Plant Omics Journal POJ

POJ 8(1):17-23 (2015) ISSN:1836-3644

Genomic sequencing using 454 pyrosequencing and development of an SSR primer database for *Lagerstroemia indica* L.

Jing Wang[1§], Xiaogang Dai[1§], Yingnan Chen[1*], Yanling Yang[2], Xinye Zhang[2], Shuxian Li[1], Tongming Yin[1]

[1]The Southern Modern Forestry Collaborative Innovation Centre, Nanjing Forestry University, Nanjing 210037, China
[2]Hubei Forestry Academy, Wuhan 430075, China

[§]These authors contributed equally to this work.
***Corresponding author: chenyingnan@njfu.edu.cn**

Abstract

Crape myrtles (*Lagerstroemia* spp.) represent a large group of woody flowing plants. Despite their high ornamental value and popularity, few genomic sequences and marker resources are available for them. *Lagerstroemia indica* is one of the most widely cultivated crape myrtle species. In this study, we partially sequenced the genome of *L. indica* using newly updated 454 sequencing technology. Over 1.2 million high-quality reads in a total length of 837.4 Mb were generated. The average read length was 679 bp. Of the reads, 779,744 (63.2%) were assembled into 65,129 contigs covering a physical length of 93.6 Mb and with N50 contig size of 1,648 bp. The contigs were used to recover microsatellites with repeat motifs of 1-6 bp. A total of 33,026 microsatellites were detected. An SSR primer database was established based on the flanking sequences of the detected microsatellites. A PCR survey of subset of these SSR primers revealed that 89.5% amplified successfully，and 66.7% of the loci were polymorphic. The polymorphic information contents of the polymorphic SSRs ranged from 0.08 to 0.79, with an average value of 0.44. This study provided valuable genomic sequences and marker resources for future genetic studies on *Lagerstroemia* species.

Keywords: Crape myrtle; genome; 454 pyrosequencing; microsatellite.

Introduction

Lagerstroemia (crape myrtles; Lythraceae) represents more than 50 species of deciduous shrubs or small trees (Brickell, 1996) that occur in temperate regions worldwide (Cai et al., 2011). They are valued in landscaping for their long summer flowering period, diverse flower colors, and ease of cultivation (Pooler, 2006). The trunks and branches are also highly attractive because of the gnarled sinuous shape and smooth bark. Among these ornamental plants, *Lagerstroemia indica* is one of the most widely cultivated species; it commonly exists in the diploid form with 2n=48 (He et al. 2014). Surveyed with *L. tomentosa*, a sister species of *L. indica*, genome size of the *Lagerstroemia* species was estimated to be up to 965 Mb (Bennett et al., 2005). Past research efforts on crape myrtle have mainly focused on traditional cross-breeding, evaluating germplasm resources, and cultivation (Pooler 2003, Yang et al. 2004, Pounders et al. 2006, Zhang et al. 2008). However, genetic studies of crape myrtle have lagged behind those of many other plants because of the lack of genomic sequences and genetic tools.

Microsatellites, or simple sequence repeats (SSRs), are important signatures in higher plant genomes. They are abundant and widely distributed in the genomes of all eukaryotes (Tóth et al. 2000). SSR loci have higher mutability than DNA sequences in the other genomic regions (Li et al. 2002) and are consequently thought to play an important role in genome evolution (Kashi et al. 1997). Additionally, SSR primers are highly transferable within species and even within genera (Castillo et al. 2008; Yin et al. 2009). Thus, microsatellites are the most powerful molecular markers for genetic analyses (Zalapa et al., 2012) and have been increasingly employed in studies on crape myrtle, including cultivar identification (Cai et al., 2011), inter-specific hybrid analysis (Pounders et al., 2007), genetic diversity assessment (Rinehart and Pounders, 2008; He et al., 2012), and genetic map construction (He et al., 2014). However, known SSR markers are very scarce for crape myrtle. Although several studies have focused on developing SSR markers for crape myrtle by conventional library–based approaches (Pounders et al., 2007; Wang et al., 2010; Cai et al., 2011; Liu et al., 2013), to date, relatively few SSR primer pairs (180 in total) have been reported and used in genetic studies of *Lagerstroemia* spp. (Pounders et al., 2007; Wang et al., 2010, 2011; Cai et al., 2011; He et al., 2012; Liu et al., 2013; He et al., 2014). Development of SSR markers for a focal organism requires genomic sequences, the availability of which is very limited for *Lagerstroemia*. As of September 2014, only 418 DNA and RNA sequences had been deposited at NCBI for members of the genus.

