



## Characterization of chloroplast DNA microsatellites from *Saccharum* spp and related species

D.M. Melotto-Passarin, E.V. Tambarussi, K. Dressano, V.F. De Martin and H. Carrer

Departamento de Ciências Biológicas,  
Escola Superior de Agricultura “Luiz de Queiroz”,  
Universidade de São Paulo, Piracicaba, SP, Brasil

Corresponding author: H. Carrer  
E-mail: [hecarrer@esalq.usp.br](mailto:hecarrer@esalq.usp.br)

Genet. Mol. Res. 10 (3): 2024-2033 (2011)  
Received August 17, 2010  
Accepted April 12, 2011  
Published September 12, 2011  
DOI <http://dx.doi.org/10.4238/vol10-3gmr1019>

**ABSTRACT.** Microsatellites, or simple sequence repeats (SSRs), and their flanking regions in chloroplast genomes (plastomes) of some species of the family Poaceae were analyzed *in silico* to look for DNA sequence variations. Comparison of the complete chloroplast DNA sequences (cpDNAs) of sugarcane (*Saccharum hybrid* cv. SP-80-3280 and *S. officinarum* cv. NCo310) and related species, *Agrostis stolonifera*, *Brachypodium distachyon*, *Hordeum vulgare* subsp *vulgare*, *Lolium perenne*, *Oryza nivara*, *O. sativa* subsp *indica*, *O. sativa* subsp *japonica*, *Sorghum bicolor*, *Triticum aestivum*, *Zea mays*, and *Z. mays* cv. B73, allowed us to examine the organization of chloroplast SSRs (cpSSRs) in genic and intergenic regions. We identified 204 cpSSRs in the sugarcane cpDNA; 22.5% were in genic regions. The *ndh*, *rps*, *trn*, and *rpl* gene clusters of the chloroplasts had the most repeats. Mononucleotide repeats were the most abundant cpSSRs in these species; however, di-, tri-, tetra-, penta-, and hexanucleotide repeats were also identified. Many base substitutions and deletions/insertions were identified in the cpSSR loci and their flanking regions. Multiple alignments of all cpSSR sequences of Poaceae species made identification of nucleotide

variability possible; repeat motifs are not uniformly distributed across the Poaceae plastomes, but are mostly confined to intergenic regions. Phylogeny was determined by maximum parsimony and neighbor-joining inference methods. The cpSSRs of these species were found to be polymorphic. It appears that individual cpSSRs in the Poaceae are stable, at least over short periods of evolutionary time. We conclude that the plastome database can be exploited for phylogenetic analysis and biotechnological development.

**Key words:** cpSSR; Plastome; Nucleotide polymorphism; Phylogeny; Molecular markers

## INTRODUCTION

Sugarcane (*Saccharum* spp) is one of the most important tropical and subtropical crops and has a complex polyploid genome of hybrid origin. The complete nucleotide sequence of the chloroplast genome of sugarcane has been already determined: it is a circular double-stranded DNA molecule and it is 141,182 bp in size, containing also a pair of inverted repeat regions of 22,794 bp each (Asano et al., 2004; Calsa et al., 2004).

The chloroplast DNA of most land plants is a circular double-stranded molecule that ranges in size from 110 to 180 kb (Calsa et al., 2004).

The plastome is widely used in plant systematic studies to infer plant phylogenies at different taxonomic levels (Rajendrakumar et al., 2007; Tambarussi et al., 2009), in part because it is slowly evolving and is assumed to be non-recombining (Clegg, 1993).

According to this, it is an important source of genetic markers for phylogenetic analysis, population-level studies, genotyping, and mapping, which can be used for genomic characterization and inter-specific comparison (Raubeson et al., 2007).

The plastomes are densely populated by microsatellites, or simple sequence repeats (SSRs), which consist of 1 to 6 nucleotides repeated in tandem.

Many studies have identified chloroplast SSRs (cpSSRs) within plastomes (Rajendrakumar et al., 2007; Raubeson et al., 2007), and these markers are sufficiently variable for phylogeographic studies within a species (Dick and Heuertz, 2008). Lack of recombination reduces homoplasy, which in turn increases the precision of phylogenetic inference in such studies (Marshall et al., 2001). cpSSRs have become one of the most widely used molecular marker systems in plant genetics and breeding. They are widely used for genetic diversity assessment, variety protection, molecular mapping, and marker assisted selection, providing an efficient tool to link phenotypic and genotypic variation.

