# Validating DNA Polymorphisms Using KASP Assay in Prairie Cordgrass (*Spartina pectinata* Link) Populations in the U.S.

Hannah Graves[1], A. L. Rayburn[1], Jose L. Gonzalez-Hernandez[2], Gyoungju Nah[3], Do-Soon Kim[3] and D. K. Lee[1]*

[1] Department of Crop Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA, [2] Plant Science Department, South Dakota State University, Brookings, SD, USA, [3] Department of Plant Science, Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul, Korea

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 study provides insight into the genomic structure of prairie cordgrass, but further analysis must be done on other cytotypes to fully understand the structure of this species. This study validates putative SNPs and confirms the potential usefulness of SNP marker technology in future breeding programs of this species.

Keywords: Prairie cordgrass, SNP, marker, *Spartina*, polymorphism, transcriptome

## 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 and 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; Kim et al., 2011; Boe et al., 2013; 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 the reproductive and geographic isolation between the cytotypes, there is likely an increase in polymorphisms and potential genetic diversity, especially within the tetraploid and octoploids cytotypes (Soltis et al., 1992; Wendel and Doyle, 2005; Hirakawa et al., 2014). There is a large amount of phenotypic variation present in all cytotypes of prairie cordgrass (Boe and Lee, 2007; Kim et al., 2012; Guo et al., 2015), but there is a lack of knowledge about the 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) (Semagn et al., 2014).

SNPs 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 functions and phenotypes (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 process was accomplished in prairie cordgrass by using a transcriptome assembly derived from multiple genotypes and tissues (Gonzalez et al., personal communication). The second and third steps are 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 separate study by Gonzalez et al. (personal communication) at South Dakota State University, 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. The assembly was developed using CLC Genomics Workbench 7.0 (Arhaus, Denmark) and annotated against the sorghum genes models. About 146,549 contigs, or transcript assemblies, of 230 bp or more with an N50 of 973 bp were used to mine over 1 million SNPs, insertions, and deletions using the variant detection function in CLC Genomics Workbench. Putative SNPs were filtered based on coverage (minimum of 100 X), a window of 80–100 bp free from additional SNPs and an allele frequency of 20–80%. Initially, nine bi-allelic SNPs were selected for analysis, associated with enzymes within the lignin biosynthesis pathway. Additional SNPs were selected without regard to putative function of the transcript assembly. A total of 121 bi-allelic SNPs were identified for use in this study (**Table 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 50 ng 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

**TABLE 1 | Summary of SNP sequences, including SNP ID, SNP sequences, and SNP alleles.**

| SNP ID | SNP sequence/allele | 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 | GATCGAGAAAAAAAA[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 |
| 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 | AGATTCTGTATTGA[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 |

*(Continued)*

**TABLE 1 | Continued**

| SNP ID | SNP sequence/allele | SNP ID | SNP sequence |
|--------|---------------------|--------|--------------|
| pcg_00084 | AAGTTCGGGGATCAGC**[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.
Bold letters are actual SNPS (SNP alleles).

conditions for all primers, except pcg_1186, was 15 min at 94°C followed by 10 cycles of 94°C for 20 s 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 s 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 s 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 s 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 prairie cordgrass using KASP, seeds and rhizomes of natural populations 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 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, 1997) 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 collection 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 and Kumar, 2000). All positions with <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 involving two individuals (PC17-109 × PC20-102) of two populations differing in morphological characteristics of potential agronomic importance 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 ~1 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. A total of 83 individuals, 70 F$_1$ individuals from PC17-109 (female) × PC20-102 (male) and 13 F$_1$ individuals from PC20-102 (female) × PC17-109 (male) were sampled. F$_1$ seeds were 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**). All of the assay results were recorded as two-letter SNP codes. To determine if the F$_1$ progeny followed segregation of a typical monohybrid cross in relation to SNP genotype, a $\chi^2$ analysis was performed using $P = 0.05$, $df = 2$, and $\chi^2$ critical value = 5.991. The observed, along with the expected genotype, was recorded for each KASP genotyping assay.

