

Development of novel EST-SSR markers for *Phyllanthus emblica* (Phyllanthaceae) and cross-amplification in two related species

Xiongfang Liu¹, Hong Ma^{1,3} , Taiqiang Li¹, Zhenghong Li¹, Youming Wan¹, Xiuxian Liu¹, and Liyong Fu^{2,3}

Manuscript received 3 February 2018; revision accepted 19 April 2018.

¹ Research Institute of Resources Insects, Chinese Academy of Forestry, Kunming 650224, People's Republic of China

² Research Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing 10091, People's Republic of China

³ Authors for correspondence: hortscience@163.com, fuly@ifrit.ac.cn

Citation: Liu, X., H. Ma, T. Li, Z. Li, Y. Wan, X. Liu, and L. Fu. 2018. Development of novel EST-SSR markers for *Phyllanthus emblica* (Phyllanthaceae) and cross-amplification in two related species. *Applications in Plant Sciences* 6(7): e1169.

doi:10.1002/aps3.1169

PREMISE OF THE STUDY: A novel set of EST-SSR markers was developed for *Phyllanthus emblica* (Phyllanthaceae) to investigate the genetic structure and gene flow, identify novel genes of interest, and develop markers for assisted breeding.

METHODS AND RESULTS: Based on the transcriptome data of *P. emblica*, 83 EST-SSR primer pairs were designed; 52 primer pairs were successfully amplified, with 20 showing polymorphisms in 90 individuals from three populations of *P. emblica*. The number of alleles per locus varied from 11 to 44. The observed and expected levels of heterozygosity for the 20 loci ranged from 0.240 to 0.868 and 0.754 to 0.933, respectively. Cross-species amplification was successful for all 20 loci in each of the two related species, *P. reticulatus* and *Leptopus chinensis*.

CONCLUSIONS: These markers will be valuable for studying the population genetics and for mining genes of *P. emblica*, and may be useful for studies of related species.

KEY WORDS EST-SSR marker; *Leptopus chinensis*; Phyllanthaceae; *Phyllanthus emblica*; *Phyllanthus reticulatus*; transcriptome.

Phyllanthus emblica L. (Phyllanthaceae) is an important medicinal and edible plant distributed across tropical and subtropical regions, and has been listed as one of the three health-related plants to be promoted for planting around the world by the World Health Organization (WHO) (Li and Zhao, 2007; Variya et al., 2016). From an ecological perspective, the tree is extremely resistant to drought and barren environments and can be used as a pioneer tree to establish forests on barren hills (Li and Zhao, 2007). The focus of recent research has been on the biochemistry and pharmacology of *P. emblica* (Variya et al., 2016), and few studies have focused on the development of molecular markers (Pandey and Changtragoon, 2012; Mawalagedera et al., 2014), which are important for genetic studies. To date, only six genomic simple sequence repeat (SSR) markers have been reported for *P. emblica* (Pandey and Changtragoon, 2012), and only one (Phyll_68) of the six genomic SSRs was found to be polymorphic in the three studied populations of *P. emblica* (percentage of polymorphic loci = 16.67%; Table 1), which are insufficient for further genetic studies.

With advances in high-throughput sequencing technologies, especially de novo transcriptome sequencing, expressed sequence tag-simple sequence repeat (EST-SSR) markers can be rapidly mined at a lower cost. They are also increasingly being used for the evaluation of genetic relationships, because they are codominant,

highly polymorphic, and are well distributed throughout the genome (Bouck and Vision, 2007). Furthermore, compared to genomic SSRs, EST-SSRs are more transferable across taxonomic boundaries (Ellis and Burke, 2007). We therefore developed 20 EST-SSR markers for *P. emblica*; these new markers have a higher level of genetic diversity than the loci previously reported in Sri Lankan populations (Mawalagedera et al., 2014). We then evaluated their transferability to two sympatric species in the Phyllanthaceae: *P. reticulatus* Poir. and *Leptopus chinensis* (Bunge) Pojark.