Most of the previous studies on microsatellite distribution were based on DNA sequence databases in which coding or gene-rich regions were overrepresented (Rajendrakumar et al., 2007). However, the availability of complete plastome sequences now permits the determination of frequencies of cpSSRs at the whole-plastome level. Such estimates should reflect the basal level of cpSSR dynamics within a species. The present paper details occurrences of cpSSRs in plastomes of Poaceae species that have been completely sequenced, including *Saccharum hybrid* cv. SP-80-3280, *S. officinarum* cv. NCo310, *Agrostis stolonifera* cultivar Penn A-4, *Brachypodium distachyon*, *Hordeum vulgare* subsp *vulgare*, *Lolium perenne*, *Oryza*

*nivara*, *O. sativa* cultivar *indica* isolate 93-11, *O. sativa* cultivar *japonica* isolate PA64S, *Sorghum bicolor*, *Triticum aestivum*, *Zea mays*, and *Z. mays* cv. B73. This has led to efficient and high-throughput *in silico* identification of cpSSR loci.

Yet, very little is known of the nature and organization of cpSSRs in genic and intergenic regions of chloroplast genomes. We, therefore, analyzed cpSSRs in these Poaceae plastomes.

The objectives of this study were to analyze the occurrence, nature, organization, and distribution of cpSSRs in the Poaceae plastomes in both coding and non-coding regions and also to determine the utility and potential of cpSSRs to be integrated into Poaceae phylogenetic analysis.

## MATERIAL AND METHODS

### Search for cpSSRs in chloroplast genome of Poaceae species

All the chloroplast genome sequences of Poaceae species were downloaded in FASTA format from <http://www.ncbi.nlm.nih.gov/sites/entrez?db=genome>. They were used for the generation of cpSSR data. The list of plastome sequences and their lengths are shown in Table 1. The cpSSRs were identified and localized using the FastPCR software (Kalendar, 2008), which identifies mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats. For further analysis, we considered only those repeats wherein the motifs were repeated as follows: mononucleotide repeats with a repeat length  $\geq 8$  nt; dinucleotide with a repeat length  $\geq 5$  nt; tri-, tetra-, penta-, and hexanucleotides with a repeat length  $\geq 3$  nt. Also, we considered  $\leq 9$  interrupting nt, the interrupted microsatellite type. The rationale for choosing a low cut-off value was that cpSSRs are often disrupted by single base substitution (Subramanian et al., 2003). The occurrence of repeats in genic and intergenic regions was identified based on the sequence annotation information available in the GenBank database.

**Table 1.** Plastome sequences and their lengths.

Plant Species	Plastome Size (bp)	Accession Number*
<i>Agrostis stolonifera</i> cultivar Penn A-4	136584	EF115543
<i>Brachypodium distachyon</i> cultivar Bd21	135199	EU325680
<i>Hordeum vulgare</i> subsp. <i>vulgare</i> cultivar Morex	136462	EF115541
<i>Lolium perenne</i>	135282	AM777385
<i>Oryza nivara</i>	134494	AP006728
<i>Oryza sativa</i> cultivar <i>indica</i> isolate 93-11	134496	AY522329
<i>Oryza sativa</i> cultivar <i>japonica</i> isolate PA64S	134551	AY522331
<i>Saccharum hybrid</i> cultivar SP-80-3280	141182	AE009947
<i>Saccharum officinarum</i> L. cultivar NCo310	141182	AP006714
<i>Sorghum bicolor</i> cultivar BTx623	140754	EF115542
<i>Triticum aestivum</i>	134545	AB042240
<i>Zea mays</i>	140384	X86563
<i>Zea mays</i> cultivar B73	140454	AY928077

\*Genbank database (<http://www.ncbi.nlm.nih.gov>).

### Phylogenetic reconstruction and sequence analysis

A distinct phylogenetic analysis was estimated by maximum parsimony and neighbor-joining inference methods with the cpSSR of the following Poaceae species: *S. hybrid* cv. SP-80-3280, *S. officinarum* cv. NCo310, *A. stolonifera* cultivar Penn A-4, *B. distachyon*, *H. vulgare* subsp

*vulgare*, *L. perenne*, *O. nivara*, *O. sativa* cultivar *indica* isolate 93-11, *O. sativa* cultivar *japonica* isolate PA64S, *S. bicolor* cv. BTx623, *Z. mays*, *Z. mays* cv. B73, and *T. aestivum*. The tree was rooted making *Z. mays* species, *S. officinarum* species, and *S. bicolor* cv. BTx623 the outgroup.

