DNA DIVERISTY IN PRAIRIE CORDGRASS (*SPARTINA PECTINATA* LINK*)* POPULATIONS INDIGENOUS TO THE UNITED STATES

BY

HANNAH ELIZABETH GRAVES

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Crop Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Master's Committee:

Associate Professor D.K. Lee, Co-Chair
Professor A. Lane Rayburn, Co-Chair
Assistant Professor Yoshie Hanzawa

ABSTRACT

Plant diversity, more specifically genetic diversity, is important for the environment and essential as an economic and social resource. Genetic variation assessment has allowed for contributions to many fields, and enables plant breeders to select plant varieties that are more suited to the current needs of the diverse agriculture system. Breeders can also monitor genetic diversity and select better crop varieties using polymorphisms in DNA. Prairie cordgrass (*Spartina pectinata* Link) has many uses such as conservation and wildlife habitat, and is currently being developed as a bioenergy crop. This species is a C4, rhizotomous, perennial grass, native to the North American Prairie. In order to utilize this species one must understand the genetic diversity found within and among native populations. Chloroplast DNA and nuclear DNA analysis provides a basic understanding into the amount of genetic variation present in this species. In this study, two highly polymorphic chloroplast regions were sequenced across 16 natural populations, in addition SNP markers were developed and tested on 38 natural populations of prairie cordgrass. Most of the variation (SNPs and indels) found in chloroplast DNA were observed in the octoploid cytotype with a few variants observed in the tetraploid cytotype, and no variants observed in the hexaploid cytotype. To delve into nuclear polymorphisms, SNP markers were developed from a complied transcriptome and were used validated putative SNPs. Also, the SNPs ability to distinguish true crosses and selfs during a breeding program was assessed. Variation identified in the nuclear transcriptome occurred equally between octoploid and tetraploid populations with less variation observed in the hexaploid population. In addition, χ^2 analysis was performed on an F_1 population and determined that the individuals followed normal Mendelian inheritance patterns and could not be

rejected for the markers chosen. This information is needed to open up the ability to understand genetic diversity of this species and for use in the development of a molecular based breeding program.

To my Mom and Dad

ACKNOWLEDGMENTS

This project would not have been possible without the help and support of a multitude of people. I would first like to thank my advisors Dr. D.K. Lee and Dr. A. Lane Rayburn for allowing me to expand my knowledge of crop science and develop many skills that I would not have if it were not for them. I also would like to thank Dr. Yoshie Hanzawa for serving on my committee and providing me with her expertise on evolutionary genetics. Many thanks to the Richard H. & Elizabeth C. Hageman Endowed Graduate Research Fellowship, Energy Bioscience Institute, and the Department of Crop Sciences at the University of Illinois Urbana-Champaign for funding me throughout this project. To my fellow graduate students, Su Min, Santanu, Joe, Jeff, Jia, and Moonsub, who have provided their support and insight, in addition to helping develop my skills as a baker by eating everything I put before them.

A big thank you goes out to my parents who raised me to follow my dreams, try my hardest, and never give up. I cannot express how thankful I am for all the love they have given to me. Thank you to Wesley, Ethan, and Abigail, for being the best brothers and sister that anyone could have ever asked for, and always keeping me on my feet. Thanks to my furry friends, Toby, Skeeter and most of all Peanut, who provided their cuteness for times when I just needed a good cuddle. I would like to express my deepest thanks to my husband for being there for me to lean on every day, and each day hereafter. Finally and most importantly, I give thanks to my Creator, God, for guiding me to this point in my life and down the future path he has planned for me.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Plant diversity is important for the environment and is essential as an economic and social resource (Henry 2005). Plant diversity can be measured on many different levels and can be evaluated using many different criteria such as phenotypic variation and genotypic variation (Henry 2005). Phenotypic variation is essential for selection and natural adaptation to specific environments, whereas genotypic variation provides a basis for understanding the genetics behind this phenotypic variation (Henry 2005). Assessing genetic variation in plant populations has allowed for contributions to many fields of studies including, evolutionary biology, conservation genetics, plant breeding, and ecological genetics (Coates & Byrne, 2005). In agriculture systems, genetic resources allow plant breeders to create plant gene combinations and select crop varieties that are better suited to the current needs of the diverse agriculture systems (Glaszmann et al., 2010). Genome wide analysis tools provide access to the thousands of polymorphisms that allow these plant breeders to broaden the capacity to monitor genetic diversity, create these plant gene combinations, and select the better crop varieties (Glaszmann et al., 2010).

Prairie cordgrass (*Spartina pectinata* Link), is currently being researched for use as a potential bioenergy crop, and has a multitude of other uses including conservation practices, wetland revegetation, streambank stabilization, wildlife habitat, and forage (Hitchcock 1950; Barkworth et al. 2007; Montemayor et al. 2008; Gonzalez-Hernandez et al. 2009; Boe et al. 2013; Kim et al. 2011; Zilverberg et al. 2014). In order to utilize this, warm season, C4, rhizotomous, perennial grass species, native to the North American Prairie, one must look at the

genetic diversity within and among natural populations. Prairie cordgrass is thought to have various levels of genetic diversity due to its wide geographic distribution within North America, and its tendency to be located in a wide range of environmental conditions such as: moderate salinity, water logged soils, drought, and mild cold temperatures (Montemayor et al. 2008; Boe et al. 2009; Gonzalez-Hernandez et al. 2009; Kim et al. 2011; Zilverberg et al. 2014). In addition to environmental adaptability, there is also the influence of polyploidy on genetic diversity of prairie cordgrass. Three cytotypes have been observed, tetraploid (2n=4x=40), octoploid (2n=8x=80) (Church 1940), and hexaploid (2n=6x=60) (Kim et al. 2010, 2012b). The tetraploid cytotype is believed to be the progenitor of the octoploid cytotype, and is likely to have increased genetic diversity between the tetra- and octoploid cytotypes, due to their reproductive and geographic isolation from each other (Wendel & Doyle 2005; Soltis et al. 2010; Hirakawa et al. 2014). Clonal/vegetative propagation usually results in little genetic variation, however, since prairie cordgrass is able to reproduce using rhizomes and by seeds, it is expected that there are variable levels of genetic variation within populations (Coates & Byrne, 2005). In addition, this species favors outcrossing due to its protogynous nature, and therefore increases the amount of genetic diversity that may be present within given populations (Gedye et al. 2012).

There are many ways to analyze genetic diversity whether looking at quantitative characteristics, chromosome rearrangements, protein variations, or nuclear and organelle DNA variation (Coates & Byrne, 2005). A multitude of phenotypic variation has been observed in all cytotypes of prairie cordgrass (Boe & Lee 2007; Kim et al. 2012b; Guo et al. 2015), but little has been reported on the amount of organelle and nuclear DNA genetic diversity. Understanding the amount of genetic diversity present in DNA will assist in the breeding of prairie cordgrass for its various uses, and will allow the utilization of molecular breeding techniques that have not be

available for this species. These techniques are important for increasing the efficiency of the breeding process by screening seeds or young plants for desired characteristics, or traits, rather than waiting for plants to come to full maturity before evaluating (Paux et al. 2012). The neutrality of molecular markers mean they are ideal for a broad range of plant population genetics, conservation genetics, and evolutionary studies (Coates & Byrne, 2005). By understanding the amount of variation within DNA, one can begin to observe genetic diversity within populations and understand the congruency of traits within that population, and begin to select the best plants that are suited for specific breeding objectives.

The objectives of this research were to (1) determine if chloroplast DNA variation exists within populations and various cytotypes of prairie cordgrass, and (2) validate SNP polymorphisms identified in the nuclear transcriptome for use in future genetic diversity analysis, and their use to analyze specific crosses of tetraploid prairie cordgrass.

CHAPTER 2

CHLOROPLAST DNA VARIATION WITHIN PRAIRIE CORDGRASS (*SPARTINA PECTINATA* LINK) POPULATIONS IN THE U.S.[1]

ABSTRACT

Chloroplast DNA (cpDNA) is most often maternally inherited and highly conserved leading to previous observation of little to no sequence variation. Comparing cpDNA haplotypes have provided valuable insight into the establishment and migration of polyploid populations. However, to use chloroplast haplotypes to their full potential intrapopulational variation needs to be addressed. In this study, cpDNA haplotype variation was surveyed within 16 natural populations of prairie cordgrass (*Spartina pectinata* Link) located east of the 100th west meridian and north of the 35th north parallel in the U.S.A. using two non-coding, polymorphic chloroplast regions. Two main clades were defined with subclades as follows: haplotype 1 and haplotype 2A and 2B. It was discovered that seven populations showed intrapopulational chloroplast genome variation. Of the total amount of variation, 95.5% occurred within the octoploid populations and 4.5% occurred within the tetraploid populations. Both variant haplotypes, 2A and 2B, were found in a larger sampling of one of the natural populations, but no variation was found in a mixed ploidy population. The intrapopulational cpDNA variation we found in this study cannot directly be related to mechanisms of introduction of the non-native populations into native populations. Therefore, this cpDNA variation could be novel natural variation that has been fixed as the octoploid populations were established and moved northwest. This analysis provides insight into determining the usefulness of indels and single nucleotide polymorphisms

[1] Reprinted with permission from H. Graves, A.L. Rayburn, S. Kim, D.K. Lee, 2015. Chloroplast DNA variation within prairie cordgrass (*Spartina pectinata* Link) populations in the U.S. This article is reprinted with the permission of the publisher and is available from http://www.jse.ac.cn using DOI: 10.1111/jse.12168

for population identification and may provide information in regards to the origin of chloroplast variation and its subsequent fixation and establishment in natural prairie cordgrass populations.

INTRODUCTION

Prairie cordgrass (PCG) (*Spartina pectinata* Link) is a warm season, C4 perennial grass native to North America that grows well in a wide range of environmental conditions and responds well to abiotic stress, such as, drought, flooding, and saline soils (Hitchcock 1950; USDA Natural Resources Conservation Service 2006; Barkworth et al. 2007; Stubbendieck et al. 2011). In addition, this species is geographically distributed throughout North America, from the eastern coast and Midwestern United States to eastern Washington and Oregon, south to Arkansas, Texas and New Mexico, and north to Canada (Hitchcock 1950; Voight & Mohlenbrock 1979; Barkworth et al. 2007). Potter et al. (1995) described *Spartina* as the most northerly-distributed C_4 perennial species in North America and Western Europe. Prairie cordgrass is also very tolerant of moderate salinity and is valued for conservation practices, wetland revegetation of irregularly flooded salt marshes, streambank stabilization, wildlife habitat, and forage (Montemayor et al. 2008; Kim et al. 2011; Boe et al. 2013; Zilverberg et al. 2014). With such a wide range of environments that prairie cordgrass has become adapted to, specific populations of this species have adapted to various ecosystems. This adaptation has resulted in genetic differences including polyploidy (Grime 1979).

Prairie cordgrass is a polyploid species that is composed of three cytotypes, tetraploid ($2n = 4x = 40$), octoploid ($2n = 8x = 80$) (Church 1940), and a hexaploid ($2n = 6x = 60$) (Kim et al., 2010, 2012a). Based on the ploidy levels, the tetraploid cytotype is considered the progenitor of

the higher ploidy populations (Wendel & Doyle 2005; Soltis et al. 2010; Hirakawa et al. 2014). All the above studies reported geographic distribution patterns with respect to cytotypes. With tetraploids, prairie cordgrass appears to be indigenous to the northeastern to mid-central US, the only known hexaploid prairie cordgrass is located in a mixed population in central IL, and octoploid prairie cordgrass range from the north-central to the south-central US (Kim et al. 2012b). This could occur by certain cytotypes being advantageous to different environmental conditions, such as cold temperature, experienced in the different regions.

Chloroplast DNA (cpDNA) is most often maternally inherited and highly conserved leading to previous observations of no variability of cpDNA within most species (Demesure et al. 1995; Hultquist et al. 1996; Ferris et al. 1997; Yannic et al. 2004; Bock 2007; Hazarika et al. 2014). This maternal mode of inheritance can make cpDNA a useful tool for genetic and phylogenetic studies (Demesure et al. 1995; Daniell 2002; Hazarika et al. 2014). Chloroplast DNA variation compared with cytotypes has provided more of an understanding into the origin and establishment of polyploids in many species (Soltis et al. 1992; Kim et al. 2013). In chloroplast genomes, sexual recombination does not occur, therefore; genetic recombination and fusion do not normally exist (Bock 2007). Because of this phenomenon, the exploration of population structure based on difference in ploidy levels using intraspecific cpDNA variation can lead to a greater understanding of the evolutionary developments that have occurred during the establishment of prairie grasslands in natural habitats (McMillan & Weiler 1959; Young et al. 2011). In addition, one can begin to delve into the possibility of using haplotypes to determine area of geographic origin (Kim et al. 2013).

Previous studies of cpDNA variation focus on inter- and intrapopulational variation of various species (Ferris et al. 1997; Mohanty et al. 2000; Fineschi et al. 2002; Lumaret et al.

2002; Palmé 2002; Yannic et al. 2004; Chen et al. 2013). These studies vary from finding very low intrapopulational variation (Chen et al. 2013) to finding high levels of intrapopulational variation (Fineschi et al. 2002). Most of these studies separated variation into haplotypes, but not all studies found correlation between haplotypes and geographic distribution (Ferris et al. 1998; Mohanty et al. 2000; Fineschi et al. 2002; Chen et al. 2013). In herbaceous and grass species, a few studies have discovered both inter- and intrapopulational variation (Hultquist et al. 1996; Young et al. 2011; Shimono et al. 2013). Hultquist et al. (1996, 1997) discovered cpDNA polymorphisms among populations of *Panicum virgatum*, a perennial grass native to North America tallgrass prairie, which differentiated populations of switchgrass into upland and lowland ecotypes. This was confirmed by Young et al. (2011). Shimono et al. (2013) reported finding intrapopulational variation that was then divided into haplotypes in *Miscanthus sinensis*.

Conventional knowledge would suggest that no intraspecific cpDNA variation exists in the genus *Spartina* (Ferris et al. 1997; Yannic et al. 2004). Yannic et al. (2004) reported finding no cpDNA variation, by using polymerase chain reaction- restriction fragment length polymorphisms (PCR-RFLPs), within species of *S. maritima,* located on the Atlantic European coast. Ferris et al. (1997) found no variation within the amphidiploid *S. anglica*, a hybrid between *S. alterniflora* and *S. maritima*, or the parents, using PCR tRNA sequences. However, Blum et al. (2007) reported intraspecific and intrapopulational variation in *S. alterniflora* indicating the cpDNA variations can exist.

Blum et al. (2007) used chloroplast haplotypes to identify populations of *S. alterniflora* of recent and/or multiple origins. For example, they observed that the *S. alterniflora* populations found in San Francisco Bay (CA, USA) contain chloroplast haplotypes found on the United States east coast; thus corroborating the proposed introduction of *S. alterniflora* to San Francisco

Bay with seeds from Maine (USA) and Virginia (USA). It was also observed that the chloroplast haplotype found in a Grays Harbor (WA, USA) population is most commonly found in a Willipa Bay (WA, USA) population; which was a result of the introduction of *S. alterniflora* to Grays Harbor, based on winds and ocean currents, occurred within the last two decades and originated from Willipa Bay. The chloroplast haplotypes located in Willipa Bay were observed on the eastern coast of the US leading to the hypothesis that *S. alterniflora* was introduced into Willipa Bay from oyster shipments that originated from New York Harbor, Chesapeake Bay, Long Island, and northern New Jersey. These observations indicate the importance of chloroplast haplotype analysis in dissecting *Spartina* origin and evolution.

