Development of EST-PCR Markers for DNA Fingerprinting and Genetic Relationship Studies in Blueberry (*Vaccinium*, section *Cyanococcus*)

Lisa J. Rowland,¹ Smriti Mehra, Anik L. Dhanaraj, Elizabeth L. Ogden, and Janet P. Slovin

United States Department of Agriculture, Agricultural Research Service, Henry A. Wallace Beltsville Agricultural Research Center, Fruit Laboratory, Building 010A, 10300 Baltimore Avenue, Beltsville, MD 20705

Mark K. Ehlenfeldt

United States Department of Agriculture, Agricultural Research Service, Blueberry and Cranberry Research Center, 125A Lake Oswego Road, Chatsworth, NJ 08019

ADDITIONAL INDEX WORDS. CAPS markers, cultivar identification, expressed sequence tags, molecular markers, STS markers

ABSTRACT. Because randomly amplified polymorphic DNA (RAPD) is the only type of molecular marker that has been used extensively in blueberry (*Vaccinium* spp.) for mapping and DNA fingerprinting of cultivars, there is a need to develop a new, robust marker system. Expressed sequence tags (ESTs) produced from a cDNA library, derived from RNA from floral buds of cold acclimated plants, were used to develop EST-PCR markers for blueberry. Thirty clones, picked at random from the cDNA library, were single-pass sequenced from the 5' and 3' ends. Thirty PCR primer pairs were designed from the ends of the best quality sequences that were generated and were tested in amplification reactions with genomic DNA from 19 blueberry genotypes, including two wild selections (the original parents of a mapping population), and 17 cultivars. Fifteen of the 30 primer pairs resulted in amplification of polymorphic fragments that were detectable directly after ethidium bromide staining of agarose gels. Several of the monomorphic amplification products were digested with the restriction enzyme *AluI* and approximately half resulted in polymorphic-sized fragments (cleaved amplified polymorphic sequences or CAPS markers). The polymorphic EST-PCR and CAPS markers developed in this study distinguished all the genotypes indicating that these markers should have general utility for DNA fingerprinting and examination of genetic relationships in blueberry. Similarity values were calculated based on the molecular marker data, and a dendrogram was constructed based on the similarity matrix. Coefficients of coancestry were calculated for each pair of genotypes from complete pedigree information. A fair correlation between similarity coefficients calculated from marker data and coefficients of coancestry was found.

Blueberry (*Vaccinium* spp.) is a high value crop which can thrive on acidic, imperfectly drained sandy soils otherwise considered worthless for agricultural crop production. The United States is the world's leading blueberry producer. However, a survey of blueberry research and extension scientists in the United States has identified lack of winter hardiness and susceptibility to spring frosts as two important genetic limitations of current cultivars (Moore, 1993). The genetic control of these traits is not well understood, and their expression undoubtedly involves the interaction of several components. Therefore, development of genetic linkage maps for blueberry populations suitable for the analysis of quantitative trait loci (QTLs) controlling cold hardiness in the acclimated state, chilling requirement, and related factors is underway (Panta et al., in press; Rowland et al., 1999).

The only type of molecular marker that has been used extensively in blueberry for mapping and DNA fingerprinting of cultivars is randomly amplified polymorphic DNA (RAPD) (Williams et al., 1990). Genetic linkage maps of relatively low density have been developed for four blueberry populations, three diploid (Panta et al., in press; Rowland and Levi, 1994; Rowland et al., 1999) and one tetraploid (Qu and Hancock, 1997). These maps are based primarily on RAPD markers, except that the maps for the diploid testcross populations, segregating for cold hardiness in the cold acclimated state and chilling requirement [described in Rowland et al. (1999) and Panta et al. (in press)], include a few inter-simple sequence repeat (ISSR) PCR markers, in addition to the RAPD markers. RAPD markers have also been used for DNA fingerprinting and for assessing genetic relationships among cultivars and wild selections of the major commercially grown types of blueberries: the highbush (Levi and Rowland, 1997), lowbush (Burgher et al., 2002), and rabbiteye types (Aruna et al., 1993, 1995).