The cpSSR sequences of each Poaceae species were concatenated, and all of them were aligned using the multiple alignment algorithm CLUSTAL W (Thompson et al., 1994), with subsequent manual correction following the guidelines of Kelchner (2000). The aligned matrix was imported into the MEGA4 software, version 4.0 (Molecular Evolutionary Genetics Analysis), for phylogenetic analysis (Tamura et al., 2007). The neighbor-joining and maximum parsimony tree was obtained from the resulting matrix using heuristic search options. Searches were made with 1000 replicates of random addition sequence (saving no more than 30 trees per replicate to reduce time spent swapping large islands of trees) with the tree bisection reconnection, branch swapping algorithm, and MulTrees (keeping multiple equally most-parsimonious trees).

Internal support was assessed using 1000 bootstrap replicates (Felsenstein, 1985). Groups with bootstrap percentages of 90-100 were considered to be strongly supported, 80-89 moderately supported and 50-79 weakly supported. Only groups with bootstrap >50 that are consistent with the strict consensus tree are shown.

## RESULTS AND DISCUSSION

### Frequency and distribution of cpSSRs in the Poaceae species database

Using the FastPCR software (Kalendar, 2008), we obtained a detailed analysis of the frequency and distribution of all mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats from the plastomes of 13 Poaceae species. From a total of 1,785,569 bp that represents these plastomes, we identified 2427 cpSSRs, with an average frequency of 1.36 cpSSR per kb, which is higher than that of *Eucalyptus* ESTs (expressed sequence tags) (0.37 SSR per kb) (Ceresini et al., 2005) and *Citrus* ESTs (0.5 SSR per kb) (Palmieri et al., 2007). This was expected since the occurrence of particular microsatellite motifs and repeats (especially the non-trimeric ones) could have implications on how the gene coding region is transcribed due to risks of frameshift mutations (Metzgar et al., 2000).

On average, 3.5% of the chloroplast genomes were found to be composed of cpSSRs. The total number of cpSSR of each species is shown in Table 2. When unit size repeat was analyzed, the mononucleotide type was the most abundant repeat in all 13 plastomes studied, a result that agrees with a study on rice chloroplasts (Rajendrakumar et al., 2007) and another one that screened perfect cpSSRs in *Nuphar advena*, *Ranunculus macranthus*, and 24 other plastomes (Raubeson et al., 2007). Our results differ from a study on the *Arabidopsis* genome in which dinucleotides were found to be most common (Cardle et al., 2000), and from a study of ESTs of barley, maize, oats, rice, and wheat, where the trinucleotides were the most frequent followed by dinucleotides (Varshney et al., 2002). Different from cpSSR, sugarcane (*Saccharum* spp) ESTs have various dinucleotide and trinucleotide motifs (Silva, 2001). Our results suggest that the mononucleotide cpSSRs are the most abundant in chloroplast genomes.

Mononucleotides were the most abundant repeats with frequent poly(A) or (T) (97.1%) followed by a (C)/(G) motif (2.9%). The (A)/(T) motif was found to be more abundant also in genic (19.4-42.3%) and intergenic (51.8-79.5%) regions. These results are in agreement with