Recent advances of massive parallel sequencing technologies, known as next generation sequencing (NGS), have drastically accelerated the collection of genome sequences in non-model plants. Very recently, Zhang et al. (2014) generated a large set of transcriptome sequences for *L. indica* using the Hiseq 2000 (Illumina, San Diego, CA, USA), which typically produces sequences of 100bp in length. In comparison, the newly updated 454 GS XL+ Titanium

pyrosequencer (454 Life Sciences, Branford, CT, USA) can generate sequence reads with an average length up to 800 bp. Long reads can significantly improve the quality of sequence assemblies, promoting the identification of large numbers of microsatellites in non-model plants. Several studies have demonstrated the efficiency of 454 pyrosequencing in the large-scale discovery of microsatellites in species such as cucumber (Cavagnaro et al., 2010), cranberry (Zhu et al., 2012), faba bean (Yang et al., 2012) and Masson pine (Bai et al., 2014). Given the need for genomic sequences and marker resources for crape myrtle, the objectives of this study were to (1) enrich the genomic sequences for crape myrtle using the newly updated 454 pyrosequencing technology; (2) characterize the genomic SSRs in *L. indica* based on the sequence assemblies; (3) establish a large SSR primer database for *Lagerstroemia*; and (4) experimentally evaluate the utility of the established SSR primers. This study provides a rich set of genomic sequences and marker resources for future studies on *Lagerstroemia* spp.

Results

Genome sequencing and assembly

A total of 1,234,265 raw reads were generated from the 454 GS FLX Titanium platform (454 Life Sciences) in total bases of 1.16 Gb. The sequencing depth was ca. 1.2× coverage of the crape myrtle genome. After removing adaptors and low-quality reads, we obtained 1,233,727 high-quality reads with a total length of 837.4 Mb. The average read length was 679 bp, and the maximum length was 1,774 bp. The resulting high-quality reads were assembled using Newbler V2.8 (Roche Diagnostics, Basel, Switzerland). Finally, 650,035 reads were assembled into 92,637 contigs with a total length of 101.27 Mb and N50 contig size of 1,648 bp (Table 1). There were 309,006 singleton reads. Among the assembled contigs, 65,169 were larger than 500 bp, with an average size of 1,438 bp. These large contigs covered 93.7 Mb of the crape myrtle genome and were further used for microsatellite mining.

SSR detection and characterization

A total of 33,026 microsatellites were identified from 20,493 contigs (31.4% of the large contigs) with MISA (http://www.pgrc.ipkgatersleben.de/misa). Among the contigs, 6,305 (9.67%) contained more than one SSR marker. The average microsatellite density was 352.5 SSRs/Mb. The microsatellites averaged 19.7 bp in length (range: 12-271 bp), and 79% were shorter than 20 bp (Fig. 1). Among the identified SSRs, tetranucleotide repeat motifs were most abundant, accounting for approximately 26.6% of all SSRs, followed by di- (25.0%) and trinucleotides (22.3%) (Table 2). Taken together, the mono-, penta- and hexa-nucleotide repeat types represented 16.9% of all SSRs. In addition, 2,982 compound SSRs, accounting for 9.0% of the total, were identified. The number of different types of microsatellites decreased sharply as microsatellite length increased (Fig. 2).

We further analyzed the nucleotide compositions of the repeat motifs in each type of microsatellite. In the mononucleotide repeats, the A/T motif was the dominant repeat unit (97.9%), whereas the G/C repeats only accounted for 2.1% of the total. As for the dinucleotide repeats, the most common repeat unit was AT/AT (49.2%) (Fig. 3). Among the trinucleotide repeats, AAG/CTT repeats were most frequent (accounting for 28.0% of all trinucleotide repeats), followed by AAT/ATT (24.7%) and AGG/CCT (14.5%) repeats. Notably, tetranucleotide repeats occurred at the highest frequency in the crape myrtle genome, and the AT-rich motifs (AAAT/ATTT, AAAG/TTTC, AATT/TTAA, ACAT/TGTA, AATC/TTAG, AAAC/TTTG, AATG/TTAC, and AGAT/TCTA) accounted for 70.4 % of the tetranucleotide repeats. AAAAT/TTTTA repeats were the most abundant pentanucleotide repeats (18.1%), and AT-rich repeat units (AAAAT/TTTTA, AAAAG/TTTTC, AAATT/TTTAA, AATAT/TTATA, AAATC/TTTAG, AATGG/TTACC) also constituted a great part of the pentanucleotide repeats (41.8%). Each of the different hexanucleotide repeat units occurred at a very low frequency.