By way of cpDNA analysis, three distinct chloroplast haplotypes have been observed in prairie cordgrass (Kim et al. 2013). These haplotypes are all associated with geographic distribution as well as with cytotypes (Kim et al. 2012b, 2013). PCG1 was associated with tetraploid populations and located to the east of the Mississippi River, whereas, PCG2 and PCG3 were associated mainly with octoploid populations and located to the west of the Mississippi River (Kim et al, 2013). It was hypothesized that indels in prairie cordgrass are population specific and could be diagnostic for geographic origin and perhaps population identification (Kim et al. 2013). However, the previous study was not designed to detect intrapopulational variation. Although such variation may be rare, if present, such variation could result in misinterpretation of the data. This information is imperative in determining the usefulness of indels and single nucleotide polymorphisms (SNP's) for population identification. In this study, two noncoding chloroplast DNA regions, identified in Taberlet et al. (1991) and Shaw et al. (2007), were used and have been determined to be the most polymorphic in prairie cordgrass, accounting for ~70% of cpDNA variation (Kim et al. 2013). These primers were used to

determine if polymorphisms occur within populations, and if so, do these polymorphisms interfere with the diagnostic potential of chloroplast haplotypes with regards to geographic distribution. The specific objectives of this study were (1) to determine if previously defined haplotypes are homogenous to geographic location, (2) to determine how much, if any, intrapopulational variation occurs within prairie cordgrass, and (3) to determine if there is a difference in the amount of intrapopulational variation occurring across ploidy levels.

MATERIALS AND METHODS

Plant material

Natural populations: To determine if cpDNA haplotype variation is found within various natural populations of prairie cordgrass, seeds of 15 natural populations were previously collected from five geographic regions of the USA as described in Kim et al. (2013) and grown at the Energy Biosciences Institute (EBI) Farm, Urbana, Illinois, USA. They consisted of 8 octoploid and 7 tetraploid populations (Table 2.1). These populations, along with the natural germplasm population 'Red River' (octoploid), were selected and tissue samples of 16 plants per population were collected from the EBI Farm and stored at -80°C until DNA extraction was performed.

Single natural population: To further examine if cpDNA haplotype variation is found within an octoploid population located in the north-central US, bulked seeds collected from a natural population in North Dakota (ND) (47^{o}27'28"N, 99^{o}04'38"W) were planted in in 11cmx11cm pots filled with Sunshine SB300 Universal soil and maintained under greenhouse conditions (16hr photoperiod at temperature of 24-26°C). Fresh, young leaf tissues from an additional 34

plants of the ND population were sampled and stored at -80°C until DNA extraction was performed. Analysis was performed on the 50 total ND plants.

Tetra- & hexaploid population: To examine if cpDNA haplotype variation can be attributed to ploidy differences, sampling was conducted from a mixed cytotype population containing hexaploids and tetraploids as described in Kim et al. (2012a). These populations are sympatric and are adapted to the same environmental conditions. Fresh, young leaf tissue from 24 plants, 12 hexaploid and 12 tetraploid, were sampled from alongside a stream bank that extends over a distance of 2.8 km, located in Champaign County, Illinois, USA (Figure 2.1). Tissue samples were stored at -80°C until DNA extraction was performed.

DNA extraction, polymerase chain reaction (PCR), and sequencing

Total genomic DNA was extracted from frozen leaf tissue using the CTAB method (Mikkilineni 1997) with slight modifications as described by Kim et al. (2013). For all samples, each amplification reaction contained 100ng template DNA, 5XGoTaq reaction buffer (Promega, Madison, WI, USA), 2 mmol/L dNTP, 5 pmol/L each primer, and 5U of GoTaq polymerase (Promega, Madison, WI, USA) for a total of 50 µL per sample. Two noncoding regions of cpDNA were amplified by using the oligonucleotide primers designed by Taberlet et al. (1991) and Shaw et al. (2007) (Table 2.2). The two primers chosen showed the highest number of informative polymorphic sites/indels based on data from Kim et al. (2013), and therefore these polymorphic sites were targeted. In addition, when the original populations were reassessed using just the two primers, a similar phylogenetic tree was obtained with all three haplotypes present indicating that these primers were the most informative with respect to haplotype. The

polymerase chain reaction (PCR) thermocycling conditions for the *trnT-trnL* region was 5 min at 94 °C followed by 30 cycles of 94 °C for 1 min, 48 °C for 1 min, and 68 °C for 2 min followed by a final extension stage at 68 °C for 10 min, whereas *3'rps16-5'trnK* was 5 min at 80 °C followed by 30 cycles of 95 °C for 1 min, 50 °C for 1 min, and 65 °C for 4 min followed by a final extension stage of 65 °C for 5 min. Electrophoresis was performed on the PCR product, in a 1.2% agarose gel using 0.5X TBE buffer, to confirm amplification, and then purified using a EZ-10 spin column PCR purification kit (Bio Basic Inc., Canada). The purified PCR product was sent to the Biotechnology Center (University of Illinois, Urbana, IL, USA) for direct sequencing. Sequences were manually corrected with BioEdit (http://www.mbio.ncsu.edu/bioedit/bioedit.html) and aligned with MEGA5.2 and trees, including five other *Spartina* species as outgroups, were produced with MEGA6 (Tamura et al. 2013). The maximum parsimony (MP) tree, inferred from 1000 replicates, was obtained using the Subtree-Pruning-Regrafting algorithm with a search level one in which the initial trees were obtained by the random addition of sequences (Felsenstein 1985; Nei & Kumar 2000). All positions with less than 95% site coverage were eliminated.

RESULTS

Natural populations

Fifteen prairie cordgrass populations, consisting of eight octoploids and seven tetraploids were analyzed for cpDNA variation (Table 2.1; Figure 2.1). Analysis was conducted based on an aligned dataset of two noncoding chloroplast regions from the fifteen prairie cordgrass populations and five additional *Spartina* species samples. This resulted in 1441-character dataset

with three SNPs, one indel, and six repeats (Table 2.3). A MP tree was constructed based on the concatenated dataset. Two main clades were defined as follows with sub-clades; haplotype 1, haplotype 2A, and haplotype 2B (Figure 2.2). The parsimony analysis generated the 10 most parsimonious trees with a length of 18, a consistency index of one, a retention index of one, and a composite index of one for all sites and parsimony-informative sites.

The variation observed was mainly found in the octoploid cytotype. The total, 18.3%, amount of variation was determined by using haplotype 1 as non-variant and haplotype 2A and 2B as variants. Of the total variation, 95.5% was found in the octoploids and 4.5% was found in the tetraploids. Out of the eight octoploid populations, five showed intrapopulational variation (62.5%). PC46-106 showed haplotype 1 and haplotype 2A, ND showed haplotype 2A and 2B, PC19-108 showed haplotype 2A and 2B, PC27-103 showed one variant as haplotype 2A, and Red River showed one variant as haplotype 2B. There were only two tetraploid populations out of eight (25%), IL99 and PC55-102, which had variation. IL99 showed one variant of haplotype 2A and PC55-102 showed one variant of haplotype 2B (Figure 2.1). Most of the variation that was found was located in the north central US, consisting of portions of North Dakota, South Dakota (SD), Iowa (IA), and Minnesota (MN) (Figure 2.1).

Single natural population

Fifty plants from the ND octoploid prairie cordgrass population were analyzed for cpDNA variation. Analysis was conducted based on an aligned dataset of two noncoding chloroplast regions from the fifty prairie cordgrass plants. This resulted in 1381-character dataset with two SNPs, one indel, and three repeats. A MP tree was constructed based on the concatenated dataset. Two main haplotypes were defined, and were correlated to the 2A and 2B haplotypes

described in the natural populations (Figure 2.3). The parsimony analysis generated the 10 most parsimonious trees with a length of 18, a consistency index of one, a retention index of one, and a composite index of one for all sites and parsimony-informative sites.

Tetra- & hexaploid population

Twenty-four total plants from the mixed cytotype population located in Sidney, Illinois were analyzed for cpDNA variation. Analysis was conducted based on an aligned dataset of 1332 characters with no SNPs, indels, or repeats. All sequences were found to be identical and were correlated to haplotype 1 described in the natural populations.

DISCUSSION

In order to better understand the extent to which cpDNA variation could exist in prairie cordgrass, natural populations from a wide range of geographical locations and environments were examined. In the natural populations, three haplotypes were observed. All three haplotypes were found in both octoploid and tetraploid populations. Only two tetraploid populations showed variation, with most of the tetraploid plants (98.2%) containing a haplotype 1. Because tetraploids are the proposed progenitor of the octoploid populations (Wendel & Doyle 2005; Soltis et al. 2010; Hirakawa et al. 2014), haplotype 1 was assigned the ancestral haplotype since it is the most frequent haplotype (Peirson et al. 2013) found in the tetraploid populations. Haplotype 1 was therefore considered the non-variant haplotype. In addition, haplotype 1 in this study corresponds to haplotype PCG1 in Kim et al. (2013). Haplotypes 2A and 2B are considered the variant haplotypes and corresponds to haplotypes PCG2, and PCG3 of

Kim et al. (2013). Due to Kim et al. (2013) only examining one plant per population, it is not possible to identify haplotype 2A and 2B as PCG2 or PCG3.

The small amount of variation within the tetraploid populations gives reason to believe that intrapopulational variation of tetraploid prairie cordgrass populations occurs at a very low rate or cpDNA of tetraploid populations may be more stable than that of the octoploid populations. The larger amount of variation within the octoploid populations may be reflective of the more recent origin of the octoploid populations and their adaptation to potentially harsher environments in the north-central portion of the U.S. One distinguishing environmental factor in the regions of cytotype adaptation may be temperature. Because prairie cordgrass emerges earlier than most other perennial grasses, the maximum and minimum temperatures during winter and early spring months are crucial. In locations where cpDNA variation was observed, ND, SD, and MN, the average maximum temperature is below 0 °C during the winter months, however, in other locations, such as OK, IL, and ME, the maximum winter temperature is above 0°C. Therefore as prairie cordgrass adapted to the norther region, the harsher environment, such as colder climate, have to be overcome and thus results in cpDNA variants. This is corroborated by Madlung (2013), stating polyploids evolve new functions that allow for ecological niche colonization and could have flexibility in responsiveness to environmental change. Alternatively, difference could be indicative of multiple octoploid population formation. In addition, it is noted that all of the variation occurs within the north-central portion of the US, mainly around the borders of three states ND, SD, IA, and MN. This location corresponds with the overlap of haplotypes PCG1 and PCG2 of Kim et al. (2013). This is also the region in which the tetraploid and octoploid cytotypes occur, and may represent the area of natural octoploid formation and stabilization, similar to observation in *Panicum virgatum* (Hultquist et al. 1997).

The large amount of variation found in octoploid populations located in the SD, IA, and MN region led to a larger study of a single natural population, ND, of prairie cordgrass. Variation in this larger study confirms the presence of cpDNA variation in this population. The haplotypes show 70 % variation occurring as haplotype 2A, and 30% variation occurring as haplotype 2B within the 50 ND plants. This amount of variation reflects the amount of variation previously observed in the natural population study. This connection confirms that the study performed on the natural populations accurately represented the haplotypes of the selected populations. Kim et al. (2013) looked solely at one plant per population; these findings demonstrate the importance of confirming if polymorphisms occur in various prairie cordgrass populations.

Intrapopulational variation can be useful in determining ploidy and geographic origin, but does have limitations. When an unknown sample falls into haplotypes 2A or 2B it can be assumed to be an octoploid population, due to the low frequency of these specific haplotypes in the tetraploid populations, and its area of segregation in the north-central region of the U.S. If an unknown sample falls into haplotype 1, the sample could be an octoploid or a tetraploid. Despite the amount of intrapopulational variation found in prairie cordgrass populations, previously defined haplotypes may be used to diagnose geographic location if a sample size of 16 or more are analyzed. These findings also bring rise to the question of if the variation that is present is due to environmental adaptation or introduction of new haplotypes into previously established populations. Doyle et al. (1990, 1999) and Soltis et al. (1989) suggest that polyploidization can occur as multiple independent events in one population, giving rise to different haplotypes within populations.

In order to determine if polyploid adaptation could be a factor in intrapopulational cpDNA variation, a mixed cytotype population from the same ecosystem was analyzed. This population was chosen because environmental pressure is eliminated as a causal factor for cpDNA variation. In addition, the tetraploids and hexaploids, within the mixed cytotype population, are reproductively isolated and therefore cannot cross with the other when reproducing, limiting the number of potential pollen donors and acceptors. In the mixed cytotype population, there was no variation found within the population consisting of tetraploids and hexaploids, which was also observed in other plant species such as *Plantago media* (Van Dijk & Bakx-Schotman 1997). Kim et al. (2010) stated that the hexaploid plants were most likely developed from seeds collected from tetraploid plants, which indicates hexaploids and tetraploids coexisting may be a result of the emergence of a neopolyploid within an ancestral population (Petit et al. 1999). Having found no variation within the mixed cytotype population compared to finding variation within the octoploid natural populations, could be explained by multiple factors. First, the mixed cytotype population was established only within the last two decade as described by Kim et al. (2013) based on the farm land history, whereas the octoploid natural populations have been in existence for a longer amount of time. This leads one to believe that the octoploid natural populations had a longer amount of time to have variation show up within the cpDNA. Harrison et al. (1997) found no chloroplast genome variation in *Fragaria* species contributing the lack of variation to relatively recent origin of the species. In addition, the location of the tetraploids and hexaploids in close proximity to one another leads one to believe that they are adapted to the same environment, and therefore the same selection pressure and could lead to no haplotype variation between the hexaploid and tetraploids. Because the

mixed cytotype population is young, this population should be watch in the future to determine if any variation occurs as the population matures.

CONCLUSIONS

This study reports the first research of intrapopulational variation of prairie cordgrass based on cpDNA analysis. Haplotype 1 appeared to be the ancestral haplotype since it is found in the ancestral tetraploid cytotype, the progenitor of the octoploid population. Haplotypes 2A and 2B may be diagnostic for octoploid prairie cordgrass populations, especially in the north-west region of this study. It was determined that of the 18.3% of variation occurring, 95.5% occurred within the octoploid populations and 4.5% occurred within the tetraploid populations. Larger sampling of one of the natural populations confirmed that the study of the natural populations is an accurate representation of the populations sampled, and therefore any future results could be deemed accurate following the same protocols. No variation was found in the mixed ploidy population leading one to believe that under the same environmental conditions, cpDNA variation is not affect by ploidy levels. This analysis provides insight into determining the usefulness of indels and SNP's for population identification and may provide insight into the origins and establishment of natural prairie cordgrass populations.