**Table 2.** Frequency (%) of the genic and intergenic cpSSRs based on motif size for each species.

	Mono		Di		Tri		Tetra		Penta		Hexa		Total
	Genic	Intergenic	Genic	Intergenic	Genic	Intergenic	Genic	Intergenic	Genic	Intergenic	Genic	Intergenic	
<i>Agrostis stolonifera</i>	11.6 (23)	40.9 (81)	0.5 (1)	6.7 (13)	8.1 (16)	18.2 (36)	1.5 (3)	10.1 (20)	0.5 (1)	1.5 (3)	0 (0)	0.5 (1)	198
<i>Brachypodium distachyon</i>	16.3 (30)	40.8 (75)	0.5 (1)	7.1 (13)	6.5 (12)	15.6 (29)	2.7 (5)	8.7 (16)	0.5 (1)	1.1 (2)	0 (0)	0.0 (0)	184
<i>Hordeum vulgare subsp. vulgare</i>	9.4 (17)	39.2 (71)	0.6 (1)	8.3 (15)	6.1 (11)	14.9 (27)	3.3 (6)	9.4 (17)	0.6 (1)	7.2 (13)	0 (0)	1.1 (2)	181
<i>Lolium perenne</i>	16.1 (30)	39.2 (73)	0.5 (1)	5.9 (11)	10.2 (19)	12.4 (23)	2.1 (4)	10.2 (19)	0.5 (1)	2.7 (5)	0 (0)	0.0 (0)	186
<i>Oryza nivara</i>	12.7 (20)	41.1 (65)	0.6 (1)	4.4 (7)	5.7 (9)	17.7 (28)	3.2 (5)	11.4 (18)	0.6 (1)	1.9 (3)	0 (0)	0.6 (1)	158
<i>Oryza sativa cv indica</i>	24.1 (39)	28.4 (46)	1.9 (3)	5.6 (9)	11.1 (18)	11.7 (19)	5.6 (9)	8.6 (14)	0.6 (1)	1.9 (3)	0 (0)	0.6 (1)	162
<i>Oryza sativa cv japonica</i>	23.7 (37)	29.5 (46)	1.9 (3)	3.2 (5)	13.5 (21)	11.5 (18)	5.1 (8)	9.0 (14)	0.0 (0)	1.9 (3)	0 (0)	0.6 (1)	156
<i>Saccharum hybrid cv SP-80-3280</i>	12.3 (25)	42.6 (87)	1.0 (2)	5.4 (11)	6.4 (13)	15.7 (32)	2.9 (6)	9.3 (19)	0.0 (0)	2.5 (5)	0 (0)	2.0 (4)	204
<i>S. officinarum cv NCo310</i>	12.3 (25)	42.6 (87)	1.0 (2)	5.4 (11)	6.4 (13)	15.7 (32)	2.9 (6)	9.3 (19)	0.0 (0)	2.5 (5)	0 (0)	2.0 (4)	204
<i>Sorghum bicolor cv BTx623</i>	13.2 (25)	45.8 (87)	0.5 (1)	7.4 (14)	3.2 (6)	15.8 (30)	2.1 (4)	7.4 (14)	0.5 (1)	3.2 (6)	0 (0)	1.0 (2)	190
<i>Triticum aestivum</i>	13.2 (25)	40.2 (76)	1.6 (3)	7.4 (14)	8.5 (16)	14.8 (28)	2.1 (4)	7.9 (15)	0.5 (1)	3.2 (6)	0 (0)	0.5 (1)	189
<i>Zea mays</i>	12.9 (27)	44.0 (92)	1.4 (3)	7.6 (16)	5.7 (12)	14.3 (30)	2.4 (5)	6.7 (14)	1.0 (2)	3.3 (7)	0 (0)	0.5 (1)	209
<i>Zea mays cv B73</i>	12.6 (26)	46.1 (95)	1.9 (4)	7.3 (15)	5.3 (11)	14.6 (30)	1.9 (4)	6.3 (13)	0.5 (1)	2.9 (6)	0 (0)	0.5 (1)	206
TOTAL													2427

Numbers within parentheses represent absolute number of microsatellites.

a study of different eukaryotic genomes, which showed that the (A)/(T) motif was more abundant than the (C)/(G) in exons (Tóth et al., 2000). In general, the adenine/thymine-rich repeat motifs are most common in cpSSRs and were also the most abundant in the family Poaceae.

Tri- and tetranucleotides were the next predominant repeats for all species, with frequencies ranging from 15.8% in *S. bicolor* cv. BTx623 to 18.2% in *A. stolonifera* and 14.6% in *Z. mays* cv. B73 to 17.7% in *O. nivara*, respectively (data not shown).

Di-, penta- and hexanucleotide repeats in the genic region were represented in proportions of 3.2 to 8.3%, 1.1 to 7.2% and 0 to 2.0%, respectively (Table 2). The (AT)/(TA) motif was the most common dinucleotide repeat with a frequency of 78.9%, and it was also the predominant repeat in the genic region, similar to liverworts, maize, and pea chloroplasts (Powell et al., 1996).

For tri- and tetranucleotides, the motifs (TAA)/(TTA) reached 23.7%, (GAA)/(TTC) 22.8%, (CTTT)/(AAAG) 25.6%, and (TTTA)/(TAAA) 19.6%, respectively, being the most frequent repeats (data not shown).

