Development of Polymorphic Genic-SSR Markers by cDNA Library Sequencing in Boxwood, Buxus spp. (Buxaceae)

Authors: Thammina, Chandra S., Olsen, Richard T., Malapi-Wight, Martha, Crouch, Jo Anne, and Pooler, Margaret R.

Development of polymorphic genic-SSR markers by cDNA library sequencing in boxwood, *Buxus* spp. (Buxaceae)[1]

Chandra S. Thammina[2,3], Richard T. Olsen[2], Martha Malapi-Wight[4], Jo Anne Crouch[4], and Margaret R. Pooler[2,5]

[2]United States Department of Agriculture, Agricultural Research Service (USDA-ARS), U.S. National Arboretum, Floral and Nursery Plants Research Unit, 10300 Baltimore Avenue, Building 010A, Beltsville, Maryland 20705 USA; [3]Department of Plant Biology and Pathology, Rutgers University, 59 Dudley Road, New Brunswick, New Jersey 08901 USA; and [4]USDA-ARS, Systematic Mycology and Microbiology Laboratory, 10300 Baltimore Avenue, Building 010A, Beltsville, Maryland 20705 USA

- *Premise of the study:* Genic microsatellites or simple sequence repeat (genic-SSR) markers were developed in boxwood (*Buxus* taxa) for genetic diversity analysis, identification of taxa, and to facilitate breeding.
- *Methods and Results:* cDNA libraries were developed from mRNA extracted from leaves of *Buxus sempervirens* 'Vardar Valley' and sequenced using the Illumina MiSeq system. Approximately 11.9 million base pairs of sequence data were examined and 845 genic-SSRs were identified, including 469 dinucleotide, 360 trinucleotide, seven tetranucleotide, one pentanucleotide, and eight hexanucleotide repeats. Primer pairs were designed for 71 selectively chosen genic-SSRs containing trinucleotide repeat motifs and were used to amplify the corresponding loci in 18 diverse boxwood accessions. Twenty-three primer pairs amplified polymorphic loci, with two to 10 alleles per locus.
- *Conclusions:* These novel polymorphic genic-SSR markers will aid in evaluating genetic diversity of boxwood germplasm and allow verification of hybrids and cultivars for breeding programs.

Key words: boxwood; Buxaceae; *Buxus sempervirens*; cDNA library sequencing; genetic diversity; genic simple sequence repeat (genic-SSR) markers; microsatellites; polymorphism.

Boxwood (*Buxus* L. spp., Buxaceae) are popular woody landscape shrubs grown for their diverse forms and broad-leaved evergreen foliage (Batdorf, 2004). The genus contains approximately 90 species originating in Africa, Eurasia, the Caribbean, and Central America (Batdorf, 2004). Boxwood plants grown in temperate zones are increasingly threatened by a destructive new blight disease caused by the ascomycete fungus *Calonectria pseudonaviculata* Henricot (syn. *Cylindrocladium pseudonaviculatum*, *Cylindrocladium buxicola*). First identified from the United Kingdom in 1994, the disease has spread throughout continental Europe, parts of western Asia, and into North America. (Ivors et al., 2012; Elmhirst et al., 2013; Gehesquière et al., 2013; Malapi-Wight et al., 2014). To date, all tested cultivated *Buxus* taxa are affected by boxwood blight, although some taxa appear to be more susceptible to the fungus than others (Henricot et al., 2008; Douglas, 2012; Lamondia, 2014). There is an urgent need to develop blight-tolerant boxwood cultivars because of the impact this disease has on landscapes and commercial growers.

