SCREENING FOR LOW-COPY NUCLEAR GENE REGIONS IN THE CUPRESSACEAE

Robert P. Adams
Biology Department, Baylor University, Waco, TX 76798
Robert_Adams@baylor.edu

ABSTRACT

Thirty two low-copy nuclear (LCN) gene regions from Steele et al. (2008) were amplified using *Hesperocyparis goveniana* (Cupressaceae) DNA. Ten of these gave successfully amplified products and these were sequenced. Four of these LCNs yielded clean sequences. Of the 4 LCNs, one proved to be a mitochondrion gene region with no variation between *Hesperocyparis* and *Juniperus*;. The second LCN, tentatively identified as Actin-11, gave 497 bp of clean sequence then appeared to be multi-copy. The third LCN, tentatively identified as part of 3-carene synthase, gave 924 bp of clean sequence in *Hesperocyparis* but appeared to be multi-copy in *Juniperus*. The fourth LCN, possibly heat shock protein 90.1, yielded 946 bp of good sequence data for *H. goveniana, J. scopulorum* and *J. virginiana* and appears to offer promise for utilization in Cupressaceae taxonomy. *Phytologia 92(1): 82-91 (April, 2010).*

KEY WORDS: Low-copy nuclear genes, Cupressaceae, *Juniperus*, *Hesperocyparis*.

The major nuclear gene region utilized for sequencing in plants is nrDNA (nuclear ribosomal DNA) that has apparently undergone concerted evolution (Baldwin et al. 1995). It is also the major nuclear region sequenced in the *Cupressaceae*. However, nrDNA (ITS) is sometimes not sufficiently variable to distinguish closely related species and varieties (Adams and Bartel, 2009). 4-coumarate CoA ligase (4CL) and abscisic acid-insensitive 3 (ABI3) are two nuclear genes that have recently been employed in *Juniperus* (Adams, 2009) and the Cupressaceae (Adams, Bartel and Price, 2009),

but neither gene region is sufficiently variable to resolve some closely related species or varieties.

Non-coding areas of chloroplast DNA (cpDNA) have been widely utilized but as Steele et al. (2008) discuss, cpDNA is of limited use in examining hybridization in the Cupressaceae with uni-parental inheritance (paternal) and chloroplast capture. Adams, Schwarzbach and Morris (2008) found only 1 bp variation in cp trnC-trnD among the 8 Caribbean *Juniperus* taxa. Most Caribbean junipers had no variation in trnC-trnD.

Recently, Steele, et al. (2008) examined 141 low-copy nuclear (LCN) gene regions in two families of rosids. From these 141 LCNs, they list 32 as most promising.

The purpose of this paper is to report on analyses of these 32 low-copy number gene regions as applied to *Juniperus* and *Hesperocyparis* (Cupressaceae) for possible phylogenetic utilization.

MATERIALS AND METHODS

Genomic DNA utilized: Juniperus ashei: Adams 6746-48, Waco, TX; *J. blancoi, Adams 6849-51*, Durango, MX; *J. scopulorum, Adams 10895-10897*, Kamas, UT; *J. virginiana*, *Adams 6753-6755,* Hewitt, TX; H*esperocyparis goveniana, Adams and Bartel 9350-9351*, Monterey, CA. All vouchers are stored at BAYLU.

One gram (fresh weight) of the foliage was placed in 20 g of activated silica gel and transported to the lab, thence stored at -20^{o} C until the DNA was extracted. DNA was extracted from leaves by use of a Qiagen mini-plant kit as per manufacturer's instructions.

PCR primers (Table 1) were synthesized by Integrated DNA Technologies (San Diego, CA).

PCR amplification Various LCNs were amplified as shown in Table 1. In general, amplifications were performed in 30 µl reactions using 6 ng of genomic DNA, 1.5 units Epi-Centre Fail-Safe Taq polymerase, 15 µl

2x buffer E (final concentration: 50 mM KCl, 50 mM Tris-HCl (pH 8.3), 200 µM each dNTP, plus Epi-Centre proprietary enhancers with 1.5 - 3.5 mM $MgCl_2$ according to the buffer used) and 1.8 µM of each primer.

