RESEARCH

CHILEANJAR
CHILEAN JOURNAL of AGRICULTURAL RESEARCH

Development and characterization of genomic simple sequence repeats for *Colocasia gigantea* (Blume) Schott using 454 sequencing

Yujing Liu[1, 3], Yu Guo[2], Deke Xing[1], and Chunlin Long[3*]

[1]Jiangsu University, College of Agricultural Equipment and Engineering, Zhenjiang 212013, PR China.
[2]BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China.
[3]Minzu University of China, College of Life and Environmental Sciences, Beijing 100081, China.
[*]Corresponding author (long@mail.kib.ac.cn).

Received: 25 July 2017; Accepted: 29 October 2017; doi:10.4067/S0718-58392018000100023

ABSTRACT

The petiole of *Colocasia gigantea* (Blume) Schott is an important agricultural and biological organ, which contains high dietary fiber, pyridoxine, and nicotinamide. However, available genomic resources of *C. gigantea* are scarce, and this restricts further genetic diversity research, linkage map construction, and marker-assisted selection in *C. gigantea*. A large-scale genomic DNA study of *C. gigantea* was conducted using the 454 sequencing technology to develop simple sequence repeats (SSRs). We identified 31 069 putative genomic *C. gigantea* SSRs, and 100 primers were randomly selected to validate their usefulness in 10 *C. gigantea* samples. The specificity of six primers yielded amplification products with expected sizes and exhibited polymorphism. The number of alleles per locus ranged from 3 to 7 alleles and the polymorphic information content (PIC) ranged from 0.561 to 0.756. The newly developed SSRs in this study should be useful tools for assessing genetic diversity, understanding population structure, and conserving and using *C. gigantea* effectively.

Key words: Microsatellite markers, 454 sequencing.

INTRODUCTION

Colocasia gigantea (Blume) Schott, a member of the monocotyledonous family Araceae, is an important food crop in Southeast Asia. It is widely planted throughout most wet tropical and subtropical regions, including Vietnam, China, and Japan, for sustainable subsistence. *Colocasia gigantea* is cultivated for its petiole, which is consumed as a vegetable (Nguyen, 2005; Ivancic et al., 2008). Sometimes individual leaves are used to make soup, and the corms and leaves are also used for medicinal purposes. *Colocasia gigantea* is now commercially cultivated in many minority regions of China and the petiole is sold in local markets. In Hawaii, besides local consumption, *C. gigantea* is exported to the rest of the United States, France, and Japan to meet the cultural nutritional needs of Vietnamese migration (Nguyen, 2005). Aroids are commonly acrid (Bown, 2000) and require plant material to be processed before it can be edible; they are usually cooked to remove acridity. Wild *C. gigantea* cannot be eaten; however, cultivated *C. gigantea* can be eaten, which suggests its high genetic diversity and long history of human selection and cultivation. Although China has a long history of consuming *C. gigantea*, the *C. gigantea* genetic resource is still scarce. Genetic studies will offer great potential to detect allelic forms of a gene and phenotypes and accelerate the progress of *C. gigantea* research.

Simple sequence repeats (SSRs) are one of the most informative and versatile DNA-based markers used in plant genetic research because of their high polymorphism, abundance, co-dominance, and short length, they include

linkage map development, quantitative trait locus (QTL) mapping, marker-assisted selection, parentage analysis, cultivar fingerprinting, genetic diversity, gene flow, and evolutionary studies (Zalapa et al., 2012). For plants without a genome sequence, developing SSR primers is a very challenging job. The traditional method of isolating microsatellites is time-consuming and cost-intensive and has a relatively low efficiency in microsatellite detection. Many studies have indicated that next-generation sequencing (NGS) is an efficient method to identify large numbers of microsatellites at a fraction of the cost and effort of traditional approaches (Angeloni et al., 2011; Ho et al., 2011; Baldwin et al., 2012; Georgi et al., 2012; Kale et al., 2012; Perry and Rowe, 2011; Sahu et al., 2012; Yang et al., 2012). The two major NGS platforms with an emergent application of SSRs are 454 and Illumina. Compared to Illumina, the advantages of 454 are longer read lengths, higher consensus accuracy, and faster run times (Egan et al., 2012; Zalapa et al., 2012); Illumina is mainly limited to organisms with available reference genomes (Wang et al., 2010). In the present study, we chose NGS (Roche 454 GS FLX$^+$) to develop SSRs and search for high-polymorphism SSRs that can be used for genetic diversity.

