Development and Multiplexed Amplification of SSR Markers for Thuja occidentalis (Cupressaceae) Using Shotgun Pyrosequencing

Authors: Xu, Huaitong, Tremblay, Francine, and Bergeron, Yves

Source: Applications in Plant Sciences, 1(5)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1200427

Applications in Plant Sciences

Applications in Plant Sciences 2013 1(5): 1200427

Primer Note

Development and multiplexed amplification of SSR markers for *Thuja occidentalis* (Cupressaceae) using shotgun pyrosequencing[1]

Huaitong Xu[2,3,4], Francine Tremblay[2], and Yves Bergeron[2]

[2]Chaire industrielle CRSNG-UQAT-UQAM en Aménagement Forestier Durable, Institut de recherche sur les forêts, Université du Québec en Abitibi-Témiscamingue, 445 Boulevard de l'Université, Rouyn-Noranda, Québec J9X 5E4, Canada; and
[3]Northwest A&F University, 3 Taicheng Road, Yangling, Shaanxi 712100, People's Republic of China

- *Premise of the study:* Sixteen novel, polymorphic, multiplexed microsatellite loci were developed for eastern white cedar (*Thuja occidentalis*) using simple sequence repeat (SSR)–enriched shotgun pyrosequencing.
- *Methods and Results:* Sixteen loci were tested on a panel of 24 individuals from different populations. The number of observed alleles ranged from four to 22. Four sets of multiplex PCR for the 16 loci were then carried out on 60 individuals of two populations from islands of FERLD Duparquet Forest, Canada. Mean number of alleles, observed heterozygosity, and expected heterozygosity were respectively 5.75, 0.594, and 0.574 for Island 58, and 5.50, 0.704, and 0.624 for Island 134.
- *Conclusions:* Four sets of multiplex microsatellite loci can be used for future genetic studies, which includes investigating genetic diversity and structure, and fragmentation and regeneration studies.

Key words: 454 GS-FLX Titanium; microsatellite marker; next-generation sequencing; population genetics; shotgun pyrosequencing; *Thuja occidentalis*.

Eastern white cedar (*Thuja occidentalis* L.) is a native, wind-pollinated conifer with a broad distribution across North America (Fowells, 1965). The species' range extends from the Gulf of St. Lawrence in the east to southeastern Manitoba in the west, and from James Bay in the north to Tennessee and North Carolina in the south (Fowells, 1965). A member of the Cupressaceae, it is also commonly called eastern arborvitae, American arborvitae, northern white cedar, Atlantic red cedar, and swamp cedar in English (USDA NRCS, 2013), and thuya occidental, cèdre, balai, cèdre blanc, thuier cèdre, and arborvitae in French (Brouillet et al., 2010). Eastern white cedar (EWC) is listed as endangered in Indiana, Massachusetts, and New Jersey, as a threatened species in Connecticut, Illinois, Kentucky, and Maryland, and of special concern in Tennessee (USDA NRCS, 2013). Genetic analyses previously conducted on EWC have been mainly based on allozyme markers (Hofmeyer et al., 2007), while highly polymorphic markers such as microsatellites have not been developed for EWC. We report on the development and characterization of microsatellite markers for EWC using shotgun pyrosequencing on a simple sequence repeat (SSR)–enriched library (Malausa et al., 2011).

[1]Manuscript received 10 August 2012; revision accepted 11 November 2012.

The authors thank V. Paul, A. Leduc, G. Trudeau, R. Libert, and S. Chauchard for help in the field and Dr. W. F. J. Parsons for corrections to the English. This research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC; STPGP 336871) to F.T., and also supported by an Industrial Innovation Doctoral Scholarship from NSERC, a BMP innovation doctoral scholarship from the Fonds de recherche du Québec–Nature et technologies (FRQNT) and Tembec, and a scholarship from the Forêt d'enseignement et de recherche du lac Duparquet (FERLD) to H.X.