METHODS AND RESULTS

Young leaf tissue was collected from five healthy plants of *P. emblica* growing in natural environments in Binchuan (25°45'59"N, 100°26'29"E, voucher specimen accession no. BC-20170622-MH; Appendix 1). These samples were immediately frozen in liquid nitrogen and stored at -80°C until use. Total RNA was extracted using the protocol described by Kumar and Singh (2012), followed by RNA purification and DNase I digestion, and then purified RNAs were fragmented into short fragments using the Ambion RNA Fragmentation Kit (Ambion, Austin, Texas, USA) according to the manufacturer's protocols. The cDNA library was prepared and sequenced using

TABLE 1. Genetic diversity of the 21 polymorphic SSR markers (including 20 newly developed markers and one previously published marker) in three populations of *Phyllanthus emblica*.^a

Locus	<i>Phyllanthus emblica</i>																
	Binchuan (N = 30)				Yuanmou (N = 30)				Yongping (N = 30)				Total (N = 90)				
	B	A	H _o	H _e ^b	F _{IS}	A	H _o	H _e ^b	F _{IS}	A	H _o	H _e ^b	F _{IS}	A	H _o	H _e	F _{IS}
PE399	0.215 [†]	11	0.286	0.754*	0.614	11	0.185	0.813*	0.768	8	0.250	0.683*	0.625	12	0.240	0.754	0.673
PE788	0.029	17	0.655	0.900*	0.259	14	0.800	0.862	0.056	16	0.571	0.928*	0.369	25	0.688	0.896	0.232
PE4618	0.184 [†]	15	0.633	0.910*	0.292	16	0.600	0.854*	0.286	9	0.619	0.829*	0.235	20	0.617	0.869	0.272
PE6781	0.283 [†]	7	0.296	0.802*	0.623	9	0.276	0.851*	0.670	10	0.263	0.900*	0.699	14	0.280	0.850	0.666
PE6950	0.041	14	0.533	0.729	0.256	16	0.444	0.818*	0.446	10	0.619	0.772	0.179	19	0.526	0.770	0.298
PE7362	0.111	10	0.517	0.868*	0.394	11	0.552	0.857*	0.345	14	0.571	0.920*	0.364	15	0.544	0.880	0.367
PE7779	0.184 [†]	12	0.552	0.916*	0.387	12	0.633	0.897*	0.282	10	0.476	0.883*	0.447	13	0.563	0.904	0.372
PE8467	0.071	9	0.643	0.747	0.124	13	0.700	0.892	0.202	13	0.714	0.868	0.157	16	0.684	0.845	0.163
PE8480	0.014	8	0.760	0.809	0.041	10	0.808	0.846	0.027	10	0.762	0.792	0.015	14	0.778	0.822	0.028
PE9600	0.020	16	0.857	0.869	-0.005	21	0.862	0.900	0.026	14	0.895	0.811	-0.133	24	0.868	0.867	-0.034
PE10156	0.071	16	0.750	0.872	0.124	17	0.600	0.834*	0.268	13	0.429	0.849*	0.483	25	0.608	0.854	0.290
PE10572	0.278 [†]	5	0.259	0.765*	0.655	7	0.379	0.797*	0.516	11	0.286	0.826*	0.646	11	0.312	0.830	0.605
PE11297	0.148 [†]	13	0.393	0.861*	0.535	16	0.448	0.898*	0.492	13	0.500	0.906*	0.434	20	0.442	0.889	0.487
PE14171	0.315 [†]	12	0.433	0.863*	0.489	9	0.500	0.711*	0.284	13	0.476	0.763*	0.361	15	0.468	0.798	0.385
PE14389	0.145	12	0.267	0.831*	0.674	13	0.467	0.785*	0.396	8	0.524	0.717	0.251	19	0.407	0.787	0.451
PE14485	0.166 [†]	13	0.414	0.874*	0.518	14	0.483	0.866*	0.433	9	0.263	0.834*	0.676	19	0.403	0.861	0.540
PE15252	0.109	11	0.643	0.843*	0.223	11	0.897	0.872	-0.047	8	0.571	0.803*	0.271	13	0.718	0.840	0.145
PE17379	0.129	20	0.690	0.937*	0.251	18	0.483	0.917*	0.464	13	0.412	0.861*	0.507	29	0.547	0.918	0.404
PE17828	0.106 [†]	26	0.852	0.951	0.087	22	0.862	0.924	0.050	19	0.857	0.889	0.012	41	0.857	0.930	0.051
PE21382	0.196 [†]	25	0.556	0.926*	0.389	24	0.517	0.926*	0.432	18	0.722	0.925*	0.197	44	0.581	0.933	0.340
Mean	0.074	14	0.549	0.830	0.347	14	0.575	0.883	0.320	12	0.539	0.808	0.340	20	0.557	0.855	0.337
Phyll_68 ^c	0.001	3	0.500	0.492	—	4	0.375	0.350	—	5	0.750	0.708	—	7	0.542	0.526	—

