

# Development and characterization of expressed sequence tag-derived simple sequence repeat markers in tropical forage legume *Stylosanthes guianensis* (Aubl.) Sw.

Xipeng Ding · Qinglin Jia · Xiaoyan Luo ·  
Long Zhang · Hanqing Cong · Guodao Liu ·  
Changjun Bai

Received: 31 January 2015 / Accepted: 7 August 2015 / Published online: 15 October 2015  
© Springer Science+Business Media Dordrecht 2015

**Abstract** Most species of *Stylosanthes* (stylo) genus are important tropical pasture legumes with very versatile, widely adapted, and productive characteristics. These legumes are commercially used in various agricultural systems in many tropical and subtropical regions. However, the few molecular markers for the stylo species limit their genetic improvement. In this study, 36,558 expressed sequence tags (ESTs) have been *de novo* assembled using Illumina paired-end sequencing in *Stylosanthes guianensis* (Aubl.) Sw. to develop simple sequence repeat (SSR) markers. We searched these ESTs for SSRs and identified 4115 SSR loci from 3643 ESTs (9.96 %). Dinucleotide and trinucleotide repeat motifs were the most abundant types (30.50 and 50.33 %, respectively), whereas tetranucleotide, pentanucleotide, and hexanucleotide motifs represented <10 % of all SSRs. The motif AG/CT was the most abundant, accounting for 21.7 % of all SSRs. Moreover, 2008 SSR markers were

developed using 1873 SSR-containing unigenes. A total of 115 EST-SSR markers located in the coding region were amplified using polymerase chain reaction to detect 29 *S. guianensis* accessions. Of these 115 markers, 96 produced reliable bands with expected sizes, and 81 markers were polymorphic, with 2–6 alleles among the 29 accessions. Analysis of the genetic diversity of all 29 accessions revealed similarity coefficients that ranged from 0.528 to 0.983. The EST-SSR markers developed in this study represent the first large-scale development of SSR markers for *S. guianensis*. These SSR markers will provide a valuable resource for genetic diversity studies, cultivar fingerprinting, construction of genetic maps, identification of quantitative trait loci for important traits, and molecular marker-assisted selection breeding in stylo species.

**Keywords** *Stylosanthes guianensis* · Transcriptome sequences · Simple sequence repeat (SSR) · Genetic diversity

**Electronic supplementary material** The online version of this article (doi:10.1007/s11032-015-0370-x) contains supplementary material, which is available to authorized users.

X. Ding · Q. Jia · X. Luo · L. Zhang · H. Cong ·  
G. Liu · C. Bai (✉)

Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences & Key Laboratory of Crop Gene Resources and Germplasm Enhancement in Southern China, Ministry of Agriculture, Danzhou 571737, Hainan, China  
e-mail: baichangjun@126.com

## Introduction

*Stylosanthes* Sw. (Fabaceae) consists of approximately 50 species, which are distributed in tropical and subtropical regions in America, Africa, and Southeast Asia (Calles and Schultze-Kraft 2010). Most species of stylo genus are highly productive,

have high protein content, and grow well in relatively poor soil (Costa 2006). Thus, these species play an important role in soil improvement, water and soil conservation, and recovery of degraded wastelands. Moreover, some species are the most important tropical legume pasture in tropical and subtropical regions (Maass and Sawkins 2004). *Styloseanthes guianensis* (Aubl.) Sw. is the most widespread and promising forage crop, and many *S. guianensis* cultivars have been bred, which produce a large amount of dry matter and retain their leaves even during dry season (Santos-Garcia et al. 2012a).

Genetic improvement and breeding of stylo have been severely hindered by lack of robust molecular markers. These markers are extremely important for the genetic analysis of existing and newly collected species, construction of specific genetic linkage map, analysis of quantitative trait loci (QTL) for important quality and agricultural traits, and molecular breeding (Chandra et al. 2011). Within the last two decades, various types of molecular markers were used for diverse purposes in stylo, for example, random amplified polymorphic DNA (RAPD), restriction fragment length polymorphisms (RFLP), amplified fragment length polymorphism (AFLP), inter simple sequence repeat (ISSR), and sequence-tagged site (STS) for genetic diversity (Nagaich et al. 2013; Liu et al. 1999; Vander Stappen et al. 2000; Nagaich and Chandra 2009); RAPD and STS for paternity analysis and genetic linkage map construction (Thumma et al. 2001; Chandra 2006); and sequence-related amplified polymorphism (SRAP) for identification of hybrids (Huang et al. 2014). Simple sequence repeat (SSR) markers, also known as microsatellites, offer several advantages because they are genome specific, highly reproducible, codominant, easily scored, highly polymorphic, and PCR-based DNA markers (Zane et al. 2002). Therefore, SSR markers are widely used in characterizing and certifying of plant materials, DNA fingerprinting, constructing linkage maps, various population genetic studies, identifying of varieties with agronomic potential, and crop-breeding programs (Kalia et al. 2011). To date, only 139 SSR markers have been reported and are available in stylo species: 44 from *S. guianensis*, 26 from *Stylosanthes macrocephala*, 23 from *Stylosanthes seabrana*, 23 from *Stylosanthes capitata* Vogel, 4 from *Stylosanthes huimilis*, 2 from *Stylosanthes hamata*, and 1 from *Stylosanthes viscosa* (Billotte et al. 1999; Vander

Stappen et al. 1999; Santos et al. 2009a, 2009b, 2009c; Chandra et al. 2011); and 16 were developed in silico using genomic and EST sequences of 11 stylo species (Mace et al. 2008). Santos-Garcia et al. (2011) tested the sets of SSR markers developed by Vander Stappen et al. (1999) and Santos et al. (2009a) in 20 accessions that were used as maternal parents in a progeny array. They found that only 5 SSR markers were polymorphic and could be used to determine outcross rates in *S. guianensis*. Results showed that the outcrossing rate in *S. guianensis* is 26 %, indicating that the species has a predominantly autogamous mixed mating system. Santos-Garcia et al. (2012a) evaluated the relationships among 150 accessions of the *S. guianensis* Brazilian germplasm collection using 20 SSR markers developed by Santos et al. (2009a). Santos-Garcia et al. (2012b) assessed the genetic diversity and population structure of 134 *S. macrocephala* accessions using 13 SSR markers developed by Santos et al. (2009b) and 192 *S. capitata* accessions using 15 SSR markers by Santos et al. (2009c). Finally, core collections were established for *S. macrocephala* and *S. capitata* including 23 and 13 accessions, respectively. The 139 SSR markers were developed from different species, so lesser SSR markers were applicable to each species in stylo. Apparently, 139 SSR markers do not meet the demand for genetic improvement and breeding, so developing more effective SSR markers in stylo is essential.

Based on the origin of the sequences used for the initial identification of SSRs, SSRs are classified into genomic SSRs (derived from genomic sequences) and EST-SSRs (derived from ESTs), and SSR markers are classified into genomic SSR markers and EST-SSR markers accordingly (Tan et al. 2012). EST-SSR markers can be used for direct gene tagging for certain important agronomic traits because of their association with coding sequences (Varshney et al. 2005). In addition, EST-SSR markers demonstrate a higher level of transferability across related species compared with genomic SSR markers, but they suffer from being less polymorphic (Ellis and Burke 2007). Transcriptome sequencing using next-generation sequencing technology provides a simple and effective method to rapidly identify SSR loci derived from expressed sequences and develop numerous EST-SSR markers (Dutta et al. 2011; Wang et al. 2014).