The remaining penta- and hexanucleotide repeats were represented by (TTTTTC)/(GAAAA) (23.1%) and (TATTA)/(TAATA) (12.8%), (GGGGAA)/(TTCCCC) (52.3%), respectively (data not shown). Interestingly, single pentanucleotide repeats were observed in *S. bicolor* cv. BTx623 [intergenic: (TAACA)/(TGTTA)], *Z. mays* [genic: (GGAAA)/(TTTTCC); intergenic: (TTCCA)/(TGGA)], *T. aestivum* [genic: (ATTAT)/(ATAAT)], *L. perenne* [intergenic: (TAGTA)/(TACTA) and (CCTTT)/(AAAGG)]. The single hexanucleotide repeats were found in *H. vulgare* subsp *vulgare* [intergenic: (TAAAAA)/(TTTTTA)] and *A. stolonifera* [intergenic: (TTTCTA)/(TAGAAA)].

In general, intergenic cpSSR in the family Poaceae were more abundant than genic cpSSR (Table 2). This also occurs in chloroplast genomes of the angiosperm families Asteraceae (Timme et al., 2007), Fabaceae (Saski et al., 2005), and Solanaceae (Daniell et al., 2006), probably due to an associated lower polymorphism of coding regions in contrast to non-coding ones. In each of these families, most of the repeats are 20-40 bp in length and they are located mostly in intergenic spacer regions and introns, although several are located in the protein-coding genes *psaA*, *psaB*, and *ycf2* (Raubeson et al., 2007). The maximum numbers of cpSSRs in the family Poaceae were found in different genes: *ndh* (NADH dehydrogenase), *rps* (ribosomal proteins), *trn* (tRNA), and *rpl* (ribosomal proteins) gene clusters.

The frequency of plastome genes with cpSSR reached a maximum of 61.5% in *O. sativa* cv. *indica*. However, this species has the smallest number of plastome genes compared with other Poaceae species (Table 3). When reviewing the largest absolute number of genes with cpSSR, we found that *O. sativa* cv. *japonica* has 45, and it is the second species that has the highest frequency of genes with cpSSR (28.3%) (Table 3).

**Table 3.** Frequency (%) of plastome genes with cpSSRs.

Poaceae Species	N° Plastome Genes	Genes with cpSSR	%Genes with cpSSR
<i>Agrostis stolonifera</i>	141	29	20.6
<i>Brachypodium distachyon</i>	133	31	23.3
<i>Hordeum vulgare</i> subsp. <i>vulgare</i>	141	25	17.7
<i>Lolium perenne</i>	130	35	26.9
<i>O. nivara</i>	165	24	14.5
<i>O. sativa</i> cv <i>indica</i>	65	40	61.5
<i>Oryza sativa</i> cv <i>japonica</i>	159	45	28.3
<i>Saccharum hybrid</i> cv <i>SP-30-3280</i>	144	26	18.1
<i>S. officinarum</i> cv <i>NCo310</i>	144	26	18.1
<i>Sorghum bicolor</i> cv <i>BTx623</i>	140	24	17.1
<i>Triticum aestivum</i>	137	34	24.8
<i>Zea mays</i>	158	30	19.0
<i>Z. mays</i> cv <i>B73</i>	158	29	18.4

The high percentage of cpSSRs with a period divisible by three is expected because of the nature of translation and how it relies on triplet codons (Table 2). It also corresponds with previous research that has shown that tri- and hexanucleotide repeats are the most common in the coding regions of eukaryotes (Goulding et al., 1996).

Of course, any particular threshold is rather arbitrary, and no consensus has been reached on which nucleotide length or repeat unit number is significant (Ellegren, 2004). However, it has been suggested that cpSSRs of length 8 nt or more (regardless of repeat motif) are prone to slip-strand mispairing (SSM, thought to be the primary mutational mechanism to affect SSRs), whereas those of lesser length are not (Rose and Falush, 1998). Elsewhere, the critical threshold is estimated at 7-10 bp (Dechering et al., 1998).

It is thought that SSRs begin as random runs of nucleotides (Levinson and Gutman, 1987; Ellegren, 2004). Any bias in mutation patterns or nucleotide composition would make certain runs more likely. Once present in a location, the repeat would then grow via SSM (Levinson and Gutman, 1987; Ellegren, 2004; Lovett, 2004). Longer SSRs lead to more stable heteroduplex intermediates, making SSM more likely (Lovett, 2004). However, longer SSRs also have higher mutation rates (Ellegren, 2004). One model of SSR evolution posits that the distribution of repeat lengths in a genome represents equilibrium between SSM and point mutations (Ellegren, 2004). In the comparison of Poaceae species, SSM and point mutations occur at about the same frequency, consistent with this hypothesis.

