Summer 2016

Determination Of Evolutionary History Of Big Bluestem Populations Through Chloroplast DNA Analysis

Tej Man Tamang
Fort Hays State University, tmtamang@mail.fhsu.edu

DETERMINATION OF EVOLUTIONARY HISTORY OF BIG BLUESTEM POPULATIONS THROUGH CHLOROPLAST DNA ANALYSIS

being

A Thesis Presented to the Graduate Faculty

of the Fort Hays State University in

Partial Fulfillment of the Requirements for

the Degree of Master of Science

by

Tej Man Tamang

B.Sc., Purbanchal University

Date_____ Approved_____
Major Professor

Approved_____
Chair, Graduate Council

This thesis for

The Master of Science Degree

by

Tej Man Tamang

has been approved by

_____

Chair, Supervisory Committee

_____

Supervisory Committee

_____

Supervisory Committee

_____

Supervisory Committee

_____

Chair, Department of Biological Sciences

ABSTRACT

Andropogon gerardii Vitman (big bluestem) is one of the most dominant and widely distributed grasses of the North American prairie. It is widely used in restoration projects for the recovery of grassland ecosystems. *A. gerardii* demonstrates genetic and adaptive variation among populations across the prairie. With the objective to understand the evolutionary relationship between the *A. gerardii* populations, two noncoding chloroplast DNA (cpDNA) spacers (*rpl32-trnL*$^{(UAG)}$ and *trnQ*$^{(UUG)}$*-rps16*) were studied. Similarly, genetic differentiation among the populations was also calculated based on the spacers. The *trnQ*$^{(UUG)}$*-rps16* spacer had more polymorphic sites than the *rpl32-trnL*$^{(UAG)}$ spacer. A phylogenetic tree based on combined cpDNA spacers generated a monophyletic tree for *A. gerardii* with a Colorado population and sand bluestem (*Andropogon hallii*) as the outgroups. The monophyletic tree was further resolved into two sub-clades. Most of the branches and nodes were well supported, with more than 70% of posterior probability values. However, the grouping of populations did not support the resolution of the phylogenetic tree with geological distribution. Analysis of molecular variance suggests there is a low level of genetic differentiation among the populations, with 90% of variation within the populations and 10% of variation among the populations. The observed high genetic variation within populations could be the result of potential gene flow, polyploidy, and the outcrossing nature of big bluestem.

ACKNOWLEDGEMENTS

This work would not have been possible without help and support of many people. First, I express my sincere gratitude to my advisor, Dr. Brian R. Maricle, for his academic guidance, support, valuable advice and encouragement, and most importantly, for providing opportunity to conduct this research. I also thank my committee members, Dr. Yashuhiro Kobayashi, Dr. William "Bill" Stark, and Dr. Loretta Johnson, for valuable suggestions. Along with this, I express my sincere thanks to all the faculty members of the Department of Biological Sciences for all their support during my academic career at Fort Hays State University.

My special gratitude goes to Dr. Yashuhiro Kobayashi for providing the technical knowledge to conduct this research. In addition, I would like to thank him for allowing me to work in his laboratory despite his busy schedule.

My sincere thanks extend to Dr. Loretta Johnson for granting the access to samples, and for arranging my accommodation during my stay at Manhattan, Kansas. In addition, I express my gratitude to her for providing the access to DNA extraction lab. Thanks to Matt Galliart and Jacob Alsdurf for their help in collection of the samples, and extraction of DNA from leaf samples.

My sincere gratitude goes to Dr. Terrell Balthazor for providing the funding for research assistantship, which supported a year of my graduate studies, and which in turn, gave me more time to focus on research. This Master's thesis was funded by KINBRE mini-grant and Kansas Academy of Science.

I thank all the graduate students of the Department of Biological Sciences for their immense help and support. Last but not least, I thank my parents for their sacrifice, advice, support, and love. This work would not have been possible without them and hence, is dedicated to my father, Saphat Singh Tamang and mother, Gyani Maiya Tamang.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF APPENDIXES

Page

PREFACE

This thesis follows the style of *Plant Systematics and Evolution*.

Introduction

The grassland of North America, known as prairie, dates back about 25 million years to Tertiary times, and has a preglacial origin (Weaver 1954). The grassland in the Great Plains extends from the Rocky Mountains east to the Mississippi river along an east-west gradient (Samson and Knopf 1994) and from central Texas to south-central Canada along a south-north gradient (Samson et al. 2004). Within the prairie, C_4 species occupy more than 80% of land from 30° to 42° N whereas C_3 species are more dominant in the area north of 42° N (Reid et al. 2005). The central grassland is divided into three types of vegetation on the basis of annual mean precipitation, namely western shortgrass prairie (260 to 375 mm), central mixed prairie (375 to 625 mm), and eastern tallgrass prairie (625 to 1200 mm) across a gradual west to east precipitation gradient (Lauenroth et al. 1999).

Urbanization and agricultural development have led to fragmentation and habitat reduction of North American Prairie (Gustafson et al. 1999). Activities like depletion of aquifers, increase in water borne chemical pollutants, and confining of running water have threatened prairie, and yet conservation efforts for prairie are being neglected (Samson and Knopf 1994). Only 4.4% of central tallgrass prairie remains (Samson et al. 2004), with more conserved patches being in the Flint Hills of Kansas and remnant prairies in Nebraska (Johnson et al. 2015). As of August 2014, about 25.6 million acres of land were enrolled in the USDA Conservation Reserve Program to restore the land from agricultural use (Stubbs 2014). The program consisted of nearly 4.3 million acres in a five state Midwest region to conserve the prairie (SCS 1990). Many restoration ecologists are relying upon a 'best guess' approach for restoring

the grassland (Broadhurst et al. 2008). This could be misguided, as study of adaptive variation and local adaptation in grassland species has been minimal (Johnson et al. 2015). Many restoration ecologists rely upon the importance of genetic variation for restoration (Falk et al. 2006), as this plays a fundamental role in the design, implementation, and expectations of many restoration projects (Falk et al. 2001).

Andropogon gerardii Vitman (big bluestem) is one of the most dominant and widely distributed grasses of the prairie (Weaver 1954; Gould 1967), and comprises 80% of biomass in native communities (Keeler and Davis 1999). *A. gerardii* is a perennial, native, drought tolerant, and warm-season grass of tallgrass prairie (Weaver 1968; Barnes 1985; Gustafson 1999). It is widely distributed in moist prairie sites, tributary valleys and ravines, slopes, and lower hillside (Barnes 1985). However, *A. gerardii* is not only confined to true prairie region. It is distributed often in small, fragmented populations in the eastern US (Tompkins et al. 2012), and extends as far as the northeastern US and Canada (USDA 2004). *A. gerardii* is a highly palatable grass to many species of livestock (Stubbendieck et al. 2011). *A. gerardii* is considered a good grass for hay-making (USDA 2004), and it provides habitat and cover for many vertebrate species (Loflin and Loflin 2006). *A. gerardii* is also used in soil conservation applications like controlling soil erosion, and consequently, it is considered a foundation species in restoration projects for the recovery of grassland ecosystems (Johnson et al. 2015).

A. gerardii demonstrates genetic and adaptive variation among populations across the prairie, as it spans across several environmental gradients (Gray et al. 2014). McMillan (1959, 1964, 1965) documented morphological and phenological variations among *A.*

gerardii ecotypes along the latitudinal gradient of the Great Plains. Johnson et al. (2015) established reciprocal gardens at four sites in Kansas and Illinois across an 1150 km precipitation gradient and examined establishment success, vegetative cover, SPAD chlorophyll absorbance, and other parameters of three *A. gerardii* ecotypes in response to different environmental variables. Mendola et al. (2015) studied variations in biomass, below ground net primary production, root C:N ratio, and root nitrogen storage of three ecotypes of *A. gerardii* at the reciprocal gardens. Furthermore, Olsen et al. (2013) reported differences in leaf anatomy across ecotypes and sites, and Caudle et al. (2014) showed that the ecotypes and cultivars from drier sites have higher chlorophyll concentration compared to more mesic populations. Gray et al. (2014) noted the distinct genetic structure between Kansas and Illinois ecotypes of *A. gerardii*. The distinct genetic, morphological, and adaptive structure of *A. gerardii* might be due to the adequate time for climatic and ecological selection pressure on populations of existing tallgrass prairie (Gray et al. 2014; Johnson et al. 2015), which was formed more than 10,000 years ago (Axelrod 1985). However, whether chloroplast DNA differs among populations of *A. gerardii* has not been investigated, a gap this study seeks to fill.

Chloroplasts, the site of photosynthesis, consist of enzymatic machinery and electron carriers required for photosynthesis (Daniell et al. 2006, Kaur et al. 2014). Chloroplasts contain DNA that resembles cyanobacterial DNA (Ku et al. 2013). Chloroplast DNA (cpDNA) is circular and contains more than 100 genes (Maliga and Bock 2011, Kaur et al. 2014), with a size ranging from 72 kb to 212 kb (Shi et al. 2012). The chloroplast genome consists of genes that code photosynthetic proteins, ATP synthase, NAD(P)H

dehydrogenase, and Rubisco. Several genes from cpDNA have been lost, which is a common pattern throughout plastid genome evolution (Clegg et al. 1994). The chloroplast genome also contains genes for tRNA, rRNA, polymerase, and open reading frames (ORFs). Broadly, cpDNA can be categorized into three categories (Ravi et al. 2008):

a. Genes coding for proteins required for photosynthesis,
b. Genes for RNA and genetic apparatus, and
c. Genes for open reading frames

The chloroplast genome has quadripartite organization (Jansen et al. 2005), in which two regions of inverted repeats (IRa and IRb) are separated by a small single copy region (SSC) and a large single copy region (LSC) (Antherton et al. 2010; Wu et al. 2012). In most cases, the two inverted repeats are mirror images of one another in terms of gene complement, and are about 25 kb in size (Shaw et al. 2007).

The chloroplast genome has been extensively studied at the molecular level to determine the phylogenetic and evolutionary relationships of plants (Clegg 1993; Kim et al. 1999). Chloroplast gene sequences can reveal relationships at various taxonomic levels, depending on the rate of change of the regions studied (Jansen et al. 2005; Zhang et al. 2009; Korotkova et al. 2011). Within cpDNA, gene and intron contents are conserved among terrestrial plants (Jansen et al. 2005). The uniparentally inherited chloroplast genome, with a lack of recombination, makes the chloroplast genome well suited for phylogenetic studies (Diekman et al. 2008; Ravi et al. 2008; Hall and Hallgrimsson 2014).