The National Boxwood Collection at the U.S. National Arboretum (USNA) contains more than 700 *Buxus* accessions, making it one of the most complete collections in the world and a valuable genetic resource for developing blight-tolerant varieties. However, genetic relationships and diversity among these accessions have not been determined. Although morphological features can be useful in determining phylogenetic relationships in Buxaceae (Carlquist, 1982; Köhler and Brückner, 1990), molecular markers are needed to distinguish among closely related accessions and to assess diversity. Van Laere et al. (2011) used amplified fragment length polymorphism (AFLP) markers to characterize and differentiate between European and Asian boxwood. In the current study, we developed and characterized 23 polymorphic genic simple sequence repeat (genic-SSR) markers to facilitate genetic diversity analysis of boxwood taxa from the National Boxwood Collection and elsewhere. Compared to AFLP markers, SSRs are multiallelic, codominant, transferable between related species, and can be used to reproducibly fingerprint organisms in different laboratories. Our objective was to generate a suite of polymorphic genic-SSRs from coding regions of the *Buxus* genome, as these markers may also be useful in analyzing the functional diversity in germplasm collections.

Table 1. Characteristics of 23 polymorphic genic-SSRs developed for *Buxus* spp.

Locus	Primer sequences (5′–3′)	Repeat motif[a]	Allele size range (bp)[b]	A[b]	T_a (°C)[c]	H_o[d]	H_e[d]	GenBank accession no.	BLAST top hit description [organism][e]
BSVV14	F: GATGAGGGAAAAGGCATTGA R: CCAGCAACCATTGGAGACTT	(GAT)5	174–180	3	58	0.300	0.265	KM660701	Predicted: protein SET [*Vitis vinifera*]
BSVV15	F: GCTCGGGAAACCCTAATGAT R: CAAGGACCGGTAACGGTAGA	(CTC)6	181–196	5	58	0.300	0.395	KM660702	Predicted: transcription factor TCP2-like [*Vitis vinifera*]
BSVV22	F: AGCGTCTAAGGGAGGAGAGG R: GCATTTTCTTTGCAAGAGCC	(TGA)7	130–145	6	58	1.000	0.775	KM660703	Hypothetical protein JCGZ 26794 [*Jatropha curcas*]
BSVV25	F: TCAGCTGATTCAACACTCGG R: AGTGTCCTCAACGTCAGCCT	(AAG)5	199–223	5	60	0.400	0.485	KM660704	Predicted: probable nucleolar protein 5-2-like [*Vitis vinifera*]
BSVV27	F: TCTCTGCTAATCCCCCTTCA R: AAGGTCCTGTTGGGAGACCT	(TTC)6	146–152	3	58	0.000	0.320	KM660705	Predicted: G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300 [*Vitis vinifera*]
BSVV30	F: CCCTCGTGAACGTCGTAAAT R: GGGGCTGACGAATAAAGTGA	(CTC)5	254–272	6	58	0.700	0.740	KM660706	Predicted: Lob domain-containing protein 38 [*Vitis vinifera*]