The PCR reaction was subjected to purification by agarose gel electrophoresis (1.5% agarose, 70 v, 55 min.). When possible, the band(s) was excised and purified using a Qiagen QIAquick gel extraction kit. The gel purified DNA band with the appropriate primer was sent to McLab Inc. (South San Francisco) for sequencing. Sequences for both strands were edited and a consensus sequence was produced using Chromas, version 2.31 (Technelysium Pty Ltd.). Alignments and NJ trees were made using MAFFT (http://align.bmr.kyushu-u.ac.jp/mafft/).

RESULTS AND DISCUSSION

Table 1 shows the forward and reverse primers for 32 most promising LCN gene regions from Steele et al. (2008).

Table 1. Primers for the 32 most promising gene regions of 141 screened by Steele et al. (2008).

Steele LCN primer pairs	sequence (5'-3'	For/ Rev
1F	GGTTTAGTGAAAATATCAGC	For
1R	TATGGTTTGAAACAAGCACCT	Rev
5F	ATCCAGAAGGAGTCCACCCTTCA	For
5R	TCCTTCTGGATGTTGTAGTCGG	Rev
6F	GATGGACAGGTGATCACCATTGG	For
6R	TAGAAGCACTTCCTGTGGA	Rev
13F	CGTCACCGTCTGCGAGATCAAC	For
13R	TACCTTGAGAAACATCCC	Rev
15F	ACGGACAAGAGCAAGCTCGATG	For
15R	TTGTAGTCTTCCTTGTTCTCAG	Rev
30F	GCTAGCCATGTCCTTGAGGT	For
30R	CAGATTTTGGGCTTGCCAAAGAT	Rev
31F	CAGATGAGCTGATCAAGACGGCC	For

31R	ACCTTTGGGCTGTCGGAGC	Rev
46F	ATGCAGATCTTCGTGAAGAC	For
46R	TCCTTCTGGATGTTGTAGTCGG	Rev
50F	TCTGAGTTCATCAGCTACCC	For
50R	GTTGTAGTCTTCCTTGTTCTCAG	Rev
51F	CCTAGCTTGATGACACCAC	For
51R	CTTGGACGTATCATGAAT	Rev
56F	GTCCTTCGCTTCAGGGCCTG	For
56R	TTTCATGATAGCAAGGTGCTCCC	Rev
57F	AAGGAGGTTTCTCATGAGTGG	For
57R	GTCTTCTTGCTCGACATGT	Rev
58F	AACATCTCCTGGATTGAGGT	For
58R	GACAATGAGGCTCTCTAC	Rev
61F	ATGGGACAAAAAGATGCTTA	For
61R	TAGAAGCACTTCCTGTGGA	Rev
64F	AACCGTCCCAACAGCATTGA	For
64R	GCACCACTGAAGCCTTGGGT	Rev
65F	TTTGTCAGTACAGTGGATCC	For
65R	CTTTGTGGGAATTGGACAT	Rev
68F	GAGTTCATGACATCAATGG	For
68R	ACCCTGGAGATCGGCATGG	Rev
69F	CATTGAGGAGAACAATGCAGG	For
69R	TGCTCAGGAGCACCTTTG	Rev
73F	CGGGTTCGAGTCCCGGCAACGG	For
73R	TGGAACCACTTCATGGCAT	Rev
81F	CGAGACCACCAAGTACTACTGC	For
81R	TCATCTTAACCATACCAG	Rev
82F	CAAGAGGTCCTTCAAGTACGC	For
82R	TCATCTTAACCATACCAG	Rev
85F	CTGTGACAATGGAACCGGAATGG	For
85R	TAGAAGCACTTCCTGTGGA	Rev
86F	ATGGTTGAGTACTTTGGTGAGCAG	For
86R	GCCCAGTCCAAGTAGAAAG	Rev
87F	GTGATCACTACATCAACAATAGC	For
87R	GTAACAGAACCATAGATCC	Rev
90F	CAGGTTCTCATTTGGGATGT	For
90R	CCTGAACACAAAGAACAGCAG	Rev