FIGURES AND TABLES

Table 2.1 Summary of plant materials used including, location, cytotype and number of plants used per population

	Population ID	Latitude	Longitude	Cytotype	Number of plants
Natural populations	PC19-108	42°19'48.21"N	96°19'37.00"W	Octoploid	16
	PC20-106	39° 3'37.98"N	96°22'53.82"W	Octoploid	16
	PC27-103	48°30'51.67"N	96°53'13.16"W	Octoploid	16
	PC31-103	40°53'5.91"N	100° 3'41.99"W	Octoploid	16
	PC40-102	36°52'25.50"N	95° 0'45.24"W	Octoploid	16
	PC46-106	42°58'1.20"N	96°49'34.73"W	Octoploid	16
	ND	47°27'26.70"N	99°05'02.59"W	Octoploid	16
	Red River	N/A	N/A	Octoploid	16
	IL99	39°44'59.76"N	88°42'02.76"W	Tetraploid	16
	PC09-103	N/A	N/A	Tetraploid	16
	PC19-109	42°12'20.60"N	96°15'5.22"W	Tetraploid	16
	PC19-110	41°47'33.84"N	96° 2'33.19"W	Tetraploid	16
	PC23-102	44°16'4.57"N	69° 1'0.65"W	Tetraploid	16
	PC29-106	37°51'12.95"N	94°18'55.53"W	Tetraploid	16
	PC55-102	44° 3'12.62"N	90° 5'23.37"W	Tetraploid	16
Single natural population	ND	47°27'26.70"N	99°05'02.59"W	Octoploid	50
Tetra- & hexaploid population	Sidney	40°00'59.6"N	88°01'26.9"W	Tetraploid Hexaploid	12 12
Outgroups	*Spartina alterniflora* (Gulf)	Big Flats Plant Material Center, NY, USA			1
	S. alterniflora (East)				1
	S. cynosurodies (East)				1
	S. patens (Sharp)	Commercial			1
	S. patens (Flageo)				1

Table 2.2 Primer sequences used for the amplification of two noncoding chloroplast DNA regions in prairie cordgrass

Chloroplast Regions	Primer Pair Sequences (5'-3')	
	Forward	Reverse
trnT-trnL	CATTACAAATGCGATGCTCT	TCTACGATTTCGCCATATC
3'rps16-5'trnK	AAAGTGGGTTTTTATGATCC	TTAAAAGCGAGTACTCTACC

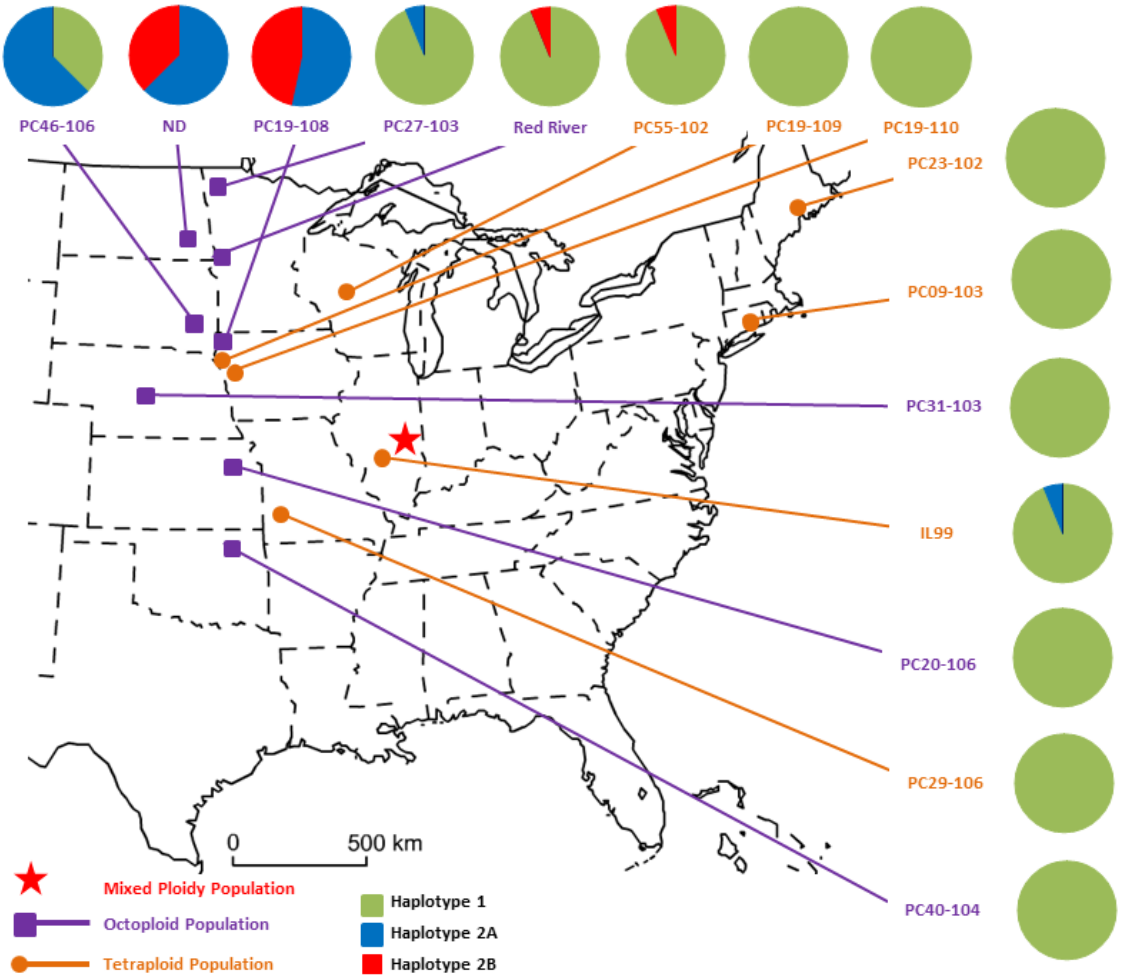

Figure 2.1 The distribution and frequency of haplotype groups of prairie cordgrass based on populations from Experiment 3 (Haplotype 1: Green, Haplotype 2A: Blue, Haplotype 2B: Red).

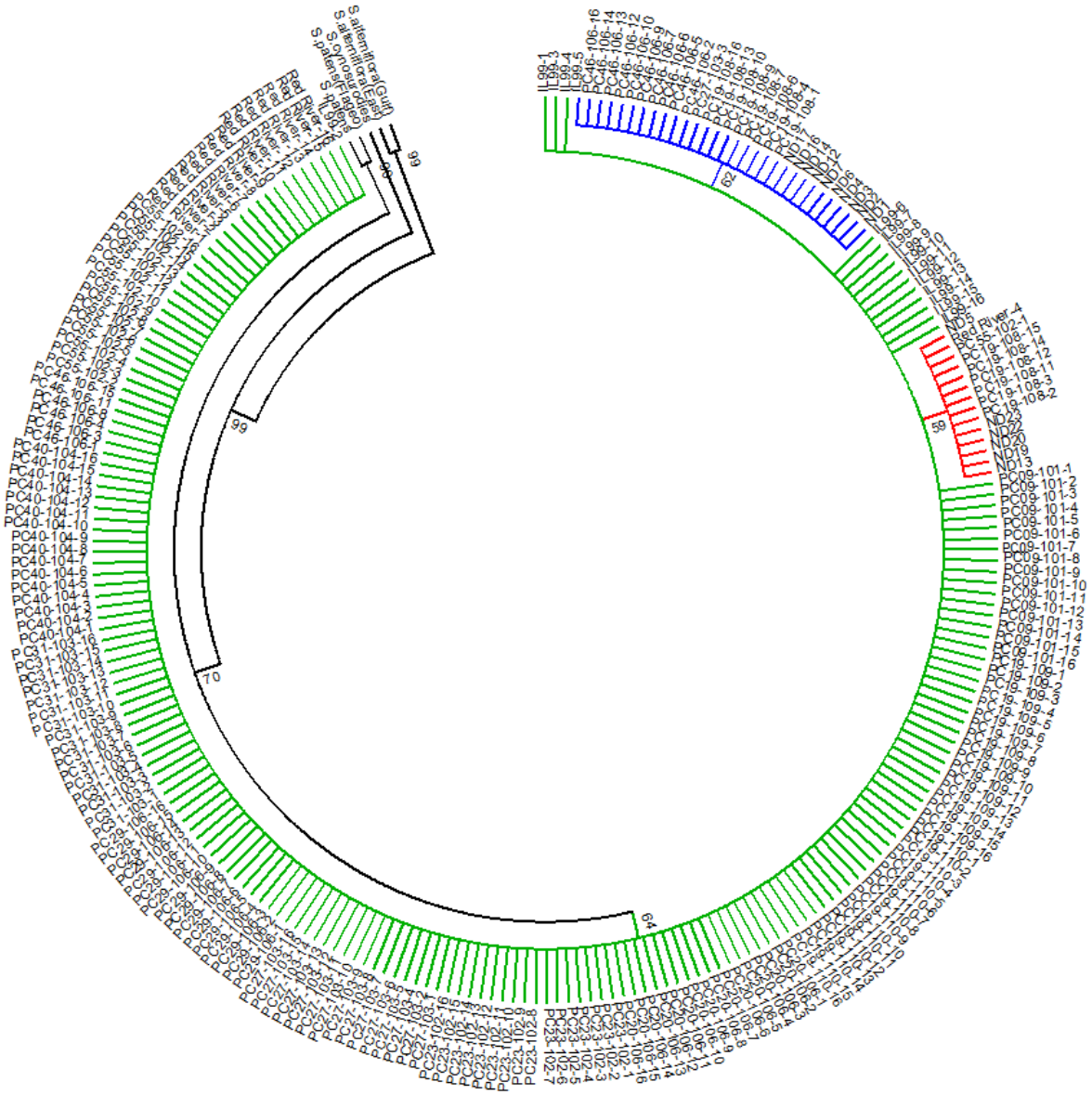

Figure 2.2 Bootstrap tree based on MP analysis of 2 combined cpDNA sequence data for 15 prairie cordgrass populations with a total of 240 plants and 5 other *Spartina* species used as outgroups. Bootstrap values are indicated on the nodes as percentages. Clades are identified as Haplotype 1 (green), Haplotype 2A (blue), and Haplotype 2B (red).

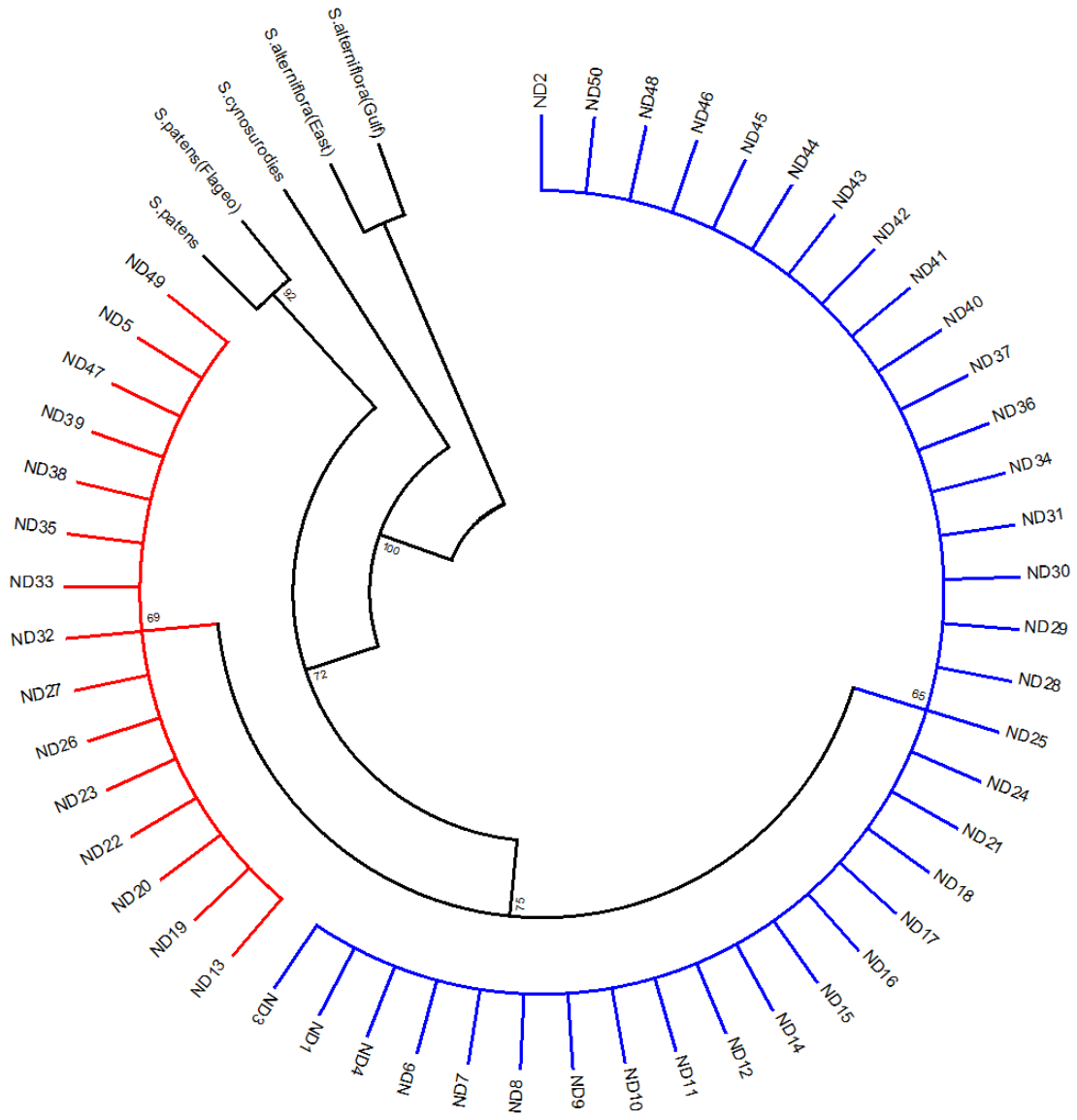

Figure 2.3 Bootstrap tree based on MP analysis of 2 combined cpDNA sequence data for the North Dakota population of prairie cordgrass with a total of 50 plants and 3 other *Spartina* species used as outgroups. Bootstrap values are indicated on the nodes as percentages. Two clades are defined as 2A (blue) and 2B (red).

CHAPTER 3

VALIDATING DNA POLYMORPHSIMS USING KASP ASSAY IN PRAIRIE CORDGRASS (*SPARTINA PECTINATA* LINK) POPULATIONS IN THE U.S.[2]

ABSTRACT

Single nucleotide polymorphisms (SNPs) are one of the most abundant DNA variants found in plant genomes and are highly efficient when comparing genome and transcriptome sequences. SNP marker analysis can be used to analyze genetic diversity, create genetic maps, and utilize marker-assisted selection breeding in many crop species. In order to utilize these technologies one must first identify and validate putative SNPs. In this study, 121 putative SNPs, developed from a nuclear transcriptome of prairie cordgrass (*Spartina pectinata* Link), were analyzed using KASP technology in order to validate the SNPs. Fifty-nine SNPs were validated using a core collection of 38 natural populations and a phylogenetic tree was created with one main clade. Samples from the same population tended to cluster in the same location on the tree. Polymorphisms were identified within 52.6% of the populations, split evenly between the tetraploid and octoploid cytotypes. Twelve selected SNP markers were used to assess the fidelity of tetraploid crosses of prairie cordgrass and their resulting F_2 population. These markers were able to distinguish true crosses and selfs. This analysis provides insight into genomic structure of this species, but further analysis must be done on other cytotypes to understand the genetic diversity of prairie cordgrass. This study validates putative SNPs and confirms the potential usefulness of SNP marker technology in future breeding programs of this species.