In recent years, an increasing number of EST datasets have become available for many organisms

(Dutta et al. 2011; Wang et al. 2014). However, relatively few ESTs are currently available in stylo species. Our laboratory has previously performed *de novo* transcriptome sequencing for *S. guianensis* ‘Reyan No. 5’ using the Illumina sequencing platform and identified a total of 36,558 unigenes (Ding unpublished). The data have been submitted to the Sequence Read Archive of the NCBI database under the accession ID PRJNA 277,095 and will serve as a public information platform for further studies in *S. guianensis*. In this study, 36,558 unigenes were used to detect SSR loci for large-scale development and characterization of EST-SSR markers. Moreover, the effectiveness of these EST-SSR markers was assessed when the genetic relationship among 29 accessions of *S. guianensis* was determined.

## Materials and methods

### Plant materials and DNA extraction

Twenty-nine *S. guianensis* accessions collected by our research group were used for the polymorphic analysis and characterization of SSR markers (Table 1). All 29 accessions were grown in the experimental fields of Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Danzhou, Hainan, China. Fresh leaves from each accession were collected for DNA extraction according to the cetyltrimethyl ammonium bromide (CTAB) method with modifications (Vander Stappen et al. 2000). The fresh leaves were ground to fine powder under liquid nitrogen in a deep frozen mortar and pestle, and then about 100 mg powder was transferred into a new tube and mixed with 750  $\mu$ L of preheated CTAB buffer [4 % CTAB, 100 mM Tris-HCl (pH 8.0), 1.5 M NaCl, 25 mM EDTA, 2 %  $\beta$ -mercapto-ethanol, 4 % Polyvinylpyrrolidone]. After incubation at 65 °C for 45 min, the samples were mixed with an equal volume of 24:1 chloroform/isopropyl alcohol and centrifuged at 12,000 rpm for 10 min. The upper phase was transferred into a new tube. Then, 10  $\mu$ L of 10 mg/mL RNase A was added into the tube, and the mixture was incubated at room temperature for 15 min. The samples were mixed with an equal volume of 24:1 chloroform/isopropyl alcohol and centrifuged at 12,000 rpm for

**Table 1** Twenty-nine *S. guianensis* accessions used for genetic diversity analysis

Code	Accessions	Abbreviation	Origin	Cluster
1	Reyan No. 5	RY5	China	I
2	CIAT11365	CI1	Columbia	I
3	Reyan No. 10	RY10	China	I
4	Reyan No. 7	RY7	China	I
5	Reyan No. 13	RY13	China	I
6	GC1578	GC1	Brazil	I
7	FM03-2	FM1	Kenya	I
8	FM9405	FM2	Kenya	I
9	ATF3308	ATF3	Australia	I
10	GC1517	GC2	Brazil	I
11	GC1557	GC3	Kenya	I
12	CIAT1283	CI2	Columbia	IV
13	GC1576	GC4	Brazil	I
14	GC348	GC5	Brazil	I
15	GC1480	GC6	Kenya	I
16	Schofield	SCH	Australia	I
17	Graham	GRA	Australia	I
18	GC1581	GC7	Brazil	II
19	CIAT350	CI3	Australia	I
20	Cook	COO	Australia	I
21	CIAT 25	CI4	Columbia	I
22	CIAT 41	CI5	Columbia	I
23	CIAT 46	CI6	Columbia	I
24	CIAT 104	CI7	Columbia	IV
25	CIAT 151	CI8	Columbia	I
26	CIAT 1015	CI9	Columbia	I
27	CIAT 1278	CI10	Columbia	IV
28	TPRC190	TPR	Ivory Coast	I
29	CIAT 10121	CI11	Columbia	III

10 min. The upper phase was transferred into a new tube containing 100  $\mu$ L of 3 M sodium acetate and 600  $\mu$ L of chilled isopropyl alcohol. The mixture was left to precipitate at -20 °C for 30 min, centrifuged at 12,000 rpm for 15 min, washed with 70 % ethanol, air-dried, and dissolved in 100 mL of TE [100 mM Tris-HCl (pH 8.0), 1 mM EDTA]. The quality and quantity of genomic DNA were estimated using 1 % agarose gel and UV spectrophotometer, respectively. The DNA concentration was adjusted to 50 ng/ $\mu$ L and stored at -20 °C until use.

## Identification of SSR loci and development of SSR markers

Total RNA from the shoots of *S. guianensis* 'Reyan No. 5' seedlings was extracted using RNAPrep Pure Plant Kit according to the manufacturer's instructions (Tiangen, Beijing, China). Both the quantity and quality of the RNA were verified using NanoDrop 2000C (Thermo Scientific) and formaldehyde 1 % agarose gel. Illumina sequencing was performed using the HiSeq™ 2000 platform according to the manufacturer's instructions (Illumina, San Diego, USA).

Using TruSeq RNA Sample Prep Kit (Illumina), poly (A) mRNA was isolated from 20 µg of total RNA and fragmented into small pieces, and then the double-stranded cDNA was synthesized using random hexamer primers. The synthesized cDNA was subjected to end-repair and tailing A. After that, sequencing adapters were ligated to the ends of these 3'-adenylated cDNA fragments. The suitable fragments were separated by agarose gel electrophoresis and purified for the PCR amplification as templates. Finally, after validating on an Agilent Technologies 2100 Bioanalyzer, the library was sequenced using Illumina HiSeq™ 2000. Raw reads were filtered by removing adapter sequences, low-quality sequences with ambiguous bases 'N', and reads with more than 10 %  $Q < 20$  bases. Finally, 36,558 unigenes were obtained by *de novo* assembly (Ding unpublished). Putative SSR loci were mined using the MicroSatellites identification tool (MISA, <http://pgrc.ipk-gatersleben.de/misa>) software within these EST databases. In this study, default criteria were used to select a minimum of 8 repeats for dinucleotide motifs, 6 repeats for trinucleotide motifs, 5 repeats for tetranucleotide motifs, and 4 repeats for pentanucleotide and hexanucleotide motifs. Mononucleotide repeat motifs and complex SSR types were excluded from the study. Primers were designed to flank the SSR loci using the public domain primer design software Batch Primer 3.0 (<http://probes.pw.usda.gov/cgi-bin/batchprimer3/batchprimer3.cgi>). The input parameters were as follows: primer length, 18–27 bp; Tm value, 57–63 °C; and estimated product size, 100–280 bp.

### Putative functional annotation

The putative function of 1873 SSR-containing sequences that are suitable for designing primers was

detected by searching for these sequences against NCBI nonredundant protein database using the BLASTX program. Each unigene was assigned to its putative functions based on the highest scores of BLASTX. Sequences with  $E$  value  $< 10^{-6}$  and scores  $> 50$  were annotated.

### Amplification and polymorphic detection of SSR markers

Polymorphism of 115 EST-SSR markers developed in this study (Supplementary Table 1) was tested and characterized using 29 *S. guianensis* accessions. The effect of genetic materials on the evaluation of polymorphism was excluded by amplifying simultaneously 12 genomic SSR markers reported by Santos et al. (2009) and Vander Stappen et al. (1999) (Supplementary Table 2). PCR amplifications and SSR assay were conducted as described by Wu et al. (2014). PCR was performed in 20 µL of reaction mixture containing 1 µL of 50 ng/µL genomic DNA, 0.4 µL of each 10 µM primer, 1.2 µL of 2.5 mM dNTPs, 0.3 µL of 5 U/µL Taq polymerase, 2.0 µL of 10× PCR buffer, and 14.7 µL of sterile distilled water. The mixture was finally overlaid with 20 µL of mineral oil. DNA was amplified under the following PCR program: initial denaturation at 94 °C for 4 min, followed by 35 cycles of denaturation at 94 °C for 45 s, annealing at 55 °C for 45 s, and elongation at 72 °C for 45 s, ending with an elongation step for 10 min at 72 °C. The amplification products were separated using 8 % denaturing polyacrylamide gel electrophoresis and then visualized by silver staining. The number and frequencies of alleles, the observed and expected heterozygosities ( $H_o$  and  $H_e$ , respectively), and Shannon index were estimated using POPGENE version 1.32 (Yeh et al. 1999). The polymorphic information content (PIC) of the SSR markers was calculated using the standard formula described by Botstein et al. (1980).