RAPD technology has been widely adopted because it is technically simple, does not require large amounts of DNA, and requires no prior knowledge of the genome being studied. However, RAPD technology has certain drawbacks such as problematic reproducibility among laboratories (Jones et al., 1997), markers that are generally limited to specific populations and therefore unsuitable for comparative mapping studies (Sunnucks, 2000), and markers with a dominant mode of inheritance. Because of these drawbacks, there is a need to develop an alternative DNA marker system for blueberry.

Expressed sequence tag (EST) databases are proving to be valuable sources of genetic markers for mapping, DNA fingerprinting, and population genetic studies for a wide variety of organisms. There are a number of advantages to using EST-based markers for genetic studies. First, they target expressed genes; thus, they should be particularly useful for QTL mapping. If an EST marker is linked to a QTL, it is possible that the gene itself, from which the EST marker was derived, controls the trait in question. Second, because they are derived from gene coding regions, which are more likely to be conserved across populations and species than noncoding regions, EST markers should be useful for comparative mapping studies. Furthermore, EST-based markers have the potential for being codominantly inherited. EST-based markers were originally

Received for publication 30 Dec. 2002. Accepted for publication 23 Apr. 2003. We would like to thank Kate Arnold for technical assistance and Kim Lewers and James Polashock for their critical reviews of the manuscript.

 $^{^{1}\}mathrm{To}$ whom correspondence should be addressed.

mapped by RFLP analysis. Recently, EST-based PCR markers have been developed for many plants, including several tree species: loblolly pine (*Pinus taeda* L.) (Temesgen et al., 2001), black spruce [*Picea mariana* (Mill.)B.S.P.] (Perry and Bousquet, 1998), Norway spruce [*Picea abies* (L.) Karst.] (Schubert et al., 2001), and sugi (*Cryptomeria japonica* D. Don) (Tsumura et al., 1997). Generally, amplification using EST-specific primers must be followed by either digestion with restriction enzymes to generate cleaved amplified polymorphic sequences or CAPS markers, heteroduplex analysis, or single-stranded conformational polymorphism (SSCP) analysis to detect polymorphisms (Cato et al., 2001).

As part of our research to better understand the genetic control of cold hardiness in blueberry, we recently began producing ESTs from highbush blueberry using a cDNA library derived from floral buds of cold-acclimated plants. Here, we report the development of 17 PCR primer pairs, derived from ESTs, that result in amplification of polymorphic-sized fragments either directly or after restriction enzyme digestion (CAPS). Furthermore, the fragments are detectable after simple ethidium bromide staining of agarose gels. These polymorphic EST-PCR and CAPS markers were screened across 19 blueberry genotypes, including two wild selections (diploid V. darrowi Camp and diploid V. corymbosum L.), that are the original parents of our mapping population, and 17 cultivars (mostly tetraploid V. corymbosum and hexaploid V. ashei Reade). A dendrogram was constructed based on genetic similarity values calculated for each pair of genotypes, and the correlation between similarity coefficients, calculated from molecular marker data, and coefficients of coancestry was evaluated.

Materials and Methods

PLANT MATERIAL. Nineteen blueberry genotypes, including 17 cultivars and two wild selections (parents of our mapping population), were evaluated in this study. The genotypes are listed in Table 1 along with their species, ploidy levels, and immediate parents. The genotypes were maintained at the USDA/ARS (Beltsville, Md.).

