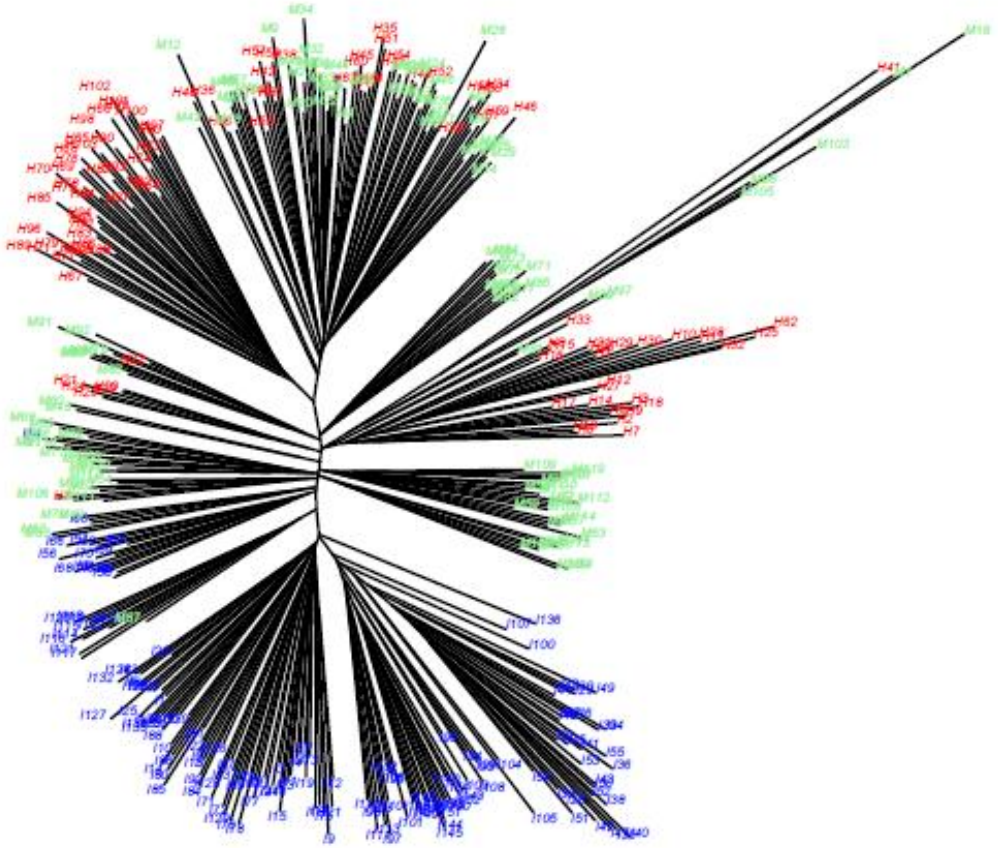

Figure 2.2 Un-rooted neighbor-joining tree of genetic dissimilarity between individuals.

The tree was built in R using the Dice coefficient of dissimilarity (Dice 1945). Individuals are color-coded according to ecotype/ecoregion: Blue=Illinois, Green=Eastern Kansas, Red= Central Kansas.

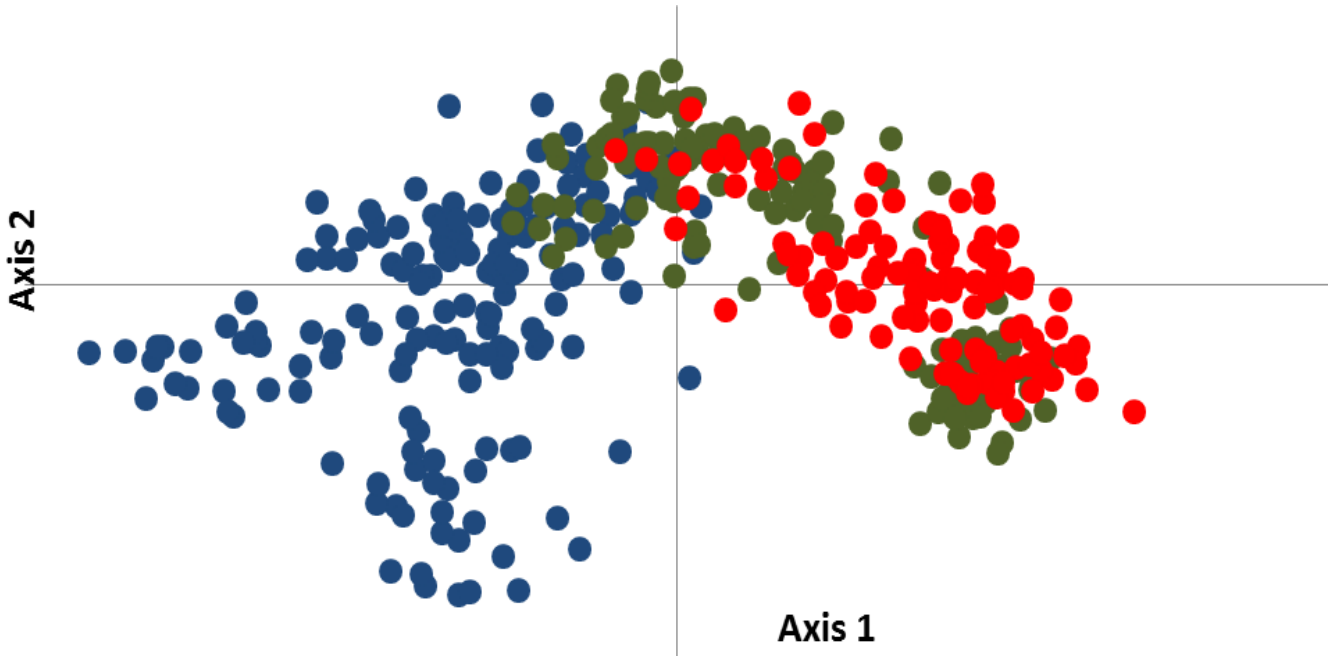

Figure 2.3 Genetic principle coordinate analysis between individuals.

Principle coordinate analysis based on the presence/absence of 387 AFLP marker loci across 378 big bluestem DNA samples. Individuals are color-coded according to ecoregion/ecotype: Blue=Central Kansas, Green= Eastern Kansas, Red= Illinois ecotype. Kansas ecotypes (red) clearly separate from Illinois ecotype (blue) within 2 PCA axes (axis 1 explaining 38% of the variation and axis 2 explaining 23% of the variation, total variation captured=61%).