This is the first study developing *C. gigantea* SSRs; it will help to evaluate genetic diversity, linkage map construction, marker-assisted selection, and other related studies of *C. gigantea* and species of Araceae.

MATERIALS AND METHODS

Plant materials and DNA isolation

The leaf samples were collected from 10 individual healthy *C. gigantea* plants (four wild and six cultivated). Detailed information of all plant materials used in the present study is listed in Table 1. For 454 sequencing, total genomic DNA was isolated from the petiole tissues using the 4×CTAB protocol (Dellaporta et al., 1983). The DNA products were then visualized by electrophoresis on SYBR Green I stained 2.0% agarose gel with *Trans* DNA marker I (Ding Guo Co., Beijing, China) to evaluate quality. DNA samples were stored at -20 °C, and DNA integrity was verified with a Bioanalyzer (2100, Agilent, Santa Clara, California, USA) prior to sequencing. Approximately 5 µg genomic DNA was sequenced with a GS FLX$^+$ platform (Roche Applied Science, Indianapolis, Indiana, USA). Sequencing was carried out according to the manufacturer's protocol in BGI Genomics (Shenzhen, China).

Polymerase chain reaction (PCR) conditions

Polymerase chain reaction (PCR) amplifications were carried out with a MyCycler Thermal Cycler (Bio-RAD, Hercules, California, USA) in a final volume of 20 µL. Each reaction tube contained 2 µL PCR buffer, 2 µL dNTP, 2 µL of each primer, 1 µL genomic DNA, 2 µL Taq DNA polymerase (Fermentas, Vilnius, Lithuania), and 10 µL distilled water. For the PCR reaction program, amplification conditions in the thermocycler were as follows: preincubation at 94 °C for 5 min followed by 35 cycles of denaturation at 94 °C for 40 s, annealing at 55 °C (depending on primer pairs) for 45 s, elongation at 72 °C for 30 s, and a final extension step at 72 °C for 5 min. Optimized SSR primers were used to amplify DNA from five plant individuals to select polymorphic microsatellite markers.

Table 1. Origin of *Colocasia gigantea* plant materials used in the present study.

Nr	Latin name	Characteristics	Part of collection	Locality
1	*C. gigantea*	Wild	Leaf	China, Xishuangbanna
2	*C. gigantea*	Wild	Leaf	China, Hunan
3	*C. gigantea*	Wild	Leaf	China, Guangxi
4	*C. gigantea*	Wild	Leaf	China, Kunming
5	*C. gigantea*	Cultivar	Leaf	China, Xishuangbanna
6	*C. gigantea*	Cultivar	Leaf	China, Xishuangbanna
7	*C. gigantea*	Cultivar	Leaf	China, Hunan
8	*C. gigantea*	Cultivar	Leaf	China, Hunan
9	*C. gigantea*	Cultivar	Leaf	China, Guangxi
10	*C. gigantea*	Cultivar	Leaf	China, Kunming

Sequence analysis and identification of simple sequence repeats (SSRs)

Before assembly, we carried out a stringent filtering process of raw sequencing reads. Only the sequences containing mitochondrial- or plastid-encoded genes were discarded. We removed 454 adaptor sequences from the raw reads obtained with the GS-FLX Titanium sequencer. Only high-quality reads that were at least 50 nucleotides in length were included in further analyses. In the present study, Newbler v2.5 (Roche/454 Life Sciences, Branford, Connecticut, USA) was used to cluster and assemble high-quality reads into a set of non-redundant contigs. Thresholds chosen for the aggregation steps were identified at 90% with a minimum overlapping length of 40 nucleotides (Li et al., 2012). We only searched for SSRs from nuclear genome using the MISA software (http://pgrc.ipk-gatersleben.de/misa/). The minimum length for each type of SSR was set as mono-nucleotide repeats ≥ 10 nucleotides, di-nucleotide repeats ≥ 12 nucleotides, tri-nucleotide repeats ≥ 15 nucleotides; tetra-nucleotide repeats ≥ 16 nucleotides, penta-nucleotide repeats ≥ 20 nucleotides, and hexa-nucleotide repeats ≥ 24 nucleotides. Oligo-nucleotide primers were designed for selected SSR loci (where repeats were located at least 50 bp from the 5' and 3' end of the sequences) using the PRIMER3 software (http://frodo.wi.mit.edu/primer3/). The parameters for primer design were preferred amplicon size of 100-200 bp, primer size 18-27 bp, and primer melting temperature of 55-60 °C; the optimum temperature was 55 °C. When none of the three criteria could be met, the priority was the melting temperature. All primers were synthesized by Invitrogen (Invitrogen Co., Shanghai, China).