Experimental analysis of SSR primers

To test PCR amplification success, we randomly selected and synthesized 48 primer pairs from our SSR primer database, 43 of which amplified successfully. We further used the 43 primer pairs to evaluate SSR polymorphism in 24 crape myrtle cultivars. Thirty-two primer pairs generated polymorphic bands, with 2-8 alleles per locus and an average of 3.9 (Table 3). Fragment size varied from 146 bp to 349 bp. The polymorphic information content (PIC) values of the tested primers ranged from 0.08 to 0.79, with an average value of 0.44.

Based on the allelic profiles generated by the polymorphic microsatellite primer pairs, we analyzed the genetic relationships of 24 crape myrtle samples. Cluster analysis grouped these cultivars into two major groups (**Fig. 4**): cluster I contained the 23 *L. indica* specimens, and cluster II contained only ZX6, a member of *Lagerstroemia limii*. Cluster I could be further divided into two major subgroups: group Ia with 15 samples; and group Ib comprising eight samples.

Discussion

The relative richness of different types of microsatellites is a prominent characteristic of a taxon. In this study, we found that tetranucleotide microsatellite repeats were the most frequent type in the crape myrtle genome. Previous studies revealed that the dominant type of SSR may vary among species. For instance, tetranucleotide repeats were also most frequent in the genomes of grapevine (Cavagnaro et al., 2010), cucumber (Cavagnaro et al., 2010), and oil camellia (Shi et al., 2013). While mononucleotide repeats predominated in the *Medicago* genome (Sonah et al., 2011) and dinucleotide repeats in the genomes of Papaya and *Prunus persica* (Shi et al., 2013). Trinucleotide repeats dominated the detected microsatellites of other species, such as faba bean (Yang et al., 2012), *Arabidopsis*, rice, and sorghum (Cavagnaro et al. 2010). In crape myrtle, AT-rich motifs occurred at a high frequency, as is commonly observed in many other dicot plants, including cucumber (Cavagnaro et al. 2010) and oil camellia (Shi et al., 2013), but not in monocots. This fact may be explained partially by the higher AT content in dicot genomes (Cavagnaro et al. 2010). The AT content in crape myrtle genome was 60.8%.

Experimental analysis of 48 SSR primer pairs in this study showed a higher rate of successful amplification (89.6%) than reported for other plants, such as cucumber (73.5%, Cavagnaro et al. 2010), faba bean (68%, Yang et al., 2012) and Masson pine (50%, Bai et al., 2014). This success rate may reflect the high quality of the sequences and genome assembly. With the updated 454 sequencing platform, we

Table 1. Parameters for the genomic sequence assemblies of crape myrtle.

Large contig (Length≥500bp)							All contigs (Length≥100bp)	
Contigs	Bases	ACS[1]	N50 contig size[2]	Largest contig Size	Q40 plus bases[3]	%Q40[4]	Contigs	Bases
65,169	93,728,145	1,438	1,648	50,509	85,297,645	91.01%	92,637	101,266,799

[1]average size of contigs longer than 500bp. [2] half of contigs of ACS or longer. [3] number of bases with a quality score > 40 among large contigs. [4] percentage of bases with a quality score > 40 among large contigs

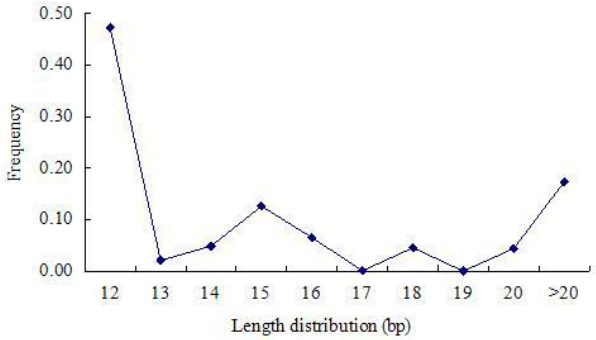

Fig 1. Length distribution and frequency of microsatellites in crape myrtle. The graph was drawn based on a total of 33,026 SSRs identified in 93.7Mb of crape myrtle genomic DNA.

