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

Brandi A. Sigmon, Robert P. Adams, Jeffrey P. Mower

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

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 grasses 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.

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.

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 micro-satellites, 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.
markers have been identified in vetiver to distinguish between fertile and non-fertile lines (Bosacchi et al., 2015). Since these non-fertile vetiver were initially derived from Sunshine and its allied cultivars, it is reasonable to assume that 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 (Door dun 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 quadruplicature 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 Ruhmlan, 2012; Wicie 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; Tah 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 interpodular 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, we report 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 fertility 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 essay 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 Dauffen, 1998). All three cultivars are non-fertile and have very similar RAPD profiles classified as a Sunshine genotype (Adams and Dauffen, 1998). Seed of two fertile vetivers (Adams 7735, PI 196257, Punjab, India, and Adams 7736, PI 213903, Allahabad, Uttar 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 EF115542), 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.
of ExTaq (TaKaRa), a high-thermal-stable polymerase for genomic DNA, 0.3 μM of each primer, 0.4 mM of dNTPs, and 0.5 units of Taq polymerase were used 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 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) and 2 μL of T4 DNA ligase 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.

### 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 T4 DNA ligase 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 grass samples (Burke et al., 2016). Differences in gene content...
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 derived in the two fertile lines and ancestral in the non-fertile vetiver. 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 vetiver.
3.3. Development of a CAPS assay to verify sterility in sunshine-derived vetiver grasses.

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 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.

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 (RAPD) revealed diagnostic banding patterns to help identify accessions (Adams and Daicorn, 1998). However, neither of these methods could reliably distinguish between non-fertile and fertile vetiver grass accessions or 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 Daicorn, 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.

4. Discussion

4.1. Applications of cpDNA markers for genotyping

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.

While 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 aboliishes the TTTAAA recognition site for the restriction enzyme DraI in most non-fertile vetiver accessions, which can be used to genotype 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.

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 (RAPD) revealed diagnostic banding patterns to help identify accessions (Adams and Daicorn, 1998). However, neither of these methods could reliably distinguish between non-fertile and fertile vetiver grass accessions or 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 Daicorn, 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 aboliishes the TTTAAA recognition site for the restriction enzyme DraI in most non-fertile vetiver grass accessions, which can be used to genotype 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.

---

**Table 3**

Polymorphisms detected between non-fertile and fertile vetiver grasses.

<table>
<thead>
<tr>
<th>Position</th>
<th>Type</th>
<th>Polymorphism</th>
<th>Genotype NF</th>
<th>Genotype F</th>
<th>Ancestral</th>
</tr>
</thead>
<tbody>
<tr>
<td>4124</td>
<td>SNP</td>
<td>G/A</td>
<td>G</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>6494</td>
<td>SNP</td>
<td>G/C</td>
<td>G</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>6974</td>
<td>SNP</td>
<td>(A)10/(A)11</td>
<td>(A)10</td>
<td>(A)11</td>
<td></td>
</tr>
<tr>
<td>8253</td>
<td>SNP</td>
<td>(T)13/(T)12</td>
<td>(T)13</td>
<td>(T)12</td>
<td></td>
</tr>
<tr>
<td>14911</td>
<td>SNP</td>
<td>(A)10/(A)12</td>
<td>(A)10</td>
<td>(A)12</td>
<td></td>
</tr>
<tr>
<td>17134</td>
<td>SNP</td>
<td>T/G</td>
<td>T</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>19245</td>
<td>SNP</td>
<td>T/A</td>
<td>T</td>
<td>A</td>
<td>T</td>
</tr>
<tr>
<td>19374</td>
<td>SNP</td>
<td>T/C</td>
<td>T</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>19843</td>
<td>SNP</td>
<td>A/T</td>
<td>A</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>33396</td>
<td>SNP</td>
<td>G/A</td>
<td>G</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>34236</td>
<td>SNP</td>
<td>C/A</td>
<td>C</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>43293</td>
<td>SNP</td>
<td>C/A</td>
<td>C</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>43669</td>
<td>SNP</td>
<td>C/T</td>
<td>C</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>43935</td>
<td>SNP</td>
<td>A/G</td>
<td>A</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>49044</td>
<td>SNP</td>
<td>(T)10/(T)11</td>
<td>(T)10</td>
<td>(T)11</td>
<td></td>
</tr>
<tr>
<td>49463</td>
<td>Indel</td>
<td>I21/D21</td>
<td>I21</td>
<td>I21</td>
<td></td>
</tr>
<tr>
<td>49509</td>
<td>SNP</td>
<td>C/A</td>
<td>C</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>55753</td>
<td>SNP</td>
<td>(T)10/(T)8</td>
<td>(T)10</td>
<td>(T)8</td>
<td></td>
</tr>
<tr>
<td>56223</td>
<td>SNP</td>
<td>(T)14/(T)13</td>
<td>(T)14</td>
<td>(T)13</td>
<td></td>
</tr>
<tr>
<td>62418</td>
<td>SNP</td>
<td>(A)10/(A)9</td>
<td>(A)10</td>
<td>(A)9</td>
<td></td>
</tr>
<tr>
<td>63970</td>
<td>Indel</td>
<td>I22/D22</td>
<td>I22</td>
<td>D22</td>
<td>D22</td>
</tr>
<tr>
<td>64599</td>
<td>SNP</td>
<td>(T)10/(T)8</td>
<td>(T)10</td>
<td>(T)8</td>
<td></td>
</tr>
<tr>
<td>72836</td>
<td>SNP</td>
<td>(A)10/(A)11</td>
<td>(A)10</td>
<td>(A)11</td>
<td></td>
</tr>
<tr>
<td>80203</td>
<td>SNP</td>
<td>(T)7/(T)8</td>
<td>(T)7</td>
<td>(T)8</td>
<td></td>
</tr>
<tr>
<td>81406</td>
<td>SNP</td>
<td>(T)13/(T)12</td>
<td>(T)13</td>
<td>(T)12</td>
<td></td>
</tr>
<tr>
<td>87014</td>
<td>SNP</td>
<td>T/G</td>
<td>T</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>108137</td>
<td>INV</td>
<td>(T)14/(A)4</td>
<td>(T)14</td>
<td>(A)4</td>
<td></td>
</tr>
<tr>
<td>134912</td>
<td>SNP</td>
<td>A/C</td>
<td>A</td>
<td>C</td>
<td>C</td>
</tr>
</tbody>
</table>

* Coordinates are for the position in the Sunshine genome.
* Types of polymorphisms include single nucleotide polymorphisms (SNPs), insertions/ deletions (indels), single nucleotide repeats (SNRs), and inversions (INV).
* Genotype of non-fertile accessions (NF).
* Genotype of fertile accessions (F).
* Ancestral state determined by comparison with other grass chloroplast genomes.

---

![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.](image)

genomes of a larger number of fertile and non-fertile accessions.

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 genes, these switchgrass ecotypes and rice subspecies are believed to have diverged from one another 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 autotetraploid 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.

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