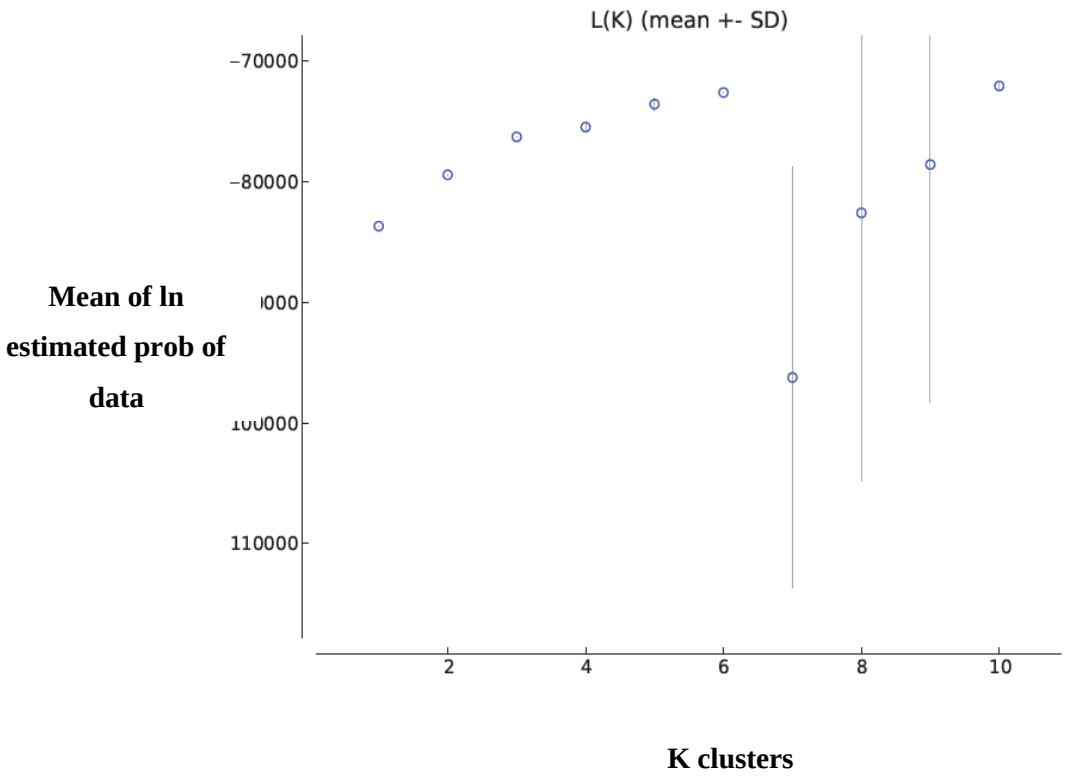

Figure 2.4 Plot of ln estimated probability of data for each K.

Results are from STRCUTURE run with burn-in=20,000 and MCMC steps=500,000. The plateau of highest likelihood at K=6 confirms delta K calculations that K=6 is indeed the most likely solution. After K=6, the variation across 10 iterations of each K increases.

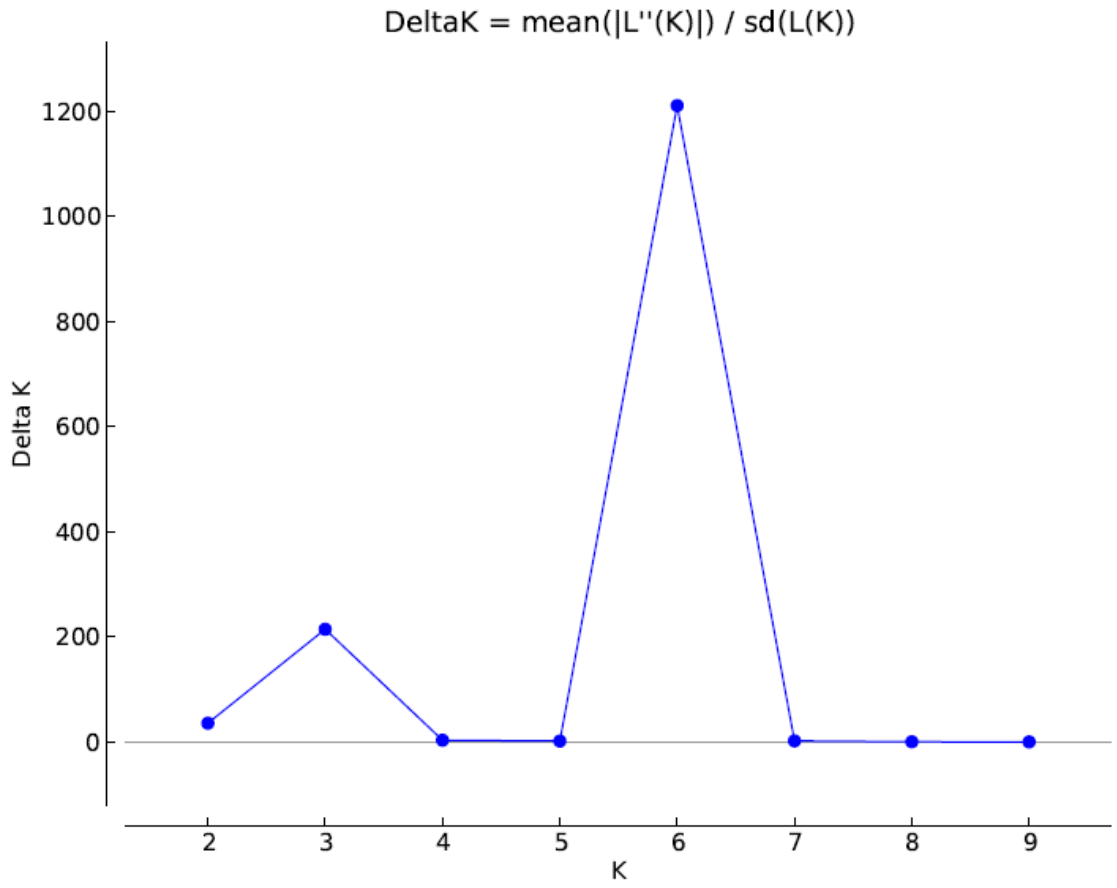

Figure 2.5 Delta K calculation plotting change in Delta K with increasing *K* clusters.

STRUCTURE analysis (20,000 burn-in and 500,000 MCMC steps) across 10 runs of *K*=1-10 was performed prior to calculation. The sharp increase at *K=6* in delta K suggests the most likely solution to be K=6 clusters. (Evanno et al. 2005).

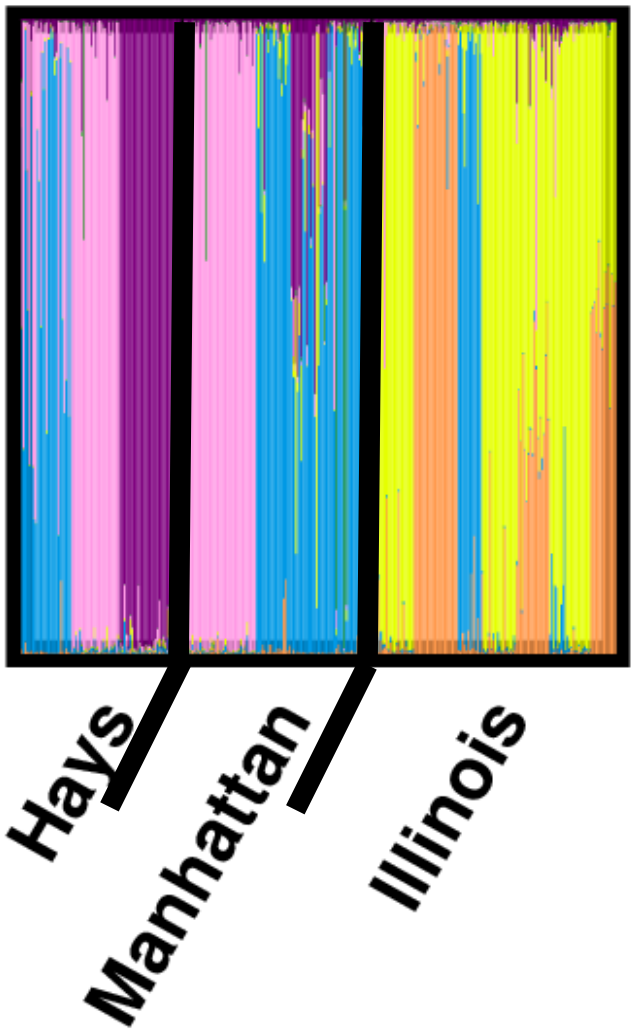

Figure 2.6 STRUCTURE barplot organized by location across environmental gradient.