Table 2. Classification of and statistics on the detected genomic SSRs.

Repeat type	Number of SSR	Percentage (%)	Density (number·Mb^{-1})	SSR types	Max. number of repeats
Mononucleotide	2,422	7.3	25.8	2	33
Dinucleotide	8,271	25.0	88.3	4	33
Trinucleotide	7,372	22.3	78.7	10	20
Tetranucleotide	8,783	26.6	93.7	33	15
Pentanucleotide	3,153	9.5	33.6	99	8
Hexanucleotide	43	0.1	0.5	38	8
Compound	2,982	9.0	31.8	2982	2
Total	33,026	100.0	352.5	3168	-

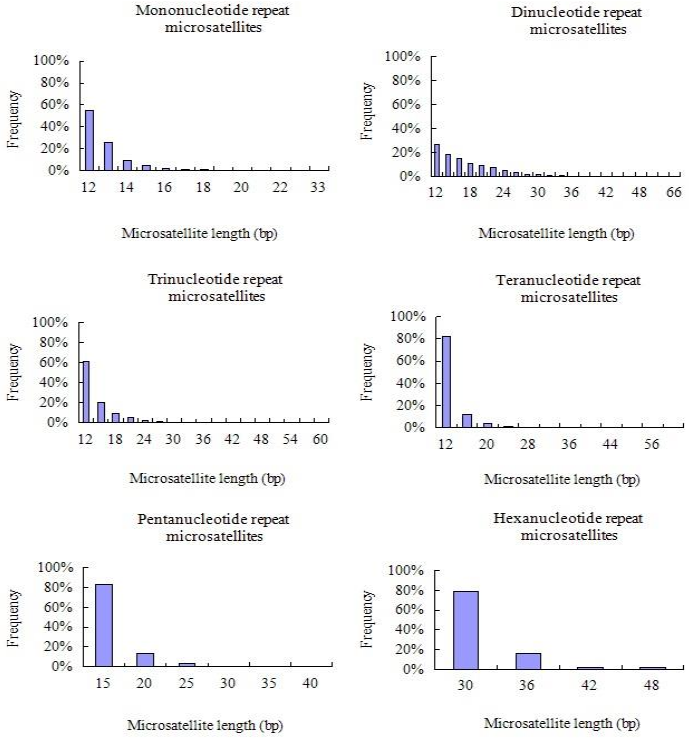

Fig 2. Length distributions of different types of microsatellites in crape myrtle. The frequency was calculated as the ratio of the number of SSRs of a given length divided by the total number SSRs in each category.

Table 3. Genetic diversity parameters of primers screened using 24 crape myrtle cultivars.

Locus	Repeat motif	Repeat length	Allele No.	PIC	Locus	Repeat motif	Repeat length	Allele No.	PIC
ZW1	4	12	3	0.31	ZW21	2	12	5	0.62
ZW3	4	12	2	0.19	ZW23	4	12	3	0.29
ZW4	2	22	7	0.49	ZW24	4	12	3	0.60
ZW5	2	18	5	0.68	ZW25	2	14	2	0.30
ZW7	2	14	3	0.44	ZW26	4	12	2	0.04
ZW8	3	15	5	0.50	ZW29	2	12	3	0.40
ZW9	3	12	2	0.50	ZW30	4	12	2	0.22
ZW10	4	12	2	0.08	ZW32	4	12	3	0.52
ZW11	4	12	2	0.15	ZW34	2	18	4	0.55
ZW12	3	15	8	0.79	ZW36	3	12	6	0.63
ZW14	3	12	4	0.64	ZW38	2	14	6	0.51
ZW15	2	12	5	0.45	ZW39	2	14	4	0.46
ZW16	2	14	6	0.69	ZW42	4	12	3	0.19
ZW17	4	12	4	0.56	ZW45	4	12	3	0.54
ZW18	3	24	5	0.58	ZW46	2	14	6	0.18
ZW20	2	12	3	0.47	ZW48	3	12	3	0.51

Note: PIC, polymorphic information content.