[4]Author for correspondence: huaitong.xu@uqat.ca

doi:10.3732/apps.1200427

METHODS AND RESULTS

Foliage of EWC individuals from 14 sites across northern Quebec (Appendix 1) was collected and maintained at −20°C before genetic analysis. Genomic DNA was extracted using the DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany). DNA extracts of 14 individuals were combined and sent on dry ice to Genoscreen (Lille, France) for microsatellite-enriched GS-FLX library construction following the methodology developed by Malausa et al. (2011). Briefly, main steps included: (1) digestion of genomic DNA with *Rsa*I (Fermentas International Inc., Burlington, Ontario, Canada); (2) enrichment of microsatellite sequences in fragmented DNA with eight types of probes (TG, TC, AAC, AAG, AGG, ACG, ACAT, ACTC), which was accomplished by using Dynabeads (Invitrogen, Carlsbad, California, USA); and (3) PCR amplification of enriched DNA with primers specific to the adapter sequences (Malausa et al., 2011). In total, 11,393 raw sequences with an average read length of 400 bp were obtained, with 2175 sequences containing microsatellite motifs. One hundred seventeen of the sequences successfully had primers designed for them using QDD software (Meglécz et al., 2010) using default parameters except optimal primer length of 22 bp (range 18–27 bp) and 50% GC content (range 40–60%).

To minimize screening costs, we initially selected 48 of 117 pairs following criteria detailed in Lepais and Bacles (2011). In brief, we restricted our selection to loci with hexa-, tetra-, and dinucleotide motif types. Dinucleotide motif was limited to AC, CA, TG, GT, AG, GA, CT, and TC types, because AT and TA types are notoriously hard to amplify (Temnykh et al., 2001). Each of the selected 48 loci was initially tested for amplification with unlabeled primers (Invitrogen) on a screening panel that included seven EWC trees collected across northern Quebec (one tree per site) (Appendix 1) and a negative control. Amplifications were carried out in a total volume of 10 μL using four 96-well Mastercycler pro S PCR systems (Eppendorf Gmbh, Wesselling-Berzdorf,

TABLE 1. Characteristics of 16 microsatellites validated for *Thuja occidentalis.*

Locus	Primer sequences (5′–3′)	GenBank accession no.	Dye	Repeat motif	Size (bp)[a]	T_a (°C)	MG	Observed		
								A	Min	Max
TO791	F: AAGAGATTTATTTGCCCTCCG	JX475983	VIC	$(CA)_{12}$	141	57	1	16	133	167
	R: ATGGTTGATGGACTCCTTGG									
TO605	F: GAATAACTTCTTCTGGGAAAGATACA	JX475984	PET	$(AC)_8$	190	59	1	10	174	196
	R: GAGGTGGAAAGAAGTGGATAAAA									
TO328	F: CCCGCAACACCTACTTGTCT	JX475985	FAM	$(TACA)_7$	215	57	1	4	203	215
	R: TGCTCCATGTTTGAAGTTGC									
TO53	F: AAATGGCCCATAAGCACAAA	JX475986	NED	$(CA)_5$	184	58	1	6	174	184
	R: GGATGTTTCCAGTTGACGGT									
TO925	F: TGTGTTTGTGGTGGCTGACT	JX475987	FAM	$(TG)_{20}$	151	58	2	22	129	217
	R: CATTCATACATTTCCCATCCA									
TO727	F: GAGATTCCTTTAAAATATTGGCAT	JX475988	VIC	$(GA)_{11}$	241	57	2	14	233	325
	R: CCCTCCCATTCCTCTTAATG									
TO659	F: TGATGCACCAATTTTCTTTGG	JX475989	PET	$(CT)_9$	191	56	2	7	181	195
	R: TGATGCACTTTAAGGTGTAGGG									
TO29	F: TGCAGTGTTAGTGGAGCAACTT	JX475990	NED	$(CA)_5$	162	57	2	14	148	186
	R: TCATTGTTTATTCCCTAAGATGGA									
TO737	F: GAGCAAGAAGGAGAGTGGGA	JX475991	PET	$(AGAT)_{11}$	124	63	3	6	102	130
	R: CCTAGGTTGCCTTGTTGTCC									
TO587	F: GTGCCAAACTTTTCAAGGTAAGA	JX475992	NED	$(CT)_8$	167	62	3	13	139	211
	R: GCAAGAGCACAAATGATCACA									
TO512	F: TGCATAACAACTCTTCTTAAATCAGC	JX475993	FAM	$(CT)_8$	194	63	3	11	146	212
	R: AGGTCCTATCTAGGTCTTAGACAACTT									
TO503	F: CTTGTCCGTCTGACATGTGTTT	JX475994	VIC	$(GA)_8$	190	55	3	12	138	202
	R: CACATAGGTTAAGGGTAGTTTCCT									
TO715	F: CATCTACATGGTCGATGATTTAAC	JX475995	VIC	$(AG)_{10}$	106	60	4	6	100	110
	R: TATCCCAAACCAGCAAAACC									
TO521	F: CAAATATGGCACCAATGCCT	JX475996	PET	$(CT)_8$	121	54	4	17	113	239
	R: CAATTTCCTCAGGTTTGGGA									
TO418	F: ATGCTTTTCTAACCCTTTTGGA	JX475997	NED	$(AC)_7$	253	61	4	8	163	255
	R: TGATCAGTTGGATTTCTAGATTGC									
TO20	F: TTTGGCTTGTAGGTGGTTTT	JX475998	FAM	$(TG)_5$	192	57	4	15	168	204
	R: CTCCATTTTGGAGTGTTGGT									