Apart from coding regions, chloroplasts contain noncoding regions that include introns and intergenic spacers (Shaw et al. 2007). Noncoding DNA sequences are considered important for their functions in the regulation of genes and for DNA replication, transcription, chromosome condensation, and chromosome pairing (Meisler 2001). Plastid noncoding regions are often used in study of molecular systematics, population genetics, and phylogeography (Shaw et al. 2007).

A single gene strategy could be problematic in identifying plants. There might be too much variation in a single gene, causing problems in alignment or too little variation, leading to no resolution at all. Thus, it is more pragmatic to use multiple markers (Zhang et al. 2009), which increases accuracy in constructing a phylogenetic tree (Doyle et al. 1992). Sequences of coding regions like *rbcL*, *matK*, *ndhF*, *rpl16*, *atpB*, and noncoding regions like *trnL* intron, *rpl16* intron, *ndhA* intron, *trnL-trnF*, *ndhF-rpl32*, *rpl32-trnL*$^{(UAG)}$, *trnQ*$^{(UUG)}$*-rps16* intergenic regions have been widely used in phylogenetic analyses, as evolutionary markers to determine relationships among plant species (Shaw et al. 2007; Ravi et al. 2008; Shaw et al. 2014). Up to 6 noncoding cpDNA regions can be used in phylogenetic analysis, with two to three markers being examined commonly, with one being minimum and 6 being maximum (Shaw et al. 2014).

The Consortium for the Barcoding of Life suggests using coding sequences of *rbcL* and *matK* genes, and the intergenic spacer *psbA-trnH* for plant barcoding (CBOL Plant Working Group 2009). However, many studies indicate the gene regions suggested by CBOL are not variable enough to resolve plant species at lower taxonomic levels, i.e., between genera and species. A study conducted by Korotkova et al. (2011) showed that

rbcL and *matK* markers are least effective in identifying operational taxonomic units. In the same study, the use of introns and noncoding sequences rather than coding sequences appeared to be better markers for study of phylogeny. However, no single noncoding cpDNA is universally best for low level molecular studies. Even though a universally best noncoding cpDNA is difficult to identify, *rpl32-trnL*$^{(UAG)}$ and *rps16-trnQ*$^{(UUG)}$ markers are considered as two of the most informative and variable noncoding regions (Shaw et al. 2014).

The objectives of this study were 1) to compare *A. gerardii* populations based on cpDNA, 2) to determine the evolutionary relationship among populations of *A. gerardii*, and 3) to test whether *rpl32–trnL*$^{(UAG)}$ and *trnQ*$^{(UUG)}$*-rps16* markers can be used to identify low levels of taxonomy. No previous studies have investigated sub-species relationships with chloroplast DNA analysis, so this study assessed the utility of these noncoding cpDNA regions at determining ecotypic phylogeny within a species.

In addition, I hypothesized the two noncoding cpDNA spacers (*rpl32-trnL*$^{(UAG)}$ and *trnQ*$^{(UUG)}$*-rps16*) would be able to resolve evolutionary relationships at the ecotype level. I also hypothesized there is genetic differentiation among the populations based on the noncoding cpDNA under study. The ability to identify ecotypes will be helpful for those who are maintaining and restoring the prairie. Moreover, study of genetic variation will provide valuable knowledge in the improvement of *A. gerardii*.

Materials and methods

Sample collection

A. gerardii populations were grown in a greenhouse facility at Kansas State University (Manhattan, KS, USA; 39.19°W; -96.58°N). A total of 24 populations of *A. gerardii* were used in the study (Figure 1). Leaf samples (3 to 6 pieces, ~2 cm each) were collected from 21 populations of *A. gerardii*, 2 cultivars of *A. gerardii*, and a sample of *Andropogon hallii* Hack (sand bluestem). From each population, leaves from 8 individuals were collected, except for *A. hallii*, where 7 individual leaf samples were collected because of lack of leaves prior to DNA extraction. The samples were freeze dried and stored at -80°C for further analysis.

DNA extraction and amplification

Total genomic DNA from freeze dried samples was collected for molecular analysis. DNA was isolated using a modified CTAB method (Doyle and Doyle 1978; Gray et al. 2014) and re-suspended in 100 µl of 10 mM Tris buffer (pH=8) + 0.003125% Triton X-100 at 4°C overnight. The extracted DNA was visualized in 0.8% agarose gel to verify the DNA was non-degraded. Furthermore, the quality and quantity of DNA was measured using a spectrophotometer (Nanodrop Technologies, Wilmington, DE).

Based on Shaw et al. (2007, 2014), the noncoding cpDNA spacers *rpl32-trnL*$^{(UAG)}$ and *trnQ*$^{(UUG)}$*-rps16* were used in this study. Primers for polymerase chain reaction (PCR) were developed using reference sequence of *Zea mays* (accession: x86563). The primers

were obtained from Sigma-Aldrich (St. Louis, MO). Primer sets used for amplification of *rpl32-trnL*$^{(UAG)}$ and *trnQ*$^{(UUG)}$*-rps16* are shown in Table 1. Amplification of selected regions was carried out in 50 µl of reaction mixture using Platinum® Pfx DNA polymerase (Thermo Fisher Scientific, Waltham, MA). The 50 µl volume of PCR reaction consisted of 5 µl of 10X amplification buffer, 1.5 µl of 10 mM dNTP mix, 2 µl of 50 mM $MgSO_4$, 2 µl of primer mix (10 µM each), 1 µl template DNA, 0.4 µl of Platinum® Pfx DNA polymerase, and 38.1 µl of autoclaved distilled water. After 3 minutes of pre-treatment at 95°C, PCR conditions for both primers were: 34 cycles of 30 seconds at 95°C, 30 seconds at 60°C, and 1 minute at 68°C, plus a final extension of 5 minutes at 68°C.

Agarose gel electrophoresis, PCR purification, and sequencing

Generation of appropriate amplicons by PCR was verified by gel electrophoresis using Tris-acetate-EDTA in 1.5% agarose gels (Agarose low EEO, Thermo Fisher Scientific, Waltham, MA). Correct amplicons were purified using QIAquick PCR purification kit (Qiagen, Hilden, Germany) following manufacturer's instructions, and sequenced at the Sequencing and Genotyping Facility of Kansas State University (Manhattan, KS) using a 3730 DNA Analyzer (Applied Biosystems, Foster City, CA).

Sequence alignment and data analysis

Nucleotide sequences were aligned in Clustal Omega (Version 1.2.1, Sievers et al. 2011) and were further refined manually using BioEdit Sequence Alignment Editor (Version 7.2.5, Hall 1999). Molecular diversity indices such as number of variable sites (S), number of haplotypes identified (h), haplotype diversity (H_d) (Nei and Tajima 1983), and

nucleotide diversity (π) (Jukes and Cantor 1969) were calculated using DnaSP (Version 2.10.01, Librado and Rozas 2009). Haplotype diversity measures the probability of difference between the two randomly chosen haplotypes (Nei 1987). Similarly, nucleotide diversity refers to the average number of nucleotide differences between two randomly chosen nucleotide sequences (Nei and Li, 1979). Both haplotype diversity and nucleotide diversity are useful diversity metrics for biodiversity assessment (Goodall-Copestake et al. 2012).

A phylogenetic tree of populations was constructed by Bayesian inference (BI) in MrBayes (Version 3.2.6, Haelsenbeck and Ronquist 2001). The Markov Chain Monte Carlo (MCMC) algorithm was run for 20 million generations with four incrementally heated chains, starting from random trees and sampling one out of every 1000 generations. The first 25% of generations were discarded as burn-in, after which the chains appeared to become stationery. A Bayesian consensus tree was constructed using the remaining trees with their respective posterior probabilities using FigTree (Version 1.4.2, Rambaut 2014). Analysis of molecular variance (AMOVA) and pairwise PhiPT (Φ_{pt}) values were estimated using GenAlEx (Version 6.5, Peakall and Smouse 2012). Significance was evaluated by 9,999 random permutations of sequences.

Results

In this study, the total combined length of the aligned sequences of the two cpDNA markers (*rpl32-trnL*$^{(UAG)}$ and *trnQ*$^{(UUG)}$*-rps16*) was 1609 bp. The length of the *rpl32-trnL*$^{(UAG)}$ spacer in *A. gerardii* was 529 bp with an A/T content of 74.3%. Similarly, the length of the *trnQ*$^{(UUG)}$*-rps16* spacer in this study was 1125 bp with an A/T content of 69.5%. The *trnQ*$^{(UUG)}$*-rps16* spacer had more polymorphic sites (78 sites) compared to the *rpl32-trnL*$^{(UAG)}$ spacer (17 sites). A potential informative character value (PIC) was calculated by dividing the number of substitutions and indels by the total number of characters. When combined, a total of 100 variable sites with a PIC value of 6.22% were found. Of the two investigated cpDNA spacers, *rpl32-trnL*$^{(UAG)}$ and *trnQ*$^{(UUG)}$*-rps16* spacers had a PIC of 3.33% and 7.18%, respectively.

A total of 68 haplotypes were identified from 191 individuals of 24 *A. gerardii* populations (Table 2). Haplotype 40 was the most common, and was shared by 48 individuals (individuals from all populations except from South Dakota). The number of haplotypes (h) and haplotype diversity (H_d) identified in *rpl32-trnL*$^{(UAG)}$, *trnQ*$^{(UUG)}$*-rps16*, and combined were 11, 59, and 68, and 0.417, 0.872, and 0.912. Similarly, nucleotide diversity (π) was calculated as 0.00127, 0.00331, and 0.00283 for *rpl32-trnL*$^{(UAG)}$, *trnQ*$^{(UUG)}$*-rps16*, and cpDNA combined. Haplotype diversity within populations ranged from 0.643 to 1.000. H_d was highest for KS-8, MO-3, and MO-5 populations with the value of 1.000, and was lowest for KS-10 with a value of 0.643 (Table 2). Similarly, the nucleotide diversity varied from 0.00126 to 0.01184. The NE-7 population had the highest nucleotide diversity (0.01184) and IL-10 had the lowest (0.00126) (Table 2).