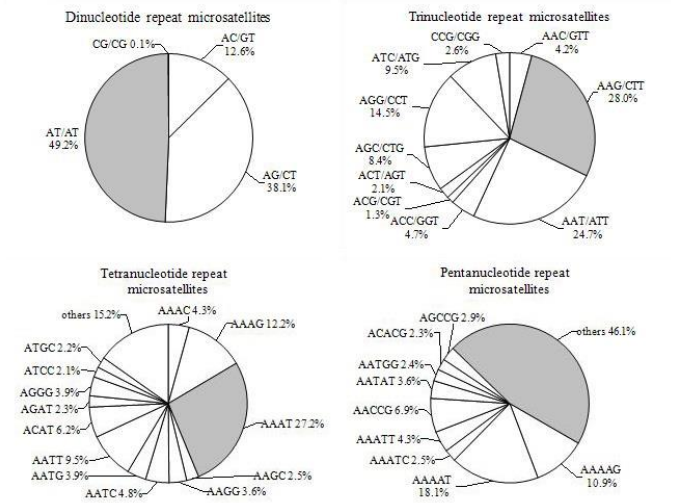

Fig 3. Proportions of different SSR types. Different microsatellite repeats are depicted in separate slices. For percentage ≤ 0.02, the corresponding slices were combined.

obtained an average read length of 679 bp, which is comparable to Sanger sequencing. Longer reads can significantly improve the sequence assembly, thus increasing the probability of developing successful SSR primers. In addition, the ratio of polymorphic SSRs (66.7%) was higher than those reported in other works (27.9%, Wang et al., 2010; 36.4%, Cai et al., 2011; 20.4%, Liu et al., 2013) in which microsatellites for crape myrtle were developed from SSR-enriched genomic libraries. Conventional library–based approaches produce SSRs with specific repeat units. In contrast, the repeat motifs developed in this study were more varied. Thus, given the numerous SSRs and the higher ratio of polymorphic primers, this study demonstrated that 454 sequencing technology is highly efficient for developing microsatellite primers. However, the average number of alleles per SSR locus (3.9) was lower than in previous reports (5.58, Wang et al. 2010; 5.58, Cai et al., 2011; 6.81, Liu et al., 2013), perhaps because relatively few individuals was sampled in this study or because the tested samples were closely related. Cluster analysis showed that ZX6 diverged from the other *L. indica* cultivars, similar to the result of Wang et al. (2011). Unlike in He et al.'s study (2012), our dendrogram indicated that the genetic relationships of the analyzed samples did not generally correspond with their geographic origins. For the samples in cluster I, only a few from the same area—such as HF1, HF2 and HF5 from Hefeng; HN2 and HN3 from Hunan; and B78 and B91 from Baokang—clustered together. Crape myrtles are important landscaping plants and are frequently introduced from one place to another; this commercial transportation confounds the origination of individual cultivars. This hypothesis was supported by the clustering results for samples from the U.S.A (M2, M5, M8, ME6), which were not closely clustered together but had closer relatives from China (**Fig. 4**).

Materials and Methods

Plant materials

All plant materials were sampled from a crape myrtle germplasm nursery that held 105 accessions from eight different places. This germplasm collection was maintained at Jiufeng Forest Farm, Wuhan, China. Accession M1-1, an elite genotype that has been widely used in several of our breeding programs, was selected for sequencing. Twenty-four accessions were selected to survey the utility of the developed

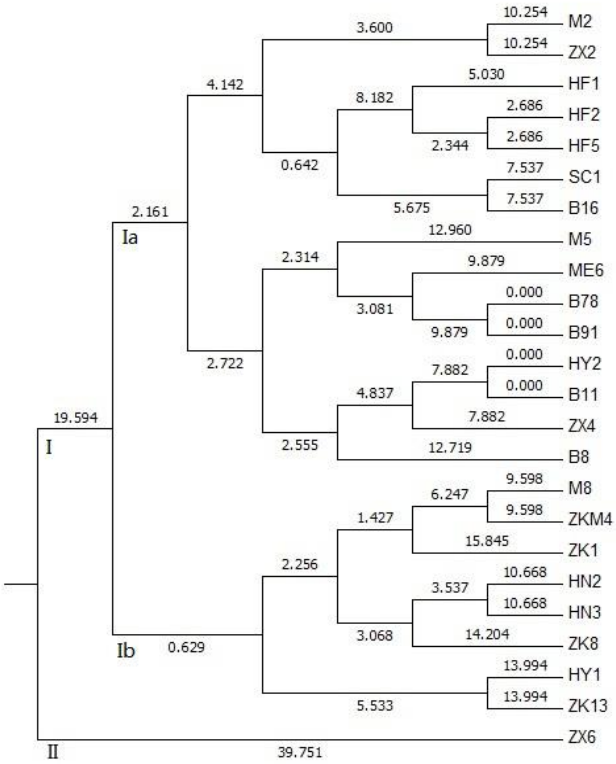

Fig 4. Cluster analysis of 24 crape myrtle cultivars.