### Determination of genetic relationships among 29 *S. guianensis* accessions

To assess the effectiveness of the EST-SSR markers developed in this study, we analyzed the genetic relationships among the 29 *S. guianensis* accessions (Table 1) using 115 of these EST-SSR markers (Supplementary Table 1). The effect of the genetic

background of these *S. guianensis* accessions on the evaluation process was excluded by including 12 genomic SSR markers (Supplementary Table 2) in the analysis. All reliable bands were scored as 'presence (1)' or 'absence (0)' among the 29 accessions of *S. guianensis* amplified by each primer pair, and a binary matrix was recorded. The data were analyzed using the Numerical Taxonomy Multivariate Analysis System (NTSYS-pc) version 2.10 software with SIMQUAL module to calculate genetic similarity (GS) coefficients based on coefficient for similarity matching (Rohlf 2000). With the GS matrix, we constructed a dendrogram using the unweighted pair group method with arithmetic average (UPGMA) to determine the genetic relationships among the 29 accessions.

## Results

### Identification of EST-SSRs in *S. guianensis* transcriptome

A total of 36,558 unigenes sequences annotated in Nr database with a mean length of 1650 bp were obtained to identify potential SSRs, which represented approximately 60.3 Mb (Ding, unpublished). Using the MISA software, 4115 SSR loci were identified from 3643 (9.96 %) unigene sequences, of which 406 (11.14 %) contained 2 SSR loci, 30 sequences contained 3 SSR loci, and 2 sequences contained 4

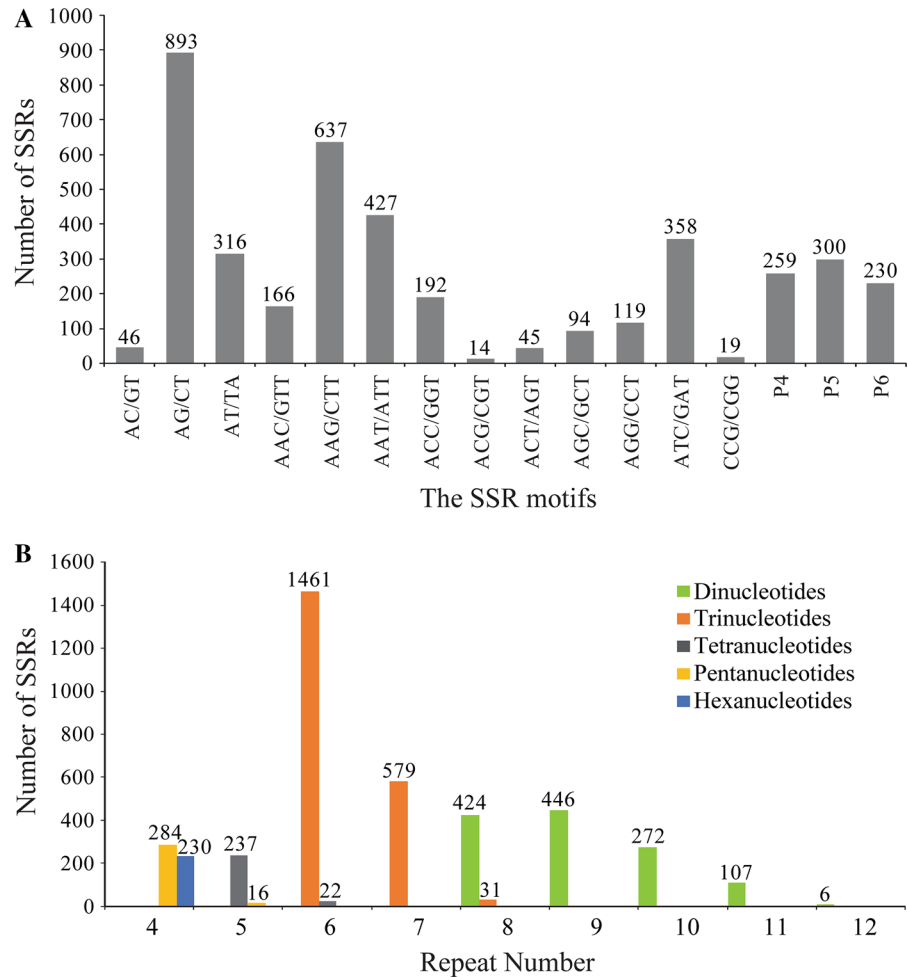
SSR loci (Table 2). The frequency of occurrence for EST-SSRs was 1 SSR per 14.7 kb of unigene sequence, or 1 SSR per 8.88 unigenes in *S. guianensis* transcriptome. Among the 4115 SSR loci, dinucleotide and trinucleotide repeat motifs were the most abundant types with frequencies of 30.50 % (1255) and 50.33 % (2071), respectively, whereas only 259 (6.29 %) tetranucleotide, 300 (7.29 %) pentanucleotide, and 230 (5.59 %) hexanucleotide motifs were identified (Table 2).

Within the developed SSR markers, 206 motif sequence types were identified. Among these, 3, 10, 17, 46, and 130 motifs contained dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide repeats, respectively. Among 13 major motifs containing dinucleotide and trinucleotide (Fig. 1a), the most abundant motif of all SSRs was AG/CT (893, 21.70 %), followed by AAG/CTT (637, 15.48 %), AAT/ATT (427, 10.38 %), ATC/GAT (358, 8.70 %), AT/TA (316, 7.68 %), ACC/GGT (192, 4.67 %), AAC/GTT (166, 4.03 %), AGG/CCT (119, 2.89 %), AGC/GCT (94, 2.28 %), AC/GT (46, 1.12 %), ACT/ATG (45, 1.09 %), CCG/CGG (19, 0.46 %), and ACG/CGT (14, 0.34 %). The remaining 193 types of motifs accounted for 19.17 % of all SSRs analyzed. The SSR lengths were mostly distributed from 16 bp to 21 bp, accounting for 89.99 % of all SSR loci, followed by 22–25 bp (10.01 %). The 18-bp long SSRs (1907, 46.34 %) were the most common. The repeat number of SSRs ranged from 4–12, and

**Table 2** Summary of EST-SSR loci and markers identified in stylo transcriptome

Searching items	Numbers	
	EST-SSR loci	EST-SSR markers
Total number of sequences examined	36,558	
Total size of sequences examined (bp)	60,321,534	
Total number of SSRs identified	4115	2051
Number of SSR-containing sequences	3643	1873
Number of sequences containing more than 1 SSR	438	167
Number of SSR markers developed		2008
Frequency of SSR loci or markers	1/14.7 kb	1/30.0 kb
Number of SSR markers containing 2 SSR loci		43
Dinucleotides	1255	502
Trinucleotides	2071	965
Tetranucleotides	259	153
Pentanucleotides	300	192
Hexanucleotides	230	153

**Fig. 1** Characterization of SSRs in *S. guianensis* transcriptome. **a** Frequency distribution of EST-SSRs based on motif types; **b** length distribution of EST-SSRs based on the number of repeat units



SSRs with 6 repeat units (1483, 36.04 %) were the most abundant followed by SSRs with 7 (579, 14.07 %), 4 (514, 12.49 %), 8 (455, 11.06 %), and 9 (446, 10.84 %) repeat units, whereas only 6 SSRs with 12 repeats were identified (Fig. 1b).