BSVV31	F: TGATGCTTCCAACAAATCCA R: GTTGCCCATGTTTCCCATAC	(CAG)5	198–216	4	58	0.300	0.545	KM660707	Hypothetical protein L484 008155 [*Morus notabilis*]
BSVV41	F: TGACCAAGAAAACAGGGAGG R: GGGTACCCACAAAGGTTGAA	(GGT)6	254–263	2	58	0.000	0.420	KM660708	Unnamed protein product [*Coffea canephora*]
BSVV56	F: ACAACGGTTTCCGTTTCTTG R: CACCACCACGGTCATGATAA	(GAT)5	181–215	9	58	0.300	0.820	KM660709	Hypothetical protein PRUPE ppa021264mg [*Prunus persica*]
BSVV59	F: CACAGCAGCAGACCATGATT R: GGAGTTCGTAGGGGGTAAGG	(GCA)5	210–219	4	60	0.300	0.410	KM660710	Uncharacterized protein LOC103334305 [*Prunus mume*]
BSVV60	F: CGACGAAGAGCTTGAAAAGG R: GGTCTCATTGCGTCGTCGAAG	(AGA)7	204–213	4	58	0.200	0.445	KM660711	Hypothetical protein JCGZ 10804 [*Jatropha curcas*]
BSVV62	F: TCTGTCTGTCGACCACCAAG R: CTTCCCACAAGGGATTACCC	(CAG)6	252–274	6	58	0.600	0.720	KM660712	Predicted: calcium-dependent protein kinase 16 [*Vitis vinifera*]
BSVV64	F: CTTTCATGATCCGTTTGGCT R: CAGGAGGCCCAAAATGTAAA	(TCA)5	212–224	3	60	0.000	0.000	KM660713	Predicted: translocase of chloroplast 159, chloroplastic-like [*Vitis vinifera*]
BSVV70	F: CCACTACAATCGCCCAATTT R: ACAGGCTTCCTCATCCATTG	(CAG)6	114–120	3	60	0.300	0.395	KM660714	Gigantea protein isoform 1 [*Theobroma cacao*]
BSVV71	F: CCCATTGTTTACAGCACCAT R: GAGATGTATTGGGCCTTGGA	(TCA)6	272–278	3	60	0.200	0.185	KM660715	Predicted: uncharacterized protein LOC100259633 [*Vitis vinifera*]
BSVV72	F: ATTGAGATGACACATCGGCA R: AAACCCACATGAAATCTCGG	(TGA)5	195–201	3	60	0.300	0.485	KM660716	Predicted: BEL 1-like homeodomain protein [*Vitis vinifera*]
BSVV74	F: GTCCTCTTGCTGCTGGGTAG R: AAGCCCTAGTCCTCCGACAT	(CCG)5	133–142	3	60	0.200	0.185	KM660717	GRAS family transcription factor [*Theobroma cacao*]
BSVV76	F: GGATTTTGAATTGGCAAGGA R: GGCGAAGTCGAATTCAGAAG	(GAG)5	100–106	3	60	0.800	0.655	KM660718	Unnamed protein product [*Vitis vinifera*]
BSVV78	F: TTGAGTTTTTCCGAGATGCC R: GATGTCTCAGCAGCAGTCCA	(TGC)6	196–221	10	60	0.900	0.840	KM660719	Clathrin assembly protein [*Ricinus communis*]
BSVV97	F: GTCGAAGTGGCAGAAGTGGT R: GTTGCTGGCCCATTGTAGTT	(CAA)6	167–182	4	60	0.400	0.545	KM660720	Hypothetical protein L484 021067 [*Morus notabilis*]
BSVV99	F: CATTGCAAGCACAGCAAAT R: ATGTATCCAGGGGTCCCTTC	(CAG)5	225–252	5	60	0.500	0.525	KM660721	Hypothetical protein RCOM 1754260 [*Ricinus communis*]
BSVV104	F: AAGCAGAAGTTGCTTTCCCA R: AGATGAATTGGCGGAATGAG	(CTG)5	167–173	3	60	0.100	0.265	KM660722	Predicted: uncharacterized protein LOC103446593 [*Malus domestica*]
BSVV106	F: AAAGCTGGTGGAGGTTGTTG R: CCCCTCATCCTTCTGTTCAA	(GCT)5	143–152	3	60	0.200	0.480	KM660723	Predicted: uncharacterized protein LOC100255482 isoform 2 [*Vitis vinifera*]