Note: A = number of alleles per locus; B = null allele frequency averaged over all populations using the Brookfield 1 equation (Brookfield, 1996); F_{IS} = inbreeding coefficient; H_e = expected heterozygosity; H_o = observed heterozygosity; N = number of individuals analyzed.

^aLocality and voucher information are provided in Appendix 1.

^bSignificant deviations from Hardy–Weinberg equilibrium after sequential Bonferroni corrections: *P < 0.001.

^cPhyll_68 was cited from Pandey and Changtragoon (2012).

[†]Loci with null alleles.

the Illumina HiSeq 4000 (Illumina, San Diego, California, USA; sequencing performed by Gene Denovo Biotechnology Company, Guangzhou, China) to produce 150-bp paired-end reads. This resulted in 72,976,514 raw reads. All raw reads have been deposited into the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA; Bioproject ID: SRR6509792). The generated raw reads were filtered by trimming adapters and removing ambiguous reads (N > 10%) and low-quality reads (more than 40% of nucleotides with Q value ≤ 10) using Trimmomatic (version 0.35; Bolger et al., 2014). Clean reads were assembled de novo into 97,628 transcripts using Trinity software (Grabherr et al., 2011) and were then clustered into 76,881 nonredundant (nr) unigenes using CD-HIT software (Fu et al., 2012); the clustered unigenes have been deposited in GenBank under the accession GGLN00000000. MISA software (Thiel et al., 2003) was used to detect microsatellites from all unigenes, with the following thresholds: six repeat units for dinucleotides, five repeat units for trinucleotides, and four repeat units for tetra-, penta-, and hexanucleotides. In all, 4934 SSR sequences were retrieved, and 83 of them with five or more di- or trinucleotide repeats were randomly selected for primer design in Primer3 software (Rozen and Skaletsky, 1999) with the following parameters: primer length of 18–27 bp, annealing temperature of 57–63°C, PCR product size of 100–280 bp, and GC content of 40–60%.

The preliminary screening of the 83 target EST-SSR primers was performed with three individuals from each of the three *P. emblica* natural populations (Appendix 1). Genomic DNA was isolated from silica-dried leaves with the Plant Genomic DNA Extraction

Kit (Tsingke Biotechnology Co. Ltd., Beijing, China) following the manufacturer's protocol. PCR amplifications were conducted with the ABI 2720 Thermal Cycler (Thermo Fisher Scientific, Waltham, Massachusetts, USA) in a 15-μL reaction mixture that contained 1 μL (10–20 ng) of genomic DNA, 7.5 μL of 2× Master Mix (Tsingke Biotechnology Co. Ltd.), 1 μL (10 pM) of forward primer, 1 μL (10 pM) of reverse primer, and 4.5 μL of ddH₂O. The PCR protocol used was as follows: an initial denaturation at 94°C for 5 min; followed by 30 cycles of denaturation at 94°C for 30 s, annealing at a temperature gradient from 57°C to 63°C (depending on the specific locus; Table 2) for 30 s, and extension at 72°C for 30 s; with final extension at 72°C for 5 min. The PCR products were electrophoresed using 1% agarose gels to determine whether amplifications were successful for the expected sizes. Of the 83 primer pairs, 52 (62.7%) produced clear amplicons of the expected size. The other primer pairs gave no product. To test for polymorphism of the 52 primers, fluorescence-based SSR genotyping was performed using multiplex-ready PCR technology (Hayden et al., 2008). For all loci, the 5' end of each forward primer was tagged with one of two fluorescent dyes (FAM or HEX; Thermo Fisher Scientific; Table 2), and multiplex PCR amplifications were performed using 30 individuals from each of three *P. emblica* natural populations (the distance between the collected individual samples within each population was at least 10 m to ensure the reliability of sampling) with the same protocol mentioned above. The fluorescently tagged PCR products were analyzed on an ABI 3730xl DNA Analyzer with a GeneScan 500 LIZ Size Standard (Thermo Fisher Scientific), and allele sizes were assessed with GeneMapper