Simple sequence repeat (SSR) marker validation and polymorphic SSR selection

Genomic DNA from *C. gigantea* was extracted as described above and diluted prior to being used as templates to select polymorphic loci. Polymerase chain reactions were also performed as described above. Products from the amplification reactions were resolved on PAGE gels consisting of 4.5% polyacrylamide and 7 M urea in 0.5×TBE buffer, and the fragments were subsequently visualized by silver staining. The statistical analyses of SSR data were performed with PowerMarker 3.25 software (http://www.powermarker.net). The allelic polymorphism information content (PIC) was calculated by the following formula: $PIC = 1-\sum(P_i)^2$ where P_i is the SSR i^{th} allele frequency (Botstein et al., 1980). Polymorphism statistics, including allele number, observed/expected heterozygosities, inbreeding coefficient, and linkage equilibrium between loci, were calculated with GENEPOP v4.2 (Raymond and Rousset, 1995).

RESULTS

Roche GS FLX$^+$ 454 sequencing and assembly

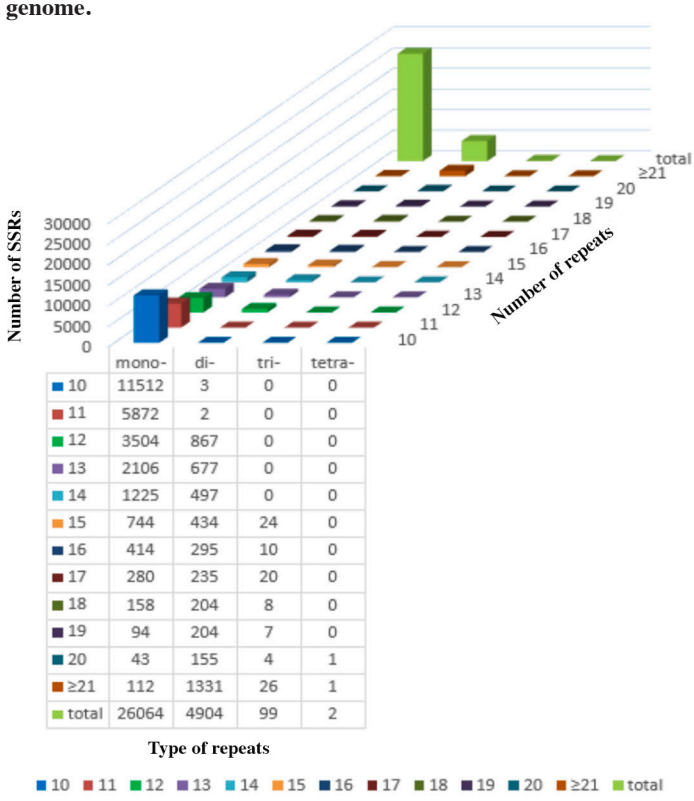
We obtained 1 232 789 reads with a total length of 621 787 375 bp. We incorporated 1 043 587 reads (84.7%) into 46 234 contigs and 630 430 singletons for a total of 676 664 unique sequences. The rest of the trimmed raw reads (189 202; ~15.3%) were excluded from the assembly because the sequences were too short, chimeric, or contained repetitive segments. The contigs had a maximum length of 71 068 nucleotides. Among all contigs, 100-200 bp was the most frequent.

Simple sequence repeat (SSR) identification and characterization

We identified 31 069 putative microsatellites. The frequencies of short tandem repeats, categorized by their unit sizes and the number of repeats, are summarized in Figure 1. Mono-nucleotide repeats (26 064, 83.9%) comprise the largest group of repeat motifs followed by di-nucleotide repeats (4904, 15.8%), while tetra-nucleotide repeats and penta-nucleotide repeats were the lowest (Figure 1).