### Phylogenetic analysis

In a landmark article that included data from multiple sources, the Grass Phylogeny Working Group (2001) examined relationships among grasses using a large and diverse as-

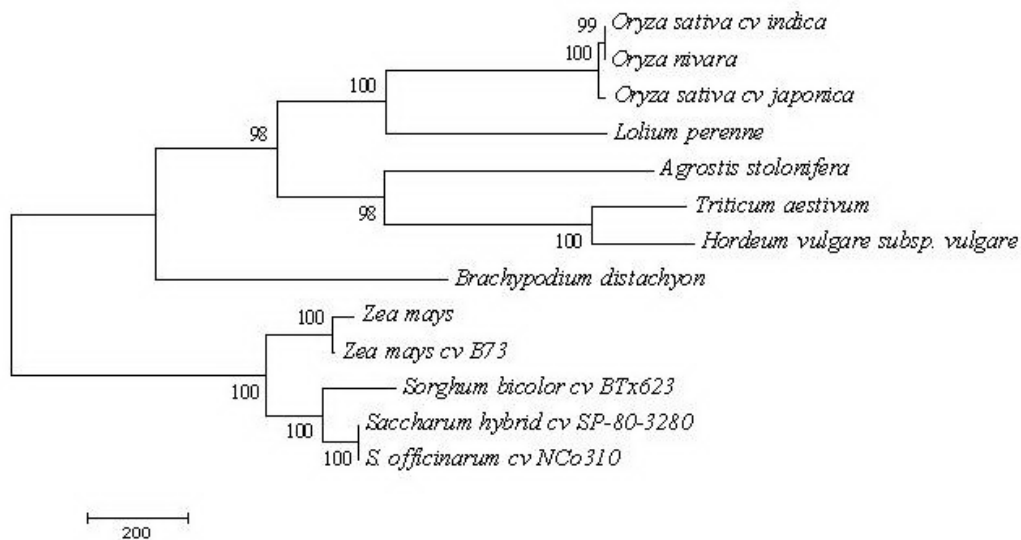
semblage of species. This study highlighted the existence of two major lineages, the BEP clade and the PACCAD clade, which together encompass the majority of grasses, each representing two major radiations ~40-50 million years ago (Bremer, 2002).

The BEP clade includes the subfamilies Bambusoideae, Ehrhartoideae, and Pooideae. *Oryza* species belongs to the subfamily Ehrhartoideae, while *T. aestivum*, *H. vulgare*, *A. stolonifera*, *L. perenne*, and *B. distachyon* are in the Pooideae.

The PACCAD clade includes several subfamilies, including the Panicoideae, a large group of mainly tropical and subtropical species, some of which are important crops worldwide, such as *Z. mays*, *S. officinarum*, and *S. bicolor*. So far, all phylogeny reconstructions of the Poaceae have used selected genes or partial regions as data. However, with cpSSRs of 13 sequenced chloroplast genomes in this family and the computer power to align them, it is possible for the first time to perform cpSSR of whole chloroplast genome phylogenetic analyses.

To examine if the cpSSR phylogenetic analysis is consistent with those based on selected genes, we employed maximum parsimony (Nei and Kumar, 2000) and neighbor-joining methods to reconstruct a Poaceae phylogeny using cpSSR of whole chloroplast sequences.

The phylogenetic analysis of 13 species of Poaceae was estimated based on the multialignment of concatenated cpSSRs. The tree was rooted making *Z. mays* species, *S. officinarum* species, and *S. bicolor* cv. BTx623 the outgroup. Of the 6175 characters from the aligned cpSSR matrix, 1511 are constant. Among the 4664 variable characters, 3454 are parsimony-informative and 1210 are parsimony-uninformative. The analysis of these data set identified one parsimonious tree. Figure 1 shows the parsimony strict consensus tree (phylogram), which produced the same topology with maximum node support from the neighbor-joining tree. The topology of the tree is in agreement with the results obtained from a larger group of species (Grass Phylogeny Working Group, 2001).



**Figure 1.** The maximum parsimonious rooted tree obtained from cpSSRs of Poaceae plastome species. Outgroup is composed by *Zea mays* species, *Saccharum officinarum* species, and *Sorghum bicolor* cv. BTx623. Numbers indicate bootstrap percentage >50%.