[2] Reprinted with permission from H. Graves, A.L. Rayburn, J. Gonzalez, G. Nah, Do-Soon Kim, D.K. Lee, 2015. Validating DNA Polymorphisms Using KASP Assay in Prairie Cordgrass (Spartina pectinata Link) Populations in the U.S. Submitted to Frontiers in Plant Science

INTRODUCTION

Prairie cordgrass (*Spartina pectinata* Link) is a native grass species of the North American Prairie that has a geographic distribution, ranging from the southern U.S. (Texas, Arkansas, and New Mexico) to northern Canada, and from the east coast through the Midwest to the western coast of the U.S. (Hitchcock 1950; Voight & Mohlenbrock 1979; Barkworth et al. 2007; Gedye et al. 2010). This species is adapted to a wide range of environmental conditions and, in addition, responds well to abiotic stresses, such as moderate salinity, water logged soils, drought, and cold tolerance (Montemayor et al. 2008; Boe et al. 2009; Gonzalez-Hernandez et al. 2009; Kim et al. 2011; Zilverberg et al. 2014; Anderson et al. 2015). Because of its wide adaptability, this warm season, C4, perennial grass is highly valued for conservation practices, wetland revegetation, streambank stabilization, wildlife habitat, forage production, and recently bioenergy feedstock production (Hitchcock 1950; Barkworth et al. 2007; Montemayor et al. 2008; Gonzalez-Hernandez et al. 2009; Boe et al. 2013; Kim et al. 2011; Zilverberg et al. 2014; Guo et al. 2015). This ability to adapt to such a wide diversity of conditions results in populations becoming adapted to specific environments, ultimately leading to genetically diverse populations. Adding to the potential genetic diversity of prairie cordgrass is polyploidy.

Prairie cordgrass is a polyploid species, composed of three cytotypes, tetraploid (2n=4x=40), hexaploid (2n=6x=60), and octoploid (2n=8x=80) (Church 1940; Kim et al. 2010, 2012). Because of reproductive and geographic isolation between the cytotypes, there is likely an increase in polymorphisms and potential genetic diversity, especially within the tetraploid and ocotploids cytotypes (Wendel & Doyle 2005; Soltis et al. 2010; Hirakawa et al. 2014). There is a large amount of phenotypic variation present in all cytotypes of prairie cordgrass (Boe & Lee

2007; Kim et al. 2012b; Guo et al. 2015), but there is a lack of knowledge about genomic structure. A few studies have revealed diversity within highly polymorphic chloroplast DNA regions observed within and among tetra- and octoploid populations (Kim et al. 2013; Graves et al. 2015). In prairie cordgrass, EST-SSR markers (Gedye et al. 2010), SSR (Gedye et al. 2012), and AFLP markers (Moncada et al. 2007) have been developed. However, these technologies may not be as cost-effective, scalable, successful, or as flexible as using single nucleotide polymorphisms (SNPs) (Segman et al. 2014).

Single nucleotide polymorphisms provide a highly efficient way to conveniently compare genomic and transcriptome sequences. Because they are one of the most abundant DNA variants found in plant genomes, SNPs are more likely to be related to specific biological function and phenotype (Rafalski 2002; Bundock et al. 2006; Salem et al. 2012). This technology has been applied in genetic diversity analysis, genetic map construction, association map analysis, and marker-assisted selection breeding in many different types of crop species (Byers et al. 2012; Saxena et al. 2012; Semagn et al. 2014; Sindhu et al. 2014; Wei et al. 2014). SNP marker technology is also utilized in high throughput genotyping, increasing the speed of the selection process by eliminating growing plants to maturity for phenotypic selection (Paux et al. 2012). In order to use SNP markers for genetic improvement there is a three-step process one must follow: (1) SNP discovery after aligning sequence reads generated by next-generation sequencing technologies for different genotypes of a given species; (2) validate SNPs to distinguish DNA polymorphisms of actual allelic variants from those of other biological phenomena such as gene duplication events; (3) SNP genotyping of germplasm collection or genetic/breeding populations (Saxena et al. 2012).

Step one of the Saxena et al. (2012) process, in prairie cordgrass, is described in the companion manuscript to this study by one of the authors of this study (Gonzalez et al. personal communication). The second and third steps have yet to be completed for polyploid prairie cordgrass. Several parameters, such as sample size, number of SNPs to be used for analysis, cost effectiveness, and the SNP genotyping platform must be considered in these analyses (Semagn et al. 2014). Many technologies exist for use in SNP genotyping analysis, but one technology performs well when it comes to adaptability, efficiency, and cost-effectiveness. Kompetitive allele-specific PCR (KASP), developed by LGC Genomics (Teddington, UK; www.lgcgenomics.com), is a PCR-based homogeneous fluorescent SNP genotyping system, which determines the alleles at a specific locus within genomic DNA (Semagn et al. 2014). The KASP technology has been utilized on other polyploid plant species, including switchgrass (LGC Genomics 2014), cotton (Byers et al. 2012), wheat (Paux et al. 2012), potato (Uitdewilligen et al. 2013), and various triploid citrus species (Cuenca et al. 2013).

In this study, SNPs, identified in the nuclear transcriptome, were converted to the KASP marker system in order to validate that these SNPs are true allelic variants. In addition, KASP markers were used in quality control analysis when making crosses, prairie cordgrass being a putative self-compatible species. The main objectives of this study were (1) to validate SNP polymorphisms identified in the nuclear transcriptome of natural populations of prairie cordgrass in the U.S. and (2) to assess the fidelity of specific tetraploid crosses and selfs, and to elucidate inheritance patterns of SNP markers.

MATERIALS AND METHODS

Development and validation of KASP genotyping assays

In a study by Gonzalez et al. at South Dakota State University (personal communication), a transcriptome of prairie cordgrass was assembled using ~1.2 billion Illumina paired-end reads from various vegetative tissues (roots, leaves, and rhizomes) under various conditions (salt stress, cold stress, and differing photoperiods) in order to obtain an abundance in diversity, with regards to the number and type of transcripts. About 146,549 contigs, or transcript assemblies, of 230 base-pairs or more with an N50 of 973 bp were used. Over 1 million SNPs, insertions, and deletions were found. Nine bi-allelic SNPs were selected for analysis, associated with enzymes within the lignin biosynthesis pathway. Additional SNPs were selected without regard to specificity of the loci. The nine lignin biosynthesis pathway SNPs by Gonzalez et al. were selected based on an allele frequency of 40-60%, and the additional random SNPs for this study were based on an allele frequency of 20-80%. An 80-100 bp window, largely free from additional variants, was required for selection. A total of 121 bi-allelic SNPs were identified for use in this study (Table 3.1). SNPs were sent for primer development to be used in KASP genotyping assays. Genotyping with KASP was performed as follows.

For all samples, each amplification reaction contained 50ng template DNA, KASP V4.0 2x Master mix standard ROX (LCG Genomics, Beverly, MA, USA) and KASP-by-Design assay mix (LGC Genomics, Beverly, MA, USA). The PCR thermocycling conditions for all primers, except pcg_1186, was 15 min at 94°C followed by 10 cycles of 94°C for 20 sec and 61°C for 1 min (dropping -0.6°C per cycle to achieve a 55°C the annealing temperature) followed by 26 cycles of 94°C for 20 sec and 55°C for 1 min. The PCR thermocycling conditions for primer

pcg_1186 was 15 min at 94°C followed by 10 cycles of 94°C for 20 sec and 65°C for 1 min (dropping -0.8°C per cycle to achieve a 57°C annealing temperature) followed by 26 cycles of 94°C for 20 sec and 57°C for 1 min. After amplification, PCR plates were read with a Spectramax M5 FRET capable plate reader (Molecular Devices, Sunnyvale, CA, USA) using the recommended excitation and emission values. Data was then analyzed using Klustercaller software (LGC Genomics. Beverly, MA, USA) to identify SNP genotypes.

Core collection analysis

In order to validate SNP polymorphisms of natural populations of prairie cordgrass using KASP, seeds and rhizomes of natural populations of prairie cordgrass were collected from across the continental U.S.A. (Kim et al. 2013) and grown at the Energy Biosciences Institute (EBI) Farm, Urbana, Illinois, USA. Individuals from 38 of these populations were selected as core collection based on geographic distribution; and two plants from each population were sampled, for a total of 76 plants (Table 3.2). Leaf tissue samples were stored at -80°C until DNA extraction was performed. Total genomic DNA was extracted from frozen leaf tissue using the CTAB method (Mikkilineni 2007) with slight modifications as described by Kim et al. (2013). Fifty-nine KASP genotyping assays out of 121 were selected and used to analyze the populations and five additional *Spartina* species samples, namely; *S. alterniflora*, *S. patens* (Flageo vt.), *S. patens* (Sharp vt.), *S. patens*, and *S. bakeri*. All of the KASP genotyping assay results were recorded as a two-letter code, or SNP code, i.e. AA, AG, GG. A DNA fingerprint was made using all the SNP genotypes creating a concatenated DNA-like sequence, which was then imported into MEGA 6 (Tamura et al. 2013) to make a phylogenetic tree. The maximum parsimony (MP) tree, inferred from 1000 replicates, was obtained using the Subtree-Pruning-Regrafting algorithm with

a search level one in which the initial trees were obtained by the random addition of sequences (Felsenstein 1985; Nei et al. 2000). All positions with less than 95% site coverage were eliminated.

F_1 cross

In order to assess the utility of the KASP marker system in confirming specific tetraploid crosses of prairie cordgrass, a reciprocal cross of two populations differing in morphological characteristics of potential agronomic importance, PC17-109 x PC20-102, was developed. PC17-109 is a tetraploid population from Illinois with a phalanx rhizome type and low seed mass, whereas PC20-102 is a tetraploid population from Kansas with a guerilla rhizome type and high seed mass. In a greenhouse, the female inflorescence was covered approximately one day prior to stigma emergence, while pollen was collected from the male parent. Pollen was directly applied to the stigmas with a brush, and rebagged until anthesis was completed. F_1 seeds were collected from crossed inflorescence and planted in greenhouse setting. Leaf tissue samples of each seedling were collected and stored at -80°C until DNA extraction was performed. Total genomic DNA was extracted from frozen leaf tissue as described previously. For the F_1 individuals, 12 KASP genotyping assays were selected based on the parental SNP genotypes (Table 3.3). All of the assay results were recorded as a SNP code. To determine if the F_1 progeny followed segregation of a typical monohybrid cross in relation to SNP genotype, a χ^2 analysis was performed using P=0.05, df=2, and χ^2 critical value=5.991. The observed, along with the expected genotype, was recorded for each KASP genotyping assay.

F_2 selfs

To assess the utility of the KASP marker system in identifying selfed individual in the tetraploid background and gauge the segregation pattern, F_2 individuals were generated and genotyped. In a greenhouse, the prairie cordgrass inflorescence was covered approximately one day prior to stigma emergence with bags constructed to view progression of inflorescence development of F_1 plants. When anthesis was reached, the bags were shaken to promote self-pollination. Bags remained until after anthesis was complete. F_2 seeds were collected and planted in a greenhouse setting. A total of eight F_1 individuals were selfed (6 F_1 of PC17-109 x PC20-102 and 2 F_1 of PC20-102 x PC17-109) and 8-11 individuals were sampled from the planted seeds of each of the selfed plants (total of 76). Leaf tissue samples were stored at -80°C until DNA extraction was performed. All 12 of the KASP genotyping assays selected to score the F_1 individuals were also tested on the F_2 individuals. All of the assay results were recorded as a SNP code as done in the F_1 analysis. All SNP codes that were not accurately identified were removed from analysis.

RESULTS

Development and validation of KASP genotyping assays

Twenty-six (21.5%) SNPs failed KASP marker development. From the remaining 95 (78.5%), 11 SNPs were found to be monomorphic when tested on the core collection DNA, resulting in 84 SNPs that were true allelic variants. Three of the 11 monomorphic markers were selected to discover if future plant samples would reveal the SNP polymorphisms previously identified in the transcriptome. From the 84 allelic variants, 56 of the most highly polymorphic SNPs were selected for further use in this study, resulting in 59 total KASP genotyping assays (Table 3.4).

Core collection analysis

The resulting data set from the DNA fingerprint contained 118 characters. There was an average of 3.8 missing character data points (SNP codes) per population. The maximum parsimony tree identified one main clade after correcting for the missing data (Figure 3.1). For 47.4% of the populations, plants sampled from the same populations were observed to form subclades; however, intrapopulational variation was observed.

Out of the 38 prairie cordgrass populations, 52.6% showed polymorphisms within populations. Of the 52.6% polymorphic populations, 50% were octoploid and 50% were tetraploid. Out of the 15 octoploid populations sampled, 66.7% of the populations showed polymorphisms between the two plants sampled and 43.5% of the 23 tetraploid populations showed polymorphisms. The average number of polymorphisms that occurred within each population was 16. In the octoploid populations, 16.4 was the average number of polymorphisms observed, and 15.5 polymorphisms were observed as the average for tetraploids.

F_1 cross

A total of 83 individuals, 70 F_1 individuals from PC17-109 (female) x PC20-102 (male) and 13 F_1 individuals from PC20-102 (female) x PC17-109 (male) were sampled. Only 6 out of 59 possible KASP genotyping assays showed both parents as homozygous SNPs but for opposite alleles. Three representative assays were selected which showed one SNP heterozygous for one parent and one SNP homozygous for the other parent, and three representative assays were selected which showed both parents as heterozygous SNPs (Table 3.3). All SNP codes that could not be accurately identified, or called, due to not appearing in one of the three genotypes in the software, were removed from the χ^2 analysis. Four individuals consistently did not satisfy

the expected heterozygous SNP genotype, with regards to KASP genotyping assays for which both parents were homozygous for opposite alleles (pcg_00050, pcg_00058, pcg_00059, pcg_000106, pcg_1186, and pcg_14142). These four individuals after being analyzed across all 12 assays, were identified as being selfs, and were removed from the χ^2 analysis (Table 3.3). Using the resulting trimmed data, the χ^2 analysis followed normal Mendelian inheritance patterns and could not be rejected for all the primers used in the χ^2 analysis (Table 3.5).

F_2 selfs

The F_1 parent genotype was identified in order to identify SNPs in which the parent was homozygous. The DNA fingerprint for the 8 F_1 parents using the 12 primers were as follows: 13-F1008: *GG, GG, AA, TC, GT, CG, CG, CT, GC, TG, GC, AA*; 13-F1011: *GG, GG, AA, TC, GT, CG, CG, CT, GC, GG, GG, AG*; 14-F1008: *GA, GA, TA, TC, GT, CG, CG, CT, GC, TT, CC, AG*; 14-F1014: *GG, GG, AA, TC, GT, CG, CG, CT, GC, TG, GC, AG*; 14-F1015: *GG, GG, AA, TC, GT, CG, CG, CT, GC, TT, CC, AG*; 14-F1042: *GG, GG, AA, TC, GT, CG, CG, CT, GC, GG, GG, AG*; 14-F1067; *GG, GG, AA, TC, GT, CG, CG, CT, GC, TG, GC, GG*; 14-F1071; *GA, GA, TA, TC, GT, CG, CG, CT, GC, TG, GC, AA*. For 3 F_1 parents that were selfed, there were F_2 progeny that did not fall into the expected homozygous parental genotype (example in Table 3.6). There were 2 F_2 progeny which consistently were identified as an unexpected genotype for 13-F1008, 1 progeny for 14-F1014, and 4 progeny for 14-F1071. Individuals that consistently fell into the heterozygous (unexpected) genotype category across multiple homozygous primers were considered outcrosses and not true F_2 selfs (Table 3.6). Most of the F_2 progeny were identified as an expected SNP genotype when considering the parental genotype.