Note: *A* = number of alleles observed; Max = maximum allele size observed during screening; MG = multiplex group; Min = minimum allele size observed during screening; T_a = annealing temperature.

[a] Product size from shotgun pyrosequencing.

Germany). Each reaction mixture contained 1 µL of DNA extract, 5 µL of 2× QIAGEN Multiplex PCR Master Mix (QIAGEN), and a final concentration of 0.2 µM for each forward and reverse primer. The PCR program consisted of an initial heat-activation step at 95°C for 15 min, 36 cycles of three-step cycling (denaturation at 94°C for 30 s, annealing at 54°C for 90 s, extension at 72°C for 60 s), and a final extension at 60°C for 30 min. A total of 2.5 µL of PCR products were visualized on 3% agarose gel (Promega Corporation, Madison, Wisconsin, USA), with electrophoretic migration performed at 100 V for 20 min on the Bio-Rad Imaging System (Bio-Rad, Montreal, Canada).

The initial test on 48 loci using agarose gel showed that 38 had amplified products, and four of 38 had three or more products in one PCR reaction (nonspecific). Among the remaining 34 loci, 16 had one or two amplification products, variable in size. Thus, we used all of them to verify polymorphisms and further design multiplex PCR. Each was tested with fluorescent dye–labeled primers (Applied Biosystems, Carlsbad, California, USA) on a panel that included 24 EWC individuals collected from 24 sites across northern Quebec (one individual per site) (Appendix 1) plus one negative control. PCR cycles were the same as those mentioned previously, except for increased annealing temperatures to achieve specific amplifications (Table 1). A total of 2 µL of 1 : 100 diluted PCR products labeled with four different dyes (6-FAM, VIC, NED, and PET; Applied Biosystems) were mixed with 8.35 µL of Hi-Di Formamide (Applied Biosystems) and 0.15 µL of GeneScan 500 LIZ Size Standard (Applied Biosystems), and sent to GenoQuebec (Montreal, Canada) for genotype reading on an ABI 3730 genetic analyzer. Results were analyzed with GeneMapper 3.7 software (Applied Biosystems).

All 16 loci showed interpretable, repeatable, and polymorphic patterns (Table 1). We used Multiplex Manager (Holleley and Geerts, 2009) to design and optimize multiplex PCRs to find annealing temperatures for each multiplex group to ensure specific amplifications and avoid complementary sequences among primers. Primer pairs were multiplexed to reduce amplification costs (Table 1). Coamplifications of all multiplexed primers were tested on two populations (30 trees per population) sampled from islands in Lake Duparquet, northwestern Quebec (Table 2). PCR cycles were the same as those mentioned previously, except for multiplexed annealing temperatures (M1 at 57°C, M2 at 56°C, M3 at 55°C, M4 at 54°C). PCR products were genotyped as previously detailed.

The number of different alleles per locus (*A*), observed heterozygosity (H_o), and expected heterozygosity (H_e) were calculated in GenAlEx version 6.2 (Peakall and Smouse, 2006). Inbreeding coefficient (F_{IS}) and Hardy–Weinberg equilibrium (HWE) tests were done in FSTAT version 2.9.3 (Goudet, 2001). Null allele presence was checked in MICRO-CHECKER (Van Oosterhout et al., 2004). Mean values for *A*, H_o, and H_e were, respectively, 5.75, 0.594, and 0.574 on Island 58, and 5.50, 0.704, and 0.624 on Island 134 (Table 2). F_{IS} ranged from −0.706 to 0.665 on Island 58, and from −0.357 to 0.194 on Island 134 (Table 2).