The most likely solution, *K=6* (six genetic groups) is shown. Dark black lines separate individual locations (regions) across the environmental gradient. Each color indicates one genetic group to which individual (represented by bars or columns) belong. Hays and Manhattan individuals consist of individuals belonging to the same genetic cluster, with admixture observed in the all locations. Illinois shares some similarity with Hays and Manhattan-derived individuals but remains distinct. Hays=Hays, Kansa (Central Kansas ecoregion), Manhattan=Manhattan, Kansas (Eastern Kansas ecoregion).

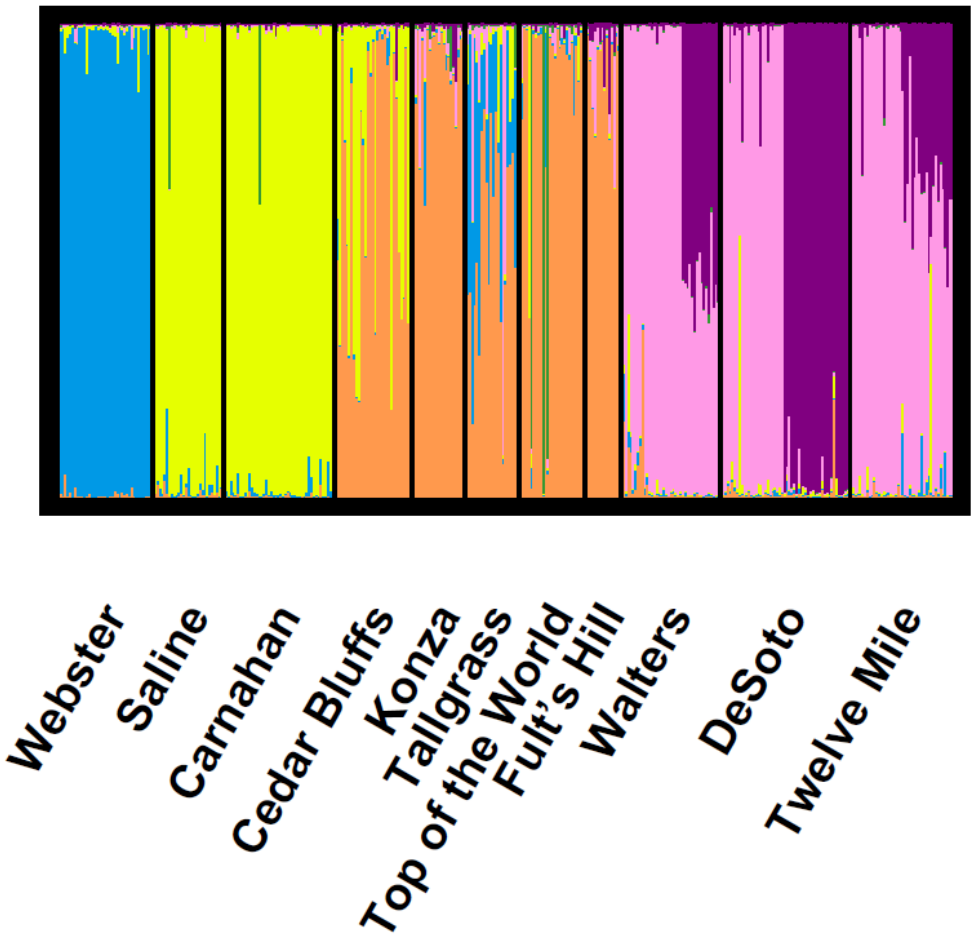

Figure 2.7 STRUCTURE barplot organized by Q (ancestry fractions).

Each color represents a genetic group or population (K) with mixed ancestry designated by mixed colors in each column of the barplot. The STRUCTURE run resulted in K=6 genetic clusters as most likely. Within each prairie (separated by thin black lines), similar allele frequencies and thus shared ancestry were observed with some admixed individuals evident. Webster, Saline, Cedar Bluff=Central Kansas ecotype; Carnahan, Konza, Tallgrass, and Top of the World = Eastern Kansas ecotype; Fult's Hill, Walters, DeSoto, and Twelve Mile=Illinois ecotype.