cDNA clones, sequence analysis, and primer design. The cDNA library for EST analysis was constructed in the unidirectional λ cloning vector Uni-ZAP (Stratagene, La Jolla, Calif.) from RNA expressed in floral buds of cold-acclimated plants of the highbush blueberry cultivar Bluecrop (Levi et al., 1999). Plaques were picked from the library at random, and in vivo excision of the pBluescript phagemids (containing the cDNA inserts) was performed according to the manufacturer's instructions (Stratagene) using the exassist/solar cell system. Plasmid DNA was isolated from bacterial cultures using the QIAprep Spin Miniprep Plasmid Kit (Qiagen Inc., Valencia, Calif.). Plasmid DNA was quantified spectrophotometrically, digested with EcoRI and XhoI restriction enzymes (New England BioLabs Inc., Beverly, Mass.) to release the inserts, and electrophoresed through 1% agarose gels to verify the presence of inserts and the quantity of DNA. Sizes of the cDNA inserts were estimated from molecular weight standards (1-kb DNA ladder or 1-kb Plus DNA ladder, Invitrogen Life Technologies, Carlsbad, Calif.) included on the gels.

Single-pass nucleotide sequencing of recombinant plasmid DNAs was performed from both ends of the cDNA inserts either by ourselves using an ABI PRISM 310 Genetic Analyzer and Big Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer Applied Biosystems, Foster City, Calif.) or by the University of Maryland, Center for Agricultural Biotechnology-DNA Sequencing Facility (College Park, Md.). For 5' end sequencing, the M13 reverse primer was used; the M13 (-21) primer was used for 3' end sequencing. Three clones were completely sequenced from both ends by the University of Florida, DNA Sequencing Core Laboratory (Gainesville, Fla.).

After trimming vector sequences, nucleotide sequences were compared against known gene sequences contained in GenBank (http://www.ncbi.nlm.nih.gov/BLAST/) using the BLASTN and BLASTX algorithms (Altschul et al., 1990). Primer pairs were designed from sequence data using the P3 website (http: //www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). If sequences were obtained from both 5' and 3' ends of the cDNA, forward and reverse primers were designed from sequences near the ends of the cDNA insert. If good sequence data (>300 bases) was obtained from only one end of the cDNA, then both forward and reverse primers were designed as far apart as possible from only that one end. Primers were synthesized by Invitrogen Life Technologies.

GENOMIC DNA EXTRACTION. Young leaves (\approx 5 g) were collected from greenhouse-grown plants of each of the genotypes used in this study. After collection, leaves were ground with dry ice in a coffee grinder and stored at -70 °C. DNA was extracted using the CTAB procedure of Doyle and Doyle (1990). DNA concentrations were estimated by agarose gel electrophoresis (1% w/v gel containing 0.5 µg·mL⁻¹ ethidium bromide) using known concentrations (15-1000 ng) of uncut lambda DNA (Invitrogen Life Technologies) as standards.

GENERATION OF EST-PCR AND CAPS MARKERS. DNA amplification reactions were performed as described previously (Levi et al., 1993) with minor modifications as described by Stommel et al. (1997). Briefly, amplification reactions were carried out at least twice in 25 µL volumes containing reaction buffer (20 mM NaCl, 50 mM Tris-HClpH9, 1% Triton-X-100, 0.1% bovine serum albumin), 1.6 mM MgCl₂, 200 µM each of dATP, dCTP, dGTP, and dTTP, 0.1 µM each of the forward and reverse EST primers, 0.7 units Taq DNA polymerase (Promega, Madison, Wis.), and 25 ng template DNA. DNA was amplified for 40 cycles in an MJ Research (Watertown, Mass.)PTC-100 thermal cycler, programmed for a 40 s denaturation step at 92 °C, 70 s annealing step at a temperature of 10 °C below the $T_m (1.0 \text{ M Na}^+)$ of the primer (forward or reverse) with the lower T_m, and 120 s extension step at 72 °C. Amplification products and molecular weight standards (1-kb DNA ladder; Invitrogen Life Technologies) were separated by electrophoresis through 1.4% agarose gels containing 0.5 µg·mL-1 ethidium bromide. Some of the monomorphic PCR products were further processed by ethanol precipitation and digestion with the restriction enzyme AluI (New England BioLabs Inc.). Digestions were performed at 37 °C for \approx 3 h using 10 units of enzyme. DNA fragments were visualized under UV light and photographed using the Eagle Eye still video system (Stratagene).