CONCLUSIONS

Shotgun pyrosequencing has proved to be effective for isolating microsatellite markers in EWC. The four sets of multiplex microsatellite loci that were developed here for the first time will facilitate future studies of population genetics in EWC, including investigating phylogeographic patterns of postglacial expansion in North America, and studying the impacts of habitat fragmentation on population genetic structure and gene flow. They will also help resolve questions regarding

TABLE 2. Results of initial primer screening in *Thuja occidentalis* samples from Lake Duparquet, Lake Duparquet Research & Teaching Forest, Quebec, Canada.

	Island 58 ($N = 30$)[a]					Island 134 ($N = 30$)[a]				
Locus	A	H_o	H_e	F_{IS}	Null alleles present (frequency)	A	H_o	H_e	F_{IS}	Null alleles present (frequency)
TO53	5.00	0.567	0.705	0.212	no	4.00	0.900	0.652	–0.366	no
TO328	4.00	0.567	0.585	0.048	no	3.00	0.667	0.491	–0.343	no
TO605	3.00	0.200	0.580	0.665*	yes (0.24)	2.00	0.367	0.433	0.169	no
TO791	12.00	0.667	0.794	0.177	no	11.00	0.800	0.839	0.063	no
TO29	7.00	0.500	0.543	0.096	no	7.00	0.700	0.712	0.034	no
TO659	4.00	0.400	0.581	0.326	yes (0.11)	6.00	0.500	0.608	0.194	no
TO727	2.00	0.833	0.486	–0.706*	no	5.00	0.900	0.686	–0.296	no
TO925	18.00	0.800	0.847	0.072	no	10.00	0.667	0.770	0.151	no
TO503	6.00	0.967	0.644	–0.487*	no	4.00	0.633	0.521	–0.199	no
TO512	5.00	0.367	0.322	–0.121	no	7.00	0.733	0.556	–0.305	no
TO587	5.00	1.000	0.712	–0.391*	no	7.00	0.867	0.710	–0.204	no
TO737	5.00	0.733	0.634	–0.139	no	4.00	0.867	0.624	–0.375	no
TO20	2.00	0.400	0.320	–0.234	no	4.00	0.933	0.676	–0.366	no
TO418	3.00	0.067	0.065	–0.009	no	3.00	0.433	0.443	0.038	no
TO521	8.00	0.933	0.777	–0.185	no	8.00	0.800	0.738	–0.067	no
TO715	3.00	0.500	0.583	0.159	no	3.00	0.500	0.529	0.072	no
Mean	5.75	0.594	0.574	—	—	5.50	0.704	0.624	—	—
SE	1.031	0.068	0.050	—	—	0.658	0.045	0.030	—	—

Note: A = number of alleles; F_{IS} = inbreeding coefficient; H_e = expected heterozygosity; H_o = observed heterozygosity.
* $P \leq 5\%$; Bonferroni correction was applied, and indicative adjusted P value for 5% nominal level was 0.0031.
[a] Geographical coordinates: Island 58 (48°26′41.4″N, 79°15′51.9″W), Island 134 (48°27′52.5″N, 79°16′19.6″W).

regeneration patterns in this species along postfire successions (Bergeron, 2000).

LITERATURE CITED

BERGERON, Y. 2000. Species and stand dynamics in the mixed woods of Quebec's southern boreal forest. *Ecology* 81: 1500–1516.

BROUILLET, L., F. COURSOL, S. J. MEADES, M. FAVREAU, M. ANIONS, P. BÉLISLE, AND P. DESMET. 2010. VASCAN, the Database of Vascular Plants of Canada. Website http://data.canadensys.net/vascan/ [accessed 8 February 2012].

FOWELLS, H. A. 1965. Silvics of forest trees of the United States. USDA Agricultural Handbook 271. U.S. Department of Agriculture, Washington, D.C., USA.

GOUDET, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Website http://www2.unil.ch/popgen/softwares/fstat.htm [accessed 8 February 2012].

HOFMEYER, P. V., L. S. KENEFIC, AND R. S. SEYMOUR. 2007. Northern white-cedar: An annotated bibliography. *Cooperative Forestry Research Unit Research Report* 01: 1–30.