#### Development of EST-SSR markers and functional annotation

EST-SSR markers were developed by designing primer pairs to flank the SSR loci using Primer 3.0 software. A total of 1873 SSR-containing unigene sequences could be used to design primer pairs. Additionally, given that the sequences that flank the SSR regions were too short or inappropriate, the remaining 1770 sequences were unsuitable for designing primer pairs. A total of 2051 SSR loci were present

in the 1873 SSR-containing sequences, of which 167 sequences contained more than 1 SSR loci (Table 2). Finally, 2008 EST-SSR markers were developed from these SSR-containing sequences, of which 115 sequences contained 2 markers, 7 sequences contained 3 markers, and 2 sequences contained 4 markers (Supplementary Table 3). Within these markers, 1965 markers including 502 dinucleotide, 965 trinucleotide, 153 tetranucleotide, 192 pentanucleotide, and 153 hexanucleotide repeat motifs harbored single SSR locus, and the other 43 markers contained two tightly linked SSR loci (Table 2, Supplementary Table 3).

These sequences were searched against GenBank using BLASTX to explore the potential function of 1873 SSR-containing sequences that could be used for designing primers. The majority of these sequences (1671, 89.22 %) matched to genes in leguminous



plants: 1070 sequences (57.13 %), *Glycine max* protein, 371 sequences (19.81 %), *Cicer arietinum* protein, 162 sequences (8.65 %), *Medicago truncatula* protein, and 68 sequences (3.36 %), *Lotus japonicus* protein (Supplementary Table 3).

#### Characterization and polymorphic analysis of EST-SSR markers

The quality and polymorphism of EST-SSR markers were evaluated by selecting 100 EST-SSR markers with trinucleotide repeats located in coding sequences and 15 EST-SSR markers with hexanucleotide repeats located in coding sequences, which were further amplified using PCR in 29 *S. guianensis* accessions. Of these 115 markers, 96 produced PCR fragments with expected sizes, so they were used as validated SSR markers for further analysis (Supplementary Fig. 1). In addition, 9 markers generated PCR fragments that were larger than expected, 5 markers generated multiple bands, and 5 markers failed to amplify bands (Supplementary Table 1). A total of 104 loci were successfully amplified by 96 validated SSR markers, of which 6 SSR markers (RM0714, RM0906, RM1641, RM1805, RM1867, and RM1950) and 1 SSR marker (RM0051) were amplified at 2 and 3 loci, respectively. Of the 104 loci, 89 were polymorphic and 15 were monomorphic among the 29 *S. guianensis* accessions (Supplementary Fig. 2). A total of 267 alleles were released by these 89 polymorphic loci, and the number of alleles ranged from 2 to 6 with an average of 3.00 per locus. Of these 89 loci, 31, 36, 14, 7, and 1 loci (RM1670) had 2, 3, 4, 5, and 6 alleles, respectively, in the 29 *S. guianensis* accessions (Table 3). All 12 genomic SSR markers produced PCR fragments, and 11 markers were polymorphic among the 29 *S. guianensis* accessions. A total of 37 alleles were released by these 11 polymorphic loci with an average of 3.56 alleles per marker (Table 3).

The  $H_o$  and  $H_e$  values of 89 polymorphic EST-SSR loci ranged from 0.00 to 0.83 (mean 0.04) and 0.03 to 0.73 (mean 0.27), respectively (Table 3). RM1340 and RM1984 loci have the highest  $H_o$  and  $H_e$ , respectively. The PIC values for 89 polymorphic EST-SSR loci ranged from 0.03 to 0.68, with mean value of 0.24 (Supplementary Fig. 3, Table 3). Furthermore, 51 loci showed PIC values smaller than 0.25, and only 2 loci showed PIC values larger than 0.50. The RM1984 locus exhibited the highest PIC value (0.68), followed

by the RM0976 locus with 0.60. PIC values for 11 polymorphic genomic SSR markers were similarly small and ranged from 0.06 to 0.79 with an average of 0.28 (Table 3). In addition, the Shannon index for 89 polymorphic EST-SSR loci ranged from 0.09 to 1.35 with an average of 0.50 (Table 3).

#### Evaluation of genetic relationship within 29 *S. guianensis* accessions

To assess the effectiveness of the newly developed EST-SSR markers, 81 polymorphic EST-SSR markers and 11 genomic SSR markers (Table 3) were used separately to evaluate the genetic diversity and relationship among 29 accessions of *S. guianensis* (Fig. 2). A total of 267 amplified alleles by EST-SSR markers and 37 alleles by genomic SSR markers were separately scored to calculate their GS using the SIMQUAL module of the NTSYS-pc software. The pairwise similarity coefficient for the EST-SSR data ranged from 0.528 to 0.983 compared with 0.514 to 1.000 for the genomic SSR markers (Supplementary Table 4). Among 406 pairwise similarity coefficients by EST-SSR markers, 103 (25.37 %) between 4 varieties (CI2, CI7, CI9, and CI11) and other varieties ranged from 0.528 to 0.654 with mean value of 0.582; and 24 (5.91 %) between GC7 and other varieties ranged from 0.751 to 0.814 with an average of 0.782. However, the remaining pairwise similarity coefficients (279, 68.72 %) were more than 0.87. The smallest GS coefficient (0.528) was found between CI2 and GC7, whereas the greatest GS coefficient was 0.983, which was observed between GC4 and COO as well as CI9 and TPR. The UPGMA cluster analysis based on the GS coefficient matrix of EST-SSR successfully differentiated all 29 accessions, whereas genomic SSR indicated 100 % similarity between few accessions (CI1, CI5, ATF3, GC5, CI4, and TPR) (Fig. 2). With EST-SSR markers, the 29 accessions could be distinctly classified into 4 clusters (Clusters I, II, III, and IV) (Fig. 2a). With 11 genomic SSR markers, the 29 accessions could be distinctly classified into 3 clusters (Clusters A, B, and C) (Fig. 2b). However, the topologies of the 2 dendrograms were quite similar. The major cluster, Cluster I, comprised 24 accessions, and Cluster II contained GC7. GC7 could be clearly separated from the other 24 accessions in Cluster A. Clusters III and C contained only 1