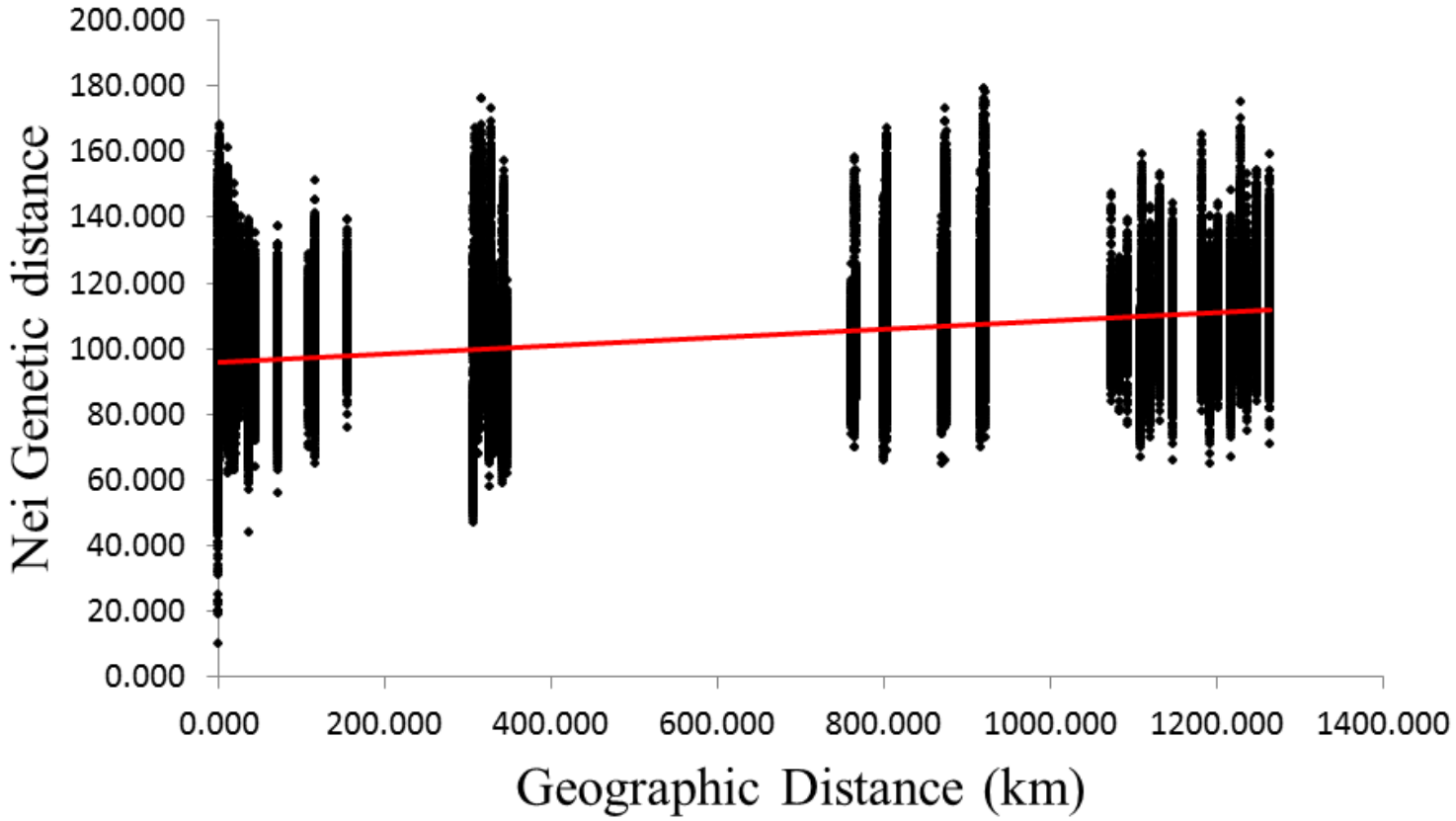

Figure 2.8 Mantel test of Nei genetic distance matrix with pairwise geographic distance matrix between prairie sites.

The results are given for 999 random permutations. Each dot represents one pairwise comparison. Geographic distance between prairie sites is given in kilometers (km) and genetic distance was calculated using Nei genetic distance (Nei and Li 1978, analogous to the Dice coefficient). The regression line fit is shown in red (R^2= 0.17, $p<0.001$).

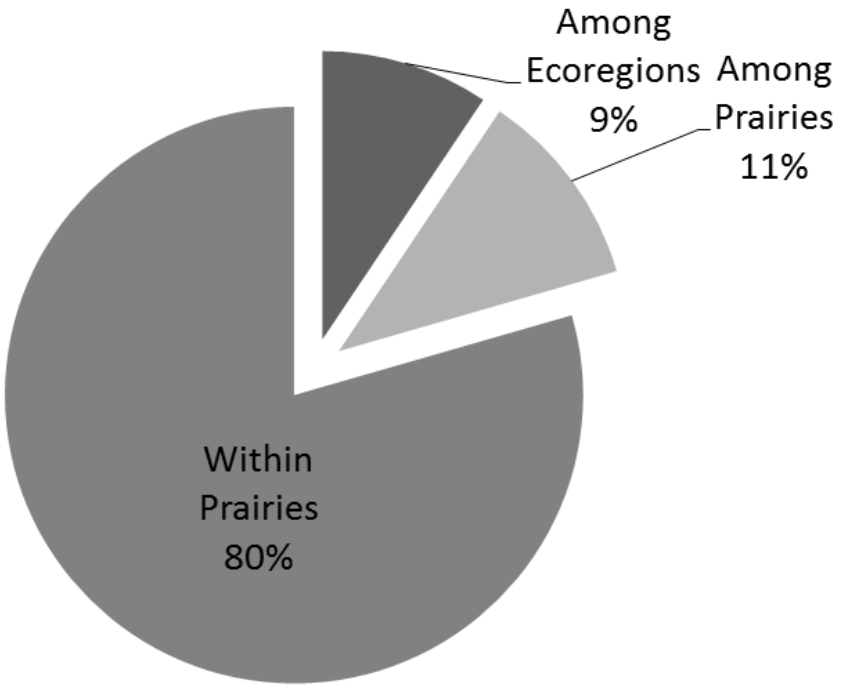

Figure 2.9 Analysis of molecular variance.

The pie chart shows the partitioning of total genetic variation within-prairies, among-prairies, and across-ecoregios (Kansas and Illinois). Central Kansas and Eastern Kansas ecoregions were grouped as an ecoregion based on population structure results. The results are shown for 999 total random permutations ($p<0.001$).

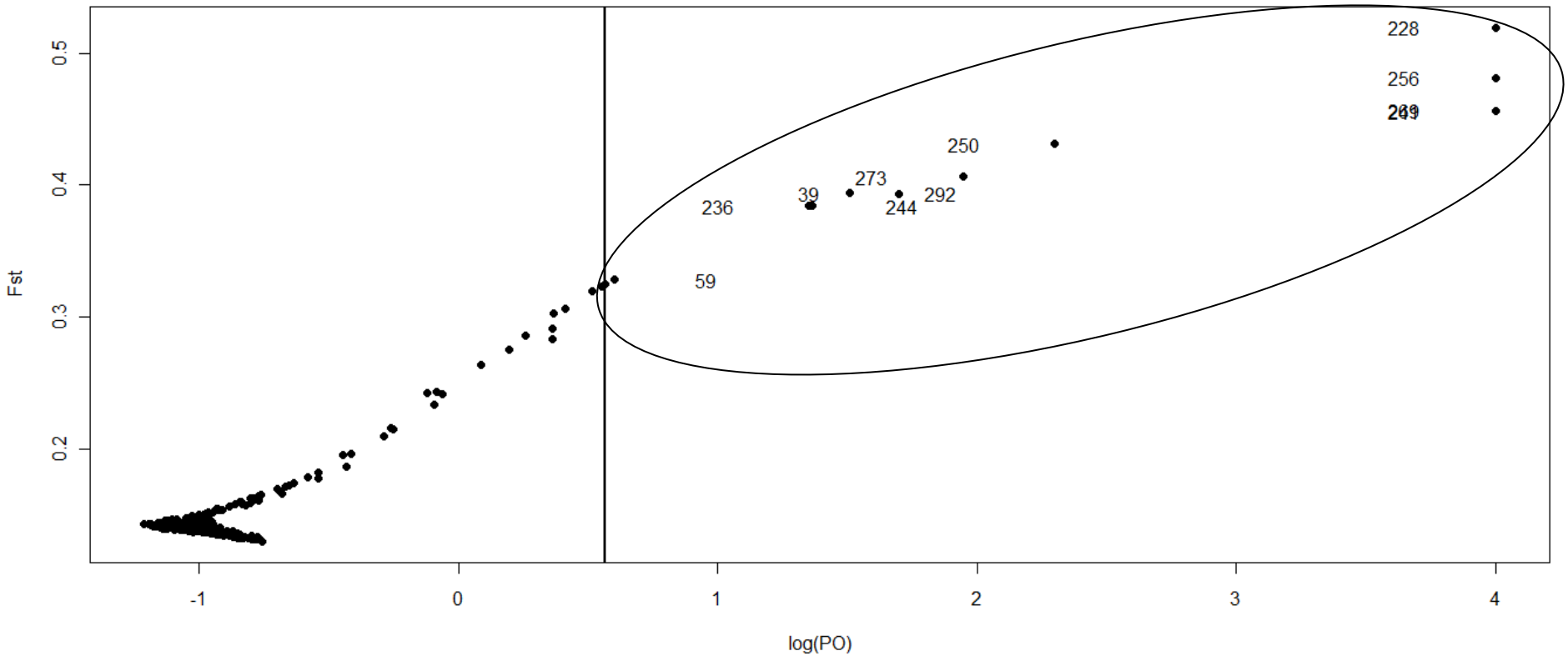

Figure 2.10 F_{ST}-outlier analysis with 387 polymorphic markers included.