DISCUSSION

In order to validate SNP polymorphisms in prairie cordgrass, 121 SNPs identified from the nuclear transcriptome were sent for KASP assay development. Among 121 SNPs, the assay success rate was 78.5% with 26 assays failing development. This is comparable with findings in the literature of success rates of 83% (Cockram et al. 2012), 88.4% (Saxena et al. 2012), and 80.9% (Semagn et al. 2014). The assays failed mainly due to paralogs within the prairie cordgrass genome. Because not all of the populations used to develop the transcriptome were in the core collection of DNA used in this study, some assays appeared as monomorphic. These selected SNPs may have been derived from the octoploid populations not present in the populations that were chosen for this study. Three monomorphic SNPs were selected for further analysis, to see if the SNPs would be identified to be polymorphic in further studies. With the failed and monomorphic assays removed, 84 putative SNPs were validated as true allelic variants, with 59 SNPs being selected for use in this study. The 59 highly polymorphic assays were selected based on the criteria that there were at least two of the three genotypes present in a large portion of the samples analyzed. These assays were tested on the 38 natural populations, creating a phylogenetic tree that resulted in one main clade. If subclades were observed, the two plants of a single population were represented in the subclade.

Within the populations, just over half of the populations showed polymorphisms, and consisted of an equal number of octoploid and tetraploid populations. This implies that that, in this study, intrapopulational polymorphisms appear not to be favorable in one cytotype over another. The average number of polymorphisms that occurred within each population did not vary between octoploid and tetraploid populations. This is different from a chloroplast DNA

study of prairie cordgrass, in which a majority of the polymorphisms was derived from the octoploid cytotype relative to the tetraploid cytotype (Graves et al. 2015).

SNPs were successfully identified in nuclear transcriptomes of prairie cordgrass and validated as allelic variants that can be used in prairie cordgrass. SNP markers were used to detect significant polymorphisms in prairie cordgrass populations collected from distinct geographic regions in the U.S. These SNP polymorphisms appear to reflect genetic relationships in prairie cordgrass, and therefore, can be used to assess genetic diversity within and among populations in future studies. Knowledge of the starting plant material SNP genotype and of any genetic variation that occurs within that population should be analyzed when breeding this species or any type of species.

The F_1 population, consisting of 83 plants, allows for the assessment of the fidelity of a specific tetraploid cross. Due to the lack of synchronization between the pollen and the ovaries, less seeds were obtained when PC20-12 was used as the female, compared to when PC17-109 was used as the female. Progeny that had SNP genotypes matching the female parent only were determined to be selfs. Of the F_1 progeny, 95.2% were identified to be hybrids. Prairie cordgrass is a protogynous outcrossing species (Gedye et al. 2012), leading to the possibility that later-maturing stigmas could have been exposed to pollen from the same female parent, resulting in 4.8% of the F_1 being selfs. The analysis of the 76 F_2 progeny obtained by selfing eight F_1 plants indicate that the SNPs, and the SNP markers chosen, could distinguish between a true selfed plant and an outcrossed plant. This is based on individuals consistently being genotyped as heterozygous rather than being homozygous as expected. Ninty-one percent of the F_2 progeny were identified as successful selfs. Because of the protogynous nature of this species, there is already a natural element working against selfing. This could explain why outcrossed

individuals were identified. It is a possibility that some of the early-maturing stigmas were exposed to pollen within the greenhouse before being bagged. This could explain why more F_2 progeny were identified as unexpected genotypes (outcrosses) than the unexpected genotype (selfs) of the F_1 progeny.

Even though prairie cordgrass is a polyploid, there is evidence that the tetraploid cytotype is an allotetraploid that may follow a disomic inheritance pattern. Two divergent copies in the *Waxy* lineages of *Spartina* genus support the allotetraploid origin of *S.pectinata* (Fortune et al. 2007). The bivalent pairing that occurs during meiosis (Church 1940; Marchant 1968a, 1968b; Bishop 2015), and the observation of disomic inheritance using genotyping-by-sequencing suggests a disomic inheritance pattern in *S.pectinata* (Crawford 2015). This hypothesis was tested in a cross between two prairie cordgrass populations, exploiting the bi-allelic nature of the KASP technology to suggest Mendelian segregation ratios in a monohybrid type cross. The analysis of the F_1 hybrids and F_2 selfs conclude that disomic inheritance of SNPs in tetraploid prairie cordgrass is in agreement with the chromosomal and genomic evidence, and a possibility in this cross (Fortune et al. 2007; Marchent 1968a, 1968b; Bishop 2015; Crawford 2015).

The primary requirement of any breeding program is to ensure that accurate crosses are made (Glaszmann et al. 2010). The small flower size of prairie cordgrass and the large number of flowers per head make it hard to perform physical emasculation. Possibilities of self-cross pollination always exist, therefore, developing a molecular way to confirm true crosses/selfs is warranted (Fang et al. 2004, Gedye et al. 2012). In prairie cordgrass, SSR markers have been developed that identified successful crosses in this protogynous species without the need for emasculation. This study also confirms that hybrids of prairie cordgrass can be created and verified with molecular markers (Gedye et al. 2012). However, utilizing SSRs can be time-

consuming, limited in number, and more expensive than SNP markers, making a way for the introduction of these newly developed and validated KASP assays.

CONCLUSION

This study reports the first research of SNP marker development for use in prairie cordgrass. The SNP markers developed from the nuclear transcriptome were tested on a core collection of DNA, and found to be polymorphic among and within populations. The amount of variation found differs from previous findings in chloroplast DNA, in which more variation was identified within the octoploid cytotype. However, one must recognize these SNP markers cover a wide range of expressed genomic DNA versus two non-coding chloroplast DNA regions, giving nucleic SNP markers an advantage in identifying random genetic variation. These markers were used to assess the validity of true crosses that were made between two different populations using F1 and F2 (selfs of F1) progeny. Utilizing the biallelic nature of the KASP system, χ^2 analysis of the F_1 samples suggests that tetraploid prairie cordgrass may follow Mendelian disomic inheritance although other modes of inheritance were not ruled out. This analysis provides insight into the genomic structure of this species, supporting the hypothesis that tetraploid prairie cordgrass is an allotetraploid. The information from this study will allow the identification of more primers, which will prove invaluable in accurately and efficiently calculating genetic diversity. However, further analysis must be done on other cytotypes to understand completely the genome structure of this species and to evaluate genetic diversity. In addition, this study confirms the usefulness of using SNP marker technology in future breeding programs of this species, and opens up the ability for the final step using SNP markers for

genetic improvement; SNP genotyping germplasm collections or genetic/breeding populations of prairie cordgrass.

FIGURES AND TABLES

Table 3.1 Summary of SNP sequences, including SNP ID, SNP sequences, and SNP alleles. *failed primers

SNP ID	SNP Sequence	SNP ID	SNP Sequence
pcg_00001	GTCCTTGAGCTCGGC**[G/A]**TCCACGTCCAAGCG	pcg_00032	GCCAGTATTGGCAAG**[A/C]**ATGCAACAATTACT
*pcg_00002	CGCCAGGTACACCGG**[C/G]**GCCGCCTGGTTAGT	pcg_00033	AAAGACTACCCTTCC**[A/C]**TATCGAATAGAGAA
pcg_00003	GTCGGCCCCGGCCTC**[A/G]**AACCACGGGACGCC	pcg_00034	ACAGCTCCGGATGAA**[A/G]**TGGTACTTGATCCG
pcg_00004	ACCCGAAGGAGAAGG**[G/T]**CGCGATGGCGCCCG	*pcg_00035	TCTTTCGACCAAGTA**[A/G]**CTCACCCAGTAGGC
pcg_00005	AAGAACAAATTTATA**[A/G]**GTTAAATACATGCA	*pcg_00036	GCTCGTGTCGATGTC**[G/A]**CCGGCGAGGTCGCT
pcg_00006	GCCAAAGGACAGATC**[A/G]**TGAATAACATGACT	pcg_00037	CGAGGTGTGATGCAC**[T/C]**AGAACGCCGCTCGT
pcg_00007	CGGAACTGAGGAACA**[A/G]**TAGCATACATGCTT	*pcg_00038	GCTCACATACCCGAC**[A/G]**GCGAACGCCAAGTC
pcg_00008	GTTCGACCGCGCGGC**[A/C]**ATCGCCGAGCTCGA	*pcg_00039	TGGGCAGGGTTGCAG**[T/C]**CACCCATGCCTCCC
pcg_00009	GAGAAGAAGAGAGTG**[A/G]**TTGCATCATTGGAC	*pcg_00040	CGCTTCTTCCGTGCC**[A/G]**GTGATGACGAGGTC
pcg_00010	GGTGCGGCTTGACAA**[T/C]**GTCACAATACAAGT	*pcg_00041	GTGTCCCCGGCCTCG**[C/T]**CGGTACACCGCCGC
pcg_00011	CTGTTTGTTAAGTGC**[A/G]**CTGAATTTGAGATT	pcg_00042	GCGGTGCTTGCCGCA**[A/G]**CCCGTACAAGGCCT
pcg_00012	GCATTCATGTTCCCA**[A/G]**TACATCCTGGCAAA	*pcg_00043	CTTCTTGAGCTTGAA**[T/C]**ACCCACTTCAGGGT
pcg_00013	ACAATCATTGTTTTT**[T/C]**GTAATTGGGGAACT	*pcg_00044	CGGGCGGTGGCCGGC**[T/C]**GGCAAGTCGACGAG
pcg_00014	CCAAATGGCAAAAAT**[T/G]**TACTCAGATTTCCA	pcg_00045	AAGTCAGTTGTTGTC**[T/A]**GCAACCCTCATCGT
pcg_00015	ACTTGATTTAGAGTC**[G/A]**GCAGACATCATTTT	*pcg_00046	TCTGTTTGATTACCA**[C/T]**GGTAAGCTCACTCA
pcg_00016	AGCGCTTCACGCGAT**[A/G]**GAGTTCTCCGAAAT	*pcg_00047	TTACCAAATACCCAG**[A/G]**TGCAGAGTTCAAGC
pcg_00017	TAGCTTTAGGTGTTG**[G/A]**GTTTCGCATCAGTA	pcg_00048	AAGCAACAACTTACT**[C/T]**GAGCAAAGTGCAAG
pcg_00018	AGAACCAACTCTTTA**[C/T]**ATCAGACTGCGTAT	pcg_00049	TTACTTTCATATAAC**[G/A]**GGATGAAGCATGCA
pcg_00019	AACAAAGACAACATG**[A/G]**CTCACGAGAAATTG	pcg_00050	CAGGGACATTCGTTT**[C/T]**GTCCTCCAAAAATA
pcg_00020	TTTGGATGTTGAACT**[G/A]**TCTCAGATGTCCTT	pcg_00051	CCCTTGAATGGCTTC**[T/C]**TTTTCTTTTGTGCA
pcg_00021	TTTGGATGTTGAACT**[G/A]**TCTCAGATGTCCTT	pcg_00052	CACCAACCACTTGTC**[A/G]**TGGTGACGCTTCGT
pcg_00022	ATGAATTTTGGCACG**[A/G]**ACTTTTTGTTTGAA	pcg_00053	CGAGGTTGATGTTTA**[T/C]**GCTCGTCGATGACG
pcg_00023	GCATCCACAAGAATG**[G/C]**CCATGAACAATTAA	pcg_00054	AAAGTATTTGTAGGA**[G/A]**ACCCCTGAGGGTTC
pcg_00024	GATCGAGAAAAAAAAA**[A/T]**TTGGATGAAGATTC	pcg_00055	CTCGCGTGGCCTTCT**[C/G]**TGTCATAAACCATG
pcg_00025	TTTGAGGAGGACGGT**[G/A]**ATGATAGCAAATCT	*pcg_00056	GGAACGTATCCTGTG**[T/C]**ATAAGGGCTCTCCG
pcg_00026	GTGAGGGATAGATTG**[T/G]**CAAGCAATGCAAGT	pcg_00057	AACTTGGTATCAGAC**[C/G]**GCCAAGGTTAAACC
pcg_00027	CCATCTAAGGTCAGG**[A/G]**TTCTAAGTTCATTC	pcg_00058	GGCACGGTAAACCTT**[T/G]**GCAAAGGTCCCTTG
pcg_00028	ACATTCTTCCGATCT**[C/A]**GGGTTTTTAACCCA	pcg_00059	TCAACCGTCTCCCCC**[G/C]**AGATGATTGTCTAA
pcg_00029	GGACCATTTGTTGTC**[A/G]**TCAAGGTTTCCCAG	pcg_00060	CACCCCACAAGACCA**[T/A]**ATGTCGGCTTTTGC
pcg_00030	GAGAGCATTGATGTC**[G/A]**CTGGCTCTTGGAAA	pcg_00061	AAATCTTTTTTTCCA**[G/T]**TATCTTTTTTCTTA
pcg_00031	AGTTAGACCTGAGAT**[T/C]**GAACATTCTGAAAA	pcg_00062	GTAATTGTTTGCAGA**[C/G]**AACTTTTCATTTGT

Table 3.1 (cont.)