Consensus trees, accompanied with posterior probability (PP), based on *rpl32-trnL*$^{(UAG)}$, *trnQ*$^{(UUG)}$*-rps16*, and combining both cpDNA spacers were generated using BI method. The tree based on the *rpl32-trnL*$^{(UAG)}$ spacer provided 2 major clades (Figure 2). Among the two clades, the smaller clade consisted of IL-1, CO-3, IL-10, KAW-1, and NE-7 populations and the remaining populations combined to make the larger clade. Three populations from Illinois, namely IL-2, IL-7, and IL-8, were in a single clade, indicating these populations might have common ancestry. The clade was supported with a posterior probability value of 95. However, the populations were not well distinguished based on their geographic distribution. The Kaw cultivar, which was developed in northeastern Kansas (Saha et al. 2013), was grouped with CO-3 and IL-10 populations.

Similarly, the *trnQ*$^{(UUG)}$*-rps16* spacer generated a tree with 2 major clades (Figure 3). The smaller clade consisted of several populations, namely IL-10, IA-2, KAW-1, KS-8, KS-10, NE-5, and OK-1. The Kaw-1 cultivar was in the same clade with the KS-10 population, from the Konza Prairie. The KS-8 population was also in the same clade. Similarly, MN-3 and MO-3 populations were clustered together with a posterior probability value of 96 within the larger clade. All Illinois populations except for IL-10 were grouped together.

The tree using combined noncoding cpDNA spacers generated a monophyletic tree with CO-3 and NE-7 populations acting as the outgroups (Figure 4). The monophyletic group was further resolved into two distinguishable sub-clades. MN-3 and MO-3 populations were closely related, supported by a posterior probability value of 100. ND-1 and OK-1 populations were in a single clade that was supported by a posterior probability of 98. Most

of the branches and nodes of the trees were well supported (posterior probability = 70% to 100%). However, some clades had lower posterior probability values of 50% to 70%.

Analysis of molecular variance (AMOVA), based on the haploid data, was calculated using GenAlEx and Φ_{pt} values were estimated. Φ_{pt}, which is similar to Fst, estimates the proportion of the variance among populations relative to the total variance. AMOVA analysis indicated significant genetic differentiation among the populations across the Great Plains (Φ_{pt}=0.099, $p<0.001$). The analysis also showed 90% of variation occurred within the populations and 10% variation among the populations. Pairwise Φ_{pt} values indicated CO-3, NE-7, and NE-8 populations had more genetic differentiation compared to other populations.

Discussion

Knowledge of phylogeny at low and high taxonomic levels has tremendous implication for understanding biological diversity and evolution. Many life science researchers take the Tree of Life as a center of many areas of biology. Consequently, placing an organism of interest in an appropriate phylogenetic context provides better knowledge of evolutionary patterns and processes (Soltis and Soltis 2003). We can understand evolution of organisms, their traits, and interactions between species through the development of robust phylogeny (Wiens 2000). Understanding phylogeny is also beneficial for crop improvement (Soltis and Soltis 2000), pest and pathogen management (Gilbert et al. 2012), germplasm conservation (Kellogg et al. 1996), medicinal plant identification (Asahina et al. 2010), and conservation of biodiversity (Yuan et al. 2011). The establishment of phylogeny provides valuable information about effects of habitat, geography, pollinators, and other ecological factors on evolution of plants (Acevedo-Rosas et al. 2004).

A. gerardii is one of the most economically important grasses of the prairie. Identification of the best suitable local ecotype(s) can contribute to the success of restoration projects. Along with this, the relationships of ecotypes might be helpful to identify the ecotypes used for the production of biofuels. Zhang et al. (2014) studied the potential of production of biofuel from *A. gerardii*, and estimated ethanol yield of 1886 L/ha. Zhang et al. (2015) compared the biofuel yield of *A. gerardii* with other native grasses, and found that 0.26 kg of ethanol with 9.4 g/L concentration can be produced from 1 kg of raw *A. gerardii*. In the same study, Zhang et al. (2015) calculated average ethanol yield

of 26.2%± 1.3% of dry mass in *A. gerardii*, compared to 21.7± 0.6%, 20.2± 1.0%, and 21.1± 1.0% of dry mass in switch grass, *Miscanthus*, and Conservation Reserve Program (CRP) grass, respectively. Identification of locally-adapted populations will be helpful for farmers or ranchers to produce forage for cattle, for production of biofuel, and for restoration efforts.

Reintroducing proper genetic material plays an important role in the success of restoration projects that are focused on restoring the ecosystem function and structure at degraded sites (Baer et al. 2014). Accordingly, two cpDNA spacers (*rpl32-trnL*$^{(UAG)}$ and *trnQ*$^{(UUG)}$*-rps16*) were used to examine genetic diversity among 24 populations of *Andropogon gerardii*. A phylogenetic tree was constructed using cpDNA spacers to study the evolutionary relationship among *A. gerardii* populations. Even though the phylogenetic trees did not distinguish populations according to their geographical distribution, they provided valuable information about relationships among populations. Additionally, the genetic differentiation within and among populations was studied. There is low level of genetic differentiation among the populations across the Great Plains, with 90% of genetic differentiation within the populations.

Genetic structure of *A. gerardii* populations

The length of the *rpl32-trnL*$^{(UAG)}$ spacer in *A. gerardii* was aberrantly small at 529 bp. The length ranges from 543-1417 bp with an average length of 1018 bp across spermatophytes (Shaw et al. 2007). But, the spacer length in *A. gerardii* is similar to that of *Zea mays* (531 bp) and *Sorghum bicolor* (522 bp). Similarly, the length of the *trnQ*$^{(UUG)}$*-rps16* spacer in

this study was 1125 bp. The length ranges from 588-1975 bp with an average length of 1046 bp across spermatophytes (Shaw et al. 2007). In addition, a total of 18, 68, and 81 one and two variants singleton variable sites were found in *rpl32-trnL*$^{(UAG)}$, *trnQ*$^{(UUG)}$*-rps16*, and cpDNAs combined, respectively. In this study, *trnQ*$^{(UUG)}$*-rps16* had higher variable sites and PIC value than the *rpl32-trnL*$^{(UAG)}$ spacer, contradicting previous results of other studies. Cires et al. (2012) showed that the *Ranunculus parnassiifolius rpl32-trnL*$^{(UAG)}$ spacer has more than 3 times more polymorphic sites compared to *trnQ*$^{(UUG)}$*-rps16.* In addition, Shaw et al. (2007) reported the *rpl32-trnL*$^{(UAG)}$ spacer to have a higher average PIC value than the *trnQ*$^{(UUG)}$*-rps16* spacer in spermatophytes. In the present study, both of the spacers were rich in A/T content with an average of 69.7%. The nucleotide composition of noncoding spacers and pseudo-genes are A/T rich as they provide low levels of functional constraints (Li 1998).

The study of genetic differentiation among and within *A. gerardii* populations could be helpful to study the spatial distribution of populations. By using combined data of *rpl-trnL* and *trnQ*$^{(UUG)}$*-rps* spacers, genetic structure of *A. gerardii* populations across the Great Plains was estimated. Although the differentiation of cpDNA was significant, differentiation was at low level (Φ_{pt}=0.0999, p<0.001). AMOVA analysis further revealed the majority of variation was within the populations (90%) compared to among populations (10%). Such high diversity within the populations of *A. gerardii* was also observed in other studies. Gustafson et al. (1999) examined genetic diversity among remnant *A. gerardii* populations in Arkansas using Random Amplified Polymorphic DNA (RAPD) analysis, and showed 89% of total genetic variation within the populations. Tompkins et al. (2012)

examined genetic diversity within and among nine *A. gerardii* populations in the Carolinas and found high genetic diversity within the populations (82.6%). Price et al. (2012) used amplified fragment length polymorphism (AFLP) to study the genetic variation of *A. gerardii* in three distinct populations, and found 86% variation within the populations. Gray et al. (2014) detected variation ranging from 84% to 92% within big bluestem ecotypes using an AFLP method.

There are several possible reasons for the observed high genetic variation within *A. gerardii* populations. In the tallgrass prairie, suitable habitats are common or continuous, which makes gene flow or dispersal common among populations. This results in low variation among the populations and high variation within the populations (Falk et al. 2001). Another reason for high genetic diversity within populations could be the ploidy level (Tompkins et al. 2012, Gray et al. 2014). *A. gerardii* demonstrates polyploidy in which hexaploidy (6x=*2n*=60; Gould 1967) and octoploidy (8x=*2n*=80; Keeler 1992) are common (Keeler 2004). There are also populations of heptaploids (7x; *2n*=70) and nonaploids (9x; *2n*=90) across the tallgrass prairie (Keeler and Davis 1999). Additionally, *A. gerardii* is highly self-incompatible (Norrmann et al. 1997), and demonstrates an obligate outcrossing nature (Gustafson et al. 1999; Price et. al 2012). This could be one of the reasons for high variation within populations in this study.

Phylogenetic relationship between the *A. gerardii* populations based on noncoding cpDNA

The phylogenetic analysis based on noncoding cpDNA (*rpl32-trnL*$^{(UAG)}$; *trnQ*$^{(UUG)}$*-rps16*) provides the first available phylogenetic construction of the *A. gerardii* populations. Both noncoding cpDNA spacer data were combined to reconstruct the phylogenetic tree of the populations, as no single noncoding cpDNA is universally best for low-level molecular studies (Shaw et al. 2014). The first notable feature of the tree constructed from combined data was the outgroups, CO-3 and NE-7. The CO-3 population was genetically different from most of the populations in this study. The reason for such differentiation could be geographical isolation based on the elevation. The CO-3 population was collected from an elevation of 1,737 m above sea level, whereas all other populations were collected from elevations below 1,000 m. The difference in the elevation might have resulted in limited gene flow. The NE-7 population was *A. hallii* (sand bluestem), collected from the Sandhills of Nebraska. *A. hallii* is a C_4 grass and closely related to *A. gerardii* cytologically and morphologically (Peters and Newell 1961). Some researchers recognize *A. hallii* and *A. gerardii* as distinct species whereas some researchers recognize them as varieties within a single species (Uchytil 1988). However, Peters and Newell (1961) demonstrated *A. gerardii* and *A. hallii* are interfertile with the hybrid showing intermediate characteristics. However, these two bluestems showed ecological divergence despite the similarities between them (Barnes 1985, 1986). *A. gerardii* is distributed in moist areas whereas *A. hallii* is distributed in the sandy regions (Barnes 1986).