Plot shows F_{ST} vs. significance (log 10 posterior odds, PO) across 387 polymorphic AFLP markers. Data was organized by ecotype (Central Kansas, Eastern Kansas, Illinois). The observed global F_{ST}-= 0.1. Eleven marker loci (ellipse) are high outliers with greater genetic differentiation than expected under neutravlity (FDR=0.05, vertical line showing singificance cut-off). These eleven candidate loci are under possible positive or diversifying (bi-directional) selection (alpha>0) wih substantial (Log(Po)≥0.5) to decisive (Log(PO)>2.0) evidence for selection.

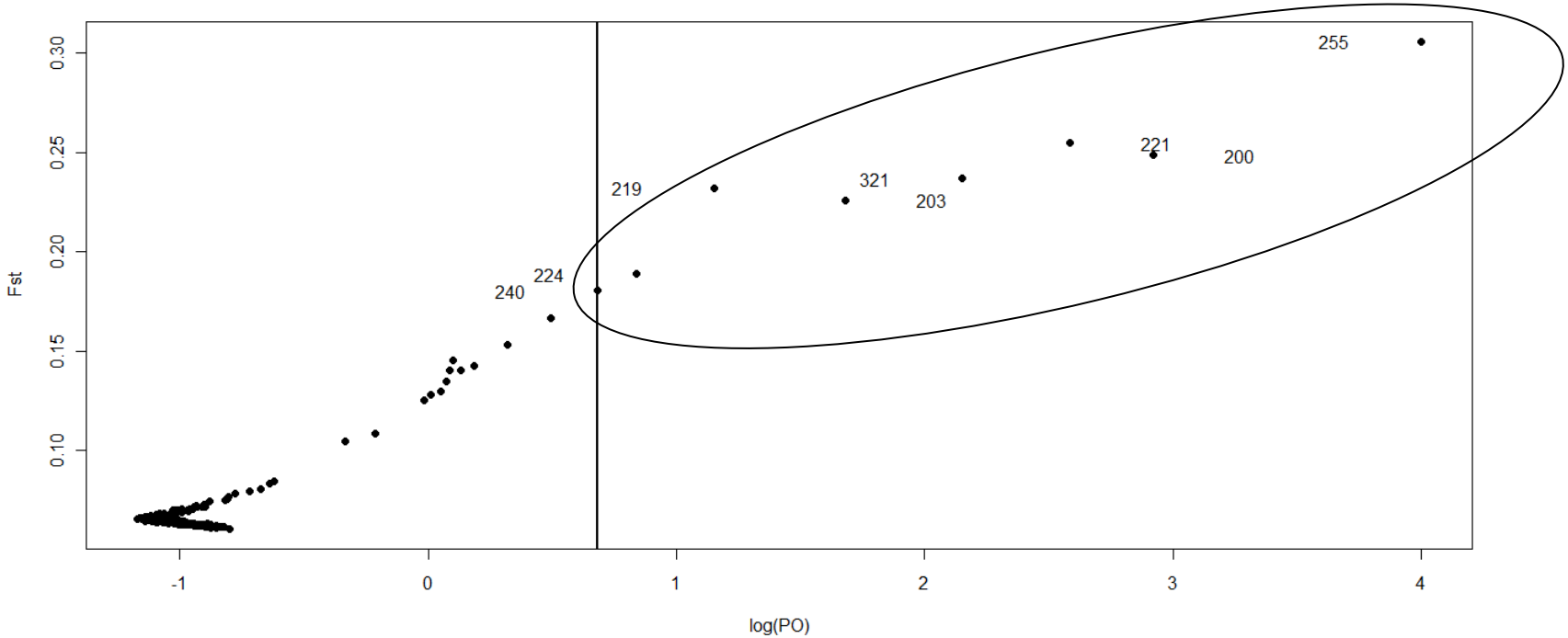

Figure 2.11 F_{ST}-outlier analysis with 325 polymorphic markers included.

Plot shows F_{ST} vs. significance (log 10 posterior odds, PO) across 325 polymorphic AFLP markers. Data was organized by ecotype (Central Kansas, Eastern Kansas, Illinois). The observed global F_{ST}- = 0.1. Eight marker loci (ellipse) are high outliers with greater genetic differentiation than expected under neutravlity (FDR=0.05, vertical line showing singificance cut-off). These eight candidate loci are under possible positive or diversifying (bi-directional) selection (alpha>0) with substantial (Log(Po)≥0.5) to decisive (Log(PO)>2.0) evidence for selection.