INHERITANCE OF MARKERS. The EST-PCR and CAPS markers were classified according to their type of inheritance, either predicted or observed. CAPS markers were classified as codominant because that is their predicted mode of inheritance. For those EST primer pairs that amplified multiple fragments (more than two fragments even in the diploid genotypes), the EST-PCR markers were simply classified as being derived from a multigene family. Mode of inheritance was determined when possible for the remaining EST-PCR markers using our diploid mapping populations. The populations have been described previously in detail (Rowland et al., 1999). Briefly, two diploid testcross populations were generated by crossing *V. darrowi* x *V. corymbosum* hybrids [Fla4B x W85-20 F₁s (#5, #6, and #10)] back to another *V. darrowi* and another *V. corymbosum* selection, NJ88-13-15 and W85-23, respectively. Segregation of

the EST-PCR markers, that were polymorphic between the original parent plants (Fla4B and W85-20), present in at least one of the three F_1 s (#5,#6, or #10), and absent in at least one of the testcross parents (NJ88-13-15 and W85-23), were followed in the appropriate subpopulations of the mapping populations. The nature of the polymorphism (presence and absence of fragments or different sizes of fragments in the parent plants), whether segregation was observed in the small F₁ population (indication of heterozygosity in the original parent plant), and the actual segregation pattern in the appropriate mapping subpopulation were all used to determine the mode of inheritance.

CLONING, SEQUENCING, AND SEQUENCE SIMILARITY ANALYSIS OF EST-PCR products. A selection of EST-PCR amplification products, including both monomorphic and polymorphic fragments, were excised from 1.4% TBE agarose gels and purified using the QIAEX II Gel Extraction Kit according to the manufacturer's instructions (Qiagen Inc.). Purified fragments were cloned using the pGEM-T Easy Vector System (Promega) according to the directions provided. Presence of inserts in the pGEM-T Easy Vector was confirmed by colony PCR. Sterile, disposable micropipetter tips were touched to white colonies and inoculates used in place of genomic DNA in PCR amplifications. Amplifications were performed as described above under the heading Generation of EST-PCR and CAPS Markers except that a 5 min denaturation step at 94 °C was included at the beginning and a 5 min elongation step at 72 °C was added at the end of the PCR program. Single-pass nucleotide sequencing of confirmed recombinant plasmid DNAs was performed from both ends of the inserts using M13 reverse and M13 (-21) primers by the University of Maryland, Center for Agricultural Biotechnology-DNA Sequencing Facility. To determine if the sequences of the PCR products were homologous to the sequences of the cDNA clones from which the EST-PCR primers were derived, sequences were compared to each other manually and by use of

BLAST algorithms, Pairwise BLAST and BLASTX or BLASTN to search GenBank.

MARKER DATA ANALYSIS. The Numerical Taxonomy and Multivariate Analysis System program package for PC (NTSYS-pc, version 2.1, Exeter Software, Setauket, N.Y.) was used to construct a similarity matrix from the EST-PCR and CAPS marker data and to perform a cluster analysis of the resulting similarity matrix. Similarity matrices were generated using the Jaccard (1908) and simple matching coefficient functions of the NTSYS-pc program. Dendrograms of the genotypes were constructed by applying the unweighted pair-group clustering method (UPGMA) to the genetic similarity matrices. Cophenetic value matrices were produced from the tree matrices to measure the goodness of fit to the similarity matrix on which each tree was based.

Coefficients of coancestry were determined by calculating the inbreeding coefficients of offspring from the hypothetical mating between each pair of blueberry genotypes. Inbreeding coefficients were calculated assuming polysomic inheritance for the tetraploids, as described by Ehlenfeldt (1994), and disomic inheritance for the diploids and hexaploids. The pedigrees of the genotypes were collected from various sources. A dendrogram was constructed by applying UPGMA to the coefficients of coancestry matrix. To determine the level of correlation between the genetic similarity values derived from the molecular marker data and the pairwise coefficients of coancestry, Pearson product-moment correlation coefficients were calculated using SAS Procedure CORR.