Note: A = number of alleles per locus; H_e = expected heterozygosity; H_o = observed heterozygosity; T_a = annealing temperature.

[a] The number of repeat units is the value for *Buxus* 'Vardar Valley', from which the original SSR markers were derived.

[b] The allele size ranges for each locus and number of alleles per locus (A) were determined based on the data from 18 boxwood accessions.

[c] The annealing temperature (T_a) was chosen based on experimental results using a gradient PCR machine to optimize PCR results. Optimization was necessary because each forward primer had an additional 18-bp sequence used for fluorescent labeling.

[d] Observed heterozygosity (H_o) and expected heterozygosity (H_e) were determined by genotyping the 10 diploid boxwood accessions. Locus BSVV64 was monomorphic for these 10 diploid accessions tested.

[e] Putative functions of homologous predicted proteins from a BLASTX search in the NCBI nonredundant database, with a threshold E-value of 1.0E-06.

METHODS AND RESULTS

Total RNA was extracted from frozen leaf tissue of *B. sempervirens* L. 'Vardar Valley' (Appendix 1) using the QIAGEN RNeasy Plant Mini Kit (QIAGEN, Valencia, California, USA). RNA was quantified using the Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, California, USA), and quality was evaluated using the QIAxcel capillary electrophoresis system (QIAGEN). cDNA libraries were constructed using the TruSeq RNA Sample Preparation LS kit (Illumina, San Diego, California, USA) following the manufacturer's protocol. Validated pooled cDNA libraries were prepared for sequencing following the Illumina protocol and sequenced using the MiSeq system on a 300-cycle MiSeq sequencing cartridge (Illumina). From a single sequencing run, 3,506,048 reads containing 0.5 Gbp of data with an average length of 140 bp per read were generated. Reads were trimmed of adapters and for quality, then assembled and mapped using the CLC Genomics Workbench version 6 software (CLC Bio,

Boston, Massachusetts, USA). The analysis yielded 2,370,726 mapped paired-end reads with an average length of 164 bp, which were assembled into a partial transcriptome of 12,027 contigs (11,912,857 bp).

The partial transcriptome assembly was mined for microsatellites using the PrimerPro Perl pipeline (http://webdocs.cs.ualberta.ca/~yifeng/primerpro/), which used the MISA algorithm (Coello Coello and Cortés, 2005) to detect tandem repeats of two to six nucleotides for at least five perfect repeat core motifs. A total of 845 SSR motifs were identified, including 469 dinucleotide, 360 trinucleotide, seven tetranucleotide, one pentanucleotide, and eight hexanucleotide repeats (sequences available from the authors). PCR primer pairs were designed using the Primer3 algorithm in the PrimerPro pipeline, with the following settings: primer length of 20 ± 2 nucleotides, GC content of 40–60%, and a PCR product size ranging from 100 to 300 bp. Trinucleotide motifs possessing unique PCR priming sites within the genome, as determined by BLASTN searches using PrimerPro, were evaluated visually for heterozygosity and mutation consistent with stepwise evolution. A total of 71 candidate markers were selected for testing from the trinucleotide SSR sites meeting these in silico criteria. PCR primers were manufactured by Integrated DNA Technologies (Coralville, Iowa, USA). The forward primers had an additional M13(–21) universal sequence (TGTAAAACGACGGCCAGT) attached to the 5′ end to allow indirect fluorescent labeling of PCR products using just one universal FAM (6-carboxy-fluorescine)–labeled M13 primer (Schuelke, 2000). These 71 primer pairs were used to amplify SSR loci in 18 boxwood accessions representing diverse species and cultivars (Appendix 1). Twenty-three of these primer pairs proved to be polymorphic and resulted in expected amplification profiles (Table 1). In addition, eight primer pairs amplified monomorphic loci, 33 primer pairs amplified multiple regions or an unexpected size product, and seven primer pairs did not amplify a product at all (data not shown).

Genomic DNA was extracted from frozen leaf tissue of 18 boxwood accessions using the QIAGEN DNeasy Plant Mini Kit and quantified using the NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, Delaware, USA). PCR was carried out in a Bio-Rad iCycler (Bio-Rad Laboratories, Hercules, California, USA). The 20-μL PCR reaction mixture contained 10 ng of template genomic DNA, 0.25 μM of each reverse and universal FAM-labeled M13(–21) primer, and 0.0625 μM of the forward primer with 1× Bioline MangoMix and 2.5 mM Bioline $MgCl_2$ (Bioline, Taunton, Massachusetts, USA). PCR profiles consisted of initial denaturation at 94°C for 5 min; followed by 30 cycles of 94°C for 30 s, optimized annealing temperature of each primer pair (Table 1) for 45 s, and 72°C for 45 s; followed by eight cycles of 94°C for 30 s, 53°C for 45 s, and 72°C for 45 s; and a final extension at 72°C for 10 min. Products were analyzed on an ABI 3730xl DNA Analyzer (Applied Biosystems, Foster City, California, USA) using 1 μL of PCR product, 10 μL of formamide (Applied Biosystems), and 0.3 μL of GeneScan 500 LIZ Size Standard (Applied Biosystems). Allele sizes and number of alleles per locus were determined using GeneMarker version 2.6.3 (SoftGenetics, State College, Pennsylvania, USA). Number of alleles per locus ranged from two to 10 with a mean of 4.34 (Table 1). The boxwood population we used included 10 diploids, four triploids, two tetraploids, and two mixoploids, as determined by flow cytometry in our laboratory (Appendix 1). Thus, we cannot report expected heterozygosities (H_e) for this population unless segregation analysis was performed to confirm dosage patterns of alleles for each locus (Dufresne et al., 2014). Instead, we treated the 10 diploids as one population and calculated the observed heterozygosity (H_o) and H_e using GenAlEx software (version 6.5; Peakall and Smouse, 2012) (Table 1). Excluding the monomorphic locus BSVV64, H_o and H_e ranged from 0.000 to 1.000 and 0.185 to 0.840 with means of 0.377 and 0.495, respectively. The contig sequences of the 23 SSR loci were subjected to a BLAST search against the National Center for Biotechnology Information (NCBI) nonredundant protein database using the BLASTX program to identify putative functions. With a threshold E-value of 1.0E-6, all 23 SSR sequences shared homology to protein sequences from dicots from diverse families (Table 1).