91F	GTGGTCTTGTCCTCAGCAGA	For
91R	TGCGTATCATCAACGAGC	Rev
94F	GAGAGCAGCGACACCATCG	For
94R	CCAGTGAGTGTCTTCAC	Rev
96F	TGGTGTGATACAAAATCAGT	For
96R	TCAGGGCTAATATCAGTA	Rev
125F	GAGCCATACAAGGGTATT	For
125R	GACCACGGTAGACAATGA	Rev
126F	TGCGCTATTTACCAGTGAGTGG	For
126R	AAAGGGGATGTTTTGTTAA	Rev
129F	TATGCCAGTGGTCGTACAAC	For
129R	TTAGAAGCATTTCCTGT	Rev
133F	ATGAAATCATTCTTGAGGAC	For
133R	CTACGCCAACAAGGTCGGCC	Rev

Each of these 32 primer pairs were utilized with *H. goveniana* DNA to determine amplification at 3 annealing temperatures (Table 2). Three of the LCN gene primers gave primer dimers (31, 81, 83) at all annealing temperatures (Table 2). Eight of the LCN gene primers gave only faint bands or smears (13, 50, 58, 65, 85, 96, 126, 129). It is worthwhile to recall that Steele et al. (2008, Appendix A) showed that most of their 141 LCN gene primers failed to amplify.

A few of the primers gave bright bands at all 3 annealing temperatures (1, 5), but generally the brightest bands were present at only one (sometimes two) of the three annealing temperatures (Table 2). Several of the LCN gene primers gave bright bands but of small size: 5 (250bp), 15 (450bp), 46 (460bp), 64 (500bp), 68 (300, 400 bp), 73 (360 bp), 90 (300bp), 91 (300bp), 125 (500bp) and 123 (400bp). Most of these were eliminated from further consideration for sequencing. It would be interesting to compare the sizes of these bands with those in the rosids by Steele et al. (2008).

Some of the LCN gene primers gave a bright band that upon closer examination or by running the gel longer turned out to be twin bands (51, 64, 68, 69, Table 2).

Table 2. Amplification results using annealing temperatures of 45ºC, 50ºC and 55ºC. Amplified products (in bold face) were prepped and sequenced. brt = bright, med = medium, and f = faint band. Band sizes were estimated using pGEM markers on a 1.6% agarose gel.

LCN	45ºC	50ºC	55ºC
1	**brt 950, 200 bp**	same as 45ºC	same as 45ºC
5	brt 250+f 350 bp	same as 45C	same as 45C
6	brt 1500+700bp	**brt 1500+700bp**	faint bands
13	faint bands	faint bands	faint bands
15	**brt 450 bp**	same as 45C	faint 400bp
30	**brt 500bp**	brt 500bp	nothing
31	primer dimer	primer dimer	primer dimer
46	brt 460+ 220 bp	**brt 460**+220bp	brt 460+ 220bp
50	smear	1 faint band	1 faint band
51	brt 500bp (2bands!)	2 faint bands	2 faint bands
56	brt 900bp	brt 900bp	**brt 900bp**
57	brt 200bp	brt 200,faint 950bp	**brt 950bp**
58	smear	nothing	nothing
61	4 bright bands	1 faint band	nothing
64	brt 500bp (2 bands)	3 brt 600-400bp	1 faint band
65	3 faint bands	nothing	nothing
68	2 brt bands, 300-400	2 faint bands	nothing
69	2 brt 600, 1 f 800bp	**brt 800bp ,2 f**	nothing
73	brt 360, f 500bp	brt 360, f 500bp	faint 360 bp
81	primer dimer	primer dimer	primer dimer
82	primer dimer	primer dimer	primer dimer
85	faint 360bp	faint 360bp	v faint 360bp
86	smear	f 600, 3brt 300bp	smear
87	**brt 650bp + smear**	faint 650bp	v faint bands
90	brt 300, f500, 200bp	f500,f300,f200bp	faint smear
91	brt 300bp + smear	f 300bp	nothing
94	**brt 500 + brt 300bp**	brt 500+m300bp	f 500+b300bp
96	smear	smear	smear
125	brt 500+m 700bp	faint	faint
126	smear	smear	smear
129	nothing	nothing	nothing
133	f800,f600,brt 400bp	m800,m600bp	faint