Results and Discussion

cDNA sequencing and gene identification. Attempts were made to sequence from the 5' and 3' ends of 30 cDNA clones that were picked at random from an unamplified cDNA library prepared from RNA from floral buds of cold acclimated 'Bluecrop' plants.

Genotype Species Ploidy Immediate parents Berkeley (BK) V. corvmbosum 4x Stanley x GS-149 (Jersey x Pioneer) Bluecrop (BC) 93.7% V. corymbosum, 6.3% V. angustifolium 4x GM-37 (Jersey x Pioneer) x CU-5 (Stanley x June) Bluegold (BG) 85.9% V. corymbosum, 14.1% V. angustifolium 4x Bluehaven x ME-US-5 (Ashworth x Bluecrop) Bluejay (BJ) V. corymbosum 4x Berkeley x Michigan 241 (Pioneer x Taylor) Blueray (BR) 93.7% V. corymbosum, 6.3% V. angustifolium 4x GM-37 (Jersey x Pioneer) x CU-5 (Stanley x June) Cooper (CP) 70.3% V. corymbosum, 25% V. darrowi, 4x G-180 [G-100 (Ivanhoe x Earliblue) x Collins] x US-75 4.7% V. angustifolium (V. darrowi Fla4B x Bluecrop) Duke (DK) 96.1% V. corymbosum, 3.9% V. angustifolium 4x G-100 (Ivanhoe x Earliblue) x 192-8 [(Berkeley x Earliblue) x (Coville x Atlantic)] Georgiagem (GG) 71.1% V. corymbosum, 25% V. darrowi, 4x G-132 (E-118 x Bluecrop) x US-75 (V. darrowi 3.9% V. angustifolium Fla4B xBluecrop) Gulfcoast (GC) 70.3% V. corymbosum, 25% V. darrowi, 4x G-180 [G-100 (Ivanhoe x Earliblue) x Collins] x US-75 4.7% V. angustifolium (V. darrowi Fla4B x Bluecrop) Nelson (NL) 96.9% V. corymbosum, 3.1% V. angustifolium 4x Bluecrop x G-107 (F-72 x Berkeley) Patriot (PT) 71.9% V. corymbosum, 28.1% V. angustifolium US 3 (Dixi x V. angustifolium Michigan Lowbush 1) x Earliblue 4x Sierra (SA) 47.6% V. corymbosum, 20% V. darrowi, 4x US 169 (self-pollination of pentaploid hybrid of V. darrowi 15% V. ashei, 15% V. constablaei, Fla4B and hexaploid V. constablaei-V. ashei hybrid US 56) x 2.4% V. angustifolium G-156 (V. corymbosum) selection Sunrise (SR) 82.8% V. corymbosum, 17.2% V. angustifolium 4x G-180 [G-100 (Ivanhoe x Earliblue) x Collins] x ME-US 6620 (E-22 x ME-US 24) Toro (TR) 93.7% V. corymbosum, 6.3% V. angustifolium 4x Earliblue x Ivanhoe Weymouth (WT) 87.5% V. corymbosum, 12.5% V. angustifolium 4x June x Cabot Climax (CX) V. ashei 6x Callaway x Ethel Tifblue (TB) V. ashei 6x Ethel x Clara Fla4B (FL) V. darrowi 2x Wild selection W85-20 (W2) V. corymbosum 2x Wild selection

Table 1. The genotypes (and their abbreviations) used in this study, along with their species, ploidy level, and immediate parents.

Table 2. Summary o	of blueberry cDNA clones	including their names,	sizes of cDNA inserts,	length of nucleotide sequence	ces, putative identification
from BLASTX	searches, and BLASTX E-	values.			