TABLE 2. Characteristics of the 20 polymorphic EST-SSR markers developed for *Phyllanthus emblica*.

Locus	Primer sequences (5'–3')	Repeat motif	Allele size range (bp)	T_a (°C)	Fluorescent dye ^a	GenBank accession no.	BLAST top hit		
							Description [organism]	GenBank accession no.	E-value ^b
PE399	F: CCGGCTATTTAGCGTGTCT R: GGCAACTTGAAGCAGAGGAC	(TC) ₉	136–213	58	FAM ¹	MG595342	No hit	—	—
PE788	F: TGGCTAAGGAAGAGCACGTT R: TTCTCCTCAACCACCAGCTT	(GAA) ₅	117–157	58	FAM ¹	MG595343	No hit	—	—
PE4618	F: CAAACGTTCTGACACGACGA R: CTCTCCAACCTGGGTCCACAT	(TTG) ₆	128–171	59	FAM ²	MG595344	No hit	—	—
PE6781	F: ATCCGAAACCATCCATGAAG R: GACATTTGTGCAATGCTGCT	(GT) ₆	155–212	58	FAM ²	MG595345	No hit	—	—
PE6950	F: AGCAGCTGCACAGCAGTCTA R: AATGCCCGGGAAGATAATTC	(GCT) ₅	110–152	58	FAM ²	MG595346	Transcription factor TCP7-like [<i>Jatropha curcas</i>]	XP_012070794.1	3.51E–12
PE7362	F: GCCCTTATCCCAGTTAACGC R: AAGCAAATGTTCCAGGATGC	(AT) ₆	123–162	58	FAM ²	MG595347	No hit	—	—
PE7779	F: ATATGTCCCGTTTCGGACAG R: TCCACATGCACTTACGAAT	(CT) ₆	226–250	59	FAM ²	MG595348	No hit	—	—
PE8467	F: TTCCAGAAATGCTCACACA R: CGGAGAAATTTCAAGAAACA	(AG) ₈	169–191	57	FAM ¹	MG595349	No hit	—	—
PE8480	F: GTTTGTGGCAGTCCGGTATT R: TGGCTCTGAAGAAATGGGTT	(GGA) ₅	111–131	57	FAM ¹	MG595350	No hit	—	—
PE9600	F: GCGGTGATAGAGAGAGTGG R: GAGGCATAGAAAGTGTGAAAT	(TCT) ₇	202–228	59	HEX ²	MG595351	No hit	—	—
PE10156	F: ATTGCTCCAATCATTGTCTCC R: ACCACTCATGCCGAAGAATC	(GA) ₆	121–207	59	HEX ²	MG595352	Sugar transporter ERD6-like 7 isoform X2 [<i>Jatropha curcas</i>]	XP_012092321.1	2.09E–31
PE10572	F: TAGAGGGTCTGTTGATGGAGG R: GTCCAAAGCTTCAAATCCGA	(GAT) ₆	234–258	59	HEX ²	MG595353	No hit	—	—
PE11297	F: CTTACGCGCTCTCTCTCTT R: ATAGGTAACGGACGCGAATG	(CTA) ₅	211–268	59	HEX ¹	MG595354	Hypothetical protein CICLE_v10016016mg [<i>Citrus clementina</i>]	XP_006449505.1	3.90E–06
PE14171	F: GGCCAATTTCAATGCATCTT R: CCTGCTGTTCAATTGCCTT	(ATT) ₆	120–216	58	HEX ¹	MG595355	Hypothetical protein JCGZ_16347 [<i>Jatropha curcas</i>]	KDP44514.1	1.38E–41
PE14389	F: ACACCTTCTTCCCTTGCTT R: CATTGCTTCGAACAAGTCCA	(CCA) ₆	115–162	58	HEX ²	MG595356	No hit	—	—
PE14485	F: GCACCAACATTTAGTTGCCA R: ACGAGGTTGCGGTCTAGAAG	(GCA) ₆	121–297	58	HEX ¹	MG595357	U-box domain-containing protein kinase family protein, putative [<i>Theobroma cacao</i>]	XP_007047770.1	1.49E–40
PE15252	F: GATCAGCCTGTGGAGGATGT R: TGAACACCAGAACCACAAA	(CAT) ₇	252–267	58	HEX ¹	MG595358	No hit	—	—
PE17379	F: TGAATTGTGATGGAAGCTGG R: CCCAAGTGATGCCAAGAAGT	(TG) ₈	193–271	58	HEX ²	MG595359	No hit	—	—
PE17828	F: CTCAGACCAACCACAGGGAT R: CGTGTGCTTTCTTCTTGT	(GAA) ₆	104–171	58	HEX ¹	MG595360	Hypothetical protein SORBIDRAFT_09g025255, partial [<i>Sorghum bicolor</i>]	XP_002441363.1	1.57E–06
PE21382	F: GGGTTTACCGTTCGAGACTT R: ACTTGATGAAGCCGATGTC	(CCG) ₇	110–190	58	FAM ¹	MG595361	40S ribosomal protein S29, partial [<i>Zea mays</i>]	DAA51657.1	3.87E–40