SNP ID	SNP Sequence	SNP ID	SNP Sequence
pcg_00063	GGAAGATATGCAACA**[C/T]**TTTGGGGAGGAAGC	pcg_00093	TACTGGGAAGAAACC**[G/A]**TTCCACTTGTCCTG
pcg_00064	GGGGATGTCACCCTT**[C/T]**CCCGGCGCGGTGAT	pcg_00094	GCTCTCCGCACACGC**[C/T]**GCCACCGCTACATC
pcg_00065	CAGCGGCAGCGACGC**[G/A]**GCGCTCCTGAGCCC	*pcg_00095	TGGTGAAAAGGTCCT**[G/C]**ATCCAGTTTGAGGA
pcg_00066	CGGCTTCGACCCGCT**[C/G]**GGCCTGGCGGAGGA	*pcg_00096	CAGGGACCGGAACCG**[G/A]**TTCCACCGGTTCAG
pcg_00067	GAGGATGTTGTCGAG**[C/T]**TTGACGTCGCGGTG	*pcg_00097	TTTTGTTACAAAATA**[C/T]**GAGCAAGCTCTGTT
*pcg_00068	CAATCCTGGAAAGGA**[C/T]**CCACTAATGTTTGT	pcg_00098	GTACAATGTCTGGGC**[C/A]**AGTACTCCTAATGG
pcg_00069	TGAAGTAACTACTAA**[A/T]**ATAGTACTGTTGTA	pcg_00099	AAAAAAAAGATGATG**[A/T]**CAGGTTACAAATTG
pcg_00070	AGGCTCTCACGATCA**[T/G]**TCCGAGTCGCTGTC	*pcg_00100	GACTCTCTACGGCTC**[C/A]**TCCAGGCTCACCGC
pcg_00071	GGCAAGGCTTTTACA**[A/C]**AAGAAGTTGTCGAG	pcg_00101	AGTACATGCAGGAGG**[G/A]**GCATTCTCTTCCTT
*pcg_00072	GGAGTACAATGGAAA**[A/G]**CTTCATGTGCCTGG	pcg_00102	CCATTTGAATCTCAA**[G/A]**GCACTGACGTGAAC
pcg_00073	TTTCCCTGGATTTGG**[C/T]**CTGGGTCTTGTTAT	pcg_00103	GCTAGCTTTTGCGCC**[C/T]**CTATACATCTTTTC
*pcg_00074	CGAGCATATAATATG**[G/A]**CCCTAAAATGATGG	pcg_00104	CGTCCTCGTCGTCTT**[C/G]**TTCCTCTGGCTGCT
pcg_00075	CGGCCGCGAGGACTC**[G/C]**CCGCTCGACATCAT	pcg_00105	GCTTGTGCTCATGGA**[T/C]**GTGGTTCACAGCCA
pcg_00076	CATCCCCACCTACGT**[C/G]**GTCGGAGTCAATGC	pcg_00106	ATTGGTGCTGTTGCT**[G/C]**GACGTGAAGCTGAC
*pcg_00077	CTCCTGCACCACCAA**[C/T]**TGCCTCGCGCCCTT	pcg_00107	AGATGACGGAGTCGG**[C/A]**GACGACGTGGGAGC
pcg_00078	ATGGAGGGACACAGC**[C/A]**GGCAAAGTGGATGT	pcg_00108	CTCTTTGCGCATGTG**[G/A]**CTCTTTTCCAGGGC
pcg_00079	AGATTCTGATATTGA**[T/C]**TTGGATGACTATTC	pcg_00109	ACTCAGACCATTTTG**[A/G]**ACCACCTCAGATGT
pcg_00080	TGCGTATATTCTCCG**[T/G]**GGTGAGACCAAAAT	pcg_00110	TATGTTATCTCAATG**[T/G]**GATCTACACCTGCA
pcg_00081	GCTCGCCCTCGCAAC**[T/A]**ATCGGATCTTGCGC	pcg_00111	GCCGACGGGATGCGG**[C/G]**CGATTCACATTTGC
*pcg_00082	CTGGCTGTAGGAATG**[G/A]**CCTTTTCACCTGAA	pcg_00112	TGACCACATGCCATG**[A/G]**GTATCAAGCCTATT
pcg_00083	TGAAGTTATGTATGA**[T/C]**CTGAGAGCTAGTGG	pcg_1186	GACCTCGCAGAACAC**[T/C]**GCAGACATGACCTC
pcg_00084	AAGTTCGGGATCAGC**[A/T]**CCGTGTATTTGGGA	pcg_13880	TCAAGTACCTCACCG**[G/A]**CGAGGCCAAGGCTT
pcg_00085	CTTCTGAAGTCGGAA**[C/A]**TGCCATCAAACTGG	pcg_14142	CACGCAGTTGGGGGC**[C/G]**AGGATGAGGACGAC
*pcg_00086	AGGAGTATCCACCTG**[G/T]**AATAACACTTGTAC	pcg_2412	CACATTGCGATTAGC**[G/A]**TATCGATCATGAAA
*pcg_00087	CAACACAATGAATCG**[T/G]**ATTGGAAAAGGAAG	pcg_37652	ACTTGAAGAGAGACG**[C/A]**ATCTGAAGGCAGAT
pcg_00088	TTTACAAATGCATAA**[A/G]**ATCTATGTTGGTAA	*pcg_38909	CACGCAGTTGGGGGC**[C/G]**AGGATGAGGACGAC
pcg_00089	CGTACCTGCAGTTCA**[T/C]**GTTCGCCTACATCT	pcg_77221	GAGCTCGCCAGGCAC**[G/T]**CTGGCTTCTGTGGC
pcg_00090	GAGGGGTAGTAAGAA**[A/G]**ACAAAGGAGACGTG	pcg_7965	TGACCAGCCGCAGCA**[G/A]**CCGCTCGTGGTAGT
pcg_00091	GGTACATAGTTTGAT**[C/T]**CACCTCCCTTCCTC	pcg_80876	TGGCGTCGTAGGTGC**[G/A]**CCACGGAGGACGCG
*pcg_00092	ATGGGAAGACAGGTT**[T/C]**GCAGCTTCATTATT		

*failed primers

Table 3.2 Summary of plant materials used including, location, cytotype and number of plants used per population

ID	Location	Ploidy	Number of samples
103 4X	IL	4X	2
9046803	NY	4X	2
IL102	IL	4X	2
IL99A	IL	4X	2
MBB4X	IL	4X	2
PC09-101	CT	4X	2
PC09-102	CT	4X	2
PC17-109	IL	4X	2
PC17-111 4X	IL	4X	2
PC19-101	IA	4X	2
PC19-103	IA	4X	2
PC19-105	IA	4X	2
PC20-102	KS	4X	2
PC20-105	KS	4X	2
PC22-101	LA	4X	2
PC23-101	ME	4X	2
PC23-104	ME	4X	2
PC29-101	MO	4X	2
PC29-104	MO	4X	2
PC34-101	NJ	4X	2
PC40-101	OK	4X	2
PC55-102	WI	4X	2
PC55-103	WI	4X	2
ND-2-51-4	ND	8X	2
PC17_111 8X	IL	8X	2
PC19-106	IA	8X	2
PC19-107	IA	8X	2
PC19-108	IA	8X	2
PC20-104	KS	8X	2
PC20-106	KS	8X	2
PC27-103	MN	8X	2
PC31-101	NE	8X	2
PC31-104	NE	8X	2
PC38-101	ND	8X	2
PC40-104	OK	8X	2
PC46-110	SD	8X	2
PCG109	SD	8X	2
Red River	MN, SD, ND	8X	2
Total			76

Table 3.3 Primers selected for use on the prairie cordgrass F_1 progeny. The first three primers indicate one parent as heterozygous and one parent as homozygous, the next six primers indicate both parents as homozygous for opposite alleles, and the last three primers indicate both parents as heterozygous. Also shown are SNP assay results for eight out of the 83 F_1 hybrids. Indicated are samples that can be identified as true crosses and selfs.

	pcg_00011	**pcg_00012**	**pcg_00024**	**pcg_00050***	**pcg_00058***	**pcg_00059***	**pcg_00106***	**pcg_1186***	**pcg_14142***	**pcg_00061**	**pcg_00062**	**pcg_7965**
PC17_109 (Parent)	GA	GA	TA	TT	GG	CC	CC	CC	GG	TG	GC	AG
PC20_102 (Parent)	GG	GG	AA	CC	TT	GG	GG	TT	CC	TG	GC	AG
13_F1001†	GG	GG	AA	CC	TT	GG	GG	TT	CC	TT	CC	AG
13_F1002	GA	GA	TA	TC	GT	CG	CG	CT	GC	GG	GG	GG
13_F1003	GA	GA	TA	TC	GT	CG	CG	CT	GC	GG	GG	AG
13_F1004	GA	GA	TA	TC	GT	CG	CG	CT	G:C	TG	GC	AA
13_F1005	GA	GA	TA	TC	GT	CG	CG	CT	GC	TT	CC	AG
13_F1006	GA	GA	TA	TC	GT	CG	CG	CT	GC	TG	GC	AG
13_F1007†	GG	GG	AA	CC	TT	GG	GG	TT	CC	GG	GG	AG
13_F1008	GG	GG	AA	TC	GT	CG	CG	CT	GC	TG	GC	AA

* Primers that can distinguish true crosses from selfed samples

† F_1 individuals that are identified as selfed samples

Table 3.4 DNA fingerprint using SNP codes

Subject ID	pcg_00006	pcg_00008	pcg_00009	pcg_00010	pcg_00011	pcg_00012	pcg_00013	pcg_00014	pcg_00015	pcg_00016	pcg_00017	pcg_00018	pcg_00021	pcg_00022	pcg_00023	pcg_00024	pcg_00025	pcg_00026	pcg_00027	pcg_00028	pcg_00029	pcg_00032	pcg_00034	pcg_00037	pcg_00042
PC19_103_1	--	AA	AA	TT	GA	GA	CT	GT	AG	GA	AG	TC	AG	AA	CG	AA	GG	GT	GA	AC	GA	CA	GG	CC	GG
PC19_103_2	--	AA	AA	TT	GA	GA	CT	GT	AG	GA	AG	TC	AG	AA	CG	AA	GG	GT	GA	AC	GA	CA	GG	CC	GG
PC20_104_1	GA	--	--	--	--	--	CT	GT	AG	GA	AG	TC	AG	GA	CG	TA	AG	GT	GA	AC	GA	CA	GA	CT	GA
PC20_104_2	GA	--	--	--	--	--	CT	GT	AG	GA	AG	TC	AG	GA	CG	TA	AG	GT	GA	AC	GA	CA	GA	CT	GA
PC34_101_1	GA	CA	--	TT	GA	GA	CC	GG	AA	AA	AA	TT	AA	AA	CC	--	GG	GG	GG	AA	GG	CC	GG	CC	GG
PC34_101_2	GA	AA	--	CT	--	GA	CT	GT	AG	--	AG	TC	AG	--	CG	TA	--	GT	--	AC	GA	CA	--	--	--
PC38_101_1	GA	CA	GA	--	AA	AA	TT	GT	GG	--	GG	CC	GG	GA	GG	TA	AG	TT	AA	CC	AA	AA	GA	CT	GA
PC38_101_2	GA	--	--	TT	GA	GA	--	GT	AG	--	AG	TC	AG	AA	CG	TA	GG	GT	GA	AC	GA	--	GG	CC	GG
Red_River_1	AA	CA	GA	CT	GA	GA	CT	GT	AG	GA	AG	TC	AG	GA	CG	TA	AG	GT	GA	AC	GA	CA	GA	CT	GA
Red_River_2	AA	AA	AA	CT	GA	GA	CT	GG	AG	AA	AG	TC	AG	AA	CG	AA	GG	GT	--	AC	GA	CA	GA	CT	GA
PC17_109	GA	CA	--	TT	GA	GA	CC	GG	AA	AA	AA	TT	AA	AA	CC	TA	GG	GG	GG	AA	GG	CC	GG	CC	GG
PC20_102	GA	AA	AA	TT	GG	GG	CC	GG	AA	AA	AA	TT	AA	AA	CC	AA	GG	GG	GG	AA	GG	CC	GG	CC	GG
S.alterniflora(East)	AA	CC	AA	TT	AA	AA	CC	--	--	GG	GG	CC	AA	AA	CC	AA	GG	GG	GG	AA	GG	CC	GG	CC	--
S.bakeri	AA	CC	AA	TT	AA	AA	CC	--	--	GG	GG	CC	AA	AA	CC	TA	GG	GG	GG	AA	GG	CC	GG	CC	--
S.patens(Sharp)	AA	CC	AA	TT	AA	AA	CC	--	--	GG	GG	CC	AA	AA	CC	AA	GG	GG	GG	AA	GG	CC	GG	CC	--
S.patens(Flageo)	AA	CC	AA	TT	AA	AA	CC	--	--	GG	GG	CC	AA	AA	CC	AA	GG	GG	GG	AA	GG	CC	GG	CC	--
S.patens	AA	CC	AA	TT	AA	AA	CC	--	--	GG	GG	CC	AA	AA	CC	AA	GG	GG	GG	AA	GG	CC	GG	CC	--

Table 3.4 (cont.)

Subject ID	pcg_00049	pcg_00050	pcg_00058	pcg_00059	pcg_00060	pcg_00061	pcg_00062	pcg_00065	pcg_00066	pcg_00078	pcg_00081	pcg_00083	pcg_00084	pcg_00085	pcg_00088	pcg_00092	pcg_00093	pcg_00098	pcg_00101	pcg_00102	pcg_00103	pcg_00104	pcg_00106
PC19_103_1	AG	TT	GG	CC	TT	GG	GG	GG	GG	AC	TT	CT	TA	--	AA	CC	GG	CC	GG	AG	TC	CC	CG
PC19_103_2	AG	TT	GG	CC	TT	GG	GG	GG	--	AC	TT	CT	TA	--	AA	CC	GG	CC	GG	AG	TC	CC	CG
PC20_104_1	GG	CC	GT	GG	AT	TG	GC	--	GC	AC	AT	TT	AA	AC	GA	CC	AG	--	GG	AG	CC	GC	--
PC20_104_2	GG	CC	GT	GG	AT	TG	GC	--	GC	AC	AT	TT	AA	AC	GA	CC	AG	--	GG	AG	CC	GC	--
PC34_101_1	AA	TT	GG	CG	TT	TG	GC	AG	GG	AC	TT	CT	TA	AA	--	CT	GG	CC	GG	AA	TC	CC	CC
PC34_101_2	--	CC	GT	--	--	TG	--	AG	GC	--	TT	CT	TA	CC	--	CT	GG	AC	AG	AG	CC	CC	CG
PC38_101_1	GG	CC	GT	GG	--	GG	GG	GG	GC	AC	TT	CT	TT	AC	GA	CT	AG	CC	AG	AG	TC	GC	CG
PC38_101_2	AA	TT	GG	CC	AT	GG	GG	GG	GC	CC	TT	CC	TT	CC	GG	TT	GG	AA	--	AG	CC	CC	CC
Red_River_1	AG	CC	GT	GG	TT	TG	GC	AG	GC	AC	AT	CT	TA	CC	GA	CT	AG	AC	AG	GG	--	GC	--
Red_River_2	AG	CC	GT	CG	AT	TG	GC	--	GC	CC	AT	CT	TA	CC	GA	CT	AG	AC	GG	AG	CC	CC	--
PC17_109	AA	TT	GG	CC	TT	TG	GC	AG	GG	CC	TT	CC	TT	AC	GG	TT	GG	AC	AG	AG	TC	CC	CC
PC20_102	AG	CC	TT	GG	TT	TG	GC	GG	GC	CC	AT	CT	TA	CC	GA	CT	AG	CC	GG	AA	CC	CC	GG
S.alterniflora(East)	GG	CC	GG	GG	AA	--	CC	GG	GC	CC	TT	TT	AA	CC	GG	TT	GG	AA	GG	GG	CC	--	CG
S.bakeri	GG	CC	GG	GG	AA	TT	CC	GG	GC	CC	TT	TT	AA	CC	GG	CC	AG	AA	GG	GG	TT	CC	CG
S.patens(Sharp)	GG	CC	GG	GG	AA	TT	CC	GG	GC	CC	TT	TT	AA	CC	GG	CC	AG	AA	GG	GG	TC	CC	GG
S.patens(Flageo)	GG	CC	GG	GG	AA	TT	CC	GG	GC	CC	TT	TT	AA	CC	GG	CC	AG	AA	GG	GG	CC	CC	GG
S.patens	GG	CC	GG	GG	AA	TT	CC	GG	GC	CC	TT	TT	AA	CC	GG	CC	GG	AA	GG	GG	CC	CC	GG

Table 3.4 (cont.)