Another notable feature in the tree was the position of the cultivars KAW-1 and IL-8. The KAW-1 cultivar was developed in northeastern Kansas, which is geographically near the origin of the KS-10 population, collected from Konza Prairie Biological Station. Despite the geography, the KAW-1 cultivar had a common ancestor with the IL-10 population, with a clade supported by a posterior probability of 100. This result is further supported by the results of pairwise AMOVA analysis, where IL-10 and KS-10 were moderately differentiated (Φ_{pt}= 0.117, $p<0.05$). Pairwise AMOVA analysis (Table 4) also demonstrated that KAW-1 and KS-10 populations were also moderately differentiated (Φ_{pt}= 0.135, $p<0.05$), but no differentiation was detected between IL-10 and KAW-1 populations (Φ_{pt}= 0.000, $p>0.05$).

Gray et al. (2014) suggested the Kansas and Southern Illinois populations consist of distinct genetic structure. However, IL-10 is an Eastern Illinois population. Similarly, the position of IL-8 is also uncertain. IL-8 is a cultivar developed by the USDA-Natural Resources Conservation Service. This cultivar is known as Rountree, initially collected near Moorhead, Iowa. Later, seeds were evaluated in different locations of Iowa, Missouri, and Southern Illinois. As indicated by the phylogenetic tree, the cultivar is grouped with the MI-1 population. However, this clade is not well supported (posterior probability = 59). Within Illinois populations, IL-8 cultivars are closely related with the IL-7 population. The IA-2 population was the only population in this study from Iowa. Interestingly, the IA-2 population is not genetically different from Rountree cultivar (Φ_{pt}=0.000, $p>0.05$). Price et al. (2012) hypothesized the cultivars were developed either by collecting seeds from a number of natural populations within a specific geographic region or mixing seeds from

plants having different traits, which could result in relationships different from those expected from geography.

The groupings of NE-5 and KS-10 (posterior probability = 88), ND-1 and OK-1 (posterior probability = 98), and MO-3 and MN-3 (posterior probability = 100) do not support the resolution of the phylogenetic tree with geographical distribution. However, this can be partially explained by the longitudinal distribution of populations. All of the most closely-related populations are at similar longitudinal degrees compared to each other. Additionally, there is a strong precipitation gradient across the Great Plains, as precipitation increases when moving from west to east (Mendola et al. 2015). The precipitation does not differ at similar longitude. In addition, abnormalities in the tree could be the result of unintentional transfer of germplasm. Zhang et al. (2011), in his study on *Panium virgatum*, hypothesized that East Coast populations were established due to transportation of prairie hay from the Sandhills of Nebraska. Similarly, abnormalities seen in this study could be attributed to lack of informative markers. Lack of resolution in developing phylogenetic tree is a widespread problem (Hughes et al. 2006). In this study, two cpDNA markers were used to assess the relationship among the *A. gerardii* populations. However, it is always better to use markers of high variability which can generate phylogeny in a higher level of resolution (Shaw et al. 2007). Another reason for abnormalities could be strong wind contributing to seed dispersal, and hence promoting gene flow. The Great Plains has both free sweeps of northerly and southerly winds that predominate as far as Texas in the south to Montana in the north (Rosenberg 1987). There

is also possibility of gene flow via pollen but, cpDNA analysis estimates gene flow via seeds as cpDNA is maternally inherited, and has haploid features (Yuan et al. 2011).

Concluding remarks and future recommendations

Despite high levels of genetic diversity within the *A. gerardii* populations, small populations are likely to be at risk of extinction because of habitat fragmentation, inbreeding depression, and stochastic processes (Yuan et al. 2011, Gray et al. 2014). Study of genetic diversity can be useful for investigating gene flow between the populations, which could affect adaptability. In addition, this kind of study could help restoration ecologists choose seeds appropriately (Gustafson et al. 2001), improve *A. gerardii* as a forage crop and a biofuel feedstock (Johnson et al. 2015), and identify unique germplasm (Price et al. 2012). Proximity in geography does not imply genetic similarity. Hence, designing the best fit for restoration projects should not be only based upon geographic closeness (Gustafson et al. 1999).

Finding evolutionary relationships within *A. gerardii* by using cpDNA is a challenge. Nuclear DNA information in combination with plastid information could be helpful in resolving the relationships at an infraspecific level, even though plastid DNA can sometimes provide more information than nuclear DNA (Domenech et al. 2014). In this study, various genetic indices based on two noncoding cpDNA spacers were studied. In addition, phylogenetic trees that reflected the evolutionary relationship among the populations were constructed. However, further study using more numbers of noncoding

cpDNA along with nuclear markers is recommended for verification, and understanding of the evolutionary relationship among *A. gerardii* populations.

Literature cited

Acevedo-Rosas R, Cameron K, Sosa V, Pell S (2004) A molecular phylogenetic study of *Graptopetalum* (Crassulaceae) based on ETS, ITS, *rpl 16*, and *trn L-F* nucleotide sequences. Am J Bot 91:1099–1104

Asahina H, Shinozaki J, Masuda K, et al. (2010) Identification of medicinal *Dendrobium* species by phylogenetic analyses using *matK* and *rbcL* sequences. J Nat Med 64:133–138. doi: 10.1007/s11418-009-0379-8

Atherton RA, McComish BJ, Shepherd LD, et al. (2010) Whole genome sequencing of enriched chloroplast DNA using the Illumina GAII platform. Plant Meth 6:22. doi: 10.1186/1746-4811-6-22

Axelrod DI (1985) Rise of the grassland biome, central North America. Bot Rev 51:163–201

Baer SG, Gibson DJ, Gustafson DJ, et al. (2014) No effect of seed source on multiple aspects of ecosystem functioning during ecological restoration: Cultivars compared to local ecotypes of dominant grasses. Evol Appl 7:323–335. doi: 10.1111/eva.12124

Barnes PW (1985) Adaptation to water stress in the big bluestem-sand bluestem complex. Ecology 66:1908–1920. http://doi.org/10.2307/2937387

Barnes PW (1986) Variation in the big bluestem (*Andropogon gerardii*)-sand bluestem (*Andropogon hallii*) complex along a local dune/meadow gradient in the Nebraska Sandhills. Am J Bot 73:172-184

Broadhurst LM, Lowe A, Coates DJ, et al. (2008) Seed supply for broadscale restoration: maximizing evolutionary potential. Evol Appl 1:587–597

Caudle KL, Johnson LC, Baer SG, Maricle BR (2014) A comparison of seasonal foliar chlorophyll change among ecotypes and cultivars of *Andropogon gerardii* (Poaceae) by using nondestructive and destructive methods. Photosynthetica 52:511–518. doi: 10.1007/s11099-014-0057-2

CBOL Plant Working Group, Hollingsworth PM, Forrest LL, et al. (2009) A DNA barcode for land plants. Proc Natl Acad Sci 106:12794–12797. doi: 10.1073/pnas.0905845106

Cires E, Cuesta C, Vargas P, Fernandez Prieto JA (2012) Unravelling the evolutionary history of the polyploid complex *Ranunculus parnassiifolius* (Ranunculaceae). Biol J Linn Soc 107:477–493. doi: 10.1111/j.1095-8312.2012.01968.x

Clegg MT (1993) Review and the study of plant evolution. Proc Natl Acad Sci 90:363–367

Clegg MT, Gaut BS, Learn GH, Morton BR (1994) Rates and patterns of chloroplast DNA evolution. Proc Natl Acad Sci 91:6795–801. doi: 10.1073/pnas.91.15.6795

Daniell H, Lee S-B, Grevich J, et al. (2006) Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analyses with other Solanaceae genomes. Theor Appl Genet 112:1503–1518. doi: 10.1007/s00122-006-0254-x

Diekmann K, Hodkinson TR, Fricke E, Barth S (2008) An optimized chloroplast DNA extraction protocol for grasses (Poaceae) proves suitable for whole plastid genome sequencing and SNP detection. PloS ONE 3:1–5. doi: 10.1371/journal.pone.0002813

Domenech B, Asmussen-Lange CB, Baker WJ, et al. (2014). A phylogenetic analysis of palm subtribe Archontophoenicinae (Arecaceae) based on 14 DNA regions. Bot J Linn Soc 175:469-481

Doyle JJ, Doyle JL (1978). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11-15

Doyle JJ, Davis JI, Soreng RJ, Garvin D, Anderson MJ (1992) Chloroplast DNA inversions and the origin of the grass family (Poaceae). Proc Natl Acad Sci 89:7722-7726

Falk DA, Knapp EE, Guerrant EO (2001) An introduction to restoration genetics. Soc Ecol Restor.

