Development and Validation of EST-SSR Markers for Fokienia hodginsii (Cupressaceae)

Authors: Ding, Mingyan, Meng, Kaikai, Fan, Qiang, Tan, Weizheng, Liao, Wenbo, et al.

Source: Applications in Plant Sciences, 5(3)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1600152

Applications in Plant Sciences 2017 5(3): 1600152

PRIMER NOTE

DEVELOPMENT AND VALIDATION OF EST-SSR MARKERS FOR *FOKIENIA HODGINSII* (CUPRESSACEAE)[1]

MINGYAN DING[2,3], KAIKAI MENG[2], QIANG FAN[2], WEIZHENG TAN[2], WENBO LIAO[2], AND SUFANG CHEN[2,4]

[2]State Key Laboratory of Biocontrol and Guangdong Key Laboratory of Plant Resources, Sun Yat-sen University, Guangzhou 510275, People's Republic of China; and [3]Shunde Polytechnic, Foshan 528300, People's Republic of China

- *Premise of the study: Fokienia hodginsii* (Cupressaceae) is a Tertiary relict evergreen conifer of the monotypic genus *Fokienia*. Polymorphic microsatellite markers were developed to investigate its genetic diversity and population structure.
- *Methods and Results:* RNA transcripts of *F. hodginsii* were sequenced and de novo assembled into 85,818 unigenes, and 1892 simple sequence repeat (SSR) markers were detected from the unigenes. A total of 273 expressed sequence tag–SSR primer pairs were designed and tested, and 129 successfully amplified. Eleven displayed clear polymorphisms in *F. hodginsii*. Amplification of these polymorphic primers across three populations of *F. hodginsii* showed the number of alleles per locus ranged from two to seven, and the expected heterozygosity per locus varied from 0.067 to 0.847. All 11 polymorphic primers amplified in *Thuja occidentalis*, while 10 amplified in *T. standishii*, *Platycladus orientalis*, and *Chamaecyparis obtusa*.
- *Conclusions:* These microsatellite markers will be useful in exploring genetic diversity of *F. hodginsii* and other conifer trees.

Key words: *Chamaecyparis obtusa*; Cupressaceae; *Fokienia hodginsii*; *Platycladus orientalis*; *Thuja occidentalis*; *Thuja standishii*.

Fokienia hodginsii (Dunn) A. Henry & H. H. Thomas (Cupressaceae), endemic to southern China, Laos, and Vietnam, is a Tertiary relict evergreen conifer of the monotypic genus *Fokienia* A. Henry & H. H. Thomas (Fu et al., 1999). It is relatively common in China and Vietnam, but many subpopulations have been decimated by legal and illegal logging, as the wood of this cupressaceous tree is much valued for construction (Thomas and Yang, 2013). Currently, *F. hodginsii* is a protected species (second degree) in China, is on the official List of Rare and Precious Flora and Fauna in Vietnam, and is listed as a Vulnerable species in the IUCN Red List of Threatened Species (Thomas and Yang, 2013).

Applying eight inter-simple sequence repeat markers, Tam et al. (2011) investigated the genetic variation of *F. hodginsii* in Vietnam and found low genetic variability at both the population and the species level. Simple sequence repeat (SSR) markers have the advantage of producing mostly codominant markers, and high-throughput sequencing technologies, such as the Illumina platform, allow for cost-effective development of these markers. In this study, 129 novel expressed sequence tag (EST)–SSR markers were developed from transcriptome data of *F. hodginsii*. Eleven of the 129 EST-SSRs demonstrated clear polymorphism among 45 individuals from three populations of *F. hodginsii* and were transferable in *Thuja occidentalis* L. (Cupressaceae), *T. standishii* (Gordon) Carrière (Cupressaceae), *Platycladus orientalis* (L.) Franco (Cupressaceae), and *Chamaecyparis obtusa* (Siebold & Zucc.) Endl. (Cupressaceae). The other 118 markers were monomorphic in nine randomly selected individuals from three populations of *F. hodginsii*. The development of these novel EST-SSR markers is valuable for conservation of *F. hodginsii* and other related conifer species.