Analysis of A/T and C/G mono-nucleotide repeats showed that the A/T motif was significantly overrepresented (71.17%) in *C. gigantea* genomic sequences (Table 2). The A/T repeats were not the only predominant mono-nucleotide repeats, but they were also the most frequent motif in the entire genome, representing 59.71% of total SSRs. Among the di-nucleotide tandem repeats, GA/TC and AG/CT were by far the most common and accounted for 31.749% and 31.505% of all di-nucleotide motifs, respectively; this was followed by AT/TA (18.94%), whereas the GT/AC repeat was extremely rare (8.667%; Table 2). The majority of tri-nucleotide repeats found in genomic sequences were TTC, which represented 17.17% of all tri-nucleotide repeats, followed by TTG (11.11%; Table 2). With respect to tetra-nucleotide repeats, we only identified ATAC in the *C. gigantea* genome (Table 2).

Figure 1. Distribution of various classes of simple repeat motifs with different numbers of repeats in the *Colocasia gigantea* genome.



	mono-	di-	tri-	tetra-
10	11512	3	0	0
11	5872	2	0	0
12	3504	867	0	0
13	2106	677	0	0
14	1225	497	0	0
15	744	434	24	0
16	414	295	10	0
17	280	235	20	0
18	158	204	8	0
19	94	204	7	0
20	43	155	4	1
≥21	112	1331	26	1
total	26064	4904	99	2

SSR: Simple sequence repeat.

Marker validation and polymorphic simple sequence repeat (SSR) selection

To assess the practicality of the microsatellites discovered *in silico*, 100 non-redundant genomic SSRs containing sequences (10 mono-, 70 di-, and 20 tri-nucleotide repeats) were randomly selected among 10 *C. gigantea* samples. Seventy-eight primers (89%) were successfully amplified. Of these amplifiable loci, 36 yielded specificity amplification. Of these 36 loci, 12 yielded products of expected sizes and 24 primers yielded at least two bands. Of these 12 markers, 6 microsatellite markers were scorable and exhibited a broad range of allelic diversity among examined *C. gigantea*. The list of six sequences containing polymorphic SSRs was deposited in GenBank (KR010361-KR010366) (Table 3). For each of the SSR markers, both the forward and reverse primer sequences and PCR product sizes of *C. gigantea* are listed in Table 3. For the six polymorphic loci, the number of alleles per locus ranged from 3 to 7 alleles. We identified 28 alleles, with a mean of 2.8 alleles per locus. The PIC ranged from 0.561 to 0.756 (Table 4).

Data access

The raw sequence data for *C. gigantea* can be accessed at the National Center for Biotechnology Information (NCBI), Sequence Read Archive (SRA) (NCBI, Bethesda, Maryland, USA) with ID SRR6206491. The sample accession number is SAMN07786806. The BioProject number is PRJNA414351.

DISCUSSION

The identification and development of molecular markers represent significant challenges, especially for organisms that have little or no genomic sequence information. We used next-generation and high-throughput sequencing technology to develop additional genomic-derived microsatellite markers in *C. gigantea*.

The genome sequences obtained in the present study were incomplete; therefore, the reported 31 069 putative genomic-derived SSRs could be underestimated. We discovered that a large proportion of mono-nucleotide repeats

Table 2. Frequencies of non-redundant simple sequence repeat motifs with respect to repeat numbers.