Falk DD, Richards CM, Montalvo AM, Knapp EE (2006) Population and ecological genetics in restoration ecology. In: Falk DA, Palmer MA, Zedler JB (eds). Foundations of Restoration Ecology, Island Press, Washington DC, pp 14–41

Gilbert GS, Magarey R, Suiter K, Webb CO (2012) Evolutionary tools for phytosanitary risk analysis: Phylogenetic signal as a predictor of host range of plant pests and pathogens. Evol Appl 5:869–878. doi: 10.1111/j.1752-4571.2012.00265.x

Goodall-Copestake WP, Tarling GA, Murphy EJ (2012) On the comparison of population-level estimates of haplotype and nucleotide diversity: a case study using the gene *cox1* in animals. Heredity 109:50–56. doi: 10.1038/hdy.2012.12

Gould FW (1967) The grass genus *Andropogon* in the United States. Brittonia 19:70-76. doi: 10.2307/2805085

Gray MM, St. Amand P, Bello NM, et al. (2014) Ecotypes of an ecologically dominant prairie grass (*Andropogon gerardii*) exhibit genetic divergence across the U.S. Midwest grasslands' environmental gradient. Mol Ecol 23:6011–6028. doi: 10.1111/mec.12993

Gustafson DJ, Gibson DJ, Nickrent DL (1999) Random amplified polymorphic DNA variation among remnant big bluestem (*Andropogon gerardii* Vitman) populations from Arkansas' Grand Prairie. Mol Ecol 8:1693–1701. doi: 10.1046/j.1365-294X.1999.00756.x

Gustafson DJ, Gibson DJ, Nickrent DL (2001) Characterizing three restored *Andropogon gerardii* Vitman (big bluestem) populations established with Illinois and non-Illinois seed: Established plants and their offspring. Proc 17th North Amer. Prairie Conference, pp. 118-124

Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Hall BK, Hallgrimsson B (2014) Strickberger's Evolution. 5th edn. Jones & Barlett Learning, Burlington, Massachusetts

Huelsenbeck JP, Ronquist F (2001). MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17:754-755

Hughes CE, Eastwood RJ, Bailey CD (2006) From famine to feast? Selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction. Philos Trans R Soc Lond B 361:211–25. doi: 10.1098/rstb.2005.1735

Jansen RK, Raubeson LA, Boore JL, et al. (2005) Methods for obtaining and analyzing whole chloroplast genome sequences. Meth Enzymol 395:348–384. doi: 10.1016/S0076-6879(05)95020-9

Johnson LC, Olsen JT, Tetreault H, et al. (2015) Intraspecific variation of a dominant grass and local adaptation in reciprocal garden communities along a US Great Plains' precipitation gradient: Implications for grassland restoration with climate change. Evol Appl 8:705–723. doi: 10.1111/eva.12281

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian Protein Metabolism. Academic Press, New York, pp 21–132

Kaur H, Singh BP, Singh H, Nagpal AK (2014) Comparative genomics of ten solanaceous plastomes. Adv Bioinform doi: 10.1155/2014/424873

Keeler KH (1992) Local polyploid variation in the native prairie grass *Andropogon gerardii*. Am J Bot 79:1229-1232. doi: 10.2307/2445049

Keeler KH, Davis GA (1999) Comparison of common cytotypes of *Andropogon gerardii* (Andropogoneae, Poaceae). Am J Bot 86:974–979. doi: 10.2307/2656614

Keeler KH (2004) Impact of intraspecific polyploidy in *Andropogon gerardii* (Poaceae) populations. Am Midl Nat 152:63–74. doi: 10.1674/0003-0031(2004)152[0063:IOIPIA]2.0.CO;2

Kellogg EA, Appels R, Mason-Gamer RJ (1996). When genes tell different stories: The diploid genera of Triticeae (Gramineae). Syst Bot 21:321–347. http://doi.org/10.2307/2419662

Kim S-C, Crawford DJ, Jansen RK, Santos-Guerra A (1999) The use of a non-coding region of chloroplast DNA in phylogenetic studies of the subtribe Sonchinae (Asteraceae: Lactuceae). Plant Syst Evol 215:85–99. doi: 10.1007/BF00984649

Korotkova N, Borsch T, Quandt D, et al. (2011) What does it take to resolve relationships and to identify species with molecular markers? An example from the epiphytic Rhipsalideae (Cactaceae). Am J Bot 98:1549–1572. doi: 10.3732/ajb.1000502

Ku C, Chung WC, Chen LL, Kuo CH (2013) the complete plastid genome sequence of Madagascar periwinkle *Catharanthus roseus* (L.) G. Don: Plastid genome evolution, molecular marker identification, and phylogenetic implications in Asterids. PLoS ONE 8:1–11. doi: 10.1371/journal.pone.0068518

Lauenroth WK, Burke IC, Gutmann MP (1999) The structure and function of ecosystems in the central North American grassland region. Gr Plains Res 9:223–259

Li JJ (1998) Studies on the origin of Chinese Mudan (tree peonies). J Beijing For Univ 20:22–26

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452. doi: 10.1093/bioinformatics/btp187

Loflin B, Loflin S (2006) Grasses of the Texas Hill Country: A Field Guide. Texas A&M University Press, Texas

Maliga P, Bock R (2011) Plastid biotechnology: food, fuel, and medicine for the 21st century. Plant Physiol 155:1501–1510. doi: 10.1104/pp.110.170969

McMillan C (1959) The role of ecotypic variation in the distribution of the central grassland of North America. Ecol Monogr 29:285-308

McMillan C (1964) Ecotypic differentiation within four North American prairie grasses. I. Morphological variation within transplanted community fractions. Am J Bot 51:1119-1128

McMillan C (1965) Ecotypic differentiation within four North American prairie grasses. II. Behavioral variation within transplanted community fractions. Am J Bot 52:55-65

Meisler MH (2001) Evolutionarily conserved noncoding DNA in the human genome: How much and what for? Gen Res 1617–1618. doi: 10.1101/gr.211401.munoglobin

Mendola ML, Baer SG, Johnson LC, Maricle BR (2015) The role of ecotypic variation and the environment on biomass and nitrogen in a dominant prairie grass. Ecology 96:2433–2445. doi: 10.1890/14-1492.1

Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press, New York

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci 76:5269–5273

Nei M, Tajima F (1983) Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. Genetics 105:207–217

Norrmann GA, Quarin CL, Keeler KH (1997) Evolutionary Implications of Meiotic Chromosome Behavior, Reproductive Biology, and Hybridization in 6X and 9X Cytotypes of *Andropogon gerardii* (Poaceae). Am J Bot 84:201-207. doi: 10.2307/2446081

Olsen JT, Caudle KL, Johnson LC, et al. (2013) Environmental and genetic variation in leaf anatomy among populations of *Andropogon gerardii* (Poaceae) along a precipitation gradient. Am J Bot 100:1957–1968. doi: 10.3732/ajb.1200628

Peakall R, Smouse PE (2012) GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537–2539. doi: 10.1093/bioinformatics/bts460

Peters LV, Newell LC (1961) Hybridization between divergent types of big bluestem, *Andropogon gerardii* Vitman, and sand bluestem *Andropogon hallii* Hack. Crop Sci 1:359-363

Price DL, Salon PR, Casler MD (2012) Big bluestem gene pools in the central and northeastern United States. Crop Sci 52:189–200. doi: 10.2135/cropsci2011.05.0280

Rambaut A (2014) FigTree: Tree figure drawing tool. http://tree.bio.ed.ac.uk/software/figtree/

Ravi V, Khurana JP, Tyagi AK, Khurana P (2008) An update on chloroplast genomes. Plant Syst Evol 271:101–122. doi: 10.1007/s00606-007-0608-0

Reid RS, Serneels S, Nyabenge M, Hanson J (2005) The changing face of pastoral systems in grass-dominated ecosystems of eastern Africa. Grasslands of the World. Rome, pp 19-76

Rosenberg, N. J. (1987). Climate of the Great Plains region of the United States. Great Plains Quarterly, Paper 344, pp 22–32

Saha MC, Bhandari HS, Bouton JH (2013) Bioenergy Feedstocks: Breeding and Genetics. John Wiley & Sons, Ames, Iowa

Samson F, Knopf F (1994) Prairie conservation in North America. Bioscience 44:418–421. doi: 10.2307/1312365

Samson FB, Knopf FL, Ostlie WR (2004) Great Plains ecosystems: past, present, and future. Wildl Soc Bull 32:6–15. doi: 10.2193/0091-7648(2004)32[6:GPEPPA]2.0.CO;2

Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. Am J Bot 94:275–288. doi: 10.3732/ajb.94.3.275

Shaw J, Shafer HL, Rayne Leonard O, et al. (2014) Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: The tortoise and the hare IV. Am J Bot 101:1987–2004. doi: 10.3732/ajb.1400398

Shi C, Hu N, Huang H, et al. (2012) An improved chloroplast DNA extraction procedure for whole plastid genome sequencing. PloS ONE 7:1–7. doi: 10.1371/journal.pone.0031468

Sievers F, Wilm A, Dineen D, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. doi: 10.1038/msb.2011.75

Soil Conservation Service (SCS) 1990 Conservation Reservation Program. U.S. Department of Agriculture, Washington, DC. National Bulletin No. 300-0-8

Soltis DE, Soltis PS (2000) Contributions of plant molecular systematics to studies of molecular evolution. Plant Mol Biol 42:45–75

Soltis DE, Soltis PS (2003) The role of phylogenetics in comparative genetics. Plant Physiol 132:1790–1800. doi: 10.1104/pp.103.022509

Stubbendieck JL, Hatch SL, Bryan NM (2011) North American Wildland Plants: A Field Guide. 2nd edn. University of Nebraska Press, Nebraska

Stubbs M (2014) Conservation Reserve Program (CRP): Status and Issues. 24.

Tompkins RD, Trapnell DW, Hamrick JL, Stringer WC (2012) Genetic variation within and among remnant big bluestem (*Andropogon gerardii*, Poaceae) populations in the Carolinas. Southeast Nat 11:455–468. doi: 10.1656/058.011.0309

Uchytil RJ (1988) *Andropogon hallii.* In: Fire effects information system, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory (Producer). http://www.fs.fed.us/database/feis/. Accessed 12 May 2016

United States Department of Agriculture (USDA) (2004) Plant Guide. USDA Website. https://plants.usda.gov/plantguide/pdf/pg_ange.pdf. Accessed 16 July 2016

Weaver JE (1954) North American Prairie. Johnsen Publishing Company, Lincoln, Nebraska

Weaver JE (1968) Prairie Plants and their Environment. University of Nebraska Press, Lincoln, Nebraska

Wiens JJ (2000) Phylogenetic analysis of morphological data. Smithsonian Institution Press, Washington, D.C.

Wu J, Liu B, Cheng F, et al. (2012) Sequencing of chloroplast genome using whole cellular DNA and Solexa sequencing technology. Front Plant Sci 3:1–7. doi: 10.3389/fpls.2012.00243

Yuan J hui, Cheng FY, Zhou SL (2011) The phylogeographic structure and conservation genetics of the endangered tree peony, *Paeonia rockii* (Paeoniaceae), inferred from chloroplast gene sequences. Conserv Genet 12:1539–1549. doi: 10.1007/s10592-011-0251-8

Zhang JM, Wang JX, Xia T, Zhou SL (2009) DNA barcoding: Species delimitation in tree peonies. Sci China, Ser C Life Sci 52:568–578. doi: 10.1007/s11427-009-0069-5

Zhang Y, Zalapa JE, Jakubowski AR, et al. (2011) Post-glacial evolution of *Panicum virgatum*: Centers of diversity and gene pools revealed by SSR markers and cpDNA sequences. Genetica 139:933–948. doi: 10.1007/s10709-011-9597-6

Zhang K, Johnson L, Yuan W, et al. (2014) Glucan yield from enzymatic hydrolysis of big bluestem as affected by ecotype and planting location along the precipitation gradient of the Great Plains. Bioenergy Res 7:799–810. doi: 10.1007/s12155-014-9477-x

Zhang K, Johnson L, Prasad PVV, et al. (2015) Comparison of big bluestem with other native grasses: Chemical composition and biofuel yield. Energy 83:358–365. doi: 10.1016/j.energy.2015.02.033

Table 1. PCR primer pairs used in the study to amplify noncoding cpDNA spacers rpl*32-trnL*$^{(UAG)}$ and *trnQ*$^{(UUG)}$*-rps16* in *A. gerardii* populations.