Note: T_a = annealing temperature.

^aPCR multiplex sets are indicated as 1 or 2.

^bE-value < 10⁻⁵.

software (version 4.1; Thermo Fisher Scientific). Number of alleles per locus, levels of observed and expected heterozygosity, and the inbreeding coefficient (F_{IS}) were calculated with POPGENE software (version 1.31; Yeh et al., 1999). Hardy–Weinberg equilibrium (HWE) and pairwise linkage disequilibrium for each population were tested with GENEPOP software (version 4.0; Rousset, 2008).

Twenty (38.5%) EST-SSR markers were found to be polymorphic in *P. emblica*. The corresponding sequences of the 20 EST-SSRs were BLASTed against the GenBank nonredundant database using BLASTX (Altschul et al., 1997) (Table 2). The number of alleles per

polymorphic locus varied from 11 to 44, with a mean of 20; levels of observed and expected heterozygosity ranged from 0.240 to 0.868 and 0.754 to 0.933, with means of 0.557 and 0.855 (Table 1). Fourteen loci in each of the three *P. emblica* natural populations showed significant deviations from HWE ($P < 0.001$; Table 1), which may be explained by a deficiency of heterozygotes in the three studied populations, whose habitats have been severely damaged by human activities based on our long-term field survey, as indicated by a relatively high inbreeding coefficient ($F_{IS} = 0.337$; Table 1). Moreover, we observed that geitonogamy mediated by wind and bees was the

TABLE 3. Genetic diversity of the 20 polymorphic EST-SSR markers developed for *Phyllanthus emblica* in *P. reticulatus* and *Leptopus chinensis*.^a