CONCLUSIONS

We have developed the first genic-SSR markers for *Buxus* species. These markers may be useful in characterizing genetic diversity and relatedness of boxwood germplasm, which could aid in our overall objective of developing blight-tolerant varieties.

LITERATURE CITED

BATDORF, L. R. 2004. Boxwood: An illustrated encyclopedia. American Boxwood Society, Boyce, Virginia, USA.

CARLQUIST, S. 1982. Wood anatomy of Buxaceae: Correlations with ecology and phylogeny. *Flora* 172: 463–491.

COELLO COELLO, C. A., AND N. C. CORTÉS. 2005. Solving multiobjective optimization problems using an artificial immune system. *Genetic Programming and Evolvable Machines* 6: 163–190.

DOUGLAS, S. M. 2012. Boxwood blight—A new disease for Connecticut and the U.S. [online]. Website http://www.ct.gov/caes/lib/caes/documents/publications/fact_sheets/plant_pathology_and_ecology/boxwood_blight-_a_new_disease_for_connecticut_and_the_u.s.__12-08-11.pdf [accessed 22 September 2014].

DUFRESNE, F., M. STIFT, R. VERGILINO, AND B. K. MABLE. 2014. Recent progress and challenges in population genetics of polyploid organisms: An overview of current state-of-the-art molecular and statistical tools. *Molecular Ecology* 23: 40–69.

ELMHIRST, J. F., B. E. AUXIER, AND L. A. WEGENER. 2013. First report of box blight caused by *Cylindrocladium pseudonaviculatum* (*C. buxicola*) in British Columbia, Canada. *Plant Disease* 97: 559.

GEHESQUIÈRE, B., S. D'HAEYER, K. T. K. PHAM, A. J. VAN KUIK, M. MAES, M. HÖFTE, AND K. HEUNGENS. 2013. qPCR assays for the detection of *Cylindrocladium buxicola* in plant, water and air samples. *Plant Disease* 97: 1082–1090.

HENRICOT, B., C. GORTON, G. DENTON, AND J. DENTON. 2008. Studies on the control of *Cylindrocladium buxicola* using fungicides and host resistance. *Plant Disease* 92: 1273–1279.

IVORS, K. L., L. W. LACEY, D. C. MILKS, S. M. DOUGLAS, M. K. INMAN, R. E. MARRA, AND J. A. LAMONDIA. 2012. First report of boxwood blight caused by *Cylindrocladium pseudonaviculatum* in the United States. *Plant Disease* 96: 1070.

KÖHLER, E., AND P. BRÜCKNER. 1990. Considerations on the evolution and chronogenesis of the genus *Buxus* (Buxaceae). *Memoirs of the New York Botanical Garden* 55: 153–168.

LAMONDIA, J. A. 2014. Management of *Calonectria pseudonaviculata* in boxwood with fungicides and less susceptible host species and varieties. *Plant Disease* 98: doi:10.1094/PDIS-02-14-0217-RE.