The ingroup is divided into two main clades: the PACCAD clade composed of *S. officinarum* species, *S. bicolor* cv. BTx623, and *Z. mays* species, and the BEP clade composed of the *Oryza* species, *L. perenne*, *T. aestivum*, *H. vulgare* subsp. *vulgare*, *A. stolonifera* cv. Penn A-4, and *B. distachyon* (Figure 1). All clades had robust bootstrap support (95-100% bootstrap support; Figure 1).

The phylogram also shows that branches in the BEP clade are much longer than those in the PACCAD clade. A similar result was found in sequence comparison and phylogenetic analysis of eight grass plastomes (Bortiri et al., 2008), and in a phylogenetic study using 61 protein-coding genes, indicating that the rates of evolution are higher in the BEP clade compared to the PACCAD species sampled here (Saski et al., 2005). However, it is possible that these slower rates do not extend to other species of the PACCAD clade, since *Z. mays*, *S. bicolor*, and *S. officinarum* are closely related, with all three belonging to the subfamily Panicoideae.

In terms of the phylogenetic utility of cpSSR variation, comparison of the Poaceae species suggests that individual cpSSRs are stable at least over relatively short periods of evolutionary time, and that they commonly vary in repeat number. This result agrees with a study that screened perfect cpSSRs in *Nuphar advena* and *Ranunculus macranthus*, and 24 other plastomes (Raubeson et al., 2007).

The availability of genome sequences in the public domain has removed technical and economic limitations, and dramatically accelerated the process of the development of SSR markers. Phylogenetic analysis based on mononucleotide repeats and flanking nucleotide sequences from the organellar genomes is possible based on *in silico* analysis (Nishikawa et al., 2005).

SSR marker technology has proved to be a dependable, rapid, and inexpensive tool for plant genotyping (Wu and Huang, 2007). SSR markers are becoming the markers of choice for plant genome analysis because of their high polymorphism and informative nature.

## CONCLUSION

In conclusion, we demonstrated a generic approach for assessing genetic variation in the Poaceae. Our study helped to identify genes possessing cpSSRs in plastomes. The repeat motifs are not uniformly distributed across the Poaceae plastomes but are mostly confined to intergenic regions. Hence, the occurrence of cpSSRs is a non-random event. A few of the genic cpSSRs have been found to be significantly different among Poaceae plastomes, which may be helpful in specific PCR markers. The high sequence conservation of the plastomes will make it easier to design primers that may work even in relatively distant species. The cpSSR markers developed in the present study could also be useful in determining the maternal origin of Poaceae species and in phylogenetic studies. CpSSR markers could be included in the analysis of multiple chloroplast regions to improve the resolution of phylogenetic studies of the species studied.

## ACKNOWLEDGMENTS

The authors thank Sanjna Shah for English review, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for financial support provided to E.V. Tambarussi (#07/05795-0), and CAPES for financial support provided to D.M. Melotto-Passarim.