## F$_2$ Self

To assess the utility of the KASP marker system in identifying selfed individuals 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 ~1 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 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

**TABLE 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 |

PC17-109 × PC20-102 and 2 $F_1$ of PC20-102 × 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 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 eleven 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 4**).

## Core Collection

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 clade after correcting for the missing data (**Figure 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.

## $F_1$ Analysis

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**). All SNP codes that could not be accurately identified or called, due to not appearing in one of the three genotypes, were removed from the $\chi^2$ analysis. Four individuals did not consistently 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 $\chi^2$ analysis (**Table 3**). Using the resulting trimmed data, the $\chi^2$ analysis indicated normal monohybrid 1:2:1 and 1:1 Mendelian inheritance patterns and could not be rejected for any of the primers (**Table 5**).

## $F_2$ Analysis

The $F_1$ parent genotype was identified in order to find SNPs that indicated the parent was homozygous (**Table 6**). For 3 $F_1$

**TABLE 3 | Primers selected for use on the prairie cordgrass F$_1$ progeny.**

| | 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 | GC | 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 |

*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.*

*[*]Primers that can distinguish true crosses from selfed samples.*

*[†]F$_1$ individuals that are identified as selfs of the PC20_102 parent.*

parents that were selfed, there were F$_2$ progeny that did not fall into the expected homozygous parental genotype (example in **Table 7**). Two F$_2$ progeny were identified consistently as unexpected offspring genotype of 13-F1008, 1 progeny of 14-F1014, and 4 progeny of 14-F1071. Individuals that consistently fell into the heterozygous (unexpected) genotype category across multiple homozygous primers were considered outcrosses and not true selfs of the F$_1$ (**Table 7**). Most of the F$_2$ progeny were identified as expected SNP genotypes when considering the parental genotype.
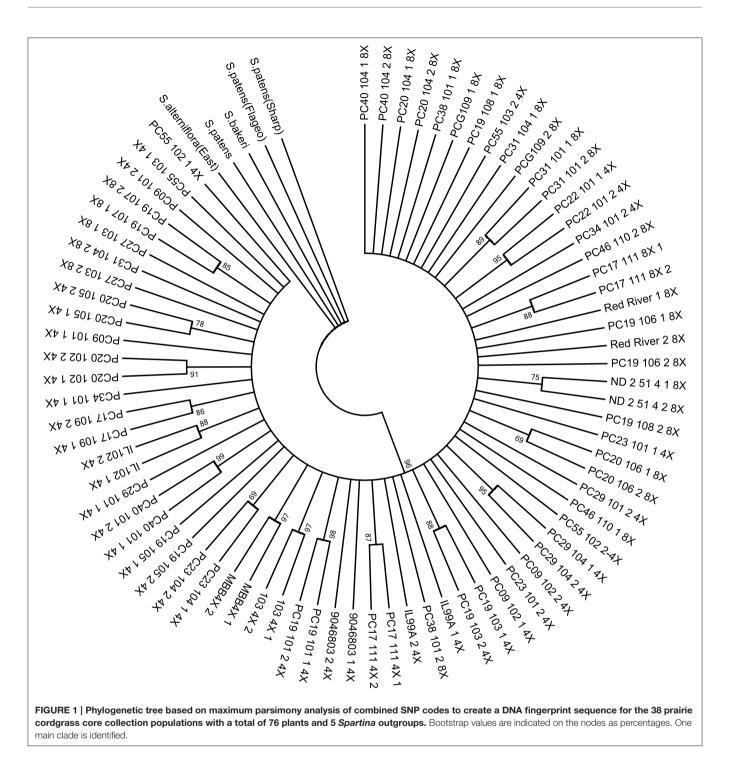
## DISCUSSION

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 core collection. Three monomorphic SNPs were selected for further analysis, to see if the SNPs would be polymorphic in future studies. With the failed and monomorphic assays removed, 84 putative SNPs were validated as true allelic variants and 59 SNPs were selected for 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 clade containing all of

the prairie cordgrass populations. If subclades were observed, the two plants of a single population were represented in the subclade.