METHODS AND RESULTS

One seedling of *F. hodginsii* was collected from Taoyuandong, Hunan Province, China (Appendix 1), and planted in a greenhouse at Sun Yat-sen University. Two months later, its fresh leaves were harvested and total RNAs were extracted immediately using a modified cetyltrimethylammonium bromide (CTAB) method (Fu et al., 2004). Then, a normalized cDNA library was constructed with the Illumina protocol and subsequently sequenced using the HiSeq 2500 Sequencing System (Illumina, San Diego, California, USA). Reads containing primers or adapters, unknown "N" bases, or more than 10% bases with a Q value < 20 were removed with NGSQCToolkit_v2.3.3 (Patel and Jain, 2012). A total of 25.1 million cleaned 125-bp paired-end reads were obtained and further de novo assembled into 85,818 unigenes using Trinity version 2.3.2 (Grabherr et al., 2011) with the default parameters. The bioproject PRJNA358452 was created in the National Center for Biotechnology Information (NCBI) database, the raw data were deposited in NCBI's Sequence Read Archive (SRA) database (SRR5127884), and the assembled unigenes were deposited in the NCBI Transcriptome Shotgun Assembly (TSA) database (GFDI00000000) after removing foreign contaminations. For these unigenes, the minimal length was 201 bp, the average length was 686 bp, and the N50 value was 1174 bp.

[1]Manuscript received 4 December 2016; revision accepted 24 January 2017.

This work was supported by the Natural Science Foundation of Guangdong Province, China (2016A030313326), the National Natural Science Foundation of China (31670189), the Special Program for Science and Technology Basic Research of the Ministry of Science and Technology of China (2013FY111500), the Science and Technology Planning Project of Guangdong Province, China (2015A030302020), the Fundamental Research Funds for the Central Universities (161gjc38), and Chang Hungta Science Foundation of Sun Yat-sen University.

[4]Author for correspondence: chsuf@mail.sysu.edu.cn

doi:10.3732/apps.1600152

TABLE 1. Characteristics of 11 polymorphic EST-SSR loci developed for *Fokienia hodginsii*.

Locus[a]	Primer sequences (5′–3′)	Fluorescent label	Repeat motif	Allele size range (bp)	GenBank accession no.	Putative function	*E*-value
F017	F: AAGACAAGATGCTCAGATCA	HEX	$(AG)_7$	194–202	KY010850	Unknown [*Picea sitchensis*]	7e-47
	R: GTGGTAGCCTAGAACTTCAT						
F049	F: CAATGTTCCTTCTGTGTCTG	TAM	$(CAG)_7$	221–239	KY010922	PREDICTED: RNA-binding protein 42 isoform X1 [*Vitis vinifera*]	3e-88
	R: TTGATACTGAGGTGCTTGAA						
F063	F: ATCACTCATCACTCGCATT	FAM	$(CTC)_7$	193–205	KY010960	Hypothetical protein AMTR_s00032p00100570 [*Amborella trichopoda*]	9e-05
	R: GGCAATCCGTCCTTGAAT						
F071	F: CTGCTGCTCAACCAACTG	HEX	$(GAC)_6$	193–269	KY010955	PREDICTED: la-related protein 6B [*Vitis vinifera*]	1e-92
	R: AGAGGAAGACGATGTACTGT						
F073	F: GGAATGTTGCCTGGAGTT	TAM	$(CTG)_6$	248–260	KY010976	PREDICTED: UBP1-associated protein 2C [*Gossypium raimondii*]	8e-86
	R: AGCCCGTAAAGATTGTTGA						
F083	F: CCACGAGAGGAGAAGGAT	FAM	$(GAG)_5...(AAG)_5$	257–269	KY010961	PREDICTED: probable anion transporter 6, chloroplastic [*Vitis vinifera*]	0.0
	R: GCTGACGCTGTTACTGAT						
F127	F: CCTTCAACTCATCATAGAATGG	FAM	$(TTC)_6$	236–239	KY010854	Not found	
	R: TGAGCCTTCACTGCTAATG						
F154	F: ACCTATGTGATCTTCAAGCA	FAM	$(TCC)_5$	258–261	KY010919	PREDICTED: uncharacterized protein At1g01500 isoform X2 [*Ziziphus jujuba*]	1e-27
	R: AGTAACTCCGAGAACAACAA						
F207	F: GAACAACCAGAATTGCCATT	FAM	$(TTC)_5$	261–273	KY011058	Hypothetical chloroplast RF1 [*Callitropsis vietnamensis*]	0.0
	R: CGACTCAAGAATCGGATCA						
F210	F: TGGAAGGAAGAAGGAAGATG	TAM	$(GTG)_5$	270–294	KY011062	PREDICTED: uncharacterized protein LOC107942469 [*Gossypium hirsutum*]	5e-47
	R: CGGACCTCATGTAAGAACTT						
F236	F: GCATCTTATCGGATTCTAGTTC	FAM	$(TCC)_5$	193–205	KY010889	PREDICTED: 1-phosphatidylinositol-3-phosphate 5-kinase FAB1B isoform X1 [*Vitis vinifera*]	1e-73
	R: GACAGGAAGCATAGAGGAAG						