MALAPI-WIGHT, M., J. B. HÉBERT, R. BUCKLEY, M. L. DAUGHTREY, N. F. GREGORY, K. RANE, S. TRIPAK, AND J. A. CROUCH. 2014. First report of boxwood blight caused by *Calonectria pseudonaviculata* in Delaware, Maryland, New Jersey and New York. *Plant Disease* 98: 698.

PEAKALL, R., AND P. E. SMOUSE. 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—An update. *Bioinformatics (Oxford, England)* 28: 2537–2539.

SCHUELKE, M. 2000. An economic method for fluorescent labeling of PCR fragments. *Nature Biotechnology* 18: 233–234.

VAN LAERE, K., D. HERMANS, L. LEUS, AND J. V. HUYLENBROECK. 2011. Genetic relationships in European and Asiatic *Buxus* species based on AFLP markers, genome sizes and chromosome numbers. *Plant Systematics and Evolution* 293: 1–11.

APPENDIX 1.   Boxwood accessions (*Buxus* spp.) evaluated in this study, from the National Boxwood Collection at the U.S. National Arboretum.

Taxa[a]	Accession no.[b]	Ploidy[c]	Source
Buxus balearica Lam.	81245	Diploid (2*x*)	Cultivated; received from Dawes Arboretum, collected from garden in Yalta, Ukraine
B. balearica	81312	Diploid (2*x*)	Cultivated; received from Woodlanders Nursery, originally from College of William & Mary
B. bodinieri H. Lév.	960187	Diploid (2*x*)	Cultivated; received from J. C. Raulston Arboretum, originally from Heronswood Nursery
B. harlandii Hance	36672-L	Diploid (2*x*)	Cultivated; collected from garden in Xi'an, Shaanxi Province, China
B. harlandii 'Richard'	81337	Mixoploid (2*x*/4*x*)	Cultivated; received from Nurseries Caroliniana
B. microphylla Siebold & Zucc. var. *japonica* (Müll. Arg.) Rehder & E. H. Wilson	17525-CJ	Tetraploid (4*x*)	Collected directly from the wild, Nikko, Tochigi Prefecture, Japan
B. microphylla var. *japonica* 'National'	7025-V	Triploid (3*x*)	Cultivated; received from Morris Arboretum
B. microphylla var. *japonica* 'Sprinter'	81341	Triploid (3*x*)	Cultivated; received from Spring Meadow Nursery
B. sempervirens L.	36365-K	Diploid (2*x*)	Collected directly from the wild, Bzyb River, Georgia (as *B. colchica*)
B. sempervirens 'Borderline'	81766	Triploid (3*x*)	Cultivated; received from North Carolina State University
B. sempervirens 'Highlander'	—	Diploid (2*x*)	Cultivated; received from Conard-Pyle
B. sempervirens 'Katerberg'	81342	Diploid (2*x*)	Cultivated; received from Conard-Pyle
B. sempervirens 'Vardar Valley'	6395	Diploid (2*x*)	Collected directly from the wild, Vardar River, Macedonia
B. sinica (Rehder & E. H. Wilson) M. Cheng var. *aemulans* (Rehder & E. H. Wilson) P. Brückner & T. L. Ming	60705-J	Diploid (2*x*)	Collected directly from the wild, Huangshan, Anhui Province, China
B. sinica var. *insularis* (Nakai) M. Cheng	35486-J	Triploid (3*x*)	Cultivated; received from the Washington Park Arboretum
B. sinica var. *insularis* 'Franklin's Gem'	—	Tetraploid (4*x*)	Cultivated; received from Conard-Pyle
B. 'Conroe'	—	Mixoploid (2*x*/4*x*)	Cultivated; received from Conard-Pyle
B. 'Green Gem'	51904-H	Diploid (2*x*)	Cultivated; received from O. E. White Arboretum

[a] Taxa include species, varieties, and cultivars. Hybrids of unknown parentage are listed by their cultivar names.
[b] National Arboretum accession numbers are given when available, with the exception of *B. bodinieri*, which is listed using its source number.
[c] Ploidy was determined using flow cytometry in our laboratory.