## REFERENCES

- Asano T, Tsudzuki T, Takahashi S, Shimada H, et al. (2004). Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: a comparative analysis of four monocot chloroplast genomes. *DNA Res.* 11: 93-99.
- Bortiri E, Coleman-Derr D, Lazo GR, Anderson OD, et al. (2008). The complete chloroplast genome sequence of *Brachypodium distachyon*: sequence comparison and phylogenetic analysis of eight grass plastomes. *BMC Res. Notes* 1: 61.
- Bremer K (2002). Gondwanan evolution of the grass alliance of families (Poales). *Evolution* 56: 1374-1387.
- Calsa JT, Carraro DM, Benatti MR, Barbosa AC, et al. (2004). Structural features and transcript-editing analysis of sugarcane (*Saccharum officinarum* L.) chloroplast genome. *Curr. Genet.* 46: 366-373.
- Cardle L, Ramsay L, Milbourne D, Macaulay M, et al. (2000). Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics* 156: 847-854.
- Ceresini PC, Silva CLSP, Missio RF, Souza EC, et al. (2005). Satellypus: analysis and database of microsatellites from ESTs of *Eucalyptus*. *Genet. Mol. Biol.* 28: 589-600.
- Clegg MT (1993). Chloroplast gene sequences and the study of plant evolution. *Proc. Natl. Acad. Sci. U. S. A.* 90: 363-367.
- Daniell H, Lee SB, Grevich J, Saski C, et al. (2006). Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analyses with other Solanaceae genomes. *Theor. Appl. Genet.* 112: 1503-1518.
- Dechering KJ, Cuelenaere K, Konings RN and Leunissen JA (1998). Distinct frequency-distributions of homopolymeric DNA tracts in different genomes. *Nucleic Acids Res.* 26: 4056-4062.
- Dick CW and Heuertz M (2008). The complex biogeographic history of a widespread tropical tree species. *Evolution* 62: 2760-2774.
- Ellegren H (2004). Microsatellites: simple sequences with complex evolution. *Nat. Rev. Genet.* 5: 435-445.
- Felsenstein J (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783.
- Goulding SE, Olmstead RG, Morden CW and Wolfe KH (1996). Ebb and flow of the chloroplast inverted repeat. *Mol. Gen. Genet.* 252: 195-206.
- Grass Phylogeny Working Group (2001). Phylogeny and subfamilial classification of the grasses (Poaceae). *Ann. Miss. Bot. Gard.* 88: 373-457.
- Kalendar R (2008). Software FastPCR: a PCR primer and Probe Design and Repeat Sequence Searching Software with Additional Tools for the Manipulation and Analysis of DNA and Protein. Available at [http://www.biocenter.helsinki.fi/bi/programs/fastpcr.htm.]. Accessed February 2, 2010.
- Kelchner SA (2000). The evolution of non-coding chloroplast DNA and its application in plant systematics. *Ann. Miss. Bot. Gard.* 87: 482-498.
- Levinson G and Gutman GA (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* 4: 203-221.
- Lovett ST (2004). Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. *Mol. Microbiol.* 52: 1243-1253.
- Marshall HD, Newton C and Ritland K (2001). Sequence-repeat polymorphisms exhibit the signature of recombination in lodgepole pine chloroplast DNA. *Mol. Biol. Evol.* 18: 2136-2138.
- Metzgar D, Bytof J and Wills C (2000). Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Res.* 10: 72-80.
- Moxon ER and Wills C (1999). DNA microsatellites: agents of evolution? *Sci. Am.* 280: 94-99.
- Nei M and Kumar Ed S (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
- Nishikawa T, Vaughan DA and Kadowaki K (2005). Phylogenetic analysis of *Oryza* species, based on simple sequence repeats and their flanking nucleotide sequences from the mitochondrial and chloroplast genomes. *Theor. Appl. Genet.* 110: 696-705.
- Palmieri DA, Novelli VM, Bastianel M, Cristofani-Yaly M, et al. (2007). Frequency and distribution of microsatellites from ESTs of *Citrus*. *Genet. Mol. Biol.* 30: 1009-1018.
- Powell W, Machray GC and Provan J (1996). Polymorphism revealed by simple sequence repeats. *Trends Plant Sci.* 1: 215-222.
- Rajendrakumar P, Biswal AK, Balachandran SM, Srinivasarao K, et al. (2007). Simple sequence repeats in organellar genomes of rice: frequency and distribution in genic and intergenic regions. *Bioinformatics* 23: 1-4.
- Raubeson LA, Peery R, Chumley TW, Dziubek C, et al. (2007). Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics* 8: 174.
- Rose O and Falush D (1998). A threshold size for microsatellite expansion. *Mol. Biol. Evol.* 15: 613-615.
- Saski C, Lee SB, Daniell H, Wood TC, et al. (2005). Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes. *Plant Mol. Biol.* 59: 309-322.
- Silva JAF (2001). Preliminary analysis of microsatellite markers derived from sugarcane expressed sequence tags (ESTs).

- Genet. Mol. Biol.* 24: 155-159.
- Subramanian S, Mishra RK and Singh L (2003). Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. *Genome Biol.* 4: R13.
- Tambarussi EV, Melotto-Passarin DM, Barbosa AL, Brigati JB, et al. (2009). *In silico* analysis of simple sequence repeats from chloroplast genomes of Solanaceae species. *Crop Breed. Appl. Biotech.* 9: 344-352.
- Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596-1599.
- Thompson JD, Higgins DG and Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680.
- Timme R, Kuehl EJ, Boore JL and Jansen RK (2007). A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: identification of divergent regions and categorization of shared repeats. *Am. J. Bot.* 94: 302-312.
- Tóth G, Gaspari Z and Jurka J (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* 10: 967-981.
- Varshney RK, Thiel T, Stein N, Langridge P, et al. (2002). *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cell Mol. Biol. Lett.* 7: 537-546.
- Wu YQ and Huang Y (2007). An SSR genetic map of *Sorghum bicolor* (L.) Moench and its comparison to a published genetic map. *Genome* 50: 84-89.