Locus	<i>Phyllanthus reticulatus</i>									<i>Leptopus chinensis</i>		
	Xishan (N = 25)			Heilongtan (N = 25)			Overall (N = 50)			Zhiwuyuan (N = 25)		
	A	H _e	H _e ^b	A	H _e	H _e ^b	A	H _e	H _e ^b	A	H _e	H _e ^b
PE399	6	0.524	0.563	5	0.429	0.559	6	0.476	0.555	1	0.000	0.000
PE788	5	0.520	0.765*	4	0.400	0.728*	5	0.460	0.741	5	0.333	0.693*
PE4618	5	0.546	0.732	5	0.318	0.534*	5	0.432	0.647	6	0.900	0.790
PE6781	5	0.522	0.778*	5	0.478	0.745*	5	0.500	0.758	1	0.000	0.000
PE6950	6	0.542	0.706	6	0.500	0.664	6	0.521	0.683	4	0.393	0.671*
PE7362	9	0.417	0.862*	9	0.417	0.835*	9	0.417	0.847	6	0.357	0.829*
PE7779	5	0.480	0.741*	5	0.400	0.716*	5	0.440	0.729	5	0.308	0.760*
PE8467	5	0.542	0.653	4	0.417	0.636*	5	0.479	0.639	5	0.385	0.769*
PE8480	6	0.409	0.806*	5	0.273	0.771*	6	0.341	0.783	7	0.517	0.812*
PE9600	7	0.818	0.672	7	0.783	0.653	7	0.800	0.655	7	0.333	0.860*
PE10156	10	0.565	0.632	10	0.542	0.642	10	0.553	0.631	8	0.321	0.820*
PE10572	3	0.522	0.405	3	0.500	0.393	3	0.511	0.394	8	0.536	0.830*
PE11297	7	0.542	0.526	7	0.520	0.510	7	0.531	0.513	10	0.393	0.849*
PE14171	4	0.429	0.617	3	0.182	0.498*	4	0.302	0.557	6	0.346	0.508
PE14389	11	0.750	0.786	10	0.708	0.762	11	0.729	0.767	5	0.429	0.761*
PE14485	4	0.280	0.608*	3	0.200	0.553*	4	0.240	0.577	7	0.269	0.789*
PE15252	3	0.783	0.561*	2	0.565	0.449	3	0.674	0.523	8	0.423	0.794*
PE17379	5	0.429	0.676*	4	0.476	0.642	5	0.452	0.653	8	0.261	0.835*
PE17828	10	0.739	0.821	10	0.708	0.833	10	0.723	0.819	7	0.333	0.631*
PE21382	6	0.522	0.800*	5	0.478	0.728*	6	0.500	0.760	10	0.345	0.817*

Note: A = number of alleles per locus; H_e = expected heterozygosity; H_o = observed heterozygosity; N = number of individuals analyzed.

^aLocality and voucher information are provided in Appendix 1.

^bSignificant deviations from Hardy–Weinberg equilibrium after sequential Bonferroni corrections: *P < 0.001.

primary pollination method for *P. emblica* under natural conditions; this is consistent with the higher inbreeding coefficient, indicating that inbreeding could exist in the populations studied. The departures from HWE may also be due to the presence of null alleles. To test this, we used MICRO-CHECKER (version 2.2.3; van Oosterhout et al., 2004) to estimate the average null allele frequency for each locus using the Brookfield 1 equation (Brookfield, 1996); the results showed that null alleles were present at 10 loci (Table 1). No significant linkage disequilibrium was observed for any pair of loci.

Cross-species application was further investigated in two Phyllanthaceae species using the same procedures described above. The 20 EST-SSR markers were amplified successfully in 50 *P. reticulatus* individuals from two natural populations and in 25 *L. chinensis* individuals from a single natural population (Appendix 1), in which only two loci in *L. chinensis* exhibited monomorphisms (Table 3).

CONCLUSIONS

We developed 20 highly polymorphic EST-SSR markers for *P. emblica*. All of these markers showed transferability in related species. These markers will be useful for investigating the population demographics, gene flow, and the genetic resource assessments of *P. emblica*. They are also of great potential to study evolutionary adaptation and genetic relationships among these related *Phyllanthus* and Phyllanthaceae species.

ACKNOWLEDGMENTS

The authors thank Gene Denovo Biotechnology Company (Guangzhou, China) for their help with bioinformatics analysis.

This study was supported by the Fundamental Research Funds for the Central Non-profit Research Institution of the Chinese Academy of Forestry (CAFYBB2016ZX003-2) and the Technology Innovation Talent Project of Yunnan Province (2016HB007).