Subject ID	pcg_00109	pcg_00110	pcg_00111	pcg_00112	pcg_7965	pcg_14142	pcg_1186	pcg_77221	pcg_13880	pcg_2412	pcg_37652
PC19_103_1	GG	GG	CC	AA	AG	GC	CT	GG	GG	GG	CC
PC19_103_2	GG	GG	CC	AA	AG	GC	CT	GG	GG	GG	CC
PC20_104_1	GA	GT	CC	AA	GG	GC	CT	TT	GG	GG	CC
PC20_104_2	GA	GT	CC	AA	--	GC	GT	TT	GG	GG	CC
PC34_101_1	GG	GG	GC	AA	GG	GG	CT	GG	GG	GG	CC
PC34_101_2	GG	GG	--	AA	--	--	TT	GG	GG	GG	CC
PC38_101_1	GA	--	CC	AA	GG	GC	TT	GG	GG	GG	CC
PC38_101_2	GG	GG	GC	AA	AG	GG	CT	GG	GG	GG	CC
Red_River_1	GG	GT	GC	AA	GG	GC	CT	GG	GG	GG	CC
Red_River_2	GG	GT	GC	AA	AG	--	TT	GG	GG	GG	CC
PC17_109	GG	GG	GC	AA	AG	GG	CC	GG	GG	GG	CC
PC20_102	GA	GT	CC	AA	AG	CC	TT	GG	GG	GG	CC
S.alterniflora(East)	GG	TT	--	AA	GG	CC					
S.bakeri	GG	TT	CC	AA	GG	CC					
S.patens(Sharp)	GG	TT	CC	AA	GG	CC					
S.patens(Flageo)	GG	TT	CC	AA	GG	CC					
S.patens	GG	TT	CC	AA	GG	CC					

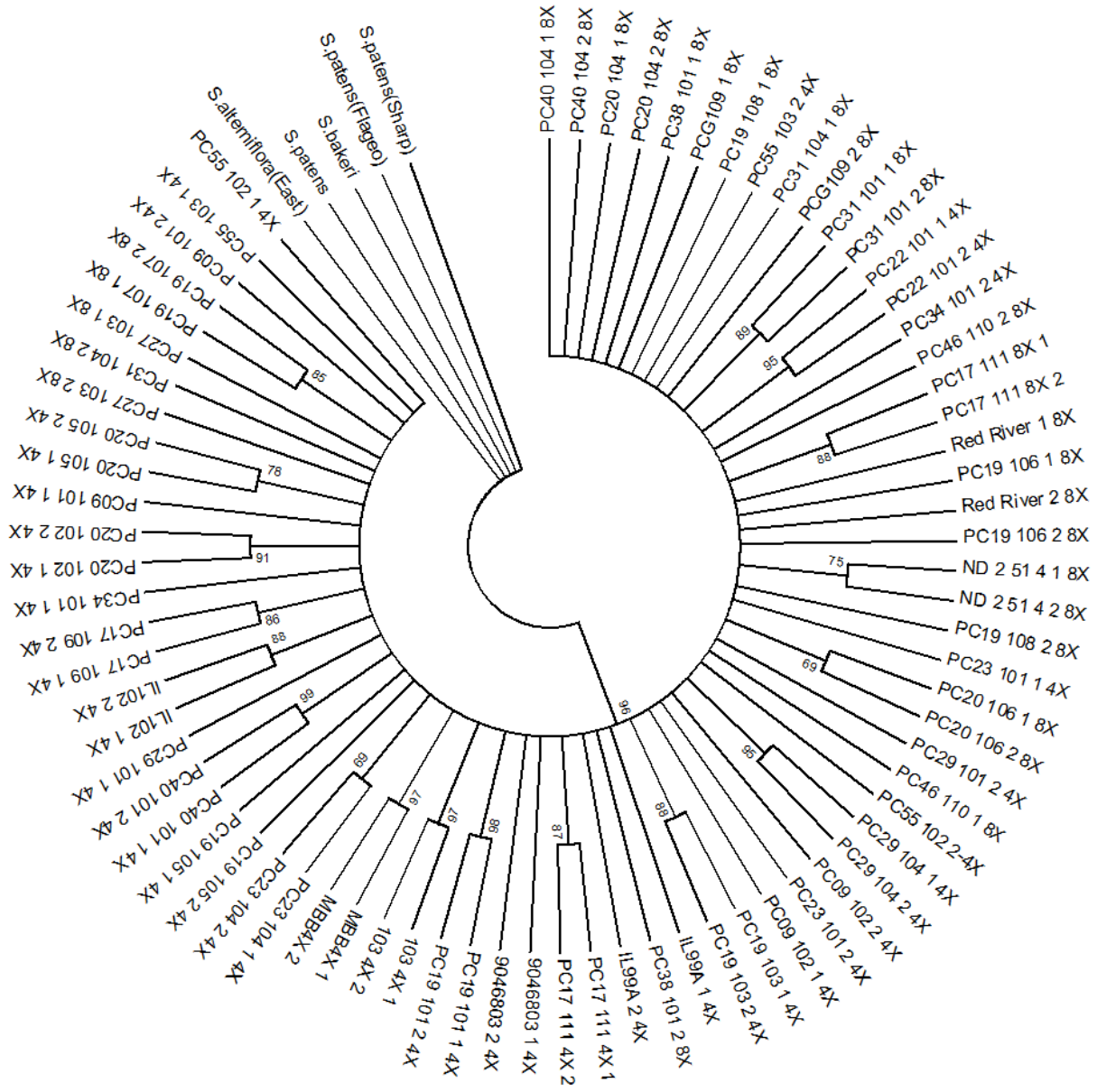

Figure 3.1 Phylogenetic tree based on maximum parsimony analysis of combined SNP codes to create a DNA fingerprint sequence for the 38 prairie cordgrass core collection populations with a total of 76 plants and 5 *Spartina* outgroups. Bootstrap values are indicated on the nodes as percentages. One main clade is identified.

Table 3.5 Summary of χ^2 analysis on F_1 progeny of a specific tetraploid prairie cordgrass cross with selfed data removed. Analysis indicates that all primers produce expected results from a monohybrid Mendelian cross. df=2, p=0.05, χ^2= 5.991

Primer ID	Observed Y Allele	Observed XY Allele	Observed X Allele	Expected Y Allele	Expected XY Allele	Expected X Allele	χ^2
pcg_00011	44	35	0	39.5	39.5	0	1.025
pcg_00012	43	35	0	39	39	0	0.821
pcg_00024	0	35	43	0	39	39	0.821
pcg_00050	0	77	0	0	77	0	0
pcg_00058	0	77	0	0	77	0	0
pcg_00059	0	78	0	0	78	0	0
pcg_00106	0	75	0	0	75	0	0
pcg_1186	0	78	0	0	78	0	0
pcg_14142	0	79	0	0	79	0	0
pcg_00061	15	37	27	19.8	39.5	19.8	3.962
pcg_00062	27	37	15	19.8	39.5	19.8	3.962
pcg_7965	20	42	17	19.8	39.5	19.8	0.544

Table 3.6 SNP assay results for F_2 individuals of two out of the eight selfed F_1 samples. Indicated are samples that can be identified as true selfs and as outcrossed.

	pcg_00011*	**pcg_00012***	**pcg_00024***	**pcg_00050**	**pcg_00058**	**pcg_00059**	**pcg_00106**	**pcg_1186**	**pcg_14142**	**pcg_00061**	**pcg_00062**	**pcg_7965***
13-F1008 (Parent)	GG	GG	AA	TC	GT	CG	CG	CT	GC	TG	GC	AA
F2:2012_13_F1_008_1	GG	GG	AA	TC	GT	CG	GG	TT	GG	GG	GG	AA
F2:2012_13_F1_008_2	GG	GG	AA	TT	GT	CG	GG	TT	GG	TG	GC	AA
F2:2012_13_F1_008_3	GG	GG	AA	TC	GT	CG	CG	CT	GC	TG	GC	AA
F2:2012_13_F1_008_4	GG	GG	AA	TC	GG	CC	CG	CT	GC	TG	GC	AA
F2:2012_13_F1_008_5†	GA	GA	TA	TT	TT	GG	CG	CC	GG	TT	CC	AG
F2:2012_13_F1_008_6†	GA	GA	TA	TC	GT	CG	CC	CC	GC	TG	CC	AG
F2:2012_13_F1_008_7	GG	GG	AA	TC	TT	GG	CG	CC	GC	TT	CC	AA
F2:2012_13_F1_008_8	GG	GG	AA	CC	TT	GG	CG	TT	GC	GG	GG	AA
F2:2012_13_F1_008_9	GG	GG	AA	TC	GG	CC	GG	CT	GC	TG	GC	AA

	pcg_00011*	**pcg_00012***	**pcg_00024***	**pcg_00050**	**pcg_00058**	**pcg_00059**	**pcg_00106**	**pcg_1186**	**pcg_14142**	**pcg_00061***	**pcg_00062***	**pcg_7965**
13-F1011 (Parent)	GG	GG	AA	TC	GT	CG	CG	CT	GC	GG	GG	AG
F2:2012_13_F1_011_1	GG	GG	AA	TC	TT	GG	GG	CT	GC	GG	GG	AG
F2:2012_13_F1_011_3	GG	GG	AA	TT	GT	CG	CG	CT	GC	GG	GG	AG
F2:2012_13_F1_011_4	GG	GG	AA	CC	GT	CG	GG	CT	GC	GG	GG	AG
F2:2012_13_F1_011_5	GG	GG	AA	TT	GT	CG	CG	CT	GC	GG	GG	GG
F2:2012_13_F1_011_6	GG	GG	AA	CC	GT	CG	CG	CC	GC	GG	GG	AA
F2:2012_13_F1_011_7	GG	GG	AA	CC	GG	CC	GG	CC	GC	GG	GG	AG
F2:2012_13_F1_011_8	GG	GG	AA	TC	GT	CG	CG	CT	GC	GG	GG	AA
F2:2012_13_F1_011_9	GG	GG	AA	TC	GT	CG	CG	CT	GC	GG	GG	AG
F2:2012_13_F1_011_10	GG	GG	AA	CC	TT	GG	CG	CC	GC	GG	GG	AG
F2:2012_13_F1_011_2	GG	GG	AA	TC	GT	CG	CG	CT	GC	GG	GG	AA

* Primers that can distinguish true selfs from outcrossed samples

† F_2 individuals that are identified as outcrossed samples

CHAPTER 4

CONCLUSION

This study reports the first research of intrapopulational variation by cpDNA analysis, and SNP marker development for use in genetic diversity analysis of prairie cordgrass. In the cpDNA analysis, three main haplotypes were identified concluding that there is variation present among and within populations in this species. Whereas, when looking at the phylogenetic tree of the SNP marker analysis one main clade with small subclades was observed. Concerning the cpDNA study, DNA variation was associated with haplotypes identified, or clades, however, in the SNP marker analysis, DNA variation was not associated with clades, or with the haplotypes in the cpDNA study. Overall, both of these studies indicate that there is inter- and intrapopulational variation present. The first portion of the study could allow for the confirmation of true crosses when using two plants with differing cpDNA haplotypes, however, the use of SNP markers to confirm true crosses is a cheaper, quicker, and more reliable procedure, in addition to observing more diversity. This is due to the number of SNPs available in prairie cordgrass versus the number of cpDNA primers, the procedure, and the location of the SNPs in the genome rather than the chloroplast DNA. Understanding the amount, and types, of genetic diversity occurring in this species allows the breeder to select the better individuals to best fit the standards required for that specific program.

It was determined in the cpDNA study that utilizing cpDNA haplotypes to determine geographic origin is a possibility when sampling an octoploid cytotype plant since two of the haplotypes found appears to occur mainly within the octoploid plants sampled. This is useful

because most of the octoploids, which had the varying haplotypes, were located in the north-west region of the Midwest. In addition, this study brought forth the suggestion that the tetraploid prairie cordgrass populations appear more stable that the octoploid populations due to the small amount of variation that was observed in the tetraploid populations compared to the octoploid populations. By sampling a larger number of plants from one of the natural populations, there is the implication that if cpDNA haplotype analysis was going to be utilized to test a new population, selecting a subset of at least 16 random samples from that population would allow the researcher to have an accurate representation of that population for future studies. Understanding this cpDNA study provided insight into the usefulness of SNP's for population identification and lead to the second study of utilizing SNP markers in prairie cordgrass breeding.

The development of SNP markers, using the KASP technology system, in prairie cordgrass will allow for a more in depth analysis of the genetic diversity present in prairie cordgrass since these SNP markers were developed from the nuclear transcriptome. The ability to test these markers on a core collection of natural populations from a wide range of environments allows future breeders to discover the best way to use the developed markers for their breeding program. In contrast to the cpDNA study, it was observed that with the 59 markers used in this study, polymorphisms exist equally between the octoploid and tetraploid prairie cordgrass populations. The capability of these markers to decipher true crosses from selfed individuals also allows future breeders of prairie cordgrass the ability to select individuals with desired characteristics to cross and be certain that they are obtaining progeny that are true crosses of those individuals with desired characteristics to further their research. Breeders can begin to compare phenotypic traits with these SNP markers looking for a connection between

specific markers and desired phenotypic traits to being a marker-assisted-selection breeding program, which could save time when it comes to growing plants to full maturity before making selections.

When looking at the impact on the breeding of prairie cordgrass, utilizing the SNP markers over cpDNA haplotype analysis may be the best option for the breeder, due to the number of SNP markers (that could be) available, the speed of the procedure, and the reliability of the procedure. In addition, developing the hypothesis that tetraploid prairie cordgrass may follow a disomic inheritance pattern unearths a need to research the genomic structure of each cytotype so that future breeders may gain a further understanding of how traits are inherited. This information will be crucial for future breeders of this, and related, species. When breeding for conservation, the environments that this species will be used for will be very diverse, mostly marginal type land. This marginal land can be made up of dry, waterlogged, or even salt laden soils. Being able to utilize molecular markers to select traits that are applicable to those marginal environments is crucial to the future breeding of prairie cordgrass. In biomass production, understanding that there is a need for prairie cordgrass that is adapted to different geographic regions is also important. Selecting populations that are developed for those regions can be difficult, but using SNP markers to develop populations that are improved to grow in these different regions can enhance the breeding programs. This information will be crucial for future breeders of this, and related, species.

REFERENCES

Anderson EK, Voigt TB, Kim SM, Lee DK. (2015). Determining effects of sodicity and salinity on switchgrass and prairie cordgrass germination and plant growth. Indus. Crops and Products. 64, 79-87.

Barkworth ME, Anderton LK, Capels KM, Long S, Piep, MB (Eds.) (2007) Manual of Grasses for North America. Utah State University Press. Logan, Utah.

Bishop J. (2015) Chromosome pairing and its relationship to fertility in cytotypes of prairie cordgrass (*Spartina pectinata* Link). University of Illinois. MS Thesis. http://hdl.handle.net/2142/88316.

Blum MJ, Bando KJ, Katz M, Strong DR (2007) Geographic structure, genetic diversity and source tracking of Spartina alterniflora. J Biogeogr 34:2055-2069.

Bock R. (2007). Structure, function, and inheritance of plastid genomes. In: Bock R ed. Cell and molecular biology of plastids. Springer-Verlag. pp 29-63.

Boe A, Lee DK (2007). Genetic variation for biomass production in prairie cordgrass and switchgrass. Crop Science, 47(3), 929-934.