Spacer	Primer Pair	
	Forward	Reverse
rpl32-tnrL$^{(UAG)}$	TTTCTGGGCAACAAGCAAACAA	AAGAGCAGCGTGTCTACCAA
TrnQ$^{(UUG)}$*-rps16*	CGCACGTTGCTTTCTACCAC	CCCTGCACGACACAAAAGTT

Table 2. Values of h (number of haplotypes), H_d (Haplotype diversity), and π (Nucleotide diversity), based on the cpDNA (*rpl32-trnL*$^{(UAG)}$ and *trnQ*$^{(UUG)}$*-rps16*) of *A. gerardii* populations.

Population	**Sample Size**	**h**	**H_d**	**π**
CO-3	8	7	0.964	0.00745
IA-2	8	7	0.964	0.00351
IL-1	8	4	0.75	0.00265
IL-2	8	6	0.893	0.00193
IL-7	8	5	0.786	0.00198
IL-8	8	5	0.786	0.0015
IL-9	8	7	0.964	0.00352
IL-10	8	6	0.893	0.00126
KAW-1	8	6	0.893	0.00236
KS-10	8	4	0.643	0.00141
KS-8	8	8	1	0.00735
MI-1	8	6	0.893	0.00378
MN-3	8	7	0.964	0.00293
MO-3	8	8	1	0.00302
MO-5	8	8	1	0.00553
MO-7	8	6	0.893	0.00285
ND-1	8	6	0.929	0.00372

NE-5	8	6	0.929	0.00208
NE-6	8	6	0.893	0.00213
NE-7	7	6	0.952	0.01184
NE-8	8	6	0.929	0.00908
OK-1	8	5	0.857	0.00211
SD-1	8	5	0.786	0.00306
WI-6	8	6	0.929	0.00199
Total	191	68	0.912	0.00283

Table 3. Results of the analysis of molecular variance (AMOVA) based on the cpDNA spacers (*rpl32-trnL*$^{(UAG)}$ and *trnQ*$^{(UUG)}$*-rps16*) for 24 populations of *A. gerardii*.

Source	df	Sum of Squares	Mean Sums of Squares	Variation (%)
Among Pops	23	482.841	20.993	10%
Within Pops	167	1866.232	11.175	90%
Total	190	2349.073		100%
Stat	**Value**	**P**		
PhiPT	0.099	0.0001		

df: degrees of freedom

Table 4. Pairwise Φ_{pt} values along with P values among populations of *A. gerardii*.

Φ_{pt} Values are below diagonal. Probability, P(rand >= data) based on 9999 permutations is shown above diagonal. Significant Φ_{pt} values (P < 0.05) are in bold font.

	CO-3	IA-2	IL-1	IL-2	IL-7	IL-8	IL-9	IL-10	KAW-1	KS-8	KS-10	MI-1	MN-3	MO-3	MO-5	MO-7	ND-1	NE-5	NE-6	NE-7	NE-8	OK-1	SD-1	WI-6
CO-3	0.000	0.018	0.029	0.005	0.024	0.009	0.012	0.000	0.001	0.006	0.001	0.065	0.026	0.026	0.015	0.036	0.036	0.000	0.001	0.430	0.141	0.009	0.004	0.001
IA-2	**0.140**	0.000	0.322	0.362	0.030	0.349	0.365	0.214	0.376	0.029	0.059	0.356	0.373	0.375	0.354	0.390	0.356	0.055	0.398	0.015	0.001	0.344	0.319	0.350
IL-1	**0.132**	0.000	0.000	0.339	0.068	0.303	0.387	0.154	0.273	0.088	0.021	0.367	0.343	0.361	0.378	0.391	0.363	0.055	0.249	0.019	0.002	0.373	0.196	0.215
IL-2	**0.197**	0.000	0.005	0.000	0.070	0.338	0.197	0.250	0.255	0.019	0.007	0.349	0.384	0.381	0.386	0.399	0.354	0.016	0.245	0.006	0.001	0.257	0.116	0.104
IL-7	**0.154**	**0.120**	0.129	0.101	0.000	0.012	0.017	0.006	0.003	0.007	0.000	0.179	0.048	0.106	0.021	0.363	0.074	0.006	0.027	0.005	0.004	0.034	0.043	0.005
IL-8	**0.176**	0.000	0.000	0.006	**0.176**	0.000	0.381	0.149	0.408	0.016	0.010	0.374	0.346	0.361	0.331	0.305	0.340	0.001	0.241	0.007	0.000	0.329	0.182	0.240
IL-9	**0.168**	0.000	0.000	0.035	**0.176**	0.000	0.000	0.078	0.169	0.029	0.008	0.356	0.395	0.236	0.366	0.243	0.374	0.006	0.281	0.011	0.000	0.332	0.136	0.298
IL-10	**0.242**	0.043	0.082	0.042	**0.165**	0.084	0.128	0.000	0.364	0.005	0.023	0.039	0.162	0.468	0.484	0.108	0.249	0.089	0.435	0.000	0.001	0.171	0.070	0.304
KAW-1	**0.207**	0.000	0.014	0.040	**0.186**	0.002	0.059	0.000	0.000	0.002	0.032	0.066	0.245	0.338	0.365	0.116	0.374	0.005	0.184	0.002	0.001	0.292	0.102	0.278
KS-8	**0.128**	**0.115**	0.105	0.134	**0.173**	**0.157**	**0.149**	**0.140**	**0.162**	0.000	0.001	0.037	0.081	0.062	0.051	0.160	0.049	0.012	0.035	0.023	0.002	0.044	0.011	0.002
KS-10	**0.318**	0.115	**0.219**	**0.222**	**0.252**	**0.229**	**0.250**	**0.117**	**0.135**	**0.292**	0.000	0.008	0.004	0.008	0.006	0.004	0.042	0.006	0.052	0.000	0.001	0.002	0.182	0.041
MI-1	**0.092**	0.000	0.000	0.000	0.060	0.000	0.000	**0.130**	0.088	0.113	**0.239**	0.000	0.400	0.379	0.340	0.384	0.394	0.009	0.196	0.021	0.004	0.311	0.141	0.124
MN-3	**0.121**	0.000	0.000	0.000	**0.115**	0.000	0.000	0.061	0.023	0.083	**0.219**	0.000	0.000	0.404	0.379	0.402	0.359	0.009	0.434	0.015	0.001	0.384	0.266	0.346
MO-3	**0.125**	0.000	0.000	0.000	0.088	0.000	0.019	0.000	0.000	0.077	**0.182**	0.000	0.000	0.000	0.386	0.436	0.373	0.016	0.270	0.007	0.001	0.409	0.126	0.269
MO-5	**0.130**	0.000	0.000	0.000	**0.124**	0.000	0.000	0.000	0.000	0.078	**0.142**	0.000	0.000	0.000	0.000	0.400	0.353	0.002	0.315	0.015	0.001	0.382	0.169	0.388
MO-7	**0.105**	0.000	0.000	0.000	0.016	0.018	0.018	0.049	0.058	0.047	**0.165**	0.000	0.000	0.000	0.000	0.000	0.420	0.125	0.383	0.016	0.001	0.390	0.383	0.186
ND-1	**0.124**	0.000	0.000	0.000	0.101	0.000	0.000	0.032	0.000	0.103	**0.125**	0.000	0.000	0.000	0.000	0.000	0.000	0.043	0.399	0.017	0.001	0.366	0.332	0.341
NE-5	**0.267**	0.103	0.153	**0.127**	**0.151**	0.201	**0.205**	0.074	**0.161**	**0.147**	**0.131**	0.179	0.145	**0.126**	**0.131**	0.047	**0.102**	0.000	0.321	0.001	0.000	0.044	0.270	0.005
NE-6	**0.206**	0.000	0.024	0.024	**0.121**	0.028	0.026	0.000	0.036	0.104	**0.067**	0.050	0.000	0.020	0.000	0.000	0.000	0.015	0.000	0.001	0.001	0.383	0.301	0.331
NE-7	0.000	**0.131**	**0.127**	**0.176**	**0.187**	**0.168**	**0.157**	**0.235**	**0.200**	**0.109**	**0.308**	**0.107**	**0.125**	**0.137**	**0.128**	**0.107**	**0.125**	**0.241**	**0.198**	0.000	0.494	0.003	0.010	0.001
NE-8	0.043	**0.249**	**0.236**	**0.298**	**0.249**	**0.287**	**0.275**	**0.330**	**0.308**	**0.216**	**0.424**	**0.211**	**0.239**	**0.235**	**0.239**	**0.215**	**0.242**	**0.345**	**0.304**	0.000	0.000	0.002	0.002	0.001
OK-1	**0.149**	0.000	0.000	0.026	**0.116**	0.000	0.000	0.037	0.017	**0.090**	**0.174**	0.004	0.000	0.000	0.000	0.000	0.000	0.085	0.000	**0.154**	**0.245**	0.000	0.290	0.331
SD-1	**0.180**	0.000	0.056	0.056	**0.103**	0.060	0.046	0.067	0.076	**0.145**	0.052	0.048	0.030	0.054	0.032	0.000	0.000	0.031	0.000	**0.160**	**0.278**	0.010	0.000	0.368
WI-6	**0.221**	0.000	0.051	0.080	**0.186**	0.044	0.010	0.024	0.027	**0.180**	**0.111**	0.074	0.010	0.019	0.000	0.038	0.000	**0.152**	0.000	**0.218**	**0.327**	0.008	0.002	0.000