[a]Annealing temperature for all loci was 52°C.

TABLE 2. Polymorphism of the 11 EST-SSRs in populations of *Fokienia hodginsii*.[a]

	TYD (n = 15)			NJ (n = 15)			MES (n = 15)		
Locus	A	H_o	H_e[b]	A	H_o	H_e[b]	A	H_o	H_e[b]
F017	3	0.667	0.522	2	0.133	0.129	3	0.400	0.503
F049	4	0.400	0.405	3	0.733	0.522	4	0.533	0.591
F063	4	0.867	0.632	3	1.000	0.683	3	0.800	0.570
F071	5	0.600	0.789	4	0.667	0.591	7	0.714	0.847
F073	2	0.067	0.067	2	0.133	0.129	2	0.214	0.198
F083	3	1.000	0.641	3	1.000	0.674	3	0.667	0.570
F127	3	0.733	0.503	2	0.400	0.331	2	0.133	0.515
F154	2	0.333	0.287	3	0.467	0.503	2	0.286	0.423
F207	2	0.067	0.467	2	0.067	0.067	2	0.133	0.239
F210	4	0.600	0.513	3	0.733	0.646	4	0.267	0.577
F236	3	1.000	0.549	3	1.000	0.549	3	1.000	0.600

Note: A = number of alleles; H_e = expected heterozygosity; H_o = observed heterozygosity; n = number of individuals collected for each population.

[a] Voucher and locality information are available in Appendix 1.

[b] No loci showed significant deviation from Hardy–Weinberg equilibrium.

To search for SSR motifs containing two to six nucleotides across these unigenes, the MISA tool (Thiel et al., 2003) was applied with the default parameters except that the settings for mononucleotide repeats were removed from analysis. The results showed that 1892 SSRs were detected in 1709 unigenes. Among them, trinucleotide repeats (54.9%) were the most common, followed by dinucleotide (41.4%), tetranucleotide (2.3%), hexanucleotide (1.1%), and pentanucleotide repeats (0.3%). Based on these unigenes containing SSR loci, we successfully designed 273 pairs of primers using Primer3 (Rozen and Skaletsky, 1999; http://frodo.wi.mit.edu/primer3). Unigenes containing these 273 SSR loci were deposited in GenBank (accession number: KY010833–KY011105).

Genomic DNA was isolated from leaves of 45 individuals from three populations of *F. hodginsii* (15 individuals for each population; Appendix 1) and 12 individuals from four other conifer species (*T. occidentalis*, *T. standishii*, *P. orientalis*, and *C. obtusa*; Appendix 1) using the CTAB method (Doyle and Doyle, 1987). All specimens are deposited at the Herbarium of Sun Yat-sen University, Guangdong, China.