Boe A, Owens V, Gonzalez-Hernandez J, Stein J, Lee DK, Koo BC. (2009). Morphology and biomass production of prairie cordgrass on marginal lands. GCB Bioenergy, 1(3), 240-250.

Boe A, Springer T, Lee DK, Rayburn AL, Gonzalez-Hernandez J (2013) Underutilized grasses. In: Saha MC, Bhandari HS, Bouton JH (eds) Bioenergy Feedstocks: Breeding and Genetics. John Wiley & Sons, Inc, pp 173-205.

Bundock PC, Cross MJ, Shapter FM, Henry RJ. (2006). Robust allele-specific polymerase chain reaction markers developed for single nucleotide polymorphisms in expressed barley sequences. Theoretical and Applied Genetics, 112(2), 358-365.

Byers RL, Harker DB, Yourstone SM, Maughan PJ, Udall JA (2012). Development and mapping of SNP assays in allotetraploid cotton. Theoretical and Applied Genetics, 124(7), 1201-1214.

Chen T, Wang X, Tang H, Chen Q, Huang X, Chen J (2013) Genetic diversity and population structure of Chinese Cherry revealed by chloroplast DNA trnQ-rps16 intergenic spacers variation. Genet Resour and Crop Evol 60:1859-1871.

Church GL (1940) Cytotaxonomic studies in the gramineae *Spartina, Andropogon* and *Panicum*. Amer Jour Bot 27:263-271.

Coates DJ & Byrne M. (2005*)* Genetic variation in plant population: assessing cause and pattern. In: Henry, R.J. (ed) Plant diversity and evolution: genotypic and phenotypic variation in higher plants. CABI Publishing, Cambridge, MA pp 139-164.

Cockram J, Jones H, Norris C, O'Sullivan DM. (2012). Evaluation of diagnostic molecular markers for DUS phenotypic assessment in the cereal crop, barley (Hordeum vulgare ssp. vulgare L.). Theoretical and Applied Genetics, 125(8), 1735-1749.

Crawford J. (2015) Linkage mapping in prairie cordgrass (*Spartina pectinata* Link) using genotyping-by-sequencing markers. University of Illinois MS Thesis. http://hdl.handle.net/2142/88128.

Cuenca J, Aleza P, Navarro L., Ollitrault P. (2013). Assignment of SNP allelic configuration in polyploids using competitive allelespecific PCR: application to citrus triploid progeny. Annals of botany, mct032.

Daniell H (2002) Molecular strategies for gene containment in transgenic crops. Nat Biotechnol 20:581-586.

Demesure B, Sodzi N, Petit R (1995) A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. Mol Ecol 4:129-134.

Doyle JJ, Doyle JL, Brown AH, Grace JP (1990) Multiple origins of polyploids in the Glycine tabacina complex inferred from chloroplast DNA polymorphism. Proc Natl Acad Sci 87:714-717.

Doyle JJ, Doyle JL, Brown AHD (1999) Origins, colonization, and lineage recombination in a widespread perennial soybean polyploid complex. Proc Natl Acad Sci 96:10741-10745.

Fang X, Subudhi PK, Venuto BC, Harrison SA. (2004). Mode of pollination, pollen germination, and seed set in smooth cordgrass (*Spartina alterniflora*, Poaceae). Int. J. of Plant Sci. *1*65, 395-401.

Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783-791.

Ferris C, King RA, Gray AJ (1997) Molecular evidence for the maternal parentage in the hybrid origin of Spartina angelica C.E. Hubbard. Mol Ecol 6:185-187.

Ferris C, King RA, Vainola R, Hewitt GM. (1998) Chloroplast DNA recognizes three refugial sources of European oaks and suggests independent eastern and western immigrations to Finland. Heredity 80:584-593.

Fineschi S, Taurchini D, Grossoni P, Petit RJ, Vendramin GG (2002) Chloroplast DNA variation of white oaks in Italy. For Ecol Manag 156:103-114.

Fortune PM, Schierenbeck KA, Ainouche AK, Jacquemin J, Wendel JF, Aïnouche ML. (2007). Evolutionary dynamics of Waxy and the origin of hexaploid *Spartina* species (Poaceae). Mol. Phylogenet. Evol. *43*, 1040-1055.

Gedye KR, Gonzalez-Hernandez JL, Owens V, Boe A. (2012). Advances towards a marker-assisted selection breeding program in prairie cordgrass, a biomass crop. Int. J. Plant Genomics 2012: doi:10.1155/2012/313545.

Gedye K, Gonzalez-Hernandez J, Ban Y, Ge X, Thimmapuram J, Sun F, ... & Owens V. (2010). Investigation of the transcriptome of prairie cord grass, a new cellulosic biomass crop. The Plant Genome, 3(2), 69-80.

Glaszmann JC, Kilian B, Upadhyaya HD, Varshney RK. (2010). Accessing genetic diversity for crop improvement. Curr Opin Plant Bio, 13, 167-173.

Gonzalez-Hernandez JL, Sarath G, Stein JM, Owens V, Gedye K, Boe A. (2009). A multiple species approach to biomass production from native herbaceous perennial feedstocks. In Vitro Cellular & Developmental Biology-Plant, 45(3), 267-281.

Graves H, Rayburn AL, Kim S, & Lee DK. (2015). Chloroplast DNA variation within prairie cordgrass (Spartina pectinata Link) populations in the US. J Syst. Evol. doi: 10.1111/jse.12168.

Grime JP (1979) Plant strategies and vegetation processes. Wiley, New York.

Guo J, Thapa S, Voigt T, Rayburn AL, Boe A, Lee DK. (2015). Phenotypic and Biomass Yield Variations in Natural Populations of Prairie Cordgrass (Spartina pectinata Link) in the USA. BioEnergy Research, 1-13.

Harrison RE, Luby JJ, Furnier GR (1997) Chloroplast DNA restriction fragment variation among strawberry (Fragaria spp.) taxa. J Am Soc Hortic Sci 122:63-68.

Hazarika TK, Hazarika BN, Shukla AC (2014) Genetic variability and phylogenetic relationships studies of genus Citrus L. with the application of molecular markers. Genet Resour Crop Ev 61:1441-1454.

Henry RJ. (2005*)*.Importance of plant diversity. In: Henry, R.J. (ed) Plant diversity and evolution: genotypic and phenotypic variation in higher plants. CABI Publishing, Cambridge, MA pp 1-5.

Hirakawa H, Shirasawa K, Kosugi S, Tashiro K, Nakayama S, Yamada, M, Isobe SN et al (2014) Dissection of the octoploid strawberry genome by deep sequencing of the genomes of Fragaria species. DNA Res 21:169-181.

Hitchcock AS (1950) Manual of the grasses of the United States (2nd ed. Rev. A. Chase ed). Dover Publications, Inc., Mineola.

Hultquist S (1996) Chloroplast DNA and nuclear DNA content variations among cultivars of switchgrass, Panicum virgatum L. Crop Sci 36:1049-1052.

Hultquist SJ, Vogel K, Lee D, Arumuganathan K, Kaeppler S (1997) DNA content and chloroplast DNA polymorphisms among switchgrasses from remnant Midwestern prairies. Crop Sci 37:595-598.

Kim S, Rayburn AL, Boe A, Lee DK (2012a) Neopolyploidy in Spartina pectinata Link: 1. morphological analysis of tetraploid and hexaploid plants in a mixed natural population. Plant Syst Evol 298:1073-1083.

Kim S, Rayburn AL, Lee DK (2010) Genome size and chromosome analyses in prairie cordgrass. Crop Sci 50:2277-2282.

Kim S, Rayburn AL, Parrish A, Lee DK (2012b) Cytogeographic distribution and genome size variation in prairie cordgrass (Spartina pectinata bosc ex link). Plant Mol Bio Rep 30:1073-1079.

Kim S, Rayburn AL, Voigt TB, Ainouche ML, Ainouche AK, Lee DK (2013) Chloroplast DNA intraspecific phylogeography of prairie cordgrass (Spartina pectinata bosc ex link). Plant Mol Bio Rep 31:1376-1383.

Kim S, Rayburn AL, Voigt TB, Parrish A, Lee DK (2011) Salinity effects on germination and plant growth of prairie cordgrass and switchgrass. Bioenerg Res 5:225-235.

LGC Genomics (2014). Organisms genotyped using KASP. [online]. Available from http://www.lgcgroup.com/LGCGroup/media/PDFs/Products/Genotyping/Species-tested-with-KASP-technology.pdf?ext=.pdf [Accessed April 28 2015].

Lumaret R, Mir C, Michaud H, Raynal V (2002) Phylogeographical variation of chloroplast DNA in holm oak (Quercus ilex L.). Mol Ecol 11:2327-2336.

Madlung A (2013) Polyploidy and its effect on evolutionary success: old questions revisited with new tools. Heredity 110: 99-104.

Marchant CJ. (1968a) Evolution in *Spartina* (Graminea). II. Chromosomes, basic to relationships and the problem of *S.* x *townsendii* agg. J. Linn. Soc. Bot. 60: 381-409.

Marchant CJ. (1968b) Evolution in *Spartina* (Graminae). III. Species chromosome numbers and their taxonomic significance. J. Linn. Soc. Bot. 60: 411-417.

McMillan C, Weiler J (1959) Cytogeography of Panicum virgatum in central North America. Am J Bot 46:590-593.

Mikkilineni V (1997) Restriction fragment length polymorphism analysis of the Illinois long-term selection chemical strains. Crop Sci. University of Illinois, Urbana. MS Thesis. pp 108-113.

Mohanty A, Martin J, Aguinagalde I (2000) Chloroplast DNA diversity within and among populations of the allotetraploid Prunus spinosa L. Theor Appl Genet 100:1304-1310.

Moncada KM, Ehlke NJ, Muehlbauer GJ, Sheaffer CC, Wyse DL, DeHaan LR. (2007) Genetic variation in three native plant species across the state of Minnesota. Crop Science, 47(6), 2379-2389.

Montemayor MB, Price JS, Rochefort L, Boudreau S (2008) Temporal variations and spatial patterns in saline and waterlogged peat fields: 1. Survival and growth of salt marsh graminoids. Environ Exp Bot 62: 333-342.

Nei M, Kumar S (2000) Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Palmé A (2002) Chloroplast DNA variation, postglacial recolonization and hybridization in hazel, Corylus avellana. Mol Ecol 11:1769-1779.

Paux E, Sourdille P, Mackay I, Feuillet C. (2012). Sequence-based marker development in wheat: advances and applications to breeding. Biotechnology advances, 30(5), 1071-1088.

Peirson JA, Dick CW, Reznicek AA (2013) Phylogeography and polyploid evolution of North American goldenrods (Solidago subsect. Humiles, Asteraceae). J Biogeogr 40:1887-1898.

Petit C, Bretagnolle F, Felber F (1999) Evolutionary consequence of diploid polyploid hybrid zones in wild species. Trends Ecol Evol 14:306-311.

Potter L, Bingham MJ, Baker MG, Long SP (1995) The potential of two perennial C4 grasses and a perennial C4 sedge as lingo-cellulosic fuel crops in N.W. Europe. Crop Establishment and Yields in E. England. Ann Bot 76:513-520.

Rafalski A. (2002). Applications of single nucleotide polymorphisms in crop genetics. Current opinion in plant biology, 5(2), 94-100.

Salem M, Vallejo RL, Leeds TD, Palti Y, Liu S, Sabbagh A, Rexroad CE, Yao J. (2012). RNA-Seq identifies SNP markers for growth traits in rainbow trout. PLoS One, 7(5), e36264.

Saxena RK, Penmetsa RV, Upadhyaya HD, Kumar A, Carrasquilla-Garcia N, Schlueter JA, ... & Varshney RK. (2012). Large-scale development of cost-effective single-nucleotide polymorphism marker assays for genetic mapping in pigeonpea and comparative mapping in legumes. DNA research, 19(6), 449-461.

Semagn K, Babu R, Hearne S, Olsen M. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. Molecular Breeding, 33(1), 1-14. doi:10.1007/s11032-013-9917-x.

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.

Shimono Y, Kurokawa S, Nishida T, Ikeda H, Futagami N (2013) Phylogeography based on intraspecific sequence variation in chloroplast DNA of Miscanthus sinensis (Poaceae), a native pioneer grass in Japan. Botany 91:449-456.

Sindhu A, Ramsay L, Sanderson LA, Stonehouse R, Li R, Condie J, ... & Sharpe AG. (2014). Gene-based SNP discovery and genetic mapping in pea. Theoretical and Applied Genetics, 127(10), 2225-2241.

Soltis DE, Soltis PS. (1989) Allopolyploid speciation in *Tragopogon:* insights from chloroplast DNA. Am J Bot. 76: 1119-1124.

Soltis DE, Soltis PS, Milligan BG (1992) Intraspecific chloroplast DNA variation: systematic and phylogenetic implications. In: Soltis, P.S. (eds) Molecular Systematics of Plants. Chapman & Hall, New York, pp 117-150.

Soltis PS, Burleigh JG, Chanderbali AS, Yoo MJ, Soltis DE (2010) Gene and Genome Duplications in Plants. In: K. Dittmar and D. Liberles (eds) Evolution after Gene Duplication. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Stubbendieck J, Hatch SL, Bryan NM (2011) North American Wildland plants [North American Range Plants] University of Nebraska Press, Lincoln, 2nd ed, pp 109.

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Bio 17:1105-1109.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Bio Evol 30: 2725-2729.

Uitdewilligen JL, Wolters AA, D'hoop BB, Borm TA, Visser RF, van Eck HJ. (2013). A Next-Generation Sequencing Method for Genotyping-by-Sequencing of Highly Heterozygous Autotetraploid Potato. Plos ONE, 8(5), 1-14. doi:10.1371/journal.pone.0062355.

United States Department of Agriculture, Natural Resources Conservation Service (2006) Prairie cordgrass plant guide. http://plants.usda.gov/plantguide/pdf/pg_sppe.pdf. Accessed 10 December 2013.

Van Dijk P, Bakx-Schotman T (1997) Choroplast DNA phylogeography and cytyptype geography in autopolyploid Plantago media. Mol Ecol 6:345-352.

Voight J, Mohlenbrock R (1979) Prairie plants of Illinois Illinois Department of Conservation Division of Forestry, Illinois, pp 230.

Wei L, Miao H, Li C, Duan Y, Niu J, Zhang T, ... & Zhang H. (2014). Development of SNP and InDel markers via de novo transcriptome assembly in Sesamum indicum L. Molecular Breeding, 34(4), 2205-2217.

Wendel J, Doyle J (2005) Polyploidy and evolution in plants. In: Henry, R.J. (eds) Plant Diversity and Evolution Genotypic and Phenotypic Variation in Higher Plants. CABI Publishing, Cambridge, MA, USA, pp 97-118.

Yannic G, Baumel A, Ainouche M (2004) Uniformity of the nuclear and chloroplast genomes of Spartina maritima (Poaceae), a salt-marsh species in decline along the Western European Coast. Heredity 93:182-188.

Young HA, Lanzatella CL, Sarath G, Tobias CM (2011) Chloroplast genome variation in upland and lowland switchgrass. PloS One 6:1-12.

Zilverberg C, Johnson WC, Archer D, Kronberg S, Schumacher T, Boe A, Novotny C (2014) Profitable prairie restoration: The EcoSun Prairie Farm experiment. J Soil Water Conserv, 69:22A-25A.