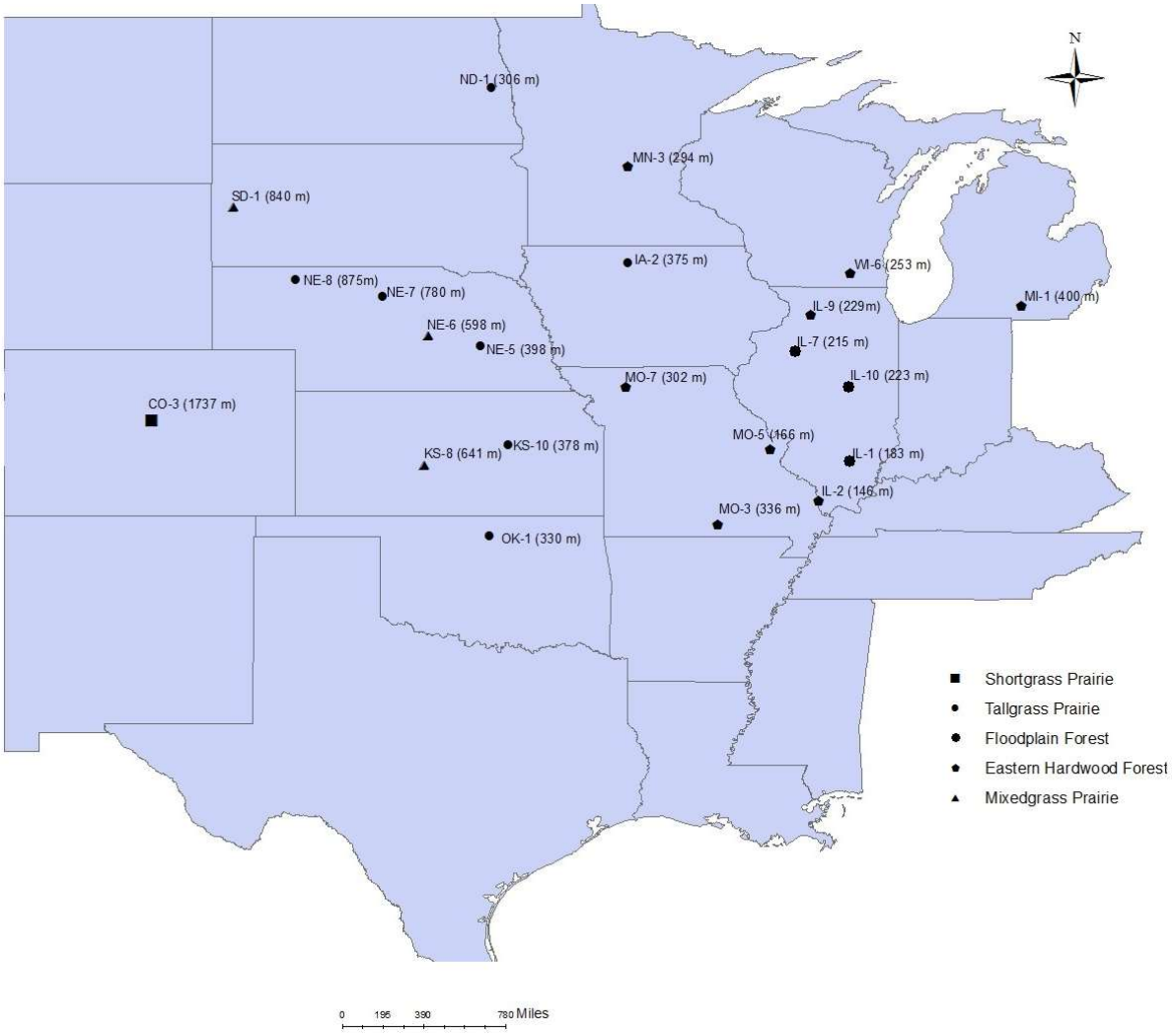


Figure 1. Map of the population locations of *A. gerardii* for cpDNA analysis, along with the elevation above sea level.

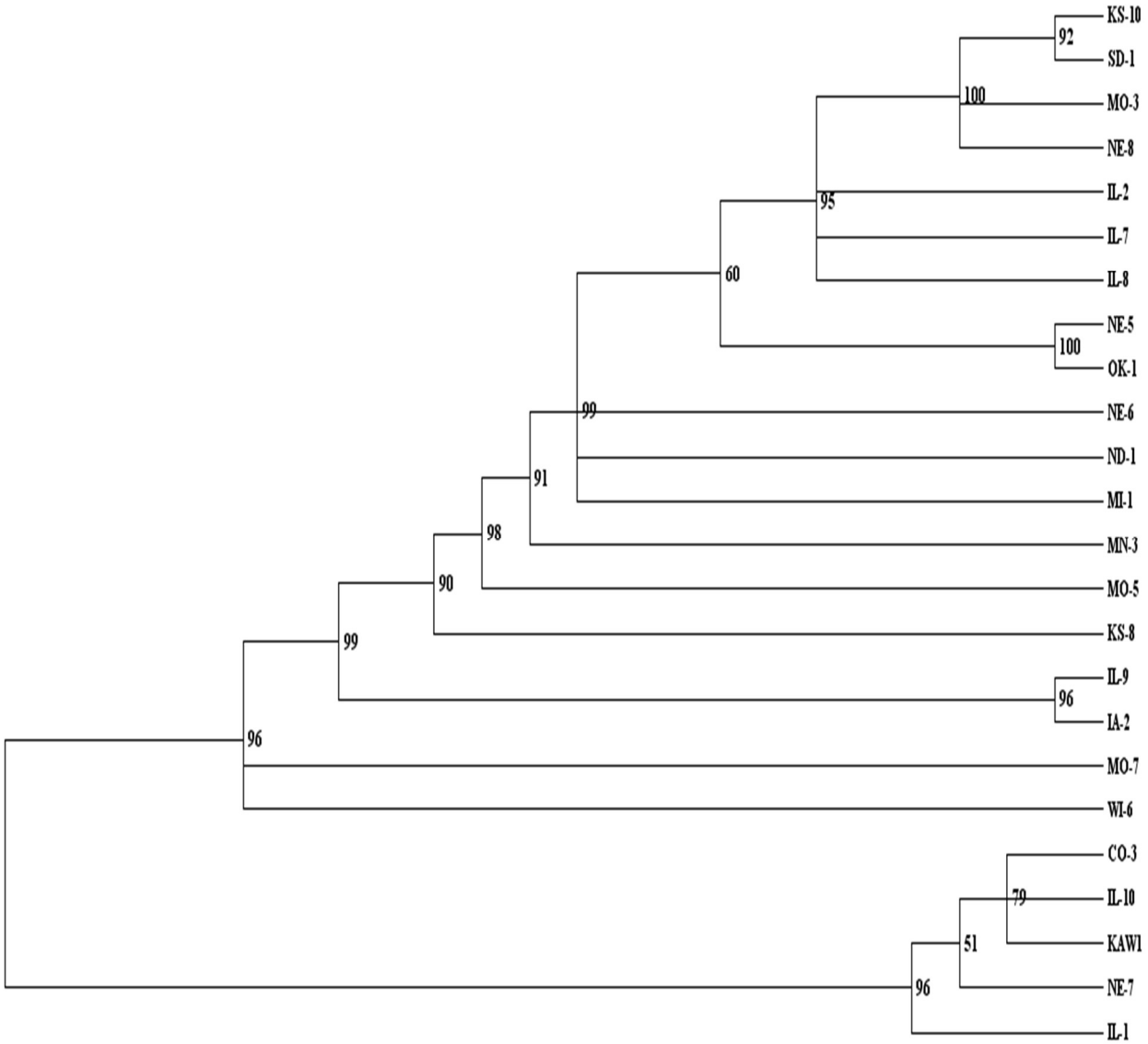


Figure 2. A Bayesian inference phylogenetic tree of populations of big bluestem (*A. gerardii*) based on the *rpl32-trnL*$^{(UAG)}$ cpDNA spacer using MrBayes 3.12 software. Numbers appearing at the nodes are posterior probabilities