Just over half of the populations showed polymorphisms within, with an equal number of octoploid and tetraploid populations. 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 there was little, if any, polymorphisms observed in 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.

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, fewer seeds were obtained when PC20-12 was used as the female, compared with crosses involving PC17-109 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

**TABLE 4 | Sampling of DNA fingerprints created using SNP codes from 59 KASP genotyping assays.**

| 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 | pcg_00049 | pcg_00050 | pcg_00058 | pcg_00059 | pcg_00060 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 | AG | TT | GG | CC | TT |
| 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 | AG | TT | GG | CC | TT |
| PC20_104_1 | GA | - | - | - | - | - | CT | GT | AG | GA | AG | TC | AG | GA | CG | TA | AG | GT | GA | AC | GA | CA | GA | CT | GA | GG | CC | GT | GG | AT |
| PC20_104_2 | GA | - | - | - | - | - | CT | GT | AG | GA | AG | TC | AG | GA | CG | TA | AG | GT | GA | AC | GA | CA | GA | CT | GA | GG | CC | GT | GG | AT |
| 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 | AA | TT | GG | CG | TT |
| PC34_101_2 | GA | AA | GA | CT | - | GA | TT | GT | AG | - | AG | TC | AG | - | CG | TA | - | GT | - | AC | GA | CA | - | - | - | - | CC | GT | - | - |
| PC38_101_1 | GA | CA | - | - | AA | AA | CT | GT | GG | - | GG | CC | GG | GA | GG | TA | AG | TT | AA | CC | AA | - | GA | CT | GA | GG | TT | GT | GG | - |
| PC38_101_2 | GA | - | GA | TT | GA | GA | - | GT | AG | GA | AG | TC | AG | AA | CG | TA | GG | GT | GA | AC | GA | CA | GG | CC | GG | AA | CC | GG | CC | AT |
| Red_River_1 | AA | CA | AA | CT | GA | GA | CT | GT | AG | AA | AG | TC | AG | AA | CG | AA | AG | GT | GA | AC | GA | CA | GA | CT | GA | AG | TT | GT | GG | TT |
| Red_River_2 | AA | AA | - | CT | GA | GA | CC | GG | AA | AA | AA | TT | AA | AA | CC | TA | GG | GT | - | AA | GA | CC | GA | CT | GA | AG | CC | GT | CG | AT |
| PC17_109 | GA | CA | AA | TT | GA | GA | CC | GG | AA | AA | AA | TT | AA | AA | CC | AA | GG | GG | GG | AA | GG | CC | GG | CC | GG | AA | CC | GG | CC | TT |
| PC20_102 | GA | AA | AA | TT | GG | GG | CC | GG | AA | AA | GG | CC | AA | AA | CC | AA | GG | GG | GG | AA | GG | CC | GG | CC | GG | AG | CC | TT | GG | TT |
| S. alterniflora (East) | AA | CC | AA | TT | AA | AA | CC | - | - | GG | GG | CC | AA | AA | CC | TA | GG | GG | GG | AA | GG | CC | GG | CC | - | GG | CC | GG | GG | AA |
| S. bakeri | AA | CC | AA | TT | AA | AA | CC | - | - | GG | GG | CC | AA | AA | CC | AA | GG | GG | GG | AA | GG | CC | GG | CC | - | GG | CC | GG | GG | AA |
| 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 | - | GG | CC | GG | GG | AA |
| 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 | - | GG | CC | GG | GG | AA |
| S. patens | AA | CC | AA | TT | AA | AA | CC | - | - | GG | GG | CC | AA | AA | CC | AA | GG | GG | GG | AA | GG | CC | GG | CC | - | GG | CC | GG | GG | AA |

| Subject ID | 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 | 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 | GG | GG | AC | TT | CT | TA | - | AA | CC | GG | CC | GG | AG | TC | CC | CG | GG | GG | CC | AA | AG | GC | CT | GG | GG | GG | CC |
| PC19_103_2 | GG | GG | GG | - | AC | TT | CT | TA | - | AA | CC | GG | CC | GG | AG | TC | CC | CG | GG | GG | CC | AA | AG | GC | CT | GG | GG | GG | CC |
| PC20_104_1 | TG | GC | - | GC | AC | AT | TT | AA | AC | GA | CC | AG | - | GG | AG | CC | GC | - | GA | GT | CC | AA | GG | GC | CT | TT | GG | GG | CC |
| PC20_104_2 | TG | GC | - | GC | AC | AT | TT | AA | AC | GA | CC | AG | - | GG | AG | CC | GC | - | GA | GT | CC | AA | - | GC | GT | TT | GG | GG | CC |
| PC34_101_1 | TG | GC | AG | GG | AC | TT | CT | TA | AA | - | CT | GG | CC | GG | AA | TC | CC | CG | GG | GG | GC | AA | GG | GG | CT | GG | GG | GG | CC |
| PC34_101_2 | TG | - | AG | GC | - | TT | CT | TA | CC | - | CT | GG | CC | AG | AG | CC | CC | CG | GG | GG | - | AA | - | - | TT | GG | GG | GG | CC |
| PC38_101_1 | GG | GG | GG | GC | AC | TT | CT | TT | AC | GA | TT | AG | AA | AG | AG | TC | GC | CG | GA | - | CC | AA | GG | GC | TT | GG | GG | GG | CC |
| PC38_101_2 | GG | GC | GG | GC | CC | AT | CC | TA | CC | GG | CT | GG | AC | - | AG | CC | CC | CC | GG | GG | GC | AA | AG | GG | CT | GG | GG | GG | CC |
| Red_River_1 | TG | GC | AG | GG | AC | AT | CT | TA | CC | GA | CT | AG | AC | AG | AG | - | - | - | GG | GT | GC | AA | GG | GC | CT | GG | GG | GG | CC |
| Red_River_2 | TG | GC | - | GC | CC | TT | CC | TT | AC | GA | TT | AG | AC | GG | AA | TC | GC | - | GG | GT | GC | AA | AG | - | TT | GG | GG | GG | CC |
| PC17_109 | TG | GC | AG | GG | CC | TT | TT | TA | CC | GG | CT | GG | CC | AG | GG | CC | CC | CC | GA | GG | CC | AA | AG | GG | CC | GG | GG | GG | CC |
| PC20_102 | TG | GC | GG | GC | CC | AT | TT | AA | CC | GG | TT | AG | AA | GG | GG | CC | CC | CG | GG | GT | - | AA | GG | CC | TT | GG | GG | GG | CC |
| S. alterniflora (East) | - | CC | GG | GC | CC | TT | TT | AA | CC | GG | CC | GG | AA | GG | GG | TT | - | CG | GG | TT | CC | AA | GG | CC | | | | | |
| S. bakeri | TT | CC | GG | GC | CC | TT | TT | AA | CC | GG | CC | AG | AA | GG | GG | TC | CC | GG | GG | TT | CC | AA | GG | CC | | | | | |
| S. patens (Sharp) | TT | CC | GG | GC | CC | TT | TT | AA | CC | GG | CC | AG | AA | GG | GG | CC | CC | GG | GG | TT | CC | AA | GG | CC | | | | | |
| S. patens (Flageo) | TT | CC | GG | GC | CC | TT | TT | AA | CC | GG | CC | AG | AA | GG | GG | CC | CC | GG | GG | TT | CC | AA | GG | CC | | | | | |
| S. patens | TT | CC | GG | GC | CC | TT | TT | AA | CC | GG | CC | GG | AA | GG | GG | CC | CC | GG | GG | TT | CC | AA | GG | CC | | | | | |