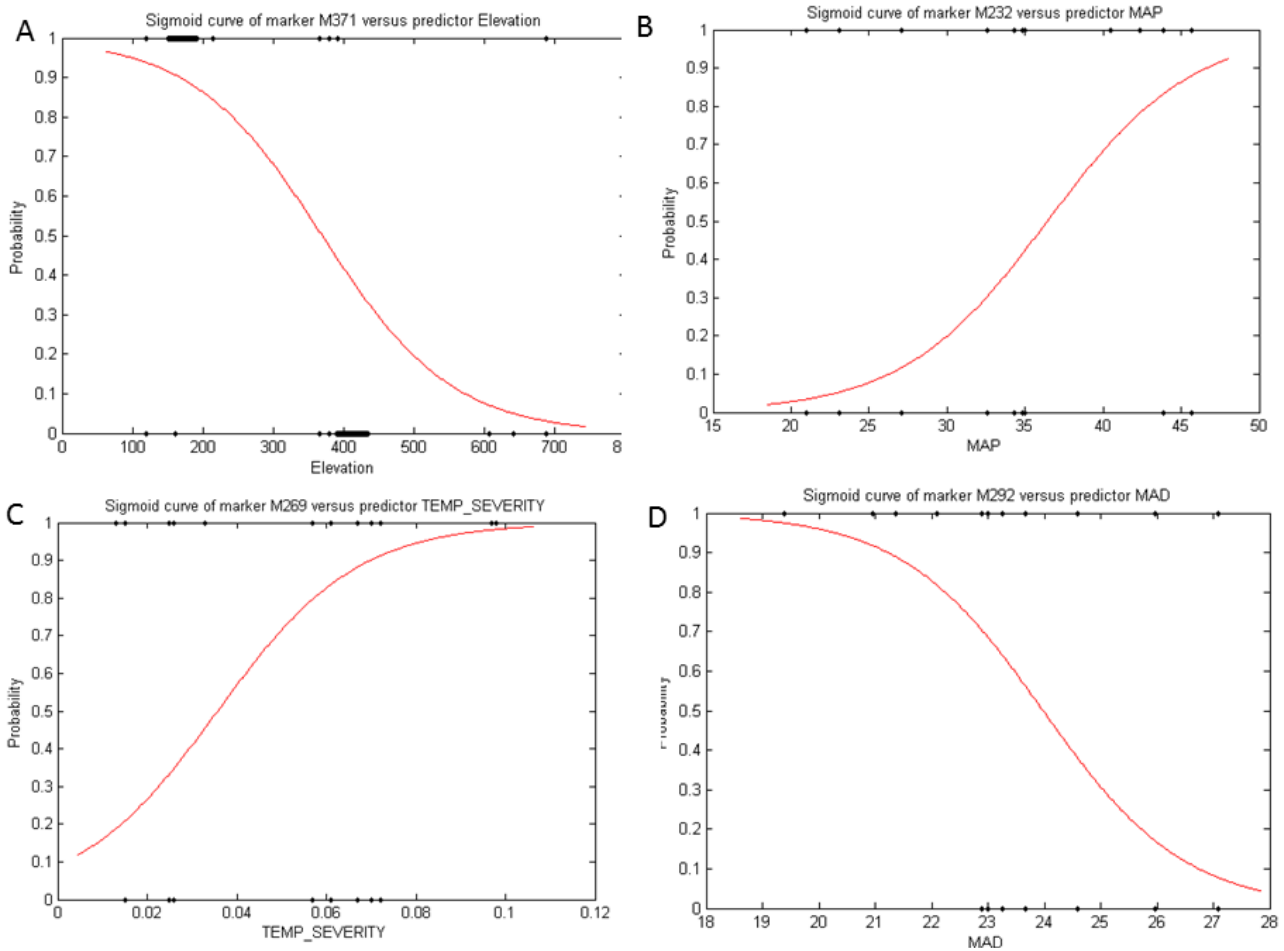

Figure 2.12 AFLP marker frequency shifts related to geo-environmental variables.

The frequency pattern plots show probability of marker (y-axis) relative to (A) elevation of prairie (m) (B) historical mean annual precipitation (cm), (C) temperature severity index (fraction of days >35˚C since 1961), and (D) mean annual diurnal temperature variation (˚C). In (A), Marker 371 is shown at high frequency with greater prairie elevation. For all logistic regressions, McFadden and Efron R^2 is greater than 0.3. Marker frequency shifts across the environmental gradient suggests important environmental pressures or conditions to AFLP marker differentiation.

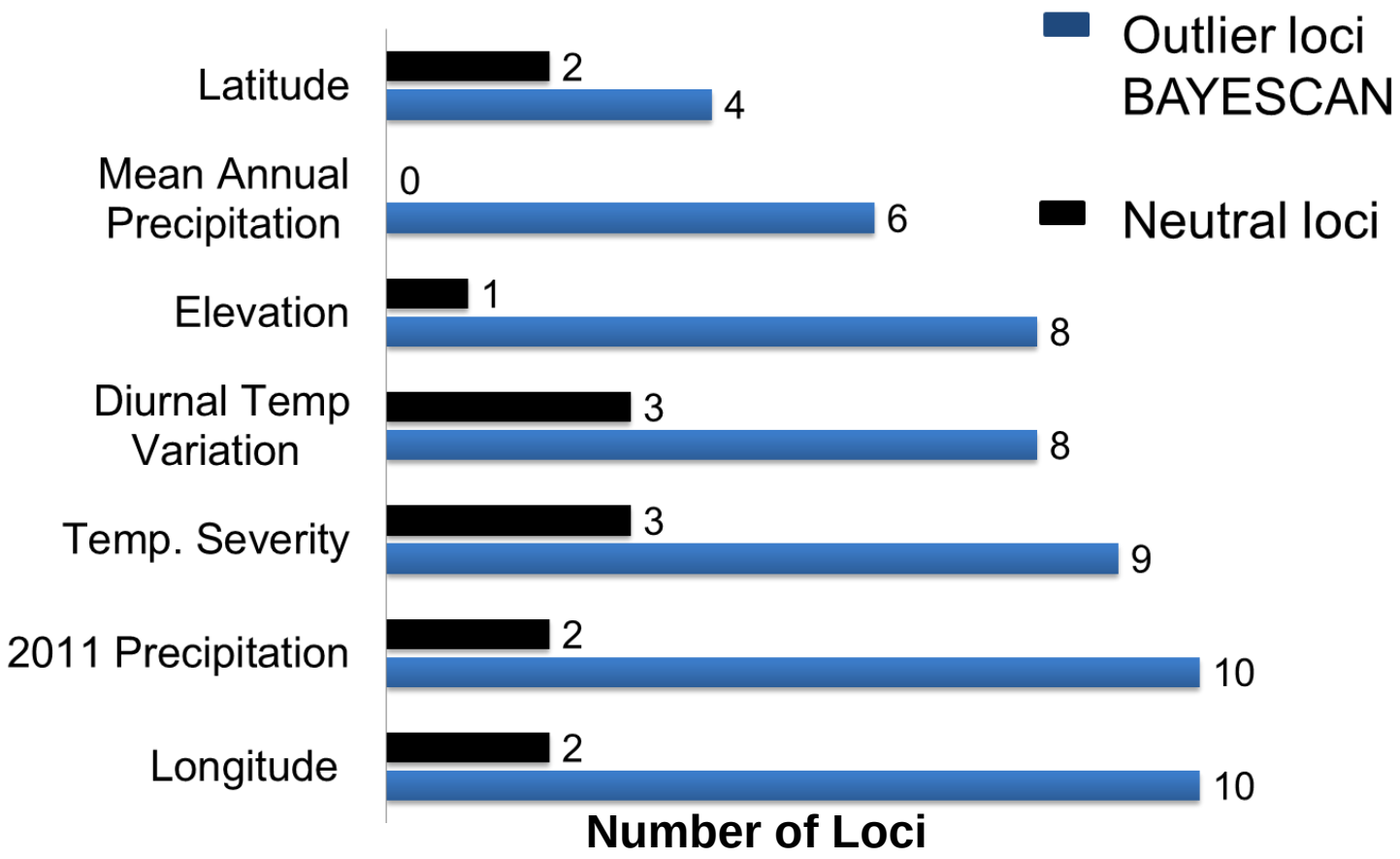

Figure 2.13 Frequency barplot of AFLP loci associated with geo-environmental variables.

Associations were generated using the MATSAM program (Joost 2007, 2007). Blue bars are representative of the number of AFLP loci which were identified as "neutral" loci in BAYESCAN which are associated with geo-environmental variables. Black bars represent the number of F_{ST}-outlier loci identified in BAYESCAN which are associated with geo-environmental variables.

Table 2.1 Environmental factors varying by ecoregion along big bluestem's dominant range.