Clone	Size of cDNA insert	Length of nucleotide	Putative	
name	(≈base pairs)	sequence (bases)	identification	E value ^z
3	1700	5'-676	DNA J protein	1×10^{-41}
4	1400	5'-611, 3'-600	Nonidentified	NSY
9	900	5'-669, 3'-549	Nonidentified	NS
10	1250	3'-728	Phospholipid-hydroperoxide glutathione peroxidase	3×10^{-59}
13	1500	5'-607	Nonidentified	NS
14	850	3'-692	Nonidentified	NS
15	1600	5'-647, 3'-740	Pyruvate decarboxylase isozyme 1	1×10^{-102}
20	850	5'-656, 3'-688	Nonidentified	NS
21	1600	5'-350, 3'-399	Nonidentified	NS
22	1500	5'-348	Dessication-responsive protein 29B	2×10^{-4}
23	2000	5'-576, 3'-622	Putative protein (Arabidopsis thaliana)	2×10^{-17}
25	600	5'-602	U6 snRNA-associated Sm-like protein	2×10^{-45}
28	1200	5'-660	Auxin-responsive protein IAA8	6×10^{-17}
30	750	5'-601	Putative protein (Arabidopsis thaliana)	1×10^{-28}
31	1200	5'-632, 3'-584	Aldehyde dehydrogenase	2×10^{-53}
32	1400	5'-660, 3'-667	Putative protein (Arabidopsis thaliana)	6×10^{-30}
33	800	5'-345, 3'-428	MAP kinase kinase	6×10^{-47}
34	1500	5'-666	Low-temperature-induced 65 kDa protein	2×10^{-13}
36	600	5'-344	Dehydrin	5×10^{-20}
39	1650	5'-330, 3'-438	Putative protein (Arabidopsis thaliana)	9×10^{-27}
42	850	5'-675, 3'-715	Putative pollen allergen, beta expansin	1×10^{-87}
43	1400	5'-667, 3'-721	Calmodulin-binding protein	6×10^{-58}
44	1400	5'-606, 3'-655	ATP synthase alpha chain	2×10^{-67}
45	1200	5'-669, 3'-659	Putative RNA helicase	7×10^{-51}
47	1500	5'-627, 3'-730	Ubiquitin activating enzyme E1	1×10^{-93}
49	2500	5'-571, 3'-557	Nonidentified	NS
51	650	Complete	Dicyanin, blue copper protein	8×10^{-27}
52	1868	Complete	ABI3-interacting protein	6×10^{-66}
53	1044	Complete	Lysophosphatidic acid acyltransferase	1×10^{-75}
148	1600	5'-611, 3'-744	Phosphatidylinositol transfer-like protein	2×10^{-99}

^zProbability of a random association with a sequence in GenBank.

^yNo significant ($p \ge 10^{-3}$) association with a sequence in GenBank.

The cDNA library represents genes that are expressed mid-winter, when plants have reached their maximum level of cold hardiness. Table 2 summarizes characteristics of these clones. Good quality single-pass sequences were obtained from both the 5' and 3' ends of 17 clones, from the 5' ends of eight clones, and from the 3' ends of two clones. Complete sequences from both the 5' and 3' directions were obtained for three clones. Eight clones did not yield good 3'end sequences apparently because of difficulty sequencing through their long poly (A) tails. The ESTs produced from the single-pass sequences were 598 bases long on average, after trimming vector sequences. To characterize the cDNAs, nucleotide sequences were compared against known nucleotide and amino acid sequences in Genbank using the network BLASTN and BLASTX algorithms. ESTs with BLASTX scores >50 and/or E values $\leq 10^{-3}$ ($p \leq 0.001$) were considered to have significant similarity to known sequences and putative biochemical functions of the cDNAs were assigned. The identified cDNAs encode a wide range of proteins including temperature stress-related proteins (DNA J, low temperatureinduced 65 kDa protein, dehydrin, dessication-induced protein, glutathione peroxidase, ubiquitin activating enzyme E1), proteins involved in signal transduction (MAP kinase kinase, calmodulinbinding protein), hormone-associated proteins (auxin-responsive protein IAA8, ABI3-interacting protein), cell wall proteins (beta expansin), and basic metabolic proteins (pyruvate decarboxylase, ATP synthase alpha chain), among others.