Repeat Motif	10	11	12	13	14	15	16	17	18	19	20	> 20	Total	%
A	4508	2083	1030	601	317	192	75	66	28	16	6	19	8941	34.30
C	924	786	624	368	159	57	9	7	2	1	0	11	2948	11.31
T	5019	2097	1094	580	348	202	130	77	32	14	3	13	9609	36.87
G	1061	906	756	557	401	293	200	130	96	63	34	69	4566	17.52
Monomeric (total)	11512	5872	3504	2106	1225	744	414	280	158	94	43	112	26064	100.00
AC	2	1	46	35	22	19	11	15	9	6	6	118	290	5.914
AG	0	0	113	100	114	93	84	71	78	74	74	414	1215	24.776
AT	0	0	186	151	93	81	33	29	9	2	2	1	587	11.970
CA	1	1	38	30	11	17	12	11	3	6	5	124	259	5.281
CT	0	0	109	71	45	29	25	13	6	8	3	21	330	6.729
GA	0	0	99	94	91	92	72	65	79	87	60	518	1257	25.632
GT	0	0	38	19	10	8	4	4	0	4	0	48	135	2.753
TA	0	0	95	91	61	48	25	10	6	4	0	2	342	6.974
TC	0	0	110	62	38	26	15	12	9	7	2	19	300	6.117
TG	0	0	33	24	12	21	14	5	5	6	3	66	189	3.854
Dimeric (total)	3	2	867	677	497	434	295	235	204	204	155	1331	4904	100.00
AAC	0	0	0	0	0	0	0	1	1	0	0	0	2	2.02
AAT	0	0	0	0	0	1	0	1	2	0	1	0	5	5.05
ACA	0	0	0	0	0	0	1	2	0	1	0	3	7	7.07
ACT	0	0	0	0	0	0	1	0	0	0	0	0	1	1.01
AGA	0	0	0	0	0	0	0	0	0	1	0	0	1	1.01
AGT	0	0	0	0	0	0	0	1	0	0	0	1	2	2.02
ATC	0	0	0	0	0	0	0	0	0	0	0	1	1	1.01
ATG	0	0	0	0	0	0	0	0	0	0	0	1	1	1.01
ATT	0	0	0	0	0	0	0	1	0	0	0	0	1	1.01
CAA	0	0	0	0	0	2	1	2	0	0	0	2	7	7.07
CAT	0	0	0	0	0	1	2	0	0	0	1	0	4	4.04
CCA	0	0	0	0	0	0	0	1	0	0	0	0	1	1.01
CTT	0	0	0	0	0	3	0	1	0	0	1	0	5	5.05
GAT	0	0	0	0	0	0	0	0	0	0	0	4	4	4.04
GGT	0	0	0	0	0	0	0	0	0	0	0	1	1	1.01
GTA	0	0	0	0	0	0	0	0	1	0	0	0	1	1.01
GTT	0	0	0	0	0	1	0	0	0	0	0	0	1	1.01
TAA	0	0	0	0	0	1	0	0	0	0	0	0	1	1.01
TAC	0	0	0	0	0	0	0	2	0	0	0	0	2	2.02
TCA	0	0	0	0	0	2	0	0	1	0	0	2	5	5.05
TCG	0	0	0	0	0	0	1	0	0	0	0	0	1	1.01
TCT	0	0	0	0	0	2	3	2	1	0	1	0	9	9.09
TGA	0	0	0	0	0	1	1	0	0	0	0	3	5	5.05
TGG	0	0	0	0	0	0	0	0	0	1	0	0	1	1.01
TGT	0	0	0	0	0	0	0	1	0	0	0	0	1	1.01
TTA	0	0	0	0	0	1	0	0	0	0	0	0	1	1.01
TTC	0	0	0	0	0	6	0	1	1	3	0	6	17	17.17
TTG	0	0	0	0	0	3	0	4	1	1	0	2	11	11.11
Trimeric (total)	0	0	0	0	0	24	10	20	8	7	4	26	99	100.00
ATAC	0	0	0	0	0	0	0	0	0	0	1	1	2	100.00
Tetrameric (total)	0	0	0	0	0	0	0	0	0	0	1	1	2	100.00

were AT-rich (Table 2). A similar occurrence was observed in *Hevea brasiliensis* (Li et al., 2012). The prevalence of genomic-derived SSRs with AT dimeric repeats has also been reported for *H. brasiliensis* (Li et al., 2012), *Pinus massoniana* (Bai et al., 2014), rice, and other plant species (Powell et al., 1996; Temnykh et al., 2001). In many studies, the AG/CT di-nucleotide repeats was the most common for the transcribed regions (Kantety et al., 2002; Xu et al., 2012); similarly, GA/TC and AG/CT di-nucleotide repeats occurred at a higher frequency in the *C. gigantea* genome in our study (Table 2). We found that GA/TC and AG/CT account for 31.749% and 31.505% of all di-nucleotide motifs, respectively. In our study, most tri-nucleotide repeats were TTC (Table 2), which is inconsistent with the reports for barley, maize, rice, sorghum (Kantety et al., 2002), and rubber (Feng et al., 2009). The most abundant tri-nucleotide motif was GGC/CCG in barley, maize, rice, and sorghum and AAC/TTG in wheat; the majority of tri-nucleotide repeats in rubber were AAG/CTT.