Based on the screening (Table 2), 10 LCN gene regions were selected for prep and sequencing for *H. goveniana*. These are highlighted in Table 2) and are LCN gene regions 1 (950 bp, 45ºC), 6 (600 bp, 50ºC), 15 (450 bp, 45ºC), 30 (500 bp, 45ºC), 46 (460 bp, 50ºC), 56 (900 bp, 55ºC), 57 (950 bp, 55ºC), 69 (800 bp, 50ºC), 87 (650 bp, 45ºC) and 94 (500 bp, 45ºC). Of course, some of these are marginal in size as it may not be practical (cost-effective) to sequence only 450 bp if one can find LCN gene regions that are 900 - 1000 bp. But it was felt that determining how many of the 10 regions would yield clean sequence data was sufficient reason to include these smaller regions in the sequencing.

Sequencing the 10 most promising 32 LCN gene regions (Table 2) gave 4 LCNs with partial or complete sequences (Table 3, 1, 6, 56, 57). LCNs 15, 30, 69, 87 and 94 appeared to be multi-copy in the *H. goveniana* sequences. For LCN 46, the prep band yield was too low to sequence.

Four of the 10 LCN gene regions seemed worthy of additional study:

LCN 1. Sequencing in *H. goveniana* gave 924 bp of nice, clean sequence but when amplifications were attempted in *Juniperus ashei*, *J. blancoi*, *J. scopulorum* and *J. virginiana* 2 bands were produced that could only be partially separated by electrophoresis to 140 min. at 70 V on a 1.6% agarose gel. The highest GenBank match for the *H. goveniana* sequence was to the *Picea glauca* FJ609175, putative 3-carene synthase gene (Hamberger, et al. 2009). The terpene synthase genes are known to be in families (Trapp and Croteau, 2001; Hamberger, et al. 2009) and appear to have arisen by gene duplication. It may be very difficult to obtain single copy genes in this family. However, by designing new, interior primers to the *H. goveniana* sequence, one might be able to obtain specific primers that would amplify single copies in *Juniperus* and other Cupressaceae species.

Table 3. Sequencing of the 10 most promising Steele LCN gene regions.

Steele LCN	Size	Sequencing result/ gene
1	924 bp	clean in *H. goveniana*, multi-copy in *Juniperus* 3-carene synthase? *Picea glauca* FJ069175
6	~1500 bp	497bp clean seq. in *H. goveniana*. then messy, multi - copy? Actin-11, *A. thaliana* U27981
	~700 bp	multi-copy, messy seq. in *H. goveniana*.
15	~450 bp	multi-copy in *H. goveniana.*
30	~500 bp	multi-copy in *H. goveniana.*
46	~460 bp	prep band too low yield to sequence.
56	871 bp	clean seq. in *H. goveniana*. and *Juniperus*. No differences between *Hesperocyparis* and *Juniperus*! NADH dehydrogenase subunit 5, mitochondrion gene *A. thaliana*, EU999005
57	946 bp	1 SNP (J. *virg-scop*); 20 SNPs (J. *virg-H. gov.*) *A. thaliana* NM_124642, heat shock protein 90.1
69	~800 bp	multi-copy in *H. goveniana.*
87	648 bp	messy, multi-copy in *H. goveniana.*
94	~500 bp	multi-copy in *H. goveniana.*

LCN 6. Two bright bands were amplified (~1500, ~700 bp, Table 2). Both bands were prepped and sequenced.

Band 1:

~1500 bp band: 497 bp of clean sequence data (forward primer) was obtained from *H. goveniana*, but the reverse primer sequence was messy indicating multi-copy. A BLAST of the 497 bp

sequence to GenBank gave the highest match to Actin-11 from *A. thaliana* U27981.