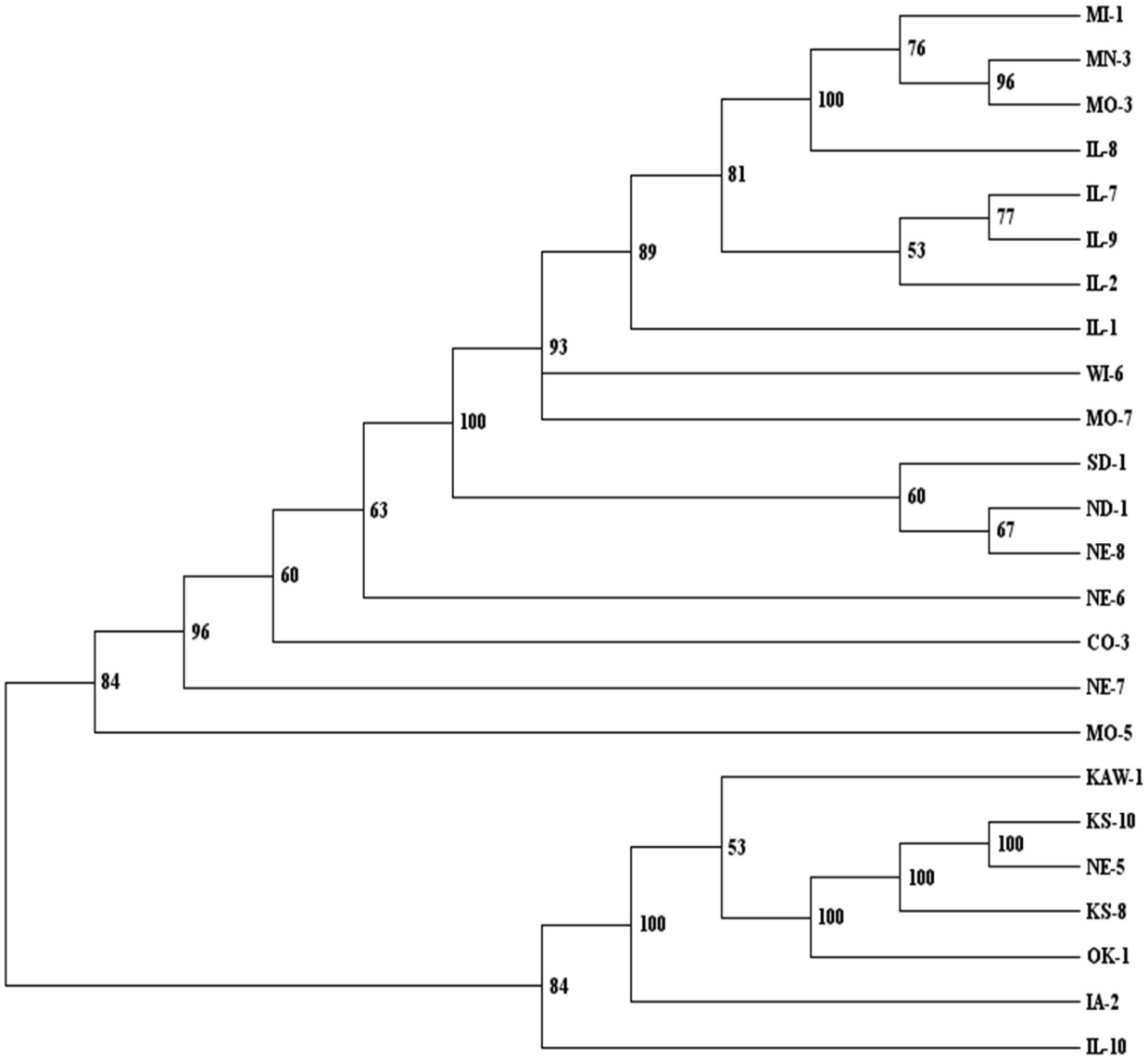


Figure 3. A Bayesian inference phylogenetic tree of populations of big bluestem (*A. gerardii*) based on *trnQ*$^{(UUG)}$*-rps16* cpDNA spacer using MrBayes 3.12 software. Numbers appearing at the nodes are posterior probabilities

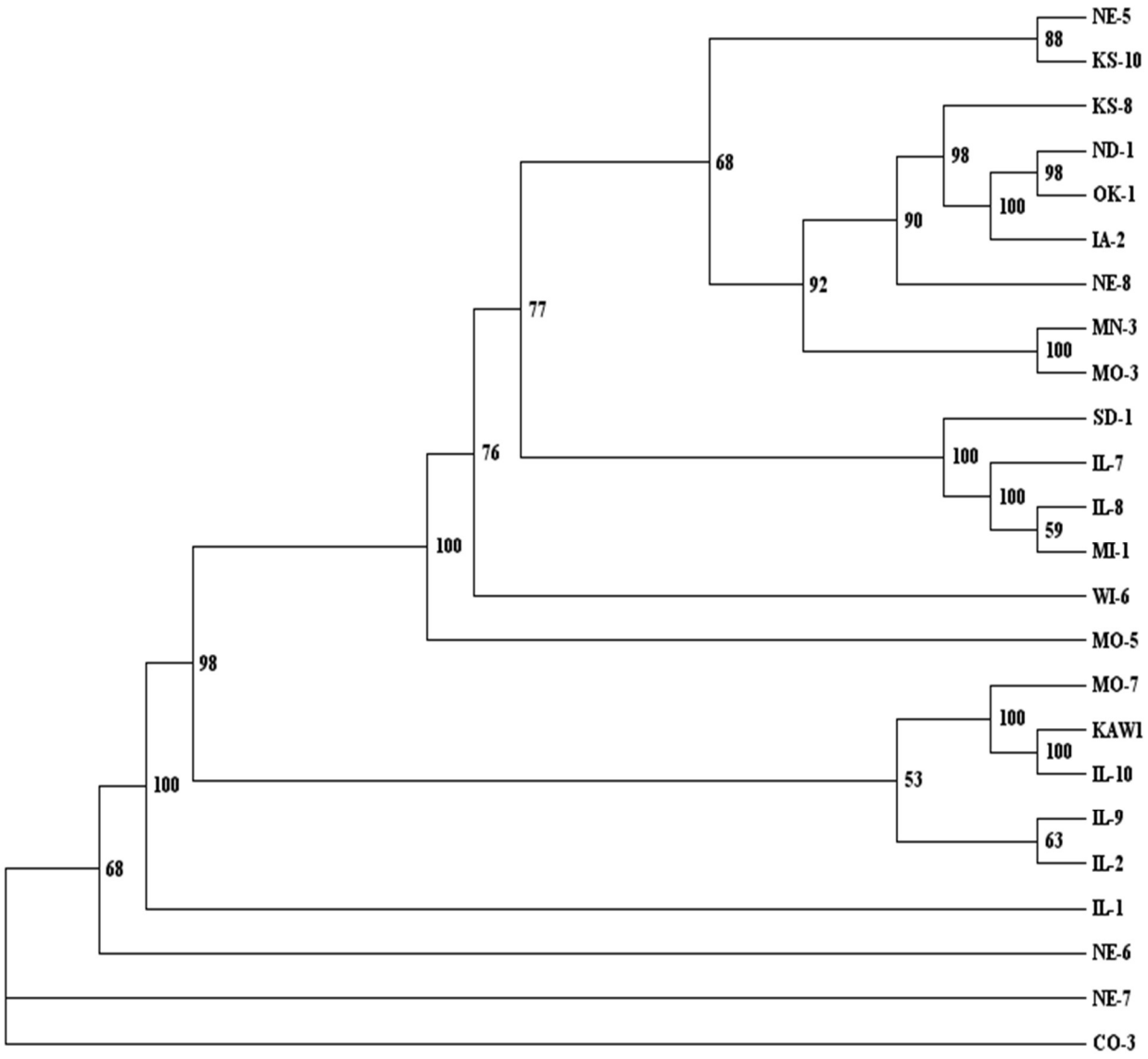
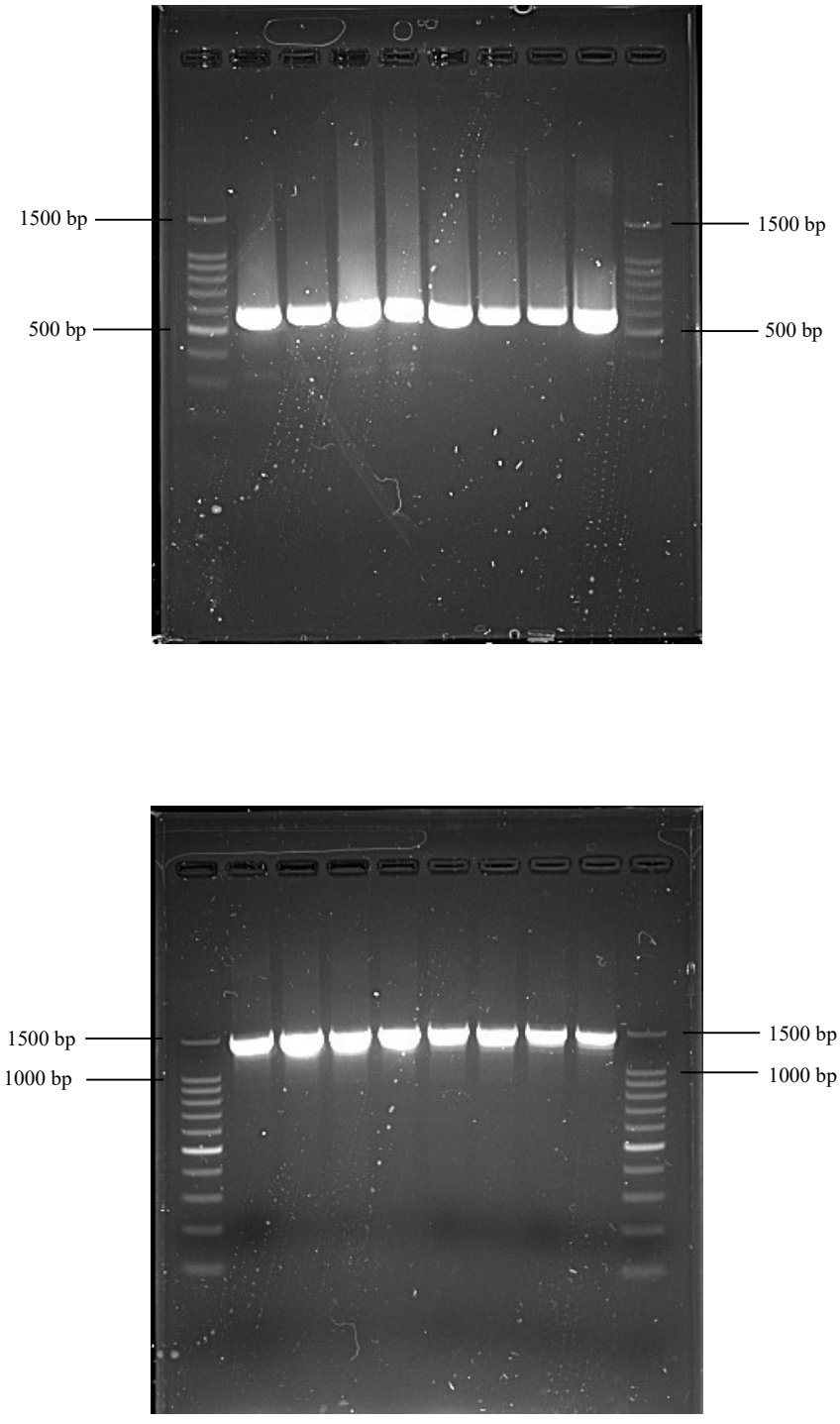


Figure 4. A Bayesian inference phylogenetic tree of populations of big bluestem (*A. gerardii*) based on both cpDNA spacers (*rpl32-trnL*$^{(UAG)}$ and *trnQ*$^{(UUG)}$*-rps16*) using MrBayes 3.12 software. Numbers appearing at the nodes are posterior probabilities.

Appendix 1. Electrophoresis gel pictures (top: *rpl32-trnL*$^{(UAG)}$ spacer and bottom: *trnQ*$^{(UUG)}$*-rps16* spacer) in *A.gerardii*.



Appendix 2. Populations of *A. gerardii* in a common garden established by Johnson et al. at Hays, KS (38.85 °W, -99.32 °N).

Photographed by Tej Man Tamang

Picture courtesy: Johnson et al.