In the first PCR trial, three individuals were randomly selected from each population to amplify the 273 pairs of primers. PCR amplifications were performed according to Fan et al. (2013), with an annealing temperature of 52°C. The PCR products were electrophoresed on a 10% polyacrylamide gel and visualized by silver nitrate staining. The band size of the amplicons was estimated by comparison with a 100-bp DNA ladder (Promega Corporation, Madison, Wisconsin, USA). The results showed that 129 primers could be successfully amplified and size polymorphism was only detected in 11 of them (Table 1), while the other 118 were monomorphic (Appendix S1). Functional annotations for these 11 polymorphic markers were performed by searching against the NCBI nonredundant protein database with the BLASTX program. The results showed that 10 had significant BLASTX hits to the protein database and two were annotated as plastid genes.

The 11 polymorphic primer sets were tested for polymorphisms in 45 individuals from three populations of *F. hodginsii*. PCR products were resolved with an internal size standard (GeneScan 500 LIZ; Applied Biosystems, Foster City, California, USA) by using an ABI 3730xl DNA Analyzer (Applied Biosystems), and the allele size was estimated using Peak Scanner Software version 1.0 (Applied Biosystems). The number of alleles, observed heterozygosity, expected heterozygosity, deviation from Hardy–Weinberg equilibrium, and linkage disequilibrium were calculated using the software POPGENE version 1.32 (Yeh et al., 1997). Null alleles were checked using the program MICRO-CHECKER version 2.2.3 (van Oosterhout et al., 2004). The results showed that the number of alleles ranged from two to seven, the observed and expected heterozygosity ranged from 0.067 to 1.000 and from 0.067 to 0.847, respectively, and no loci showed significant deviation from Hardy–Weinberg equilibrium (Table 2). Statistical tests for linkage disequilibrium showed that 21 of the 165 tests showed significant results ($P < 0.05$), and 15 of them are found in the population MES (Appendix S2). Signs of null alleles were detected on locus F207 in population TYD and loci F127 and F210 in population MES.

Finally, 12 individuals of four other conifer species (*T. occidentalis*, *T. standishii*, *P. orientalis*, and *C. obtusa*) were used to test the transferability of the 129 loci by performing PCR amplification and inspecting their PCR products on 1% agarose gels. Among the 118 monomorphic loci, 88, 86, 78, and 78 can be amplified in *T. standishii*, *T. occidentalis*, *P. orientalis*, and *C. obtusa*, respectively (Appendix S3). All of the 11 polymorphic loci can be amplified in *T. occidentalis*, and 10 of them can be amplified in the three other species (Table 3).

CONCLUSIONS

In this study, we sequenced the transcriptome of *F. hodginsii* and developed 129 novel EST-SSR markers for the species, which showed high transferability in four other conifer trees. Eleven loci demonstrated polymorphism in 45 individuals of *F. hodginsii*, while the other 118 primers were monomorphic among 12 randomly selected individuals.

Testing of polymorphism of 129 loci was extremely limited as only three populations were collected; higher levels of polymorphism may be revealed with the survey of additional populations. Future population genetic studies of *F. hodginsii* based on the 11 polymorphic EST-SSR markers will provide fundamental genetic information for developing conservation strategies. These 129 novel EST-SSR markers may also be useful for genetic studies of other conifer trees.

LITERATURE CITED

DOYLE, J. J., AND J. L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

FAN, Q., S. F. CHEN, M. W. LI, S. Y. HE, R. C. ZHOU, AND W. B. LIAO. 2013. Development and characterization of microsatellite markers from the transcriptome of *Firmiana danxiaensis* (Malvaceae s.l.). *Applications in Plant Sciences* 1: 1300047.