DEVELOPMENT OF EST-PCR AND CAPS MARKERS AND DNA FINGERPRINTING. Thirty PCR primer pairs were designed from the ends of the available nucleotide sequence data. The primer pairs were tested in amplification reactions with DNA from the 19 blueberry genotypes. Fifteen primer pairs resulted in amplification of polymorphic fragments that were detectable directly after ethidium bromide staining of 1.4% agarose gels (Table 3). An example is shown in Fig. 1A. Eleven primer pairs resulted in amplification of monomorphic fragments (derived from ESTs 10, 14, 20, 25, 30, 32, 33, 42, 44, 52, and 53). An example is shown in Fig. 1B. The remaining five primer pairs did not work well (3, 9, 13, 36, and 41), yielding faint bands that would have been difficult to score. For those primer pairs that worked well and yielded either a single fragment or a major fragment plus some secondary fragments, the single or major amplification product was, in all cases, of the expected size (calculated from the sequence data and size of the cDNA inserts) or larger than expected, indicating the presence of introns. None of the major products were smaller than expected. Of the eleven monomorphic amplification products, four (derived from primer pairs 20, 33, 44, and 52) were digested with the restriction enzyme AluI and the digestion products were analyzed on 1.4% agarose gels. Amplification products from one other primer pair (45), that originally yielded only one polymorphic fragment present in one genotype along with a monomorphic fragment present in all genotypes, were also digested with AluI. AluI was chosen because

it is a four-base cutter, it cuts DNA at 37 °C, and it is inexpensive. Of these five digestions, three (from primer pairs 44, 45, and 52) resulted in polymorphisms (Table 3).

The type of inheritance, predicted or observed, of the polymorphic EST-PCR and CAPs markers is summarized in Table 3, along with the primer sequences and annealing temperatures. All the annealing temperatures were in the range of 52 to 60 °C. About half of the primer pairs yielded multiple fragments (more than two fragments even in the diploid genotypes). Because this study did not enable us to determine which fragments were derived from different alleles and which were derived from different genes, these markers were simply classified as being derived from a multigene family. A few primer pairs (21, 43, and 45) yielded one polymorphic-sized fragment (present in one parent and absent in the other) and one monomorphic fragment (present in both parent plants of the mapping population). Segregation patterns of these fragments in the mapping population could be explained by both a dominant and codominant model of inheritance; thus they were classified as not verifiable (see footnote to Table 3). Polymorphic markers resulting from restriction enzyme digestion were classified as codominant because that is the predicted mode of inheritance for CAPS markers if a product is amplified in both parental genotypes (44/*Alu*I, 45/*Alu*I, and 52/*Alu*I). Four primer pairs (4, 34, 49 and 51) yielded one fragment in one of the parents of the mapping population and no fragments in the other parent. Segregation of these markers in the mapping population was consistent with a dominant mode of inheritance (present or absent). One primer pair (47) yielded one fragment that was polymorphic among some of the genotypes but not between the parent plants of the mapping population; thus, the mode of inheritance for this fragment was also classified as not verifiable.

Amplification using the 17 primer pairs listed in Table 3, followed by digestion with *Alu*I for primer pairs 44, 45, and 52, resulted in a total of 37 polymorphic fragments (without digestion) and seven different digestion profiles. With these markers, all genotypes were distinguishable from each other. Of these fragments and digestion profiles, a subset of the most easily scored markers that were sufficient for all the possible distinctions was identified. This subset

Table 3. Summary of EST primers that yielded scorable polