Industrial Crops & Products 108 (2017) 629–635

Contents lists available at ScienceDirect

Industrial Crops & Products

journal homepage: www.elsevier.com/locate/indcrop

Complete chloroplast genome sequencing of vetiver grass (*Chrysopogon zizanioides*) identifies markers that distinguish the non-fertile 'Sunshine' cultivar from other accessions

Brandi A. Sigmon[a], Robert P. Adams[b], Jeffrey P. Mower[a,c,*]

[a] Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, 68583, USA
[b] Biology Department, Baylor University, Waco, TX, 76798, USA
[c] Center for Plant Science Innovation, University of Nebraska-Lincoln, Lincoln, NE, 68588, USA

ARTICLE INFO

Keywords:
Vetiver grass
Chrysopogon zizanioides
Sunshine
Chloroplast markers
Plastid genome

ABSTRACT

Non-fertile accessions of vetiver grass (*Chrysopogon zizanioides* (L.) Roberty) have been utilized in many parts of the world for environmental remediation and erosion control, but fertile plants can become noxious weeds. Due to the global interest in the environmental benefits of sterile vetiver, unique polymorphisms are needed to distinguish non-fertile from fertile plants. In this research, the chloroplast (cp) genomes of three non-fertile and two fertile vetivers were sequenced. The size, structure, and content of the vetiver cp genomes were typical of other grasses within the Andropogoneae. However, the total cp polymorphism rate of 0.02% was 5- to 7-fold lower, suggesting a recent divergence or slower rate of evolution in non-fertile and fertile vetivers relative to other grasses. In total, 28 polymorphisms—including 14 single nucleotide polymorphisms (SNPs), 11 microsatellites, two small indels, and one microinversion—were identified that distinguished the sterile Sunshine from fertile accessions. In a broader survey of Poaceae cp genomes, one of the SNPs was used to develop a cleaved amplified polymorphic sequence (CAPS) assay to uniquely identify Sunshine genomes. In a panel of 26 fertile and sterile vetivers, the four Sunshine and six additional non-fertile vetivers were distinguished from all other accessions. Two other sterile accessions were not detected by the assay, suggesting that sterility has originated more than once in vetiver. These variations in cp genome sequence can be used to assess sterility in other accessions through sequencing or by the newly developed CAPS assay.

1. Introduction

Vetiver grass (*Chrysopogon zizanioides* (L.) Roberty, syn. *Vetiveria zizanioides* (L.) Nash) has fragrant roots from which is extracted an important essential oil called 'Oil of Vetiver' (Adams et al., 2003, 2004; Weyerstahl et al., 1996, 1997, 2000a, b, c). Non-fertile vetiver grasses have been widely used in the tropics and subtropics as a low-cost, proven alternative for environmental management, including for soil and water management, phytoremediation, and atmospheric carbon sequestration (Chiu et al., 2005; Lavania and Lavania, 2009; National Research Council, 1993; Wilde et al., 2005). In particular, vetiver hedges provide an effective living dam against erosion (National Research Council, 1993), and this technique is now in use in more than 100 countries. The use of non-fertile vetiver lines for environmental management is preferred because of the potential invasiveness of fertile vetiver grasses, which are considered noxious weeds (National Research Council, 1993).

Although the origin of the non-fertile vetiver is not known, it most likely originated in the area from India to Vietnam, where its fragrant roots have been used for centuries for mats and perfumes (National Research Council, 1993). The botanical and agronomic literature distinguishes two broad complexes of vetiver: 'North India' and 'South India'. The 'North India' complex comprises wild, fully fertile populations across the Ganges plain from Pakistan to Bangladesh, while the 'South India' complex includes cultivated, non-fertile lines grown from cuttings for their essential oil in South India, Sri Lanka, Indochina, the Malay Archipelago and introduced for perfumery use (as a fixative) to Reunion Island, thence into Haiti, and around the world. General morphological differences between 'North India' and 'South India' vetiver grass include variation in leaf shape, plant and root architecture, plant height, and oil quality (Chakrabarty et al., 2015). Differences in gene expression of several transcription factor families including ERF, MYB, B3, bHLH, bZIP, and WRKY may underlie these morphological and biochemical differences between the vetiver grass morphotypes

* Corresponding author at. Center for Plant Science Innovation, University of Nebraska-Lincoln, Lincoln, NE, 68588, USA.
E-mail address: jpmower@unl.edu (J.P. Mower).

http://dx.doi.org/10.1016/j.indcrop.2017.07.029
Received 28 January 2017; Received in revised form 12 July 2017; Accepted 14 July 2017

(Chakrabarty et al., 2015). Previously, an examination of 121 accessions of pantropical vetiver revealed that 86% appeared to be clones of a single non-fertile type (Adams and Dafforn, 1998), named Sunshine after a collection site in Sunshine, Louisiana, USA. Plants from Haiti and Reunion, used for commercial essential oil production, clustered with the Sunshine group, indicating that these commercial cultivars are also derived from Sunshine (Adams and Dafforn, 1998). Analysis of additional collections of Thailand cultivars revealed that Sunshine and its allied cultivars form the bulk of vegetatively propagated cultivars in the world (Adams et al., 1999).

Despite clear environmental applications, many countries are reluctant to utilize vetiver until it can be certified as non-fertile. Concerns over the real threat of introducing an invasive alien species limits its broader potential use by governments, agencies and programs. Although DNA fingerprinting (RAPDs) has been used to identify fertile and non-fertile vetiver, the technique is beset with problems of band homology (Rieseberg, 1996) and is difficult to reproduce among labs (Penner et al., 1993). Thus, a DNA sequence method is critically needed for certification of non-fertile vetiver. Chloroplast (cp) genomes are ideal candidates for developing such an assay because they contain a variety of both highly conserved and variable regions which have been frequently used to investigate relationships among plants (e.g., Ruhfel et al., 2014; Zhong et al., 2014), as DNA barcodes for plant identification (C.B.O.L. Plant Working Group, 2009; Nock et al., 2011), or for finding polymorphisms between closely related subspecies or ecotypes in plants (Doorduin et al., 2011; Melodelima and Lobreaux, 2013; Tang et al., 2004; Young et al., 2011). Most seed plant cp genomes have a highly conserved structure consisting of a single circular chromosome with a quadripartite structure, which includes a large single copy region (LSC), a small single copy region (SSC), and two copies of inverted repeats (IRb and IRa), which separate the LSC and SSC regions (Jansen and Ruhlman, 2012; Wicke et al., 2011).

Vetiver is a member of the grass family (Poaceae) within the Andropogoneae (Estep et al., 2014; Welker et al., 2016), a tribe that diversified approximately 20 million years ago (Vicentini et al., 2008) and includes important crops such as maize (*Zea mays* L.), sugarcane (*Saccharum officinarum* L.), and sorghum (*Sorghum bicolor* (L.) Moench). To date, many cp genomes have been sequenced and annotated from the Poaceae, revealing a high degree of conservation in gene content and organization (Asano et al., 2004; Bosacchi et al., 2015; Nah et al., 2015; Tang et al., 2004; Young et al., 2011; Welker et al., 2016), but enough polymorphisms exist to distinguish between closely related species, subspecies, and ecotypes. In rice (*Oryza sativa* L.), 72 single nucleotide polymorphisms (SNPs) and 27 insertions/deletions (indels) were discovered that distinguish between *japonica* and *indica* subspecies (Tang et al., 2004). When the cp genome was sequenced in representative individuals from both lowland and highland ecotypes in switchgrass (*Panicum virgatum* L.), distinctive polymorphisms including 116 SNPs and 46 indels were found to distinguish between the two ecotypes, which could be used for interploidal comparisons (Young et al., 2011). Recently in maize, cp genome sequencing was used to identify three polymorphisms that distinguish between cytoplasmic male sterile (CMS) lines and fertile lines (Bosacchi et al., 2015). Since markers have been identified within cp genomes of closely related grass subspecies, ecotypes and maize lines, it is reasonable to assume that polymorphic sites may also be identified in vetiver chloroplasts in order to distinguish between fertile and non-fertile accessions.

In this study are reported the complete and annotated cp genome sequences of five vetiver individuals, including three non-fertile cultivars ('Sunshine', 'Capitol', and 'Huffman') and two fertile accessions collected from distant sites (Punjab and Allahabad) in northern India. Because these non-fertile cultivars were initially derived from Sunshine and propagated vegetatively thereafter, it was hypothesized that there will be one or more cp-based markers that associate with the Sunshine-derived sterility phenotype. The five cp genomes were aligned and examined to determine the number and location of all polymorphisms, including microsatellite repeat polymorphisms, indels, and SNPs. Several of these polymorphisms were then verified experimentally and tested in a wider panel of vetiver accessions. From these results, an assay was developed that may differentiate between Sunshine-type non-fertile accessions from other vetiver grasses. The development of genetic resources distinguishing fertile and non-fertile vetiver should allow for future genetic mapping studies, conservation of germplasm, and analyses of population structure of fertile vetiver.

2. Methods

2.1. Plant materials

Rooted plant materials for *Adams 7749* cv Sunshine, *Adams 8029* cv Huffman and *Adams 8048* cv Capitol were obtained from Dr. Tomas Ayala-Silva (Curator, National Germplasm Repository, Subtropical Horticulture Research Station, Miami, FL) from the original clones donated by RP Adams (Adams and Dafforn, 1998). All three cultivars are non-fertile and have very similar RAPD profiles classified as a Sunshine genotype (Adams and Dafforn, 1998). Seed of two fertile vetivers (*Adams 7735*, PI 196257, Punjab, India, and Adams 7736, PI 213903, Allahabad, Utter Pradesh, India) were obtained from the USDA (Griffin, GA). Plants were grown in the greenhouse from seed (fertile accessions) and slips (non-fertile accessions).

2.2. cpDNA extraction and sequencing

Mature leaf material (155–160 mg, fresh wt.) was taken from each of three non-fertile and two fertile vetiver individuals (Table 1). Total DNA was extracted using the Qiagen DNeasy mini kit by grinding leaves in liquid nitrogen and then sequenced at BGI (Shenzhen, China) using the Illumina HiSeq2500 platform. For each sample, ~6 M pairs of 125 bp reads were generated from an 800 bp library.

2.3. Genome assembly and annotation

To assemble the cp genomes, raw reads generated at BGI were *de novo* assembled using Velvet version 1.1.06 (Zerbino and Birney, 2008). Multiple iterations of Velvet were run using different pairwise combinations of Kmer values (51, 61, 71, 81, 91) and expected coverage values (50, 100, 200, 500, 1000, 2000). For each run, minimum coverage was set to 10% of expected coverage and scaffolding was turned off. Chloroplast contigs were identified in each assembly by using BlastN with the plastid genome from *Sorghum bicolor* (GenBank accession number EF_115542), a close relative of vetiver, as the query sequence. For each genome, contigs from the best three assemblies, which maximized the average length of chloroplast contigs, were aligned manually, and the consensus of these three assemblies was taken as the final sequence. All polymorphisms identified among the five sequenced samples were verified by mapping raw Illumina reads to the genome at the location of each polymorphic site.

Genome sequences were initially annotated using DOGMA (Wyman et al., 2004). Putative start and stop codons, as well as intron and exon boundaries, were then checked manually using BlastN searches against the *Sorghum bicolor* cp genome sequence. A graphical map of the annotated cp genome was generated using OGDRAW (Lohse et al., 2013). Annotated cp genome sequences were deposited in GenBank under accession numbers KY610123 (Sunshine) and KY610124 (PI 196257).

2.4. cpDNA marker identification and verification

Polymorphisms distinguishing fertile from non-fertile vetiver were obtained by examining multiple sequence alignments of the cp genomes. To ensure the accuracy of distinguishing mutations, a subset of polymorphisms was verified experimentally through Sanger sequencing of PCR amplicons. Primers (Table 2) were designed using Primer3

Table 1
Vetiver accessions used in this study.

Adams ID	Accession	Source	RAPD Type[a]	Fertile	CAPS[b]
7720	VET-RGG-PA-B	Panama	PA	N	–
7749[c]	VET-MRL-001	Louisiana, USA	S	N	–
7775	VET-SJC-2	Zomba, Malawi	O	N	–
7951	VET-RN-001	Colombo, Sri Lanka	SL	N	–
8029[c]	VET-MB-01	Florida, USA	S	N	–
8048[c]	VET-LW-0001	Louisiana, USA	S+	N	–
8076	VET-JM-PV1	Puerto Viejo, Costa Rica	CR	N	–
8244	VET-TGML-001	Spain	O	N	+
8245	VET-TGAVC-002	Spain	S-	N	–
8246	VET-TGKN-003	Spain	KR	N	+
8248	VET-TGSVB-005	Spain	O	N	–
8249	VET-IMZ-AGA	Lilongwe, Malawi	O	N	–
7713	VET-PT-1C	Queensland, Australia	GrA	Y	+
7723	VET-BANG-B001	Bangladesh	Ib	Y	+
7724	VET-BANG-B002	Bangladesh	Ib	Y	+
7735[c]	PI 196257	Punjab, India, USDA	Ib	Y	+
7736[c]	PI 213903	Allahabad, India, USDA	Ib	Y	+
7737	PI 271633	India, USDA	Ib	Y	+
7739	PI 538753	India, USDA	Ib	Y	+
7752	VET-K-Dtp-1	Orissa, India	Ib	Y	+
7981	VET-UCL-027	Lucknow, India, CIMAP	Gb	Y	+
7982	VET-UCL-040	Lucknow, India, CIMAP	G + b	Y	+
7987	PI 554617	Utter Pradesh, India, USDA	Gb	Y	+
8400	PI 537061	Yamuma River, India, USDA	–	Y	+
8401	PI 538754	India, USDA	–	Y	+
8402	PI 538755	India, USDA	–	Y	+

[a] Random Amplified Polymorphic DNA.
[b] Cleaved Amplified Polymorphic Sequence.
[c] Accessions chosen for chloroplast genome sequencing.

Table 2
Primers used for experimental verification of chloroplast DNA markers.

Name	Sequence	Position[a]	T_m (°C)
19264F	CCCTTAGGGACCTTTGGCTA	19264	60.4
19905R	TGCGACTGGAAATCCACTTT	19905	60.6
43496F	TCTGTTCCAGAGCCTATCCCTA	43496	60.2
44055R	TCAGAAAATTCGAACGAAGGA	44055	59.8
49261F	ATGCACAAGAAAGGGTCAGG	49261	60.1
49778R	GGCTTCATTTCGATTTTCCA	49778	60.0
107834F	ATCATTTTCTAGCGGCAACG	107834	60.2
108482R	TGCTTGAGCACTGCTTCCTA	108482	59.9

[a] Coordinates are for the position in the Sunshine genome.

version 0.4.0 (Untergasser et al., 2012), and T_m values were calculated using the nearest-neighbor method (SantaLucia, 1998). PCR products were amplified in a total volume of 20 µL containing ~20 ng of genomic DNA, 0.3 µM of each primer, 0.4 mM of dNTPs, and 0.5 units of ExTaq (TaKaRa), a high-fidelity polymerase. The amplification was carried out with an initial denaturation step at 94 °C for 3 min followed by 35 cycles of denaturation for 30 s at 94 °C, primer annealing for 1 min at 54 °C, and then product extension for 1 min at 72 °C. A final extension step at 72 °C for 5 min was then performed under standard PCR conditions. The amplified fragments were sequenced and trace files were assembled and examined using MacVector version 14.5.2 (MacVector, Inc.).

To determine whether fertile or non-fertile genotypes contained the derived polymorphic state, ancestral states were manually deduced by aligning the vetiver cp genomes to other grass cp genome sequences from *Sorghum bicolor* (NC_008602), *Sorghum timorense* (Kunth) Buse (NC_023800), *Saccharum officinarum* (NC_006084), *Miscanthus sinensis* Andersson (NC_028721), *Zea mays* (NC_001666), *Coix lacryma-jobi* L. (NC_013273), *Setaria italica* (L.) P. Beauv. (NC_022850), *Panicum virgatum* (NC_015990), and *Oryza sativa* (NC_027678). The state in vetiver that matched the outgroup grass species was considered ancestral, while the alternative state was considered derived.

2.5. Genotyping of vetiver cpDNA

Smaller quantities (10–100 mg) of dried leaf samples were collected from a wider panel of 26 vetiver grasses (Table 1). Genomic DNA was extracted using DNeasy Plant Mini Kit (QIAGEN) following the manufacturer's instructions. In order to amplify the target region of the cp genome, the 49261F and 49778R primers (Table 2) were used according to the PCR conditions described previously. The PCR products were examined by 1% TAE agarose electrophoresis with ethidium bromide prior to restriction enzyme digestion. The restriction enzyme digest was carried out by combining 10 µL of amplified PCR products with 1 µL of *DraI* restriction enzyme (Thermo Scientific), 2 µL of Tango buffer supplied by the manufacturer, and 18 µL of deionized sterile water. The reaction mixture was mixed and incubated at 37 °C for 1.5 h. The *DraI* enzyme was then deactivated by incubating at 65 °C for 20 min. Digestion results were examined by 2% TAE agarose electrophoresis stained with ethidium bromide.

3. Results

3.1. Chloroplast genome size, content, and organization

The complete cp genome for the three non-fertile vetivers is 139,971 bp in size while the cp genome for the two fertile vetiver accessions is 139,972 bp. The minimal size difference belies a larger number of indels and single nucleotide microsatellites, described in detail below. The cp genomes of all three non-fertile accessions were identical in sequence, and they included an LSC region of 81,954 bp in length and an SSC region of 12,543 bp separated by two inverted repeats (IR) of 22,737 bp (Fig. 1). The two fertile vetiver accessions were also identical to one another, and the lengths of the SSC and IR were identical to the non-fertile accessions, whereas the LSC was one base pair longer. All five genomes contain the same complement of 79 protein-coding genes, four rRNA genes, 29 tRNA genes, and 19 introns.

The overall size and content of the vetiver cp genome is typical of other sampled grasses (Burke et al., 2016). Differences in gene content

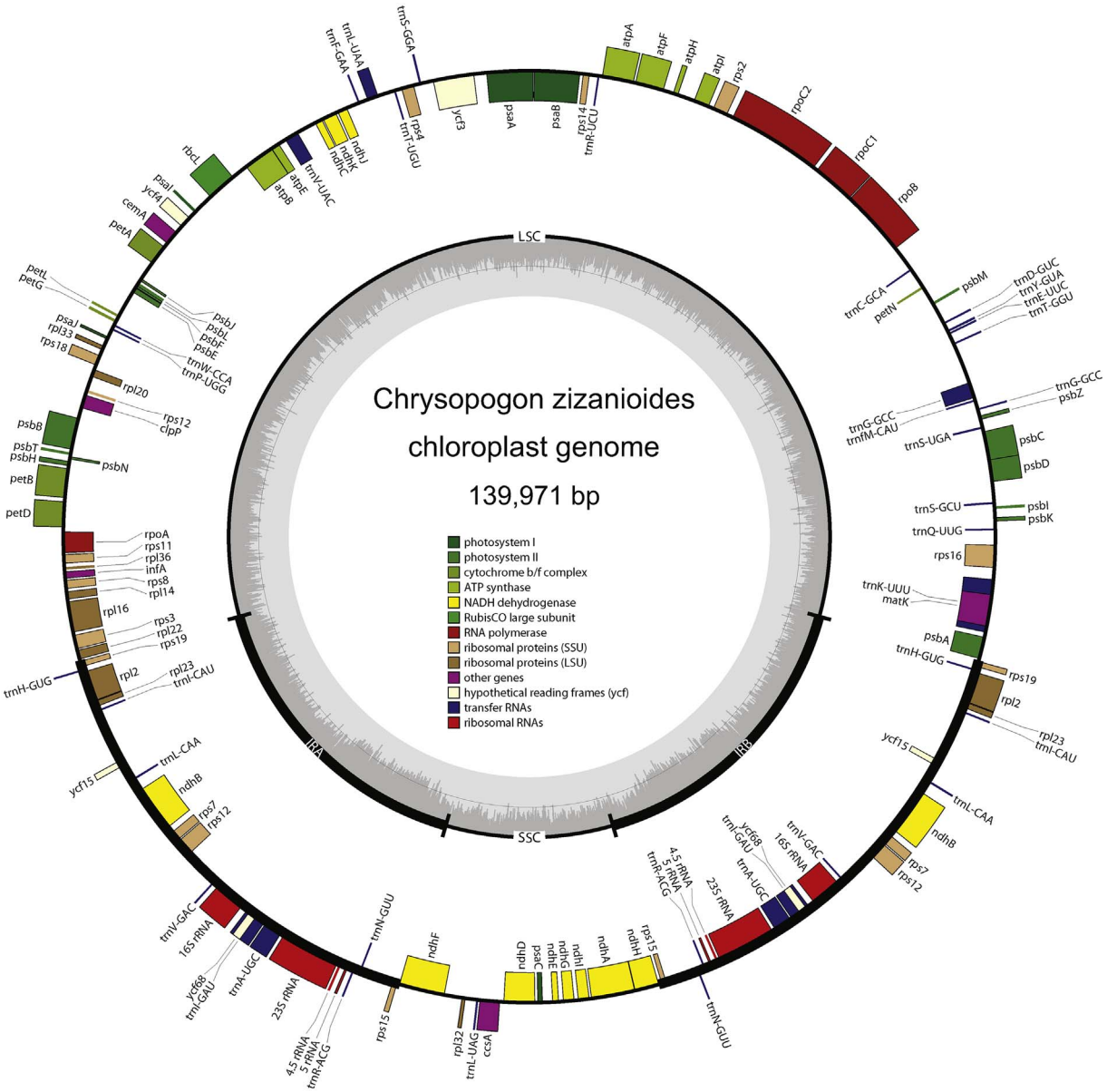

Fig. 1. Chloroplast genome map of *C. zizanioides* var. 'Sunshine.' The inner circle indicates GC content and the locations of the inverted repeats (IRb, and IRa), the large single copy (LSC) and the small single copy (SSC) regions. The outer circle displays the relative positions of the genes. Genes on the outside are transcribed counter-clockwise whereas those inside are transcribed clockwise. Genes are colored-coded by class, according to the legend.

are due mostly to differences in the annotation of the hypothetical genes *ycf15* and *ycf68*. Complete opening reading frames for both hypothetical genes are predicted for vetiver as well as many other closely related grasses (Asano et al., 2004; Bosacchi et al., 2015; Nah et al., 2015; Tang et al., 2004; Young et al., 2011; Welker et al., 2016). In other species, however, either or both genes contain frameshifting mutations in their coding sequences and are not annotated as functional genes.

3.2. Polymorphisms distinguishing fertile and non-fertile vetivers

An alignment of the five completed vetiver cp genomes was examined to identify polymorphisms that distinguish non-fertile from fertile vetiver. A total of 14 SNPs, 11 mononucleotide microsatellite length polymorphisms, two small indels, and one microinversion were identified, all of which were located in intergenic regions (Table 3). Of the 28 polymorphisms, 26 were located in the LSC region, while the microinversion was located in the SSC and one SNP was located in the two copies of the IR (positions 87014 and 134912). Sanger sequencing of PCR amplicons (primers listed in Table 2) was used to examine seven of the polymorphisms, which experimentally confirmed the polymorphic calls from high throughput sequencing. Compared with other studies of polymorphism in the plastid genomes of grasses (Bosacchi et al., 2015; Tang et al., 2004; Young et al., 2011), the vetiver cp genomes exhibit a comparatively low rate of total polymorphism of 0.02%. The rate for SNPs was 0.01% whereas the rate for microsatellites was lower at 0.008% and the indel rate was lowest at 0.001%.

To determine the most reliable polymorphisms for genotyping of Sunshine-type non-fertile lines, derived polymorphisms unique to the Sunshine cp genomes were identified from an alignment of cp genomes from vetiver and other grasses (Table 3). Nine polymorphisms were identified where the non-fertile vetiver contained the derived polymorphism and the genotype of the fertile accession matched other grass species. These markers could be used for positive identification of non-fertile Sunshine cultivars by direct sequencing. Six of the polymorphisms were derived in the two fertile lines and ancestral in the non-fertile

Table 3
Polymorphisms detected between non-fertile and fertile vetiver grasses.

Position[a]	Type[b]	Polymorphism	Genotype NF[c]	Genotype F[d]	Ancestral[e]
4124	SNP	G/A	G	A	G
6494	SNP	G/C	G	C	T
6974	SNR	$(A)_{10}/(A)_{11}$	$(A)_{10}$	$(A)_{11}$	–
8253	SNR	$(T)_{13}/(T)_{12}$	$(T)_{13}$	$(T)_{12}$	–
14911	SNR	$(A)_{11}/(A)_{12}$	$(A)_{11}$	$(A)_{12}$	–
17134	SNP	T/G	T	G	–
19245	SNP	T/A	T	A	T
19374[f]	SNP	T/C	T	C	C
19843	SNP	A/T	A	T	A
33396	SNP	G/A	G	A	G
34236	SNP	C/A	C	A	C
43293	SNP	C/A	C	A	C
43609[f]	SNP	C/T	C	T	T
43935[f]	SNP	A/G	A	G	G
49044	SNR	$(T)_{10}/(T)_{11}$	$(T)_{10}$	$(T)_{11}$	–
49463	Indel	I_{21}/D_{21}	D_{21}	I_{21}	–
49509[f]	SNP	C/A	C	A	A
55753	SNR	$(T)_9/(T)_8$	$(T)_9$	$(T)_8$	–
56223	SNR	$(T)_{14}/(T)_{13}$	$(T)_{14}$	$(T)_{13}$	–
62418[f]	SNR	$(A)_{10}/(A)_9$	$(A)_{10}$	$(A)_9$	$(A)_9$
63970[f]	Indel	I_{22}/D_{22}	I_{22}	D_{22}	D_{22}
64599	SNR	$(T)_{10}/(T)_9$	$(T)_{10}$	$(T)_9$	–
72836	SNR	$(A)_{12}/A_{13}$	$(A)_{12}$	$(A)_{13}$	–
80203	SNR	$(T)_7/(T)_8$	$(T)_7$	$(T)_8$	–
81406	SNR	$(T)_{12}/(T)_{13}$	$(T)_{12}$	$(T)_{13}$	–
87014[f]	SNP	T/G	T	G	G
108137[f]	INV	$(T)_4/(A)_4$	$(T)_4$	$(A)_4$	$(A)_4$
134912[f]	SNP	A/C	A	C	C

[a] Coordinates are for the position in the Sunshine genome.
[b] Types of polymorphisms include single nucleotide polymorphisms (SNPs), insertions/deletions (indels), single nucleotide repeats (SNRs), and inversions (INV).
[c] Genotype of non-fertile accessions (NF).
[d] Genotype of fertile accessions (F).
[e] Ancestral state determined by comparison with other grass chloroplast genomes.
[f] Derived polymorphisms in non-fertile accessions that are most reliable for genotyping.

lines, so they are not reliable for assessing sterility but they could be explored for their potential to genotype fertile vetiver. Some polymorphisms, particularly mononucleotide microsatellites, were too polymorphic to determine the ancestral state and therefore should be considered less reliable for genotyping sterility or fertility.

3.3. Development of a CAPS assay to verify sterility in sunshine-derived lines of vetiver

Of the nine derived polymorphisms for the non-fertile Sunshine accessions, one SNP affected the occurrence of a recognition site for the restriction enzyme *DraI* (Fig. 2a). This SNP therefore provided an opportunity to develop a cleaved amplified polymorphic sequence (CAPS) assay for genotyping Sunshine-type sterility in situations where DNA sequencing is impractical. The *DraI* recognition site is lost from the non-fertile Sunshine and Sunshine-derived accessions and therefore their amplified fragment does not get cleaved, while the enzyme cuts cpDNA from other accessions that retain the recognition site. To examine the applicability of this CAPS assay, a panel of 26 fertile and non-fertile vetiver accessions (Table 1) was collected that exhibit a variety of nuclear genotypes based on previous RAPD analysis (Adams and Dafforn, 1998), including four Sunshine-type non-fertile lines (S, S+, and S-RAPD profiles), eight additional non-fertile lines (CR for Costa Rica, KR for cv Karnataka, PA for Panama, SL for Sri Lanka, and O for other lines, each with a unique RAPD profile), and 14 lines reported as fertile (Ib, Gb, G + b, and GrA RAPD profiles).

This CAPS assay of the *DraI* site was used to test for Sunshine-based sterility in the panel of 26 vetiver accessions (Fig. 2b). As expected, *DraI* did not cut any of the four accessions with Sunshine-like RAPD genotypes, while it did cut all of the 14 fertile accessions. In addition to

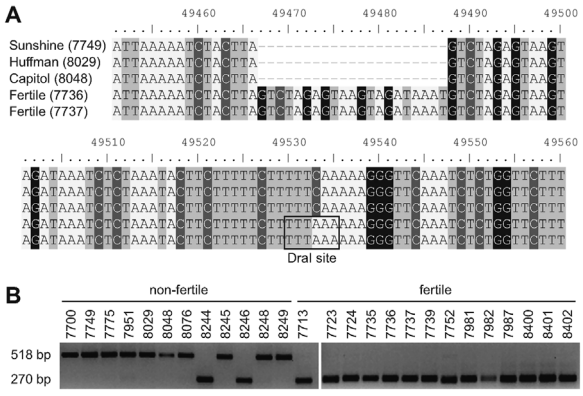

Fig. 2. Genotyping using *DraI* restriction enzyme. Complete cp genomes were aligned in order to find polymorphisms that may distinguish between fertile and non-fertile vetiver grasses (A). A derived C/A SNP located at position 49509 abolishes the TTTAAA recognition site for the restriction enzyme *DraI* in most non-fertile vetiver accessions, which can be used to genotype via the CAPS assay (B). Genotyping results of a panel of 26 vetiver accessions. Digestion with *DraI* results in two fragments, which are 269 bp and 270 bp in length, resulting in a shift in fragment size when analyzed with agarose gel electrophoresis.

Sunshine genotypes, the assay did not cleave six of the eight additional non-fertile lines, indicating that the restriction site is also lost from these lines. Consistent with this CAPS result, principal component analysis of RAPD profiles showed that the nuclear genomes of these six non-fertile accessions were closely allied with Sunshine (Adams et al., 1999). For non-fertile accessions 8244 and 8246, the *DraI* assay cleaved their product, which demonstrates that the restriction site is present. Notably, these two accessions were also reported to be highly divergent relative to other non-fertile vetiver accessions based on RAPD profiling (Adams et al., 1999), suggesting that their sterility is unrelated to Sunshine. Overall, analysis of the CAPS assay on these 26 accessions provides evidence that non-fertile vetiver lines with Sunshine-based sterility can be reliably distinguished from other accessions.

4. Discussion

4.1. Applications of cpDNA markers for genotyping

Despite the importance of vetiver throughout much of the tropics and subtropics for soil erosion control, water management, and the production of essential oil, relatively few genetic resources have been developed. Flow cytometry enabled the broad categorization of vetiver grasses into four groups based on chromosome size and DNA content (Lavania, 1985). Later, DNA fingerprinting (RAPDs) revealed diagnostic banding patterns to help identify accessions (Adams and Dafforn, 1998). However, neither of these methods could reliably distinguish between non-fertile and fertile vetiver grass accessions nor determine the number of independent origins of sterility. The absence of a method to verify sterility has hindered the practice of employing non-fertile vetiver for environmental management in many countries, despite the numerous benefits that vetiver offers. In this study, genetic resources were developed that provide a first step in helping to genotype non-fertile lines of Sunshine-type sterility, which represents the large majority of non-fertile vetiver accessions that are used in practice (Adams and Dafforn, 1998; Adams et al., 1998a, b). These resources may facilitate the adoption of vetiver for soil environmental management purposes in these countries by reducing the threat of inadvertently introducing fertile vetiver as a noxious weed.

Through complete cp genome sequencing, nine derived polymorphisms that are associated with Sunshine-like sterility were identified, and a marker-based genotyping assay based on one of these polymorphisms was developed. The SNP at position 49509 disrupts a recognition site for the restriction enzyme *DraI* in most non-fertile

'South Indian' vetiver grasses including all four Sunshine-based accessions and six additional non-fertile accessions which are likely to be derived from Sunshine, whereas the restriction site is retained in all 14 tested fertile vetiver accessions as well as two divergent non-fertile accessions that are presumably not derived from Sunshine. Thus, this cp marker shows promise as a candidate for uniquely identifying Sunshine-type sterile lines, although more extensive testing is needed. Importantly, it needs to be stressed that proper positive and negative controls must be included when genotyping unknown vetiver grass using the *DraI*-based CAPS assay. From the ancestral state reconstruction (Table 3), it is known that related grasses including maize, sorghum, and sugarcane also contain the *DraI* site, and therefore any of these species may be used as positive cutting controls for the CAPS assay. For a negative cutting control, the widely available non-fertile vetiver Sunshine should be included in genotyping experiments. Finally, until more extensive testing is performed using the assay, it is strongly recommended to perform secondary verification of Sunshine-type sterility by using the developed primers (Table 2) to amplify and directly sequence the additional derived markers in the cp genome.

From a broader perspective, the use of cpDNA markers for genotyping has several advantages over nuclear markers. cpDNA markers are particularly useful for genotyping polyploids since gene copy number and allele frequencies are affected by ploidy, which can complicate genotyping using nuclear markers. For example, cpDNA RFLP was successfully used to genotype upland and lowland switchgrass ecotypes with differences in ploidy level (Hultquist et al., 1996). Since many genera within the Andropogoneae contain species that vary in ploidy level (Estep et al., 2014), cpDNA markers would be ideal for genotyping these grasses. Also, genotyping using cp markers is more efficient and robust than using nuclear markers, which is due to the much higher copy number of cp genomes relative to the nuclear genome following standard DNA extraction (Bendich, 1987; Kuroiwa et al., 1981).

4.2. Evolution of vetiver cp genomes

The vetiver cp genome is not appreciably different from other closely related grasses based on genome size, structure, and content (Asano et al., 2004; Bosacchi et al., 2015; Nah et al., 2015; Tang et al., 2004; Young et al., 2011; Welker et al., 2016). Most apparent differences, such as the number of tRNA or protein-coding genes, are due to differences in annotation. The most notable difference was the lower polymorphism rates of 0.01% for SNPs and 0.001% for indels in the vetiver cp genome in comparison with other grass studies focusing on inter-subspecific or inter-ecotype polymorphisms. For example, the inter-subspecific polymorphism rate is 0.05% for SNPs and 0.02% for indels between rice subspecies *indica* and *japonica* (Tang et al., 2004) and the inter-ecotype polymorphism rates for switchgrass were slightly higher at 0.07% for SNPs and 0.03% for indels (Young et al., 2011).

Based on molecular clock estimates using polymorphisms at synonymous sites in protein-coding coding genes, these switchgrass ecotypes and rice subspecies are believed to have diverged from one other 200,000–845,000 years ago (Young et al., 2011). In vetiver, all of the cp polymorphisms were located in intergenic regions, meaning that there were no synonymous polymorphisms that could be used to estimate vetiver divergence times. Nevertheless, the 5- to 7-fold lower SNP polymorphism rate in vetiver compared with switchgrass ecotypes and rice subspecies suggests that the examined vetiver lines shared a common ancestor more recently. Alternatively, the rate of cp mutation may be lower in vetiver relative to other grasses. Timing the origins of Sunshine and other non-fertile vetiver grass will depend on future genetic studies that assess the natural diversity in the cp and nuclear genomes of a larger number of fertile and non-fertile accessions.

4.3. Multiple origins of sterility

In the panel of 26 vetiver grasses, a total of 12 different non-fertile lines were evaluated using the CAPS assay. From the assay results, all four Sunshine lines (7749, 8029, 8048, 8245) plus six additional non-fertile accessions (7720, 7775, 7951, 8076, 8248, 8249) did not cleave in the assay, consistent with loss of the restriction site. In contrast, two of the non-fertile accessions (8244 and 8246) have retained the *DraI* restriction site, resulting in a cleaved fragment in the CAPS assay. In agreement with the results from the CAPS assay, a previous study using principal component analysis of RAPD profiles (Adams et al., 1998a, 1998b) also placed the six non-fertile lines that lack the cut site as close allies with Sunshine, whereas the two non-fertile lines that contained the cut site were more divergent RAPD genotypes. Together, the RAPD and CAPS assays suggest that the two divergent non-fertile genotypes may have acquired sterility independently of Sunshine and its allies, although further analysis of additional markers in their cp and nuclear genomes are needed to test this hypothesis. Previous work has shown that approximately 86% of non-fertile vetiver accessions are clones, identical to Sunshine based on RAPDs (Adams and Dafforn, 1998). The *DraI*-based CAPS data lends support to these data as most of the other non-fertile lines have genetically similar cpDNA, suggesting that these also originate from southern India.

Given the increased production and superior quality of essential oil, there has been increasing interest in developing tetraploid vetiver from 'North India' fertile accessions (Lavania, 1988). Although these autotetraploid vetiver grasses are sterile, these accessions would genotype as fertile using cpDNA markers. Therefore, these markers should not be used to test recently developed polyploid vetiver of northern Indian and should only be used to verify Sunshine-like vetiver grass. Also, the broad applicability of cpDNA markers and CAPS assays depends upon extensive testing of wild vetiver populations throughout their native ecological range. In this study, 14 wild fertile vetiver grasses were included, and all of these accessions genotyped as fertile based on cpDNA markers. However, this sampling is limited and must be expanded to assess the broader applicability of these cp markers.

Author contributions

RPA conceived the study. BAS, RPA, and JPM designed the experiments. BAS and JPM performed the analyses. BAS, RPA, and JPM analyzed and interpreted the results. BAS, RPA, and JPM wrote the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

The authors gratefully acknowledge Amy K. TeBeest for extracting DNA for cp genome sequencing and Tomas Ayala-Silva and the USDA, Griffin, GA for providing vetiver seeds and rooted plants. This work was supported in part by the National Science Foundation award [grant number MCB 1125386] and by Baylor University award [grant number BU 0324512].

References

Adams, R.P., Dafforn, M.R., 1998. Lessons in diversity: DNA sampling of the pantropical vetiver grass uncovers genetic uniformity in erosion control germplasm. Diversity 13, 27–28.

Adams, R.P., Zhong, M., Srifah, P., Sangduen, N., 1998a. DNA genetic diversity of *Vetiveria zizanioides* (Poaceae). Phytologia 85, 85–95.

Adams, R.P., Zhong, M., Turuspekov, Y., Dafforn, M.R., Veldkamp, J.F., 1998b. DNA fingerprinting reveals clonal nature of *Vetiveriazizanioides* (L.) Nash, Gramineae and sources of potential new germplasm. Mol. Ecol. 7, 813–818.

Adams, R.P., Pandey, R.N., Dafforn, M.R., James, S.A., 2003. Vetiver DNA-fingerprinted cultivars: effects of environment on growth, oil yields and composition. J. Ess Oil Res. 15, 363–371.
Adams, R.P., Habte, M., Park, S., Dafforn, M.R., 2004. Preliminary comparison of vetiver root essential oils from cleansed (bacteria- and fungus-free) versus non-cleansed (normal) vetiver plants. Biochem. Syst. Ecol. 32, 1137–1144.
Asano, T., Tsudzuki, T., Shimadi, H., Kadowaki, K., 2004. Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: a comparative analysis of four monocot chloroplast genomes. DNA Res. 11, 93–99.
Bendich, A.J., 1987. Why do chloroplasts and mitochondria contain so many copies of their genomes? Bioessays 6, 279–282.
Bosacchi, M., Gurdon, C., Maliga, P., 2015. Plastid genotyping reveals uniformity of cms-T maize cytoplasms. Plant Physiol. 169, 2129–2137.
Burke, S.V., Wysocki, W.P., Zuloaga, F.O., Craine, J.M., Pires, J.C., Edger, P.P., Mayfield-Jones, D., Clark, L.G., Kelchner, S.A., Duvall, M.R., 2016. Evolutionary relationships in panicoid grasses based on plastome phylogenomics (Panicoideae; Poaceae). BMC Plant Biol. 16, 140.
C.B.O.L. Plant Working Group, 2009. A DNA barcode for land plants. Proc. Natl. Acad. Sci. U. S. A. 106, 12794–12797.
Chakrabarty, D., Chauhan, P.S., Chauhan, A.S., Indoliya, Y., Lavania, U.C., Nautiyal, C.S., 2015. De novo assembly and characterization of root transcriptome in two distinct morphotypes of vetiver, Chrysopogon zizanioides (L.) Roberty. Sci. Rep. 5 (18630).
Chiu, K.K., Ye, Z.H., Wong, M.H., 2005. Enhanced uptake of As, Zn, and Cu by *Vetiveria zizanioides* and *Zea mays* using chelating agents. Chemosphere 60, 1365–1375.
Doorduin, L., Gravendeel, B., Lammers, Y., Ariyurek, Y., Chin-A-Woeng, T., Vrieling, K., 2011. The complete chloroplast genome of 17 individuals of pest species *Jacobaea vulgaris*: SNPs, microsatellites and barcoding markers for population and phylogenetic studies. DNA Res. 18, 93–105.
Estep, M.C., McKain, M.R., Diaz, D.V., Zhong, J., Hodge, J.G., Hodkinson, T.R., Layton, D.J., Malcomber, S.T., Pasquet, R., Kellogg, E.A., 2014. Allopolyploidy, diversification, and the miocene grassland expansion. Proc. Natl. Acad. Sci. U. S. A. 111, 15149–15154.
Hultquist, S.J., Vogel, K.P., Lee, D.J., Arumuganathan, K., Kaeppler, S., 1996. Chloroplast DNA and nuclear DNA content variations among cultivars of switchgrass, *Panicum virgatum* L. Crop Sci 36, 1049–1052.
Jansen, R.K., Ruhlman, T.A., 2012. Plastid genomes of seed plants. Genomics of Chloroplasts and Mitochondria. Springer, Netherlands, pp. 103–126.
Kuroiwa, T., Suzuki, T., Ogawa, K., Kawano, S., 1981. The chloroplast nucleus: distribution, number, size, and shape, and a model for the multiplication of the chloroplast genome during chloroplast development. Plant Cell Physiol. 22, 381–396.
Lavania, U.C., Lavania, S., 2009. Sequestration of atmospheric carbon into subsoil horizons through deep-rooted grasses −vetiver grass model. Curr. Sci. 97, 618–619.
Lavania, U.C., 1985. Nuclear DNA and karyomorphological studies in vetiver (*Vetiveria zizanioides* L.) Nash. Cytologia 50, 177–185.
Lavania, U.C., 1988. Enhanced productivity of the essential oil in the artificial autopolyploid of vetiver (*Vetiveria zizanioides* L. Nash). Euphytica 52, 271–276.
Lohse, M., Drechsel, O., Kahlau, S., Bock, R., 2013. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res. 41, W575–W581.
Melodelima, C., Lobreaux, S., 2013. Complete *Arabis alpina* chloroplast genome sequence and insight into its polymorphism. Meta. Gene. 1, 65–75.
Nah, G., Im, J., Lim, S.H., Kim, K., Choi, A.Y., Yook, M.J., Kim, S., Kim, C., Kim, D., 2015. Complete chloroplast genomes of two *Miscanthus* species. Mitochondrial DNA Part A 27, 4359–4360.
National Research Council, 1993. Vetiver Grass. A Thin Green Line Against Erosion. National Academy Press, Wash. D.C.
Nock, C.J., Waters, D.L.E., Edwards, M.A., Bowen, S.G., Rice, N., Cordeiro, G.M., Henry, R.J., 2011. Chloroplast genome sequences from total DNA for plant identification. Plant Biotechnol. J. 9, 328–333.
Penner, G.A., Bush, A., Wise, R., Kim, W., Domier, L., Kasha, K., Laroche, A., Scoles, G., Molnar, S.J., Fedak, G., 1993. Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. PCR Methods Appl. 2, 341–345.
Rieseberg, L.H., 1996. Homology among RAPD fragments in interspecific comparisons. Mol. Ecol. 5, 99–105.
Ruhfel, B.R., Gitzendanner, M.A., Soltis, P.S., Soltis, D.E., Burleigh, J.G., 2014. From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evol. Biol. 14, 23.
SantaLucia, J., 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc. Natl. Acad. Sci. U. S. A. 95, 1460–1465.
Tang, J., Xia, H.A., Cao, M., Zhang, X., Zeng, W., et al., 2004. A comparison of rice chloroplast genomes. Plant Physiol. 135, 412–420.
Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G., 2012. Primer3-new capabilities and interfaces. Nucleic Acids Res. 40, e115.
Vicentini, A., Barber, J.C., Aliscioni, S.S., Giussani, L.M., Kellogg, E.A., 2008. The age of the grasses and clusters of origins of C4 photosynthesis. Glob. Chang Biol. 14, 2963–2977.
Welker, C.A.D., Souza-Chies, T.T., Longhi-Wagner, H.M., Peichoto, M.C., McKain, M.R., Kellogg, E.A., 2016. Multilocus phylogeny and phylogenomics of *Eriochrysis* P. Beauv. (Poaceae-Andropogoneae): Taxonomic implications and evidence of interspecific hybridization. Mol. Phylogenet. Evol. 99, 155–167.
Weyerstahl, P., Marschall, H., Splittgerber, U., Wolf, D., 1996. New sesquiterpene ethers from vetiver oil. Liebigs Ann. 1996, 1195–1199.
Weyerstahl, P., Marschall, H., Splittgerber, U., Wolf, D., 1997. New *cis*-eudesm-6-ene derivatives from vetiver oil. Liebigs Ann. 1997, 1783–1787.
Weyerstahl, P., Marschall, H., Splittgerber, U., Wolf, D., 2000a. Analysis of the polar fraction of Haitian vetiver oil. Flav. Fragr. J. 15, 153–173.
Weyerstahl, P., Marschall, H., Splittgerber, U., Wolf, D., 2000b. 1,7-Cyclogermacra-1(10), 4-dien-15-al, a sesquiterpene with a novel skeleton, and other sesquiterpenes from Haitian vetiver oil. Flav. Fragr. J. 15, 61–83.
Weyerstahl, P., Marschall, H., Splittgerber, U., Wolf, D., Surburg, H., 2000c. Constituents of Haitian vetiver oil. Flav. Fragr. J. 15, 395–412.
Wicke, S., Schneeweiss, G.M., de Pamphilis, C.W., Muller, K.F., Quandt, D., 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol. Biol. 76, 273–297.
Wilde, E.W., Brigmon, R.L., Dunn, D.L., 2005. Phytoextraction of lead from firing range soil by Vetiver grass. Chemosphere 61, 1451–1457.
Wyman, S.K., Jansen, R.K., Boore, J.L., 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20, 3252–3255.
Young, H.A., Lanzatella, C.L., Gautam, S., Tobias, C.M., 2011. Chloroplast genome variation in upland and lowland switchgrass. PLoS One 6, e23980.
Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829.
Zhong, B., Xi, Z., Goremykin, V.V., Fong, R., Mclenachan, P.A., Novis, P.M., Davis, C.C., Penny, D., 2014. Streptophyte algae and the origin of land plants revisited using heterogeneous models with three new algal chloroplast genomes. Mol. Biol. Evol. 31, 177–183.