Table 3. Characterization of six newly developed polymorphism genomic-derived simple sequence repeats (SSRs) in *Colocasia gigantea*.

Primer	Sequence (5'-3')	Repeat motif	PCR product size (bp)	Ta (°C)	GenBank accession number
H4311UH01A8IX1	F: ATAGCTTCATGTGCATACCC R: CCATCTCTCCACCGTTACTA	$(GA)^{16}$	127	55	KR010361
H4311UH01DFSZR	F: TTGTTCATCCTACGGATTTC R: ATGTTTCCATGCATTGTCTT	$(TG)^{28}$	127	55	KR010362
H4311UH01EM3GH	F: TCGTAATTAACCGTAACGAAA R: CAAGAGTGTTCGTCTTGGAC	$(TC)^{18}$	178	55	KR010363
H4311UH01DQIBR	F: CAAGTTTGTGGGAAATTCAT R: ATCAAGTACAAAGAGGCAACA	$(TTG)^{15}$	169	55	KR010364
H4311UH01DCJBJ	F: GAAGAAGAGCCACCAAGAAT R: GTTTAAGAATGCGCCACTT	$(CA)^{19}$	172	55	KR010365
H4311UH01EW29I	F: TTCCTCTCTTTCTTCCTCCTT R: GAGAGTGTGGGGGAAAAG	$(CA)^{16}$	102	55	KR010366

Ta: Annealing temperature.

Table 4. Genetic diversity parameters for 10 *Colocasia gigantea*.

	Number of alleles	PIC value
P8	4	0.615
P29	3	0.732
P37	6	0.756
P41	7	0.621
P59	3	0.561
P73	5	0.689

PIC: Polymorphic information content.

In the present study, the rest of the 24 amplifiable markers yielded at least two bands that could not be reliably scored; this could perhaps be the result of primer design or heterozygosis of *C. gigantea* in the tested population or an amplification area containing larger introns. This first set of microsatellite markers developed for *C. gigantea* will be useful for assessing genetic diversity and understanding the population structure of wild populations of *C. gigantea* and the Araceae family in general.

CONCLUSION

Colocasia gigantea is a very important vegetable in China that contains high dietary fiber, pyridoxine, and nicotinamide. Prior to this study, none of the public databases provided sequencing information for *C. gigantea*. We adopted the 454 sequencing technology to analyze the *C. gigantea* genome, characterized it by de novo sequencing without the presence of a reference genome, and developed a set of polymorphic simple sequence repeats (SSRs).

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation of China (31600254), National Science Foundation of Jiangsu Province of China (BK20150491), Start-up Fund for Advanced Talents of Jiangsu University (14JDG150), School of Agricultural Equipment Engineering at Jiangsu University, and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD, [2014]37).

REFERENCES

Angeloni, F., Wagemaker, C.A., Jetten, M.S., Op den Camp, H.J., Janssen-Megens, E.M., Francoijs, K.J., et al. 2011. De novo transcriptome characterization and development of genomic tools for *Scabiosa columbaria* L. using next-generation sequencing techniques. Molecular Ecology Resources 11:662-674. doi:10.1111/j.1755-0998.2011.02990.x.

Bai, T.D., Xu, L.A., Xu, M., and Wang, Z.R. 2014. Characterization of masson pine (*Pinus massoniana* Lamb.) microsatellite DNA by 454 genome shotgun sequencing Tree. Genetics and Genomes 10:429-437. doi:10.1007/s11295-013-0684-y.

Baldwin, S., Pither-Joyce, M., Wright, K., Chen, L., and McCallum, J. 2012. Development of robust genomic simple sequence repeat markers for estimation of genetic diversity within and among bulb onion (*Allium cepa* L.) populations. Molecular Breeding 30(3):1401-1411. doi:10.1007/s11032-012-9727-6.

Botstein, D., White, R.L., Skolnick, M., and Davis, R.W. 1980. Construction of genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics 32(3):314-331.

Bown, D. 2000. Aroids: Plants of the Arum family. Vol. 2. Timber Press, London, UK.