**Table 3** Characterization of polymorphic SSR markers

Markers code	Na	$H_o$	$H_e$	PIC	Shannon index	Markers code	Na	$H_o$	$H_e$	PIC	Shannon index
RM0003	3	0.00	0.24	0.23	0.48	RM1314	2	0.00	0.24	0.21	0.40
RM0041	3	0.00	0.24	0.23	0.48	RM1324	3	0.03	0.19	0.18	0.38
RM0051A	3	0.03	0.52	0.40	0.77	RM1328	5	0.07	0.48	0.44	0.93
RM0051B	3	0.00	0.42	0.35	0.69	RM1335	4	0.03	0.27	0.26	0.56
RM0051C	3	0.00	0.51	0.41	0.80	RM1340	3	0.83	0.59	0.50	0.95
RM0053	2	0.03	0.03	0.03	0.09	RM1371	2	0.00	0.24	0.21	0.40
RM0094	2	0.00	0.40	0.32	0.59	RM1379	3	0.28	0.44	0.40	0.78
RM0107	4	0.03	0.22	0.21	0.48	RM1439	2	0.03	0.03	0.03	0.09
RM0113	3	0.00	0.46	0.40	0.78	RM1521	2	0.00	0.07	0.06	0.15
RM0180	3	0.00	0.13	0.13	0.30	RM1557	3	0.10	0.25	0.23	0.49
RM0203	4	0.03	0.27	0.26	0.56	RM1587	3	0.03	0.31	0.27	0.54
RM0208	3	0.00	0.24	0.23	0.48	RM1615	3	0.00	0.29	0.27	0.55
RM0249	3	0.00	0.54	0.48	0.93	RM1619	4	0.07	0.16	0.16	0.38
RM0255	2	0.00	0.24	0.21	0.40	RM1636	2	0.00	0.19	0.17	0.33
RM0272	2	0.00	0.19	0.17	0.33	RM1641A	4	0.03	0.27	0.26	0.56
RM0396	2	0.00	0.19	0.17	0.33	RM1641B	3	0.03	0.22	0.20	0.42
RM0398	2	0.00	0.19	0.17	0.33	RM1670	6	0.03	0.43	0.42	0.96
RM0436	2	0.00	0.19	0.17	0.33	RM1771	4	0.00	0.35	0.33	0.72
RM0466	3	0.00	0.29	0.27	0.55	RM1781	2	0.00	0.07	0.06	0.15
RM0479	2	0.00	0.07	0.06	0.15	RM1805A	4	0.10	0.54	0.48	0.96
RM0522	2	0.14	0.13	0.12	0.25	RM1805B	3	0.00	0.59	0.50	0.95
RM0530	3	0.07	0.07	0.07	0.17	RM1813	3	0.00	0.35	0.32	0.64
RM0561	2	0.00	0.24	0.21	0.40	RM1834	5	0.03	0.36	0.34	0.77
RM0593	2	0.00	0.29	0.25	0.46	RM1841	2	0.00	0.07	0.06	0.15
RM0640	3	0.07	0.27	0.24	0.49	RM1843	5	0.10	0.40	0.38	0.85
RM0651	2	0.03	0.03	0.03	0.09	RM1859	3	0.00	0.25	0.23	0.50
RM0652	2	0.00	0.07	0.06	0.15	RM1867A	3	0.00	0.13	0.13	0.30
RM0671	2	0.03	0.03	0.03	0.09	RM1867B	4	0.00	0.25	0.24	0.54
RM0674	2	0.00	0.07	0.06	0.15	RM1879	4	0.00	0.51	0.45	0.92
RM0714A	2	0.00	0.13	0.12	0.25	RM1881	3	0.03	0.07	0.07	0.17
RM0714B	3	0.00	0.30	0.28	0.58	RM1917	5	0.10	0.16	0.16	0.41
RM0735	3	0.00	0.24	0.23	0.48	RM1927	3	0.00	0.39	0.36	0.72
RM0792	3	0.00	0.29	0.27	0.55	RM1938	3	0.00	0.25	0.23	0.50
RM0861	3	0.07	0.07	0.07	0.17	RM1950A	2	0.00	0.19	0.17	0.33
RM0898	2	0.00	0.19	0.17	0.33	RM1950B	3	0.00	0.30	0.28	0.58
RM0906A	2	0.00	0.19	0.17	0.33	RM1968	4	0.03	0.27	0.26	0.56
RM0906B	2	0.00	0.19	0.17	0.33	RM1981	3	0.07	0.46	0.41	0.81
RM0975	3	0.00	0.19	0.18	0.40	RM1984	4	0.00	0.73	0.68	1.35
RM0976	5	0.03	0.65	0.60	1.23	Mean	3.00	0.04	0.27	0.24	0.50
RM1005	5	0.10	0.43	0.41	0.89	<b>SG03D1</b>	2	0.04	0.27	0.23	0.43
RM1034	3	0.00	0.52	0.42	0.81	<b>SG03G2</b>	4	0.04	0.28	0.27	0.60
RM1039	3	0.14	0.16	0.15	0.34	<b>SG03G8</b>	2	0.00	0.24	0.21	0.40
RM1058	2	0.00	0.33	0.27	0.51	<b>SG01D7</b>	3	0.10	0.10	0.10	0.24
RM1100	4	0.07	0.25	0.23	0.50	<b>SSR1-2a</b>	4	0.66	0.49	0.41	0.84



**Table 3** continued

Markers code	Na	$H_o$	$H_e$	PIC	Shannon index	Markers code	Na	$H_o$	$H_e$	PIC	Shannon index
RM1110	4	0.00	0.30	0.28	0.62	<b>SSR1-3</b>	2	0.00	0.19	0.17	0.33
RM1116	2	0.03	0.03	0.03	0.09	<b>SSR1-9</b>	7	0.07	0.83	0.79	1.79
RM1147	3	0.00	0.30	0.28	0.58	<b>SSR1-24</b>	2	0.00	0.07	0.06	0.15
RM1182	2	0.03	0.03	0.03	0.09	<b>SSR2-43</b>	4	0.04	0.22	0.21	0.48
RM1239	4	0.03	0.37	0.33	0.69	<b>SSR4-5</b>	2	0.00	0.07	0.06	0.15
RM1240	2	0.14	0.19	0.17	0.33	<b>SSR4-51</b>	5	0.14	0.65	0.58	1.21
RM1270	5	0.03	0.52	0.49	1.06	<b>Mean</b>	3.36	0.10	0.31	0.28	0.60

The bold indicates genomic SSR markers

accession, CI11. Clusters IV and C comprised 3 accessions, namely CI2, CI7, and CI10.

## Discussion

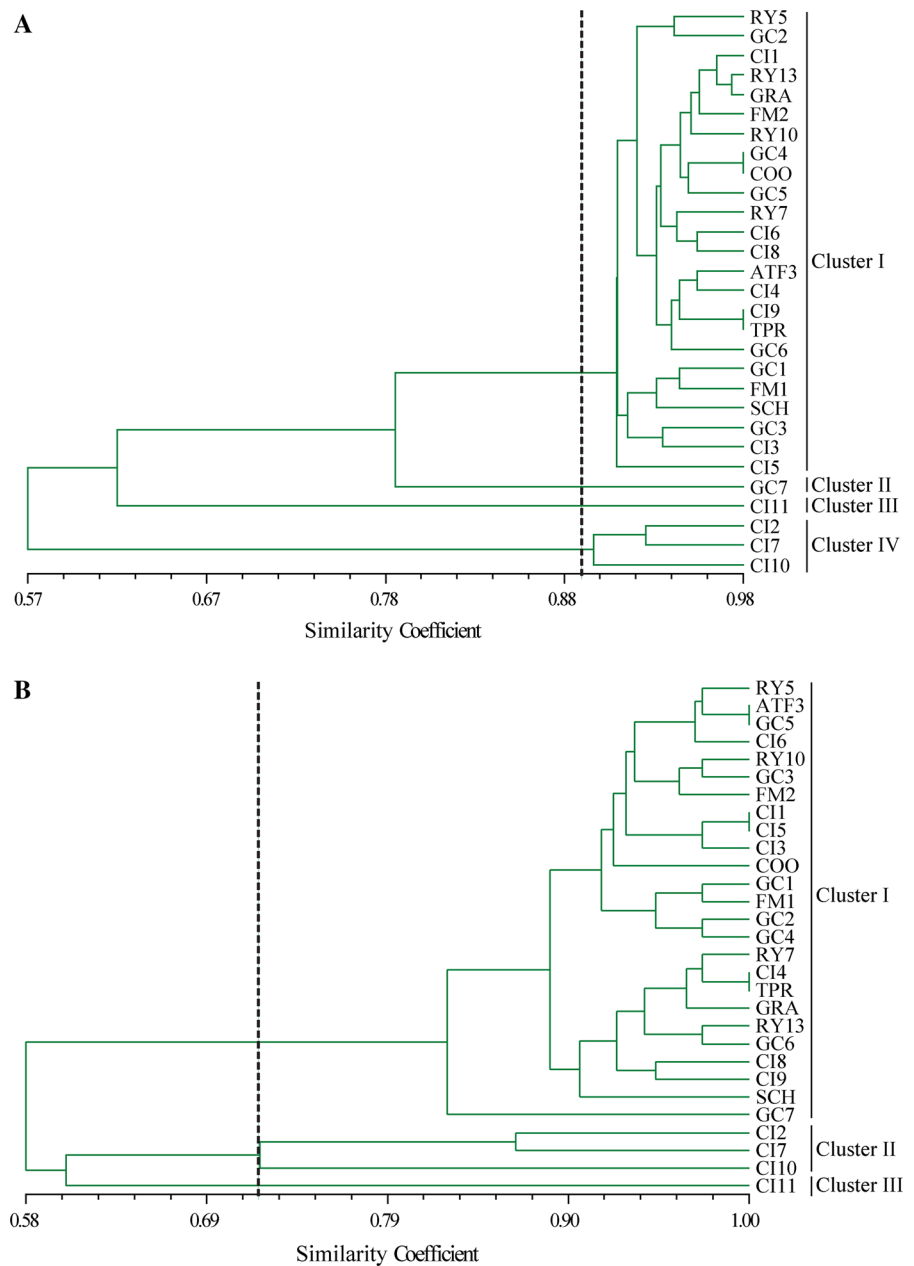
Identification and characterization of EST-SSRs in *S. guianensis* transcriptome

Transcriptome sequencing using next-generation sequencing technology provides a simple and effective method to obtain large EST sequences that will enable large-scale development of functional SSR markers. An increasing number of EST-SSR markers, based on transcriptome sequences, have been developed in more species, such as alfalfa, sheepgrass, bermudagrass, and pigeonpea (Liu et al. 2013a, b; Chen et al. 2013; Tan et al. 2012; Dutta et al. 2011). In stylo, only 139 SSR markers have been reported, which is not sufficient for the demand of effective genetic improvement and molecular marker-assisted selection (MAS) breeding. This work is the first study in which a large number of effective EST-SSR markers have been developed based on transcriptome sequences in *S. guianensis*. These EST-SSR markers will provide useful molecular tools for genetic improvement and MAS breeding of *Stylosanthes* spp.

The frequency and distribution of SSR-containing ESTs in different plant species are significantly influenced by the diversity in search criteria, size of the dataset, and database-mining tools (Varshney et al. 2005). In this study, approximately 9.96 % of 36,558 ESTs contained at least 1 SSR, which was similar to the reported values for *Gossypium hirsutum* (9.56 %)

and *Medicago truncatula* (10.62 %; Kumpatla and Mukhopadhyay 2005), higher than that for pigeonpea (7.6 %; Dutta et al. 2011), orchardgrass (7.0 %; Bushman et al. 2011), bermudagrass (6.4 %; Tan et al. 2012), and sheepgrass (4.12 %; Chen et al. 2013), but lower than that for tea (16.7 %; Tan et al. 2013) and coffee (18.5 %; Aggarwal et al. 2007). The frequency of EST-SSRs was 14.7 kb in *S. guianensis* ESTs (Table 2). This value was comparable with the frequency reported in *Arabidopsis* (13.8 kb) and poplar (14.0 kb), lower than that in cotton (20.0 kb) and ramie (19.3 kb; Liu et al. 2013a, b), but much higher than that in tea (2.4 kb; Tan et al. 2013), rice (3.4 kb), wheat (5.4 kb), and gerbera (5.6 kb; Kantety et al. 2002). In *S. guianensis*, trinucleotide repeat (50.33 %) was the most abundant motif (Table 2), followed by dinucleotide (30.50 %). This finding was consistent with earlier reports in alfalfa (Liu et al. 2013a, b), soybean (Xin et al. 2012), sheepgrass (Chen et al. 2013), radish (Wang et al. 2012), and sweet-potato (Wang et al. 2011), in which the trinucleotide repeat was reported as the most common motif. However, dinucleotide repeat was the most abundant motif in many reports for tea, pigeonpea, tree peony, and coffee (Tan et al. 2013; Dutta et al. 2011; Wu et al. 2014; Aggarwal et al. 2007). The different relative abundances of dinucleotide and trinucleotide repeats may be partly attributed to the search criteria and database-mining tools in different studies. For example, in *Jatropha curcas*, when the minimum repeat unit criteria of dinucleotide were increased from five to eight, with five units retained for trinucleotide, the abundance of dinucleotide repeats decreased from 66.0 to 32.0 %, but that of trinucleotide repeats

**Fig. 2** Dendrogram plot for 29 accessions based on cluster analysis of SSR markers. **a** Dendrogram plot for 29 accessions based on cluster analysis of EST-SSR markers; **b** dendrogram plot for 29 accessions based on cluster analysis of genomic SSR markers. The vertical dashed line indicates the level of genetic similarity at which clusters were delineated



increased from 23.5 to 46.8 % (Yadav et al. 2011). Thus, the abundance of EST-SSR loci and motifs is significantly influenced by different search criteria and mining tools used for SSR search.

In this study, among the dinucleotide motifs (Fig. 1a), AG/CT was the most dominant motif, accounting for 21.70 % of all EST-SSRs, followed by AT/TA (7.68 %). Similar results were found in the majority of plants, except for several species, such as

*Lycopersicon pennellii*, in which AT/TA is the most abundant dinucleotide motif with higher frequency of 23.6 % than AG/CT (Kumpatla and Mukhopadhyay 2005). GC/CG was the rarest dinucleotide repeat in most plants reported: 2.38 %, sheepgrass (Chen et al. 2013); 0.76 %, *J. curcas* (Yadav et al. 2011); 0.19 %, tree peony (Wu et al. 2014); 0.08 %, tea (Tan et al. 2013); 0.03 %, radish (Wang et al. 2012); lower than 1.07 %, barley; 0.28 %, maize; 0.04 %, rice (Kantety

et al. 2002); 0.01 %, soybean (Xin et al. 2012); and 0.01 %, alfalfa (Liu et al. 2013a, b). No GC/CG repeat motifs were identified among the 36,558 ESTs in *S. guianensis*. Similarly, no GC/CG repeat motifs were observed in sweetpotato (Wang et al. 2011) and ramie (Liu et al. 2013a, b). The most abundant trinucleotide repeat motif was AAG/CTT (15.48 %) in *S. guianensis* (Fig. 1a), which was consistent with the results in most dicotyledonous plants (Kumpatla and Mukhopadhyay 2005). By contrast, the most frequent trinucleotide repeat motif was AAC/TTG in wheat, followed by CCG/CGG, and CCG/CGG in rice, barley, sorghum, and sheepgrass (Kantety et al. 2002; Chen et al. 2013). However, the CCG/CGG was very rare in dicotyledonous plants. Thus, the abundance of CCG/CGG motif was a specific feature in monocots, which may be due to high GC content.

#### Development and polymorphism analysis of EST-SSR markers

In this study, a total of 2008 EST-SSR markers were developed from *Stylosanthes* ESTs that had been assembled in our previous transcriptome sequencing study (Ding unpublished). To evaluate the quality of the EST-SSR markers designed in this study, 115 markers located in coding sequences were amplified, and 96 markers produced clear bands with expected size (83.34 % success rate). Changes in the lengths of SSRs will result in the sequential changes of proteins, which dictate virtually every reaction in the cell and thus are directly responsible for observable characteristics, when the EST-SSR is located in coding region. So, the polymorphism of EST-SSR markers located in coding sequences of genes responsible for important agronomic traits has high relativity with phenotype. A total of 115 EST-SSR markers located in coding sequences, which showed the higher potential for functional markers, were chosen from markers with trinucleotide or hexanucleotide motifs because trinucleotide and hexanucleotide repeats primarily occurred within the coding sequences (Wu et al. 2014). The success ratio (83.48 %) was similar to that in sweetpotato (84.6 %, Wang et al. 2011) but higher than those in pigeonpea (71.24 %; Dutta et al. 2011), barley (69.8 %), alfalfa (62.96 %; Liu et al. 2013a, b), tea (59.9 %; Tan et al. 2013), and tree peony (47.3 %; Wu et al. 2014) and lower than that in ramie (98 %, Liu et al. 2013a, b). A total of 104 loci were successfully

amplified among 96 validated markers, of which 89 (85.6 %) were polymorphic and indicated a high level of intraspecific polymorphism within *S. guianensis* accessions. The average number of alleles for polymorphic EST-SSR loci was 3.00, which was lesser than that for genomic SSR markers (3.56) (Table 3), and similar results were found in wheat (Eujayl et al. 2001) and rice (Cho et al. 2000). This phenomenon may be due to the fact that the expressed sequences are more conserved compared with noncoding sequences.

The PIC value of each polymorphic SSR locus, which was determined by the number and frequency distribution of alleles within the population, was used to assess the informativeness level of SSR loci. In this investigation, among 89 polymorphic EST-SSR loci, the majority of the SSR loci (87, 97.75 %) exhibited low PIC value (PIC < 0.50), which indicated a low-to-moderate level of informativeness (Supplementary Fig. 3; Table 3). Only 2 SSR loci significantly showed a high degree of PIC. Meanwhile, PIC values for 11 polymorphic genomic SSR markers were similarly small and ranged from 0.06 to 0.79 with an average of 0.28 (Table 3). This characteristic may be due to the use of a genotype set including 24 accessions with close genetic relationship, in which the pairwise similarity coefficients ranged from 0.867 to 0.983 with an average of 0.925, out of 29 *S. guianensis* accessions observed (Supplementary Table 4).

#### Genetic diversity analysis by EST-SSR markers

In the present investigation, 29 *S. guianensis* accessions (Table 1) were selected as sample material to characterize the developed EST-SSR markers. Moreover, genetic diversity analysis among the 29 accessions was performed by UPGMA (Fig. 2). Using 96 validated EST-SSR markers, the result of cluster analysis among these accessions revealed that the highest GS was between GC4 and COO collected from Brazil and Australia, respectively, as well as between CI9 and TPR collected from Columbia and Ivory Coast, respectively. The genetic distance among these accessions was inconsistent with their geographical origin, which may be due to material exchanges between different states and regions. The similarity coefficient for the EST-SSR data ranged from 0.528 to 0.983 with a high average of 0.925 compared with 0.514–1.000 for the genomic SSR markers (Supplementary Table 4). The high similarity in *Stylosanthes*

genus was also reported by Chandra et al. (2011) and Tang et al. (2009). The result of cluster analysis using EST-SSR markers was consistent with that using genomic SSR markers (Fig. 2), so the EST-SSR markers developed in this study are useful for genetic diversity in *S. guianensis*. Finally, the 29 accessions were distinctly classified into 4 clusters based on EST-SSR markers (Fig. 2a). This result agreed with the classification of *S. guianensis* proposed by Brandão et al. (1985), according to whom *S. guianensis* consists of 4 botanical varieties: *S. guianensis* var. *guianensis*, *S. guianensis* var. *canescens*, *S. guianensis* var. *microcephala*, and *S. guianensis* var. *pauciflora*.

#### Potential application of EST-SSR markers in stylo

In recent years, an increasing number of SSR markers as an important tool for genetic improvement and breeding have been developed and extensively applied in phylogenetic analysis, genetic diversity analysis, genetic linkage mapping, and MAS breeding in different plants. In this investigation, large-scale development of EST-SSR markers was performed using next-generation sequencing technology in *S. guianensis*. A total of 2008 EST-markers were developed, of which 115 markers were selected to evaluate the quality and polymorphism of the EST-SSR markers newly developed in this study. A total of 96 markers successfully amplified their target sequences (83.48 % success ratio), and 81 markers showed polymorphisms with 2–6 alleles among 29 accessions (70.43 % polymorphism ratio). In addition, the 81 polymorphic EST-SSR markers facilitated the determination of the genetic diversity and relationships among the 29 *S. guianensis* genotypes. The set of EST-SSR markers developed in this study will facilitate the global genetic improvement of *Stylosanthes* species.

The expressed sequences are more conserved compared with noncoding sequences, so EST-SSR markers demonstrate a higher level of transferability across related species compared with genomic SSR markers (Ellis and Burke 2007). A total of 139 SSR markers were reported from different species, so lesser SSR markers were applicable to each species in stylo genus. The large-scale EST-SSR markers developed from *S. guianensis* in this study will be used as general molecular markers among other stylo species. However, given their association with coding sequences, EST-SSR markers can be used for direct gene tagging

for certain important agronomic traits. When EST-SSRs are located in coding sequences of functional genes, the corresponding EST-SSR markers will be closely related with certain important agronomic traits (Varshney et al. 2005). These EST-SSR markers will be useful for selecting and pyramiding agriculturally valuable alleles in stylo species. In this study, 115 EST-SSR markers were developed and characterized in coding regions of genes. Of these 115 EST-SSR markers, 81 were polymorphic, with 2–6 alleles among the 29 stylo accessions analyzed (Table 3). The EST-SSR markers developed in this study will provide a valuable resource for genetic mapping, QTL analysis for important traits, and MAS breeding in stylo species.

**Acknowledgments** This work was supported by the Natural Science Foundation of Hainan Province (313067), the Fundamental Scientific Research Funds for CATAS-TCGRI (1630032013002), and Modern Agriculture Talents Program.

#### References

- Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Singh L (2007) Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. *Theor Appl Genet* 114:359–372
- Billotte N, Lagoda PJ, Risterucci AM, Baurens FC (1999) Microsatellite-enriched libraries: applied methodology for the development of SSR markers in tropical crops. *Fruits* 54:277–288
- Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314–331
- Brandão M, Costa NM, Schultze-Kraft R (1985) Pauciflora: uma nova variedade de *Stylosanthes guianensis* (Aub.) Sw. *Anais do 36 Congresso Nacional de Botanica, Curitiba: Sociedade Botanica do Brasil* 235–241
- Bushman BS, Larson SR, Tuna M, West MS, Hernandez AG, Vullaganti D, Gong G, Robins JG, Jensen KB, Thimmapuram J (2011) Orchardgrass (*Dactylis glomerata* L.) EST and SSR marker development, annotation, and transferability. *Theor Appl Genet* 123:119–129
- Calles T, Schultze-Kraft R (2010) *Stylosanthes falconensis* (Leguminosae, Papilionoideae, Dalbergieae), a new species endemic to Venezuela. *Kew Bull* 65:73–76
- Chandra A (2006) A partial genetic map of *Stylosanthes* developed using inter-specific F<sub>2</sub> cross. *Cytologia* 71:353–358
- Chandra A, Tiwari KK, Nagaich D, Dubey N, Kumar S, Roy AK (2011) Development and characterization of microsatellite markers from tropical forage *Stylosanthes* species and