**FIGURE 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.

consistently being genotyped as heterozygous (outcrossed) rather than being homozygous (selfed) as expected. Ninety-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. There is also a possibility that some of the early-maturing stigmas were exposed to pollen in the greenhouse before bagging. This could explain

why more $F_2$ progeny were identified as unexpected genotypes (outcrosses) than the expected genotype (selfs) of the $F_1$ progeny.

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,

TABLE 5 | Summary of $\chi^2$ analysis on $F_1$ progeny of a specific tetraploid prairie cordgrass cross with selfed data removed.

| Primer ID | Observed X Allele | Expected X Allele | Observed Y Allele | Expected Y Allele | Observed XY Allele | Expected XY Allele | $\chi^2$ |
|---|---|---|---|---|---|---|---|
| **1:1 TEST** | | | | | | | |
| pcg_00011 | 0 | 0 | 44 | 39.5 | 35 | 39.5 | 1.025 |
| pcg_00012 | 0 | 0 | 43 | 39 | 35 | 39 | 0.821 |
| pcg_00024 | 43 | 39 | 0 | 0 | 35 | 39 | 0.821 |
| pcg_00050 | 0 | 0 | 0 | 0 | 77 | 77 | 0 |
| pcg_00058 | 0 | 0 | 0 | 0 | 77 | 77 | 0 |
| pcg_00059 | 0 | 0 | 0 | 0 | 78 | 78 | 0 |
| pcg_00106 | 0 | 0 | 0 | 0 | 75 | 75 | 0 |
| pcg_1186 | 0 | 0 | 0 | 0 | 78 | 78 | 0 |
| pcg_14142 | 0 | 0 | 0 | 0 | 79 | 79 | 0 |
| **1:2:1 TEST** | | | | | | | |
| pcg_00061 | 27 | 19.8 | 15 | 19.8 | 37 | 39.5 | 3.962 |
| pcg_00062 | 15 | 19.8 | 27 | 19.8 | 37 | 39.5 | 3.962 |
| pcg_7965 | 17 | 19.8 | 20 | 19.8 | 42 | 39.5 | 0.544 |

*Analysis indicates that all primers produce expected results from a monohybrid Mendelian cross. df = 2, p = 0.05, critical $\chi^2$ = 5.991.*

TABLE 6 | SNP assay results for the $F_1$ progeny used to determine SNP codes that could indicate true selfs in $F_2$ progeny.

| Parent | 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 | 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* |

*Indicates homozygous SNPs.*

1940; Marchant, 1968a,b; Bishop, 2015) and the observation of disomic inheritance using genotyping-by-sequencing (Crawford, 2015) both suggest a disomic inheritance pattern in *S. pectinata*. 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 (Marchant, 1968a,b; Fortune et al., 2007; 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-pollination always exist and, therefore, developing a molecular way to confirm true crosses

from 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. 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. 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 differs from previous findings based on chloroplast DNA, which identified the octoploid cytotype as the

**TABLE 7 | SNP assay results for F$_2$ individuals of two out of the eight selfed F$_1$ samples.**

| | 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 |

*Indicated are samples that can be identified as true selfs and as outcrossed.*
*\*Primers that can distinguish true selfs from outcrossed samples.*
*[†]F$_2$ individuals that are identified as outcrossed samples.*

most variable. However, one must recognize these SNP markers cover a wide range of expressed genomic DNA vs. 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, $\chi^2$ 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. However, further analysis must be done on other cytotypes to completely understand the genome structure of this species and to evaluate genetic diversity. In addition, this study underlines the usefulness of using SNP marker technology in future breeding programs of prairie cordgrass, and opens up the ability for the final step using SNP markers in genotyping germplasm collections or genetic/breeding populations of prairie cordgrass.