Dellaporta, S.L., Wood, J., and Hicks, J.B. 1983. A plant DNA minipreparation: Version II. Plant Molecular Biology Reporter 1(4):19-21. doi:10.1007/BF02712670.

Egan, A.N., Schlueter, J., and Spooner, D.M. 2012. Applications of next-generation sequencing in plant biology. American Journal of Botany 99:175-185. doi:10.3732/ajb.1200020.

Feng, S., Li, W., Huang, H., Wang, J., and Wu, Y. 2009. Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*). Molecular Breeding 23:85-97. doi:10.1007/s11032-008-9216-0.

Georgi, L., Herai, R.H., Vidal, R., Carazzolle, M.F., Pereira, G.G., Polashock, J., et al. 2012. Cranberry microsatellite marker development from assembled next-generation genomic sequence. Molecular Breeding 30:227-237. doi:10.1007/s11032-011-9613-7.

Ho, C.W., Wu, T.H., Hsu, T.W., Huang, J.C., Huang, C.C., and Chiang, T.Y. 2011. Development of 12 genic microsatellite loci for a biofuel grass, *Miscanthus sinensis* (Poaceae). American Journal of Botany 98:e201-e203. doi:10.3732/ajb.1100071.

Ivancic, A., Roupsard, O., Garcia, J.Q., Melteras, M., Molisale, T., Tara, S., et al. 2008. Thermogenesis and flowering biology of Colocasia gigantea, Araceae. Journal of Plant Research 121:73-82. doi:10.1007/s10265-007-0129-5.

Kale, S.M., Pardeshi, V.C., Kadoo, N.Y., Ghorpade, P.B., Jana, M.M., and Gupta, V.S. 2012. Development of genomic simple sequence repeat markers for linseed using next-generation sequencing technology. Molecular Breeding 30:597-606. doi:10.1007/s11032-011-9648-9.

Kantety, R.V., La Rota, M., Matthews, D.E., and Sorrells, M.E. 2002. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Molecular Biology 48:501-510. doi:10.1023/A:1014875206165.

Li, D., Deng, Z., Qin, B., Liu, X., and Men, Z. 2012. De novo assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.) BMC Genomics 13:192. doi:10.1186/1471-2164-13-192.

Nguyen, L.T. 2005. Bac há (*Colocasia gigantea* [Blume] Hook. F.) in the culinary history of Vietnamese-Americans. Economic Botany 59:185-190.

Perry, J.C., and Rowe, L. 2011. Rapid microsatellite development for water striders by next-generation sequencing. Journal of Heredity 102:125-129. doi:10.1093/jhered/esq099.

Powell, W., Machray, G.C., and Provan, J. 1996. Polymorphism revealed by simple sequence repeats. Trends in Plant Science 1:215-222. doi:10.1016/1360-1385(96)86898-1.

Raymond, M., and Rousset, F. 1995. GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. Journal of Heredity 86:248-249.

Sahu, B., Patel, A., Sahoo, L., Das, P., Meher, P., and Jayasankar, P. 2012. Rapid and cost effective development of SSR markers using next generation sequencing in Indian major carp, *Labeo rohita* (Hamilton, 1822). Indian Journal of Fisheries 59(3):21-24.

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S., and McCouch, S. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. Genome Research 11:1441-1452. doi:10.1101/gr.184001.

Wang, Z., Fang, B., Chen, J., Zhang, X., Luo, Z., Huang, L., et al. 2010. De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweetpotato (*Ipomoea batatas*). BMC Genomics 11:726. doi:10.1186/1471-2164-11-726.

Xu, W., Yang, Q., Huai, H., and Liu, A. 2012. Development of EST-SSR markers and investigation of genetic relatedness in tung tree. Genetics and Genomes 8:933-940. doi:10.1007/s11295-012-0481-z.

Yang, H., Tao, Y., Zheng, Z., Li, C., Sweetingham, M.W., and Howieson, J.G. 2012. Application of next-generation sequencing for rapid marker development in molecular plant breeding: A case study on anthracnose disease resistance in *Lupinus angustifolius* L. BMC Genomics 13:318. doi:10.1186/1471-2164-13-318.

Zalapa, J.E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E., et al. 2012. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. American Journal of Botany 99:193-208. doi:10.3732/ajb.1100394.