SSR primers. The number of selected trees from each geographical origin was determined by $n \times \frac{24}{105}$, where n was the total number of trees from the same place in this germplasm. The originations and accession numbers of the sampled individuals are listed in Table S1. In April, 2013, young leaves were collected for DNA extraction.

DNA isolation and sequencing library construction

Total genomic DNA was extracted from each sample using a DNeasy Plant Mini kit (Qiagen, Valencia, CA). Quality and quantity of DNA were checked using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE) and agarose gel electrophoresis. The extracted DNA was dissolved in 10× EB buffer (Qiagen) and stored at −80°C until use.

For genome sequencing, approximately 2 μg of DNA from M1-1 was fragmented by nebulization using compressed nitrogen gas according to the GS-FLX+ library preparation protocol (Roche, Indianapolis, IN, USA). Both ends of the DNA fragments were blunt-ended and ligated to DNA adaptors (Roche). DNA fragments larger than 350 bp were then purified using AMPure beads (Beckman Coulter, Brea, CA, USA) and sizing solution (Roche). Quality and quantity of the library was evaluated by using an Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and TBS 380 Fluorometer (Turner Biosystems, Vernon Hills, IL, USA), respectively. Finally, the library was diluted into 1×10^7 molecules/μL in TE Buffer (Roche).

Shotgun sequencing and assembly

The constructed library was titrated by small volume emulsion PCR (emPCR) and then pooled into two libraries for large volume emPCR (Christensen et al., 2013). Amplified products were then denatured to generate single-stranded DNA that was sequenced on a 454 GS-FLX Sequencer (454 Life Sciences). One full run of genomic pyrosequencing was performed with the newly updated XL+ Titanium chemistry kit (Roche) and two-region gasket according to the manufacturer's standard protocol. The XL+ Titanium kit increased the average read length to 800 bp.

After filtering low-quality reads (quality score ≤ 39) and reads less than 50bp in length, the remaining high-quality reads were assembled using Newbler V2.8 (Roche) with the following settings: seed step, 12; seed length, 16; seed count, 1; minimum overlap length, 40; minimum overlap identity, 95%; alignment identity score, 2; alignment difference score, −3 (Chen et al., 2014). The assembled sequences were deposited in the "Crape myrtle assembly" database archived at http://115.29.234.170/Database/Crape_myrtle.

Microsatellite detection and primer development

The Perl script MISA (http://www.pgrc.ipkgatersleben.de/misa) was used to identify microsatellites in the assembled genomic sequences. The minimum number of repeat units was 12 for mononucleotides, six for dinucleotides, four for trinucleotides, three for tetra- and pentanucleotides, and five for hexanucleotides. For compound repeat types, the maximum size of the interruption between two different SSR loci was 100 bp. SSR primers were designed using Primer Premier 5.0 software (Premier Biosoft International, Palo Alto, CA, USA). The expected amplicons ranged from 100 bp to 300 bp. The primer sequences were deposited in the "SSR primer database of crape myrtle" archived at: http://115.29.234.170/Database/Crape_myrtle.

PCR conditions

We randomly selected and synthesized 48 SSR primers from our database to experimentally evaluate this resource. PCR was carried out in a volume of 15μL, including 30ng DNA template, 1.5μL 10× PCR buffer (Mg^{2+} plus), 0.15μL BSA (10mM), 0.4μL dNTP (2.5mM), 1μL dUTP (2.5μM), 1U Taq DNA polymerase (Lifefeng, Shanghai, China), and 2μL of primer mix (forward and reverse,10μM). PCR was performed in an ABI-9700 Thermocycler (Applied Biosystems, Foster City, CA, USA) using a touchdown program with initial denaturation at 94°C for 4 min; then 94°C for 30s, 60°C for 30s, and 72°C for 30s, followed by 10 cycles at decreasing annealing temperatures of 1°C per cycle; 25 cycles of 30s at 94°C, 30s at 50°C, 30s at 72°C, and a final extension at 72°C for 10min.