- analysis of genetic variability and cross-species transferability. *Genome* 54:1016–1028
- Chen S, Huang X, Yan X, Liang Y, Wang Y, Li X, Peng X, Ma X, Zhang L, Cai Y et al (2013) Transcriptome analysis in sheepgrass (*Leymus chinensis*): a dominant perennial grass of the Eurasian Steppe. *PLoS ONE* 8:e67974
- Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theor Appl Genet* 100:713–722
- Costa N (2006) Revisão do género *Stylosanthes*. PhD thesis, Universidade Técnica de Lisboa, Portugal, 470 pp
- Dutta S, Kumawat G, Singh BP, Gupta DK, Singh S, Dogra V, Gaikwad K, Sharma TR, Raje RS, Bandhopadhy TK et al (2011) Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea (*Cajanus cajan* (L.) Millspaugh). *BMC Plant Biol* 11:17
- Ellis JR, Burke JM (2007) EST-SSRs as a resource for population genetic analyses. *Heredity* 99:125–132
- Eujayl I, Sorrells M, Baum M, Wolters P, Powell W (2001) Assessment of genotypic variation among cultivated durum wheat based on EST-SSRs and genomic SSRs. *Euphytica* 119:39–43
- Huang C, Liu G, Bai C, Wang W, Tang J (2014) Application of SRAP markers in the identification of *Stylosanthes guianensis* hybrids. *Mol Biol Rep* 41:5923–5929
- Kalia RK, Mai MK, Kalia S, Singh R, Dhawan AK (2011) Microsatellite markers: an overview of the recent progress in plants. *Euphytica* 177:309–334
- Kantety RV, La Rota M, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol Biol* 48:501–510
- Kumpatla SP, Mukhopadhyay S (2005) Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. *Genome* 48:985–998
- Liu CJ, Musial JM, Thomas BD (1999) Genetic relationships among *Stylosanthes* species revealed by RFLP and STS analyses. *Theor Appl Genet* 99:1179–1186
- Liu T, Zhu S, Fu L, Tang Q, Yu Y, Chen P, Luan M, Wang C, Tang S (2013a) Development and characterization of 1,827 expressed sequence tag-derived simple sequence repeat markers for ramie (*Boehmeria nivea* L. Gaud). *PLoS ONE* 8:e60346
- Liu Z, Chen T, Ma L, Zhao Z, Zhao PX, Nan Z, Wang Y (2013b) Global transcriptome sequencing using the Illumina platform and the development of EST-SSR markers in autotetraploid alfalfa. *PLoS ONE* 8:e83549
- Maass BL, Sawkins M (2004) History, relationships and diversity among *Stylosanthes* species of commercial significance. In: Chakraborty S (ed) High-yielding anthracnose-resistant *Stylosanthes* for agricultural systems. Australian Centre for International Agricultural Research (ACIAR), Canberra, pp 9–26
- Mace ES, Varshney RK, Mahalakshmi V, Seetha K, Gafoor A, Leeladevi Y, Crouch JH (2008) In silico development of simple sequence repeat markers within the aescynomenoid/dalbergoid and genistoid clades of the *Leguminosae* family and their transferability to *Arachis hypogaea*, groundnut. *Plant Sci* 174:51–60
- Nagaich D, Chandra A (2009) Assessment of genetic diversity and identification of informative molecular markers for germplasm characterization in Caribbean Stylo (*Stylosanthes hamata*). *J Plant Biochem Biotechnol* 18:257–260
- Nagaich D, Tiwari KK, Srivastva N, Chandra A (2013) Assessment of genetic diversity and morpho-physiological traits related to drought tolerance in *Stylosanthes scabra*. *Acta Physiol Plant* 35:3127–3136
- Rohlf FJ (2000) NTSYS-pc. Numerical taxonomy and multivariate analysis system, version 2.10. Exeter Software, New York
- Santos MO, Karia CT, Resende RM, Chiari L, Jungmann L, Zucchi MI, Souza AP (2009a) Isolation and characterization of microsatellite loci in the tropical forage legume *Stylosanthes guianensis* (Aubl.) Sw. *Conserv Genet Resour* 1:43–46
- Santos MO, Sasaki RP, Ferreira TH, Resende RM, Chiari L, Karia CT, Faleiro FG, Jungmann L, Zucchi MI, Souza AP (2009b) Polymorphic microsatellite loci for *Stylosanthes macrocephala* Ferr. et Costa, a tropical forage legume. *Conserv Genet Resour* 1:481–485
- Santos MO, Sasaki RP, Chiari L, Resende RM, Souza AP (2009c) Isolation and characterization of microsatellite loci in tropical forage *Stylosanthes capitata* Vogel. *Mol Ecol Resour* 9:192–194
- Santos-Garcia MO, Resende RM, Chiari L, Zucchi MI, Souza AP (2011) Mating systems in tropical forages: *Stylosanthes capitata* Vog. and *Stylosanthes guianensis* (Aubl.) Sw. *Euphytica* 178:185–193
- Santos-Garcia MO, Karia CT, Resende RM, Chiari L, Vieira ML, Zucchi MI, Souza AP (2012a) Identification of *Stylosanthes guianensis* varieties using molecular genetic analysis. *AoB Plants* 2012:pls001
- Santos-Garcia MO, Toledo-Silva G, Sasaki RP, Ferreira TH, Resende RM, Chiari L, Karia CT, Carvalho MA, Faleiro FG, Zucchi MI et al (2012b) Using genetic diversity information to establish core collections of *Stylosanthes capitata* and *Stylosanthes macrocephala*. *Genet Mol Biol* 35:847–861
- Tan C, Wu Y, Taliaferro CM, Anderson MP, Tauer C, Samuels T (2012) Development of simple sequence repeat markers for bermudagrass from its expressed sequence tag sequences and preexisting sorghum SSR markers. *Mol Breed* 29:23–30
- Tan LQ, Wang LY, Wei K, Zhang CC, Wu LY, Qi GN, Cheng H, Zhang Q, Cui QM, Liang JB (2013) Floral transcriptome sequencing for SSR marker development and linkage map construction in the tea plant (*Camellia sinensis*). *PLoS ONE* 8:e81611
- Tang Y, Wu Z, Liu G, Yi K (2009) Research advances in germplasm resources of *Stylosanthes*. *Chin Bull Bot* 44:752–762
- Thumma BR, Naidu BP, Chandra A, Cameron DF, Bahnisch LM, Liu C (2001) Identification of causal relationships among traits related to drought resistance in *Stylosanthes scabra* using QTL analysis. *J Exp Bot* 52:203–214
- Vander Stappen J, Weltjens II, Volckaert G (1999) Microsatellite markers in *Stylosanthes guianensis*. *Mol Ecol* 8:514–517
- Vander Stappen J, Weltjens I, Gama Lopez S, Volckaert G (2000) Genetic diversity in Mexican *Stylosanthes humilis*

- as revealed by AFLP, compared to variability of *S. humilis* accessions of South American origin. *Euphytica* 113:145–154
- Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. *Trends Biotechnol* 23:48–55
- Wang Z, Li J, Luo Z, Huang L, Chen X, Fang B, Li Y, Chen J, Zhang X (2011) Characterization and development of EST-derived SSR markers in cultivated sweetpotato (*Ipomoea batatas*). *BMC Plant Biol* 11:139
- Wang S, Wang X, He Q, Liu X, Xu W, Li L, Gao J, Wang F (2012) Transcriptome analysis of the roots at early and late seedling stages using Illumina paired-end sequencing and development of EST-SSR markers in radish. *Plant Cell Rep* 31:1437–1447
- Wang Z, Yu G, Shi B, Wang X, Qiang H, Gao H (2014) Development and characterization of simple sequence repeat (SSR) markers based on RNA-sequencing of *Medicago sativa* and in silico mapping onto the *M. truncatula* genome. *PLoS ONE* 9(3):e92029
- Wu J, Cai C, Cheng F, Cui H, Zhou H (2014) Characterisation and development of EST-SSR markers in tree peony using transcriptome sequences. *Mol Breed* 34:1853–1866
- Xin D, Sun J, Wang J, Jiang H, Hu G, Liu C, Chen Q (2012) Identification and characterization of SSRs from soybean (*Glycine max*) ESTs. *Mol Biol Rep* 39:9047–9057
- Yadav HK, Ranjan A, Asif MH, Mantri S, Sawant SV, Tuli R (2011) EST-derived SSR markers in *Jatropha curcas* L.: development, characterization, polymorphism, and transferability across the species/genera. *Tree Genet Genomes* 7:207–219
- Yeh FC, Yang RC, Boyle T (1999) POPGENE, version 1.32: the user friendly software for population genetic analysis. Molecular Biology and Biotechnology Centre, University of Alberta, Edmonton
- Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Mol Ecol* 11:1–16