Factors varying across location / ecoregion (Hays, KS=Central Kansas; Manhattan, KS=Eastern KS; Carbondale, IL) in the U.S. Midwest include soil type, elevation (m), mean annual precipitation (MAP, cm) calculated in 2010 and historically since 1961, precipitation of the driest year since 1961, average and 2010 growing degree days (GDD), potential evapotranspiration from lakes (PET, cm/year) based on free water surface evaporative demand (Koelliker personal comm.), and an aridity index.

Location (Ecoregion)	Soil Type	Elevation (m)	MAP, 2010 (cm)	MAP, since 1961 (cm)	Ppt of driest year (cm)	Average GDD	2010 GDD	PET (cm/yr)	Aridity
Hays, KS (Central KS) Ellis County	Roxbury Silt-loam	603	50.11	58.22 (± 13.13)	36.27	3799	3237	139	81
Manhattan, KS (Eastern KS) Riley County	Sandy-loam	315	67.82	87.15 (± 20.04)	39.16	4156	3205	127	41
Carbondale, Illinois (Illinois) Jackson County	Silt-loam	127	66.95	116.73 (± 24.76)	67.38	4087	3597	99	-18

Table 2.2 Prairie collection site including AFLP sample size and prairie size in acres.

Ecoregion (Location)	Collection Site	County	AFLP Sample Size	Prairie Size (acres)
Central Kansas (Hays, KS)	Webster Res.	Rooks	40	880
	Saline Expt Range	Ellis	30	2,400
	Cedar Bluffs Res	Trego	33	1,100
Eastern Kansas (Manhattan, KS)	Carnahan Cove St. Pk	Pottawatomie	47	245
	Konza Prairie	Riley/Geary	22	3,487
	Tallgrass Nat. Pk	Chase	23	10,894
	Top of the World Pk	Riley	28	150
Illinois (Illinois)	Desoto	Jackson	55	7 to 24
	Twelve Mile	Effingham, Fayette, and Marion	43	NA
	Fults	Monroe	15	528
	Walters	Jasper	42	NA

Table 2.3 Prairie environmental variables used in AFLP marker-environmental correlation (SAM) analyses.

Prairie weather site for weather data is included as well as supplementary weather sites used from NOAA database. Environmental descriptors include location (GPS coordinates), prairie elevation (m), 2011 cumulative seasonal precipitation (during bluestem growing season from April-August), 2011 mean annual precipitation (cm, 2011 MAP), long-term mean annual precipitation (cm, LT MAP), a temperature severity index (fraction of days since 1961 above 35˚C), and mean annual diurnal temperature variation (from average daily max temp-min temp) since 1961 (MAD, ˚C). NA= Carnahan Cove site 2011 MAP was omitted as data was unavailable.

REGION	PRAIRIE	WEATHER SITE(S)	PRAIRIE LONG (W)	PRAIRIE LAT (N)	PRAIRIE ELEVATION (m)	2011 SEASONAL PPT (cm)	2011 MAP (cm)	LT MAP (cm)	TEMP SEVERITY INDEX	MAD (˚C)
CENTRAL KS	Webster Res	Webster Dam	99.32	39.24	606.00	99.67	58.47	56.62	0.054	-2.73
	Saline Experimental Range	Ellis 12, Hays1S, Ellsworth	99.14	39.02	641.00	99.11	51.64	66.36	0.040	-4.11
	Cedar Bluffs Res	Cedar Bluff Dam	99.46	38.45	688.00	97.66	38.13	51.42	0.055	-3.35
EASTERN KS	Konza Prairie	Manhattan 6 SW	96.36	39.05	366.00	99.19	79.88	85.34	0.037	-5.00
	Tallgrass Prairie NP	Tallgrass Prairie	96.33	38.25	392.00	97.16	74.32	79.89	0.067	-5.00
	Carnahan Cove St. Park	Westmoreland, Wamego	96.38	39.20	389.00	50.47	NA	84.11	0.032	-4.86
	Top of the World Park	Tuttle Creek, KSU	96.37	39.13	379.00	99.39	84.61	85.70	0.039	-4.63
ILLINOIS	DeSoto Prairie	Murphysboro, Carbondale	89.14	37.51	119.00	95.28	175.62	111.82	0.014	-5.50
	Twleve Mile Prairie	Kinmundy, Salem	88.50	38.46	160.00	97.69	148.72	103.76	0.014	-6.13
	Fults Hill Prairie	Sparta, Prairie Du Rocher	89.48	37.58	215.00	95.45	120.42	107.43	0.018	-5.91
	Walters Prairie	Newton, Charleston	88.09	38.59	150.00	98.02	141.45	99.14	0.008	-7.01

Table 2.4 Pairwise Nei's unbiased genetic distances between 11 prairies.

Nei's unbiased genetic distance = -1 * Ln (Nei Identity) (Nei and Li 1978). Prairies are color-coded by ecotype/ecoregion: Blue= Illinois, Green= Eastern Kansas, Red= Central Kansas.

	DES	FUL	TM	WAL	CAR	KON	TAL	TOW	CDB	SAL	WEB
DES	0.000										
FUL	0.054	0.000									
TM	0.012	0.048	0.000								
WAL	0.019	0.044	0.008	0.000							
CAR	0.047	0.071	0.058	0.060	0.000						
KON	0.053	0.023	0.048	0.043	0.056	0.000					
TAL	0.042	0.034	0.038	0.033	0.035	0.022	0.000				
TOW	0.051	0.030	0.050	0.047	0.043	0.012	0.020	0.000			
CDB	0.048	0.038	0.058	0.057	0.029	0.028	0.025	0.020	0.000		
SAL	0.053	0.071	0.066	0.066	0.004	0.054	0.037	0.043	0.026	0.000	
WEB	0.068	0.076	0.066	0.064	0.032	0.056	0.034	0.048	0.049	0.035	0.000

Table 2.5 Ecoregion-specific AFLP marker loci.

Selective Primer Set	Band Length (bp)	Marker ID	Private Band Ecoregion
FAM-E-AAA+M-CGC	135	M34	Hays / Manhattan
FAM-E-AAA+M-CGC	257	M118	Hays / Manhattan
FAM-E-AAA+M-CGC	378	M193	Hays / Manhattan
HEX-E-ACC+M-CTG	363	M365	Hays / Manhattan
FAM-E-AAA+M-CGC	365	M187	Manhattan
FAM-E-AAA+M-CGC	419	M210	Manhattan
FAM-E-AAA+M-CGC	443	M217	Manhattan
HEX-E-ACC+M-CTG	323	M349	Manhattan
HEX-E-ACC+M-CTG	366	M366	Manhattan
HEX-E-ACC+M-CTG	394	M374	Manhattan
FAM-E-AAA+M-CGC	367	M188	Illinois / Manhattan
FAM-E-AAA+M-CGC	374	M191	Illinois / Manhattan
HEX-E-ACC+M-CTG	265	M326	Illinois / Manhattan
HEX-E-ACC+M-CTG	273	M330	Illinois / Manhattan
HEX-E-ACC+M-CTG	275	M331	Illinois / Manhattan
HEX-E-ACC+M-CTG	280	M333	Illinois / Manhattan
HEX-E-ACC+M-CTG	294	M338	Illinois / Manhattan
HEX-E-ACC+M-CTG	310	M344	Illinois / Manhattan
HEX-E-ACC+M-CTG	318	M347	Illinois / Manhattan
HEX-E-ACC+M-CTG	327	M351	Illinois / Manhattan
HEX-E-ACC+M-CTG	337	M357	Illinois / Manhattan
HEX-E-ACC+M-CTG	372	M368	Illinois / Manhattan
HEX-E-ACC+M-CTG	395	M375	Illinois / Manhattan
HEX-E-ACC+M-CTG	298	M339	Illinois
HEX-E-ACC+M-CTG	349	M361	Illinois
HEX-E-ACC+M-CTG	400	M376	Illinois
HEX-E-ACC+M-CTG	453	M385	Illinois

Table 2.6 Analysis of molecular variance statistical summary.