LITERATURE CITED

- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. H. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* 25: 3389–3402.
- Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.
- Bouck, A., and T. Vision. 2007. The molecular ecologist's guide to expressed sequence tags. *Molecular Ecology* 16: 907–924.
- Brookfield, J. 1996. A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology* 5: 453–455.
- Ellis, J. R., and J. M. Burke. 2007. EST-SSRs as a resource for population genetic analyses. *Heredity* 99: 125–132.
- Fu, L. M., B. F. Niu, Z. W. Zhu, S. T. Wu, and W. Z. Li. 2012. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28: 3150–3152.
- Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644–652.
- Hayden, M. J., T. M. Nguyen, A. Waterman, and K. J. Chalmers. 2008. Multiplex-Ready PCR: A new method for multiplexed SSR and SNP genotyping. *BMC Genomics* 9: 80.
- Kumar, A., and K. Singh. 2012. Isolation of high quality RNA from *Phyllanthus emblica* and its evaluation by downstream applications. *Molecular Biotechnology* 52: 269–275.
- Li, Q. M., and J. L. Zhao. 2007. Genetic diversity of *Phyllanthus emblica* populations in dry-hot valleys in Yunnan. *Biodiversity Science* 15: 84–91 [in Chinese with English abstract].

- Mawalagedera, S. M. U. P., P. Janathani, S. W. M. B. Dunuwille, G. A. D. Perera, C. K. Weebadde, D. S. A. Wijesundara, and S. D. S. S. Sooriyapathirana. 2014. DNA marker analysis reveals genomic diversity and putative QTL associated with drupe traits in *Phyllanthus emblica*. *Ceylon Journal of Science* 43: 31–46.
- Pandey, M., and S. Changtragoon. 2012. Isolation and characterization of microsatellites in a medicinal plant, *Phyllanthus emblica* (Euphorbiaceae). *American Journal of Botany* 99: e468–e469.
- Rousset, F. 2008. GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* 8: 103–106.
- Rozen, S., and H. Skaletsky. 1999. Primer3 on the WWW for general users and for biologist programmers. In S. Misener and S. A. Krawetz [eds.], *Methods in molecular biology*, vol. 132: Bioinformatics methods and protocols, 365–386. Humana Press, Totowa, New Jersey, USA.
- Thiel, T., W. Michalek, R. K. Varshney, and A. Graner. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* 106: 411–422.
- van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills, and P. Shipley. 2004. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4: 535–538.
- Variya, B. C., A. K. Bakrania, and S. S. Patel. 2016. *Emblia officinalis* (Amla): A review for its phytochemistry, ethnomedicinal uses and medicinal potentials with respect to molecular mechanisms. *Pharmacological Research* 111: 180–200.
- Yeh, F. C., R. C. Yang, and T. Boyle. 1999. POPGENE version 1.31: Microsoft Windows-based freeware for population genetic analysis, quick user guide. Centre for International Forestry Research, University of Alberta, Edmonton, Alberta, Canada.

APPENDIX 1. Locality and voucher information for the populations of *Phyllanthus emblica*, *P. reticulatus*, and *Leptopus chinensis* used in this study.

Species	Voucher no. ^a	N	Collection locality ^b	Geographic coordinates	Altitude (m)
<i>Phyllanthus emblica</i> L.	YP-2017621-MH	30	Yongping	25°26'53"N, 99°20'43"E	1486
	BC-2017622-MH	30	Binchuan	25°45'57"N, 100°26'28"E	1742
	YM-2017623-MH	30	Yuanmou	25°40'09"N, 101°49'07"E	1246
<i>Phyllanthus reticulatus</i> Poir.	XS-2017826-MH	25	Xishan	24°57'44"N, 102°37'52"E	2172
	HL-2017827-MH	25	Heilongtan	25°08'26"N, 102°44'24"E	1990
<i>Leptopus chinensis</i> (Bunge) Pojark.	ZW-2017903-MH	25	Zhiwuyuan	25°07'26"N, 102°44'30"E	1921

Note: N = number of individuals sampled.

^aVoucher specimens deposited at the Herbarium of the Kunming Institute of Botany, Chinese Academy of Sciences (KUN), Kunming, China.

^bCollection locality in Yunnan, China.