## ACKNOWLEDGMENTS

# REFERENCES

Anderson, E. K., Voigt, T. B., Kim, S. M., and Lee, D. K. (2015). Determining effects of sodicity and salinity on switchgrass and prairie cordgrass germination and plant growth. *Ind. Crops Prod.* 64, 79–87. doi: 10.1016/j.indcrop.2014.11.016

Barkworth, M. E., Anderton, L. K., Capels, K. M., Long, S., and Piep, M. B. (eds.). (2007). *Manual of Grasses for North America*. Logan, UT: Utah State University Press.

Bishop, J. (2015). *Chromosome Pairing and its Relationship to Fertility in Cytotypes of Prairie Cordgrass (Spartina pectinata Link)*. MS Thesis, University of Illinois. Available online at: http://hdl.handle.net/2142/88316

Boe, A., and Lee, D. K. (2007). Genetic variation for biomass production in prairie cordgrass and switchgrass. *Crop Sci.* 47, 929–934. doi: 10.2135/cropsci2006.05.0323

Boe, A., Owens, V., Gonzalez–Hernandez, J., Stein, J., Lee, D. K., and Koo, B. C. (2009). Morphology and biomass production of prairie cordgrass on marginal lands. *GCB Bioenerg.* 1, 240–250. doi: 10.1111/j.1757-1707.2009.01018.x

Boe, A., Springer, T., Lee, D. K., Rayburn, A. L., and Gonzalez-Hernandez, J. (2013). "Underutilized grasses," in *Bioenergy Feedstocks: Breeding and Genetics*, eds M. C. Saha, H. S. Bhandari, and J. H. Bouton (Oxford, UK: John Wiley & Sons, Inc.), 173–205.

Bundock, P. C., Cross, M. J., Shapter, F. M., and Henry, R. J. (2006). Robust allele-specific polymerase chain reaction markers developed for single nucleotide polymorphisms in expressed barley sequences. *Theor. Appl. Genet.* 112, 358–365. doi: 10.1007/s00122-005-0137-6

Byers, R. L., Harker, D. B., Yourstone, S. M., Maughan, P. J., and Udall, J. A. (2012). Development and mapping of SNP assays in allotetraploid cotton. *Theor. Appl. Genet.* 124, 1201–1214. doi: 10.1007/s00122-011-1780-8

Church, G. L. (1940). Cytotaxonomic studies in the gramineae *Spartina, Andropogon* and *Panicum. Am. J. Bot.* 27, 263–271. doi: 10.2307/2436892

Cockram, J., Jones, H., Norris, C., and O'Sullivan, D. M. (2012). Evaluation of diagnostic molecular markers for DUS phenotypic assessment in the cereal crop, barley (*Hordeum vulgare* ssp. *vulgare* L.). *Theor. Appl. Genet.* 125, 1735–1749. doi: 10.1007/s00122-012-1950-3

Crawford, J. (2015). *Linkage Mapping in Prairie Cordgrass (Spartina pectinata Link) Using Genotyping-by-Sequencing Markers*. MS thesis, University of Illinois. Available online at: http://hdl.handle.net/2142/88128

Cuenca, J., Aleza, P., Navarro, L., and Ollitrault, P. (2013). Assignment of SNP allelic configuration in polyploids using competitive allelespecific PCR: application to citrus triploid progeny. *Ann. Bot.* 111, 731–742. doi: 10.1093/aob/mct032

Fang, X., Subudhi, P. K., Venuto, B. C., and Harrison, S. A. (2004). Mode of pollination, pollen germination, and seed set in smooth cordgrass (*Spartina alterniflora*, Poaceae). *Int. J. Plant Sci.* 165, 395–401. doi: 10.1086/382810