HOLLELEY, C. E., AND P. G. GEERTS. 2009. Multiplex Manager 1.0: A cross-platform computer program that plans and optimizes multiplex PCR. *BioTechniques* 46: 511–517.

LEPAIS, O., AND C. F. E. BACLES. 2011. Comparison of random and SSR-enriched shotgun pyrosequencing for microsatellite discovery and single multiplex PCR optimization in *Acacia harpophylla* F. Muell. ex Benth. *Molecular Ecology Resources* 11: 711–724.

MALAUSA, T., A. GILLES, E. MEGLÉCZ, H. BLANQUART, S. DUTHOY, C. COSTEDOAT, V. DUBUT, ET AL. 2011. High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Molecular Ecology Resources* 11: 638–644.

MEGLÉCZ, E., C. COSTEDOAT, V. DUBUT, A. GILLES, T. MALAUSA, N. PECH, AND J.-F. MARTIN. 2010. QDD: A user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics (Oxford, England)* 26: 403–404.

PEAKALL, R., AND P. E. SMOUSE. 2006. GenAlEx 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288–295.

TEMNYKH, S., G. DECLERCK, A. LUKASHOVA, L. LIPOVICH, S. CARTINHOUR, AND S. MCCOUCH. 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.

USDA, NRCS. 2013. The PLANTS database. National Plant Data Team, Greensboro, North Carolina, USA. Website http://plants.usda.gov [accessed 27 March 2013].

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.

Appendix 1. Voucher information for *Thuja occidentalis* samples. All samples were preserved at the Institut de recherche sur les forêts, Université du Québec en Abitibi-Témiscamingue, Canada.

No.	Site	Latitude	Longitude	Location	Country	Year of collection
1	MZ1	49°52′31.44″N	74°23′34.224″W	Chibougamau	Canada	2007
2	MZ2	49°54′32.976″N	74°19′21.396″W	Chibougamau	Canada	2007
3	MZ3	49°57′12.636″N	74°13′44.688″W	Chibougamau	Canada	2007
4	MZ4	49°38′30.336″N	74°20′2.58″W	Chibougamau	Canada	2007
5	MZ5	48°55′39.792″N	78°53′8.808″W	James Bay	Canada	2007
6	MZ6	49°25′23.412″N	79°12′39.492″W	James Bay	Canada	2007
7	MZ7	49°51′30.708″N	78°36′25.956″W	James Bay	Canada	2007
8	MZ8	49°53′0.564″N	78°38′45.78″W	James Bay	Canada	2007
9	MZ9	49°51′21.924″N	78°38′41.496″W	James Bay	Canada	2007
10	DZ1	48°32′24.72″N	78°38′30.696″W	Abitibi	Canada	2007
11	DZ2	48°28′12.54″N	79°27′8.46″W	Abitibi	Canada	2007
12	DZ3	48°28′47.244″N	79°26′12.624″W	Abitibi	Canada	2007
13	DZ4	48°25′53.796″N	79°24′6.588″W	Abitibi	Canada	2007
14	DZ5	48°15′6.656″N	78°34′29.208″W	Abitibi	Canada	2007
15	DZ6	48°25′51.636″N	79°23′2.976″W	Abitibi	Canada	2007
16	DZ7	48°12′4.752″N	79°25′8.796″W	Abitibi	Canada	2007
17	CZ1	47°25′45.192″N	78°40′42.528″W	Témiscamingue	Canada	2007
18	CZ2	47°25′0.084″N	78°40′55.704″W	Témiscamingue	Canada	2007
19	CZ3	47°23′44.052″N	78°43′53.904″W	Témiscamingue	Canada	2007
20	CZ4	47°20′2.18″N	79°23′33.396″W	Témiscamingue	Canada	2007
21	CZ5	47°18′39.96″N	78°30′55.8″W	Témiscamingue	Canada	2007
22	CZ6	47°27′14.22″N	78°35′15.54″W	Témiscamingue	Canada	2007
23	CZ7	47°25′8.184″N	78°40′42.384″W	Témiscamingue	Canada	2007
24	CZ8	47°24′56.844″N	78°42′41.94″W	Témiscamingue	Canada	2007
25	IS58	48°26′41.4″N	79°15′51.9″W	Abitibi	Canada	2008
26	IS134	48°27′52.5″N	79°16′19.6″W	Abitibi	Canada	2008