FU, L. G., Y. F. YU, AND F. ALJOS. 1999. Cupressaceae Bartlett. *In* Z. Y. Wu, P. H. Raven, and D. Y. Hong [eds.], Flora of China, vol. 4, 62–77. Science Press, Beijing, China, and Missouri Botanical Garden Press, St. Louis, Missouri, USA.

FU, X. H., S. L. DENG, G. H. SU, Q. L. ZENG, AND S. H. SHI. 2004. Isolating high-quality RNA from mangroves without liquid nitrogen. *Plant Molecular Biology Reporter* 22: 197.

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.

PATEL, R. K., AND M. JAIN. 2012. NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. *PLoS ONE* 7: e30619.

TABLE 3. Cross-amplification of 11 *Fokienia hodginsii* EST-SSR markers in other conifer trees.

Species	N	F017	F049	F063	F071	F073	F083	F127	F154	F207	F210	F236
Thuja occidentalis	3	+	+	+	+	+	+	+	+	+	+	+
Thuja standishii	3	+	+	+	+	+	+	+	+	+	—	+
Platycladus orientalis	3	+	+	+	—	+	+	+	+	+	+	+
Chamaecyparis obtusa	3	+	+	+	+	+	+	+	+	+	—	+

Note: + = primers could be successfully amplified in all individuals; — = primers could not be amplified in any individual; N = number of individuals.

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.

Tam, N. M., N. T. P. Trang, and N. T. Hoa. 2011. Genetic diversity of an endangered species, *Fokienia hodginsii* (Cupressaceae). *African Journal of Biotechnology* 10: 15838–15844.

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.

Thomas, P., and Y. Yang. 2013. *Fokienia hodginsii*. The IUCN Red List of Threatened Species 2013: e.T32351A2815809. Website http://dx.doi.org/10.2305/IUCN.UK.2013-1.RLTS.T32351A2815809.en [accessed 15 January 2017].

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.

Yeh, F. C., R. C. Yang, T. J. B. Boyle, Z. H. Ye, and J. X. Mao. 1997. POPGENE: User-friendly shareware for population genetic analysis. Molecular Biology and Biotechnology Center, University of Alberta, Canada. Website https://sites.ualberta.ca/~fyeh/popgene.html [accessed 10 February 2017].

Appendix 1. Voucher and location information for specimens used in this study. All specimens are deposited at the Herbarium of Sun Yat-sen University, Guangdong, China.

Species	Voucher	Population	Collection locality	Geographic coordinates	*N*
Fokienia hodginsii (Dunn) A. Henry & H. H. Thomas	TYD20141203[a]	TYD	Yanling, Hunan, China	26°28′32.25″N, 114°02′48.28″E	1
Fokienia hodginsii	Ding20150502[b]	TYD	Yanling, Hunan, China	26°28′30.48″N, 114°02′52.20″E	15
Fokienia hodginsii	Ding20150601[b]	MES	Xingan, Guangxi, China	25°47′19.32″N, 110°34′26.04″E	15
Fokienia hodginsii	Ding20150501[b]	NJ	Nanjing, Fujian, China	24°30′17.10″N, 117°18′52.25″E	15
Thuja occidentalis L.	Ding201608001[c]	PEK	Beijing, China	39°59′53.34″N, 116°12′11.95″E	3
Thuja standishii (Gordon) Carrière	Ding201608002[c]	PEK	Beijing, China	39°59′53.34″N, 116°12′11.95″E	3
Platycladus orientalis (L.) Franco	Ding201608003[c]	PEK	Beijing, China	39°59′53.34″N, 116°12′11.95″E	3
Chamaecyparis obtusa (Siebold & Zucc.) Endl.	Ding201608004[c]	PEK	Beijing, China	39°59′53.34″N, 116°12′11.95″E	3

Note: *N* = number of individuals.

[a] Sample used in transcriptome sequencing.

[b] Samples used for initial PCR amplification trials and detailed polymorphism evaluation.

[c] Samples used for transferability tests; all collected species were cultivated.