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.2307/2408678

Fortune, P. M., Schierenbeck, K. A., Ainouche, A. K., Jacquemin, J., Wendel, J. F., and Aïnouche, M. L. (2007). Evolutionary dynamics of Waxy and the origin of hexaploid *Spartina* species (Poaceae). *Mol. Phylogenet. Evol.* 43, 1040–1055. doi: 10.1016/j.ympev.2006.11.018

Gedye, K., Gonzalez-Hernandez, J., Ban, Y., Ge, X., Thimmapuram, J., Sun, F., et al. (2010). Investigation of the transcriptome of prairie cord grass, a new cellulosic biomass crop. *Plant Genome* 3, 69–80. doi: 10.3835/plantgenome2010.06.0012

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

Glaszmann, J. C., Kilian, B., Upadhyaya, H. D., and Varshney, R. K. (2010). Accessing genetic diversity for crop improvement. *Curr. Opin. Plant Bio.* 13, 167–173. doi: 10.1016/j.pbi.2010.01.004

Gonzalez-Hernandez, J. L., Sarath, G., Stein, J. M., Owens, V., Gedye, K., and Boe, A. (2009). A multiple species approach to biomass production from native herbaceous perennial feedstocks. *In Vitro Cell. Dev. Biol.-Plant* 45, 267–281. doi: 10.1007/s11627-009-9215-9

Graves, H., Rayburn, A. L., Kim, S., and Lee, D. K. (2015). Chloroplast DNA variation within prairie cordgrass (*Spartina pectinata*) populations in the U.S. *J. Syst. Evol.* doi: 10.1111/jse.12168. [Epub ahead of print].

Guo, J., Thapa, S., Voigt, T., Rayburn, A. L., Boe, A., and Lee, D. K. (2015). Phenotypic and biomass yield variations in natural populations of prairie cordgrass (*Spartina pectinata* Link) in the USA. *BioEnerg. Res.* 8, 1371–1383. doi: 10.1007/s12155-015-9604-3

Hirakawa, H., Shirasawa, K., Kosugi, S., Tashiro, K., Nakayama, S., Yamada, M., et al. (2014). Dissection of the octoploid strawberry genome by deep sequencing of the genomes of *Fragaria* species. *DNA Res.* 21, 169–181. doi: 10.1093/dnares/dst049

Hitchcock, A. S. (1950). "Spartina Scherb. Cordgrass" in *Manual of the Grasses of the United States, 2nd Edn. Rev*, ed A. Chase (Mineola: Dover Publications, Inc.), 508–514.

Kim, S., Rayburn, A. L., Boe, A., and Lee, D. K. (2012). Neopolyploidy in *Spartina pectinata* Link: 1. morphological analysis of tetraploid and hexaploid plants in a mixed natural population. *Plant Syst. Evol.* 298, 1073–1083. doi: 10.1007/s00606-012-0617-5

Kim, S., Rayburn, A. L., and Lee, D. K. (2010). Genome size and chromosome analyses in prairie cordgrass. *Crop Sci.* 50, 2277–2282. doi: 10.2135/cropsci2010.03.0140

Kim, S., Rayburn, A. L., Voigt, T. B., Ainouche, M. L., Ainouche, A. K., and Lee, D. K. (2013). Chloroplast DNA intraspecific phylogeography of prairie cordgrass (*Spartina pectinata* bosc ex link). *Plant Mol. Bio. Rep.* 31, 1376–1383. doi: 10.1007/s11105-013-0619-7

Kim, S., Rayburn, A. L., Voigt, T. B., Parrish, A., and Lee, D. K. (2011). Salinity effects on germination and plant growth of prairie cordgrass and switchgrass. *Bioenerg. Res.* 5, 225–235. doi: 10.1007/s12155-011-9145-3

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

Marchant, C. J. (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. doi: 10.1111/j.1095-8339.1968.tb00096.x

Marchant, C. J. (1968b). Evolution in *Spartina* (Graminae). III. species chromosome numbers and their taxonomic significance. *J. Linn. Soc. Bot.* 60, 411–417. doi: 10.1111/j.1095-8339.1968.tb00097.x

Mikkilineni, V. (1997). *Restriction Fragment Length Polymorphism Analysis of the Illinois Long-Term Selection Chemical Strains. Crop Sci.* MS thesis, University of Illinois, Urbana, 108–113.