Polymorphism evaluation and cluster analysis

Allele scoring was performed using GeneMapper v3.7 (Applied Biosystems). Polymorphic differences among the tested individuals were used to calculate the PIC associated with each SSR marker using the following formula: $PIC=1-\sum_{i=1}^{n} P_i^2$, where n is the total number of alleles detected for a SSR marker, and P_i is the frequency of the *i*th allele (Smith et al., 1997).

The program POPGENE (V1.31, Yeh et al., 1997) was used to estimate Nei's genetic distance (1972). Cluster analysis was conducted with a UPGMA algorithm, using the NEIGHBOR program in PHYLIP (V3.5c, Felsenstein, 1989); and a dendrogram was constructed using the program MEGA6 (Tamura et al., 2013).

Conclusions

In the present study, a large number of SSR markers were developed for *L. indica* using the newly updated 454 sequencing technology, which produced sequence reads with an average length comparable to that of Sanger sequencing. Based on sequence assemblies, we established a database containing numerous SSR primers, far more the total number of microsatellite markers reported thus far. An experimental survey revealed a high amplification successful rate and polymorphic ratio for the derived SSRs. These novel SSR markers, together with the extensive genome sequence information, will greatly accelerate genetic studies on crape myrtles, such as cultivar identification, genetic linkage map construction, QTL mapping, and molecular breeding.

Acknowledgments

This work was funded by the 948 Project (2014-4-13) and projects of Hubei province, China (2008CDA076 and 201020622240). This research was also supported by Jiangsu Planned Projects for Postdoctoral Research Funds (1402011B), the Program for Innovative Research Team in Universities of Jiangsu Province, the Educational Department of China, and the Priority Academic Program Development (PAPD) Program at Nanjing Forestry University, Nanjing, China.

References

Bai TD, Xu LA, Xu M, Wang ZR (2014) Characterization of masson pine (*Pinus massoniana* Lamb.) microsatellite DNA by 454 genome shotgun sequencing. Tree Genet Genomes. 10: 429-437.

Bennett MD, Leitch IJ (2005) Nuclear DNA amounts in angiosperms: progress, problems and prospects. Ann Bot. 95: 45-90.

Brickell C (1996) Encyclopedia of garden plants. Macmillan Press, New York, 250-252.

Cai M, Pan HT, Wang XF, He D, Wang XY, Wang XJ, Zhang QX (2011) Development of novel microsatellites in *Lagerstroemia indica* and DNA fingerprinting in Chinese *Lagerstroemia* cultivars. Sci Hortic. 131: 88-94.

Castillo A, Budak H, Varshney RK, Dorado G, Graner A, Hernandez P (2008) Transferability and polymorphism of barley EST-SSR markers used for phylogenetic analysis in *Hordeum chilense*. BMC Plant Biol. 8: 97.

Cavagnaro PF, Senalik DA, Yang L, Simon PW, Harkins TT, Kodira CD, Huang S, Weng Y (2010) Genome-wide characterization of simple sequence repeats in cucumber (*Cucumissativus* L.). BMC Genomics. 11: 569.

Chen Y, Mao Y, Liu H, Yu F, Li S, Yin T (2014) Transcriptome analysis of differentially expressed genes relevant to variegation in peach flowers. PloS One. 9: e90842.

Christensen KA, Brunelli JP, Lambert MJ, DeKoning J, Phillips RB, Thorgaard GH (2013) Identification of single nucleotide polymorphisms from the transcriptome of an organism with a whole genome duplication. BMC Bioinformatics. 14: 325.

Felsenstein J (1989) PHYLIP-Phylogeny Inference Package. Cladistics. 5: 164-166

He D, Liu Y, Cai M, Pan HT, Zhang QX, Wang XY, Wang XJ (2012) Genetic diversity of *Lagerstroemia* (Lythraceae) species assessed by simple sequence repeat markers. Genet Mol Res. 11: 3522-3533.

He D, Liu Y, Cai M, Pan HT, Zhang QX (2014) The first genetic linkage map of crape myrtle (*Lagerstroemia*) based on amplification fragment length polymorphisms and simple sequence repeats markers. Plant Breeding. 133: 138-144.

Kashi Y, King D, Soller M (1997) Simple sequence repeats as a source of quantitative genetic variation. Trends Genet. 13:74-78.

Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol. 11: 2453-2465.

Liu Y, He D, Cai M, Tang W, Li XY, Pan HT, Zhang QX (2013) Development of microsatellite markers for *Lagerstroemia indica* (Lythraceae) and related species. Appl Plant Sci. 1:1200203.