AFLP marker data with 11 prairie sites entered as assumed populations. Regions are Kansas ecoregion and Illinois ecoregions based on population structuring observed in STRUCTURE. In analyses, control samples were included. $p<0.001$. df= degrees of freedom, SS= sum of squares, MS= mean squares.

Source of variation	df	SS	MS	Estimated Variance	Percent Total variance
Among Regions	1	1320.239	1320.239	5.120	9%
Among Prairies	9	2387.695	265.299	6.081	11%
Within Prairies	401	17386.880	43.359	43.359	80%
Total	411	21094.813		54.560	100%

Table 2.7 Significant associations of AFLP markers with geo-environmental predictor variables

For all models, logistic regression analyses in MATSAM were performed (Joost 2007, 2008). Geo-environmental predictor variables, in order of number of significant associations with AFLP markers included: prairie longitude (W), prairie annual precipitation in 2011 (cm), prairie temperature severity given by fraction of days over 35˚C since 1961, mean annual diurnal temperature variation given by average daily max. temperature – min temperature since 1961 (˚C), prairie elevation (m), prairie mean annual precipitation since 1961 (cm), and prairie latitude (N). Marker ID associated with the geo-environmental predictor variable is given in parentheses. Associations were deemed significant if it was significant in both Wald and G tests and had McFadden and Efron pseudo R^2>3.0.

Geo-environmental Predictor Variable	Number of Significant AFLP Marker Correlations (Marker ID)
Prairie Longitude	10 (228, 232, 237, 250, 252, 256, 263, 275, 292, 371)
Annual Precipitation in 2011	10 (228, 232, 237, 242, 250, 252, 256, 275, 292, 371)
Historical Temperature Severity	9 (228, 232, 237, 250, 256, 269, 275, 336, 371)
Mean Annual Diurnal Temperature Variation	8 (228, 232, 237, 275, 292, 301, 336, 371)
Prairie Elevation	8 (228, 232, 237, 250, 256, 275, 292, 371)
Historical Mean Annual Precipitation	6 (228, 232, 250, 256, 275, 371)
Prairie Latitude	4 (242, 252, 292, 336)

Chapter 3 - Summary and Future Directions

Big bluestem (*Andropogon gerardii* Vitman) is a vital ecological-dominant of the rapidly disappearing tallgrass prairie ecosystem; thus, studies of ecotypic genetic variation within the species will help to inform prairie restoration practices and land management currently underway in the U.S. Midwest. The foundation behind this research was based on greenhouse and field reciprocal garden experiments in which significant phenotypic differentiation of big bluestem occurring naturally across a sharp precipitation gradient was observed (Johnson et al. in prep). Given phenotypic diversity of ecotypes along this gradient, we predicted differentiation of ecotypes. ***The specific objectives of this research project were to:***

1) probe the extent of neutral and non-neutral genetic differentiation and diversity across the genome of big bluestem ecotypes using an AFLP genome scan approach
2) determine whether large-scale geographic isolation leading to population structuring can be detected across the Midwest environmental gradient
3) discover whether environmental conditions of the native locations of big bluestem prairie sites are associated with AFLP marker loci frequency patterns and thus may generate hypotheses regarding important environmental selective pressures to ecotype differentiation.

Based on our study of neutral genetic diversity, which developed 387 AFLP marker loci for big bluestem, it was discovered that:

- Kansas (Central and Eastern) and Illinois ecotypes of big bluestem are able to be genetically discriminated based on AFLP marker presence/absence, although Kansas ecotypes have significant overlap and thus cannot be genetically discriminated given this analysis. This trend in the data was strengthened by confirmation with different measures: marker presence/absence, Dice coefficient of genetic distance, and marker allele frequencies.

- There is a weak correspondence (R^2=0.17, $p<0.001$) between increasing geographic distances across prairie sites/populations with increasing Nei's unbiased genetic distance. This suggests ecotypes are differentiated also based on factors other than geographic isolation (founder effects, subdivision) or that results are limited by the geographic scales this study investigated (relatively large distances between ecoregions with very small distances between prairies within ecoregions). A different sampling design may find better support for isolation by distance.

- Population structure is evident among Kansas vs. Illinois prairie ecoregions across the U.S. Midwest environmental gradient. Based on similar allele frequencies, six genetic clusters were identified. Individuals from Kansas prairies, however, largely share membership in three genetic clusters. Thus, allele frequency divergence exists between Kansas and Illinois ecotypes.

- Despite population structure and genetic differentiation, high within-prairie diversity still exists, suggesting high genetic variation is maintained through the outcrossing nature of big bluestem or the large polyploidy genome of big bluestem, or a combination of both.

- Eight AFLP outlier-loci were found to be highly differentiated (substantial to decisive evidence for selection for AFLP loci) relative to the species overall species F_{ST} (F_{ST}=0.1). These AFLP marker loci were detected when prairies were entered according to ecoregion into BAYESCAN, and thus represent loci that are highly differentiated across the three ecotype groups.

- All marker loci identified to be under possible divergent selection were also significantly associated with environmental predictors (R^2>0.3 for all marker-environment associations, significant Wald and G test results), namely mean annual precipitation (7 markers associated) and longitude (8 markers associated). The marker loci under divergent selection and linked to geo-environmental parameters represented about half (57%) of the total number of markers linked to the environmental cline. By nature, environmental variables are correlated and some commentary on this correlation remains to be made.

- Given our findings, we can make the prediction that the superior Central Kansas ecotype (which appears to be locally adapted to drier environments in the field), if planted in restorations in Illinois, has the potential to be genetically "swamped" due to differences in allele frequencies between Kansas and Illinois ecoregions. However, genetic testing on field tests is needed to make substantial recommendations for restoring the tallgrass prairie ecosystem.

This current neutral marker study could benefit from development of sequence-based markers (RAD-tags, for example) such that more detailed studies of sequence divergence across the genome between the Kansas ecotypes and Illinois ecotype may be studied. Furthermore, it is recommended that fine-scale and broad-range sampling for genetic diversity of big bluestem be connected more extensively. Such a study could lead to a better understanding of the population genetic processes leading to the genetic structuring of big bluestem populations along an important environmental gradient found here.