Moncada, K. M., Ehlke, N. J., Muehlbauer, G. J., Sheaffer, C. C., Wyse, D. L., and DeHaan, L. R. (2007). Genetic variation in three native plant species across the state of Minnesota. *Crop Sci.* 47, 2379–2389. doi: 10.2135/cropsci2007.02.0082

Montemayor, M. B., Price, J. S., Rochefort, L., and 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. doi: 10.1016/j.envexpbot.2007.10.004

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

Paux, E., Sourdille, P., Mackay, I., and Feuillet, C. (2012). Sequence-based marker development in wheat: advances and applications to breeding. *Biotechnol. Adv.* 30, 1071–1088. doi: 10.1016/j.biotechadv.2011.09.015

Rafalski, A. (2002). Applications of single nucleotide polymorphisms in crop genetics. *Curr. Opin. Plant Biol.* 5, 94–100. doi: 10.1016/S1369-5266(02)00240-6

Salem, M., Vallejo, R. L., Leeds, T. D., Palti, Y., Liu, S., Sabbagh, A., et al. (2012). RNA-Seq identifies SNP markers for growth traits in rainbow trout. *PLoS ONE* 7:e36264. doi: 10.1371/journal.pone.0036264

Saxena, R. K., Penmetsa, R. V., Upadhyaya, H. D., Kumar, A., Carrasquilla-Garcia, N., Schlueter, J. A., et al. (2012). Large-scale development of cost-effective single-nucleotide polymorphism marker assays for genetic mapping in pigeonpea and comparative mapping in legumes. *DNA Res.* 19, 449–461. doi: 10.1093/dnares/dss025

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

Sindhu, A., Ramsay, L., Sanderson, L. A., Stonehouse, R., Li, R., Condie, J., et al. (2014). Gene-based SNP discovery and genetic mapping in pea. *Theor. App. Genet.* 127, 2225–2241. doi: 10.1007/s00122-014-2375-y

Soltis, D. E., Soltis, P. S., and Milligan, B. G. (1992). "Intraspecific chloroplast DNA variation: systematic and phylogenetic implications," in *Molecular Systematics of Plants,* ed P. S. Soltis (New York, NY: Chapman & Hall), 117–150.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Bio. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Uitdewilligen, J. L., Wolters, A. A., D'Hoop, B. B., Borm, T. A., Visser, R. F., and van Eck, H. J. (2013). A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PLoS ONE* 8:e62355. doi: 10.1371/journal.pone.0062355

Voight, J., and Mohlenbrock, R. (1979). *Prairie Plants of Illinois*. Springfield, IL: Illinois Department of Conservation Division of Forestry.

Wei, L., Miao, H., Li, C., Duan, Y., Niu, J., Zhang, T., et al. (2014). Development of SNP and InDel markers via *de novo* transcriptome assembly in *Sesamum indicum* L. *Mol. Breed.* 34, 2205–2217. doi: 10.1007/s11032-014-0174-4

Wendel, J., and Doyle, J. (2005). "Polyploidy and evolution in plants," in *Plant Diversity and Evolution Genotypic and Phenotypic Variation*

*in Higher Plants*, ed R. J. Henry (Cambridge, MA: CABI Publishing), 97–118.

Zilverberg, C., Johnson, W. C., Archer, D., Kronberg, S., Schumacher, T., Boe, A., et al. (2014). Profitable prairie restoration: the EcoSun Prairie Farm experiment. *J. Soil Water Conserv.* 69, 22A–25A. doi: 10.2489/jswc.69.1.22a