Nei M (1972) Genetic distance between populations. Am Nat. 106: 283-292.

Pooler MR (2003) Molecular genetic diversity among 12 clones of *Lagerstroemia fauriei* revealed by AFLP and RAPD markers. Hortscience. 38: 256-259.

Pooler MR (2006) 'Arapaho' and 'Cheyenne' *Lagerstroemia*. Hortscience 41: 855-856.

Pounders C, Reed SM, Pooler MR (2006) Pollination biology of *Lagerstroemia indica* and several interspecific hybrids. Hortscience. 413: 575-578.

Pounders C, Rinehart T, Sakhanokho H (2007) Evaluation of interspecific hybrids between *Lagerstroemia indica* and *L. speciosa*. Hortscience. 42:1317-1322.

Rinehart TA, Pounders CT (2008) Estimating diversity among *Lagerstroemia* species and hybrids using SSR markers. Acta Hort. 885: 285-290.

Shi J, Dai X, Chen Y, Chen J, Shi J, Yin T (2013) Discovery and experimental analysis of microsatellites in an oil woody plant *Camellia chekiangoleosa*. Plant Syst Evol. 299: 1387-1393.

Smith J, Chin E, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresovich S, Ziegle J (1997) An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPs and pedigree. Theor Appl Genet. 95: 163-173.

Sonah H, Deshmukh RK, Sharma A, Singh VP, Gupta DK, Gacche RN, Rana JC, Singh NK, Sharma TR (2011) Genome-wide distribution and organization of microsatellites in plants: an insight into marker development in *Brachypodium*. PloS one. 6: e21298.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 30: 2725-2729.

Tóth G, Gáspári Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res. 10: 967-981.

Wang XW, Dean D, Wadl P, Hadziabdic D, Scheffler B, Rinehart T, Cabrera R, Trigiano R (2010) Development of microsatellite markers from crape myrtle (*Lagerstroemia* L.). Hortscience. 45: 842-844.

Wang XW, Wadl PA, Pounders C, Trigiano RN, Cabrera RI, Scheffler BE, Pooler M, Rinehart TA (2011) Evaluation of genetic diversity and pedigree within crape myrtle cultivars using simple sequence repeat markers. J Am Soc Hortic Sci. 136: 116-128.

Yang T, Bao S, Ford R, Jia T, Guan J, He Y, Sun X, Jiang J, Hao J, Zhang X, Zong X (2012) High-throughput novel microsatellite marker of faba bean via next generation sequencing. BMC Genomics. 13: 602.

Yang YL, Zhang YD, Zhang XY, Lei XH (2004) Study on genetic diversity of wild *Lagerstroemia* from Baokang. J Huazhong Agri Univ. 23: 667-670.

Yeh FC, Yang RC, Boyle TBJ, Ye ZH, Mao JX (1997) POPGENE, the user-friendly shareware for population genetic analysis. Molecular biology and biotechnology centre, University of Alberta, Canada.

Yin TM, Zhang XY, Gunter LE, Li SX, Wullschleger SD, Huang MR, Tuskan GA (2009) Microsatellite primer resource for *Populus* developed from the mapped sequence scaffolds of the Nisqually-1 genome. New Phytol. 181: 498-503.

Zalapa JE, Cuevas H, Zhu H, Steffan S, Senalik D, Zeldin E, McCown B, Harbut R, Simon P (2012) Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. Am J Bot. 99: 193-208.

Zhang J, Wang LS, Gao JM, Shu QY, Li CH, Yao J, Hao Q, Zhang JJ (2008) Determination of anthocyanins and exploration of relationship between their composition and petal coloration in crape myrtle (*Lagerstroemia* hybrid). J Integr Plant Biol. 50: 581-588.

Zhang Z, Wang P, Li Y, Ma L, Li L, Yang R, Ma Y, Wang S, Wang Q (2014) Global transcriptome analysis and identification of the flowering regulatory genes expressed in leaves of *Lagerstroemia indica*. DNA Cell Biol. 33: 680-688.

Zhu H, Senalik D, McCown BH, Zeldin EL, Speers J, Hyman J, Bassil N, Hummer K, Simon PW, Zalapa JE (2012) Mining and validation of pyrosequenced simple sequence repeats (SSRs) from American cranberry (*Vaccinium macrocarpon* Ait.). Theor Appl Genet. 124: 87-96.