Band 2:

~700 bp band: 340 bp of clean sequence obtained from the forward primer for *H. goveniana* with the balance of the sequence messy. This is probably a multi-copy gene.

LCN 56. 871 bp of clean sequence was obtained from *H. goveniana, J. ashei* and *J. virginiana.* No differences were found between *Hesperocyparis* and *Juniperus*! A blast of GenBank matched to NADH dehydrogenase subunit 5, a mitochondrion gene of *A. thaliana*, EU999005 with a 90% similarity. This is an easy section to amplify, but there was no variation between these two Cupressaceae genera. LCN 56 seems of limited use.

LCN 57: 946 bp of clean sequence data was obtained from *H. goveniana* and *J. virginiana.* These species differ by 20 SNPs. Amplification from *J. ashei, J. blancoi* and *J. scopulorum* gave messy sequences. New internal primers were designed based on the sequences from *H. goveniana* and *J. virginiana. Juniperus scopulorum* was successfully sequenced using these primer pairs:

STLvirg57F40	ATGCTTCCTTCTACAAGAGT
STLvirg57R957	TCATTATCCTCTCCATGTT
and	
STLvirg57F40	ATGCTTCCTTCTACAAGAGT
STLvirg57R965	CTGAGCCTTCATTATCCT

Juniperus scopulorum and *J. virginiana* differed by only 1 bp SNP. A blast of the H. goveniana sequence in GenBank matched heat shock protein 90.1, *A. thaliana* NM_124642. This may be the most promising LCN gene region discovered in this study and may prove useful in phylogenetic studies in the Cupressaceae, but it does not appear promising for infra-specific analysis due to the low variation found between closely related *J. scopulorum* and *J. virginiana* (1 SNP).

ACKNOWLEDGEMENTS

Special thanks to Randy Linder for furnishing the detailed information on the 32 most promising LCNs (from Steele et al., 2008).

This research was supported in part with funds from Baylor University. Thanks to Tonya Yanke for lab assistance.

LITERATURE CITED

Adams. R. P. 2009. Variation among the smooth leaf margin *Juniperus* of Mexico: Analyses of nrDNA, 4CL, ABI3 and petN-psbM SNPs. Phytologia 91(3): 569-578.

Adams, R. P. and J. A. Bartel. 2009. Infraspecific variation in *Hesperocyparis goveniana* and *H. pygmaea*: ISSRs and terpenoid data. Phytologia 91(2): 277-286.

Adams, R. P., J. A. Bartel and R. A. Price. 2009. A new genus, *Hesperocyparis*, for the cypresses of the new world. Phytologia 91(1): 160-185.

Adams, R. P., A. E. Schwarzbach and J. A. Morris. 2008. The evolution of Caribbean *Juniperus* (Cupressaceae): Terpenoids, RAPDs, and DNA SNPs data Phytologia 90(1): 103-119.

Baldwin, B. C., M. J. Sanderson, M. F. Wojciechowski, C. S. Campbell and M. J. Donoghue. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Ann. Missouri Bot. Gard. 82: 247-277.

Hamberger, B., D. Hall, M. Yuan, C. Oddy, B. Hamberger, C. I. Keeling, C. Ritland, K. Ritland and J. Bohlmann. 2009. Targeted isolation, sequence assembly and characterization of two white spruce (*Picea glauca*) BAC clones for terpenoid synthase and cytochrome P450 genes involved in conifer defense reveal insights into conifer genome. BMC Plant Biology 9: 106 (13 pgs), http://www.biomedcentral.com/1471-2229/9/106

Steele, P. R., M. Guisinger-Bellian, C. R. Linder and R. K. Jansen. 2008. Phylogenetic utility of 141 low-copy nuclear regions in taxa at different taxonomic levels in two distantly related families of rosids. Molec. Phylogenetics Evol. 48: 1013-1026.

Trapp, S. C. and R. B. Croteau. 2001. Genomic organization of plant terpene synthases and molecular evolutionary implications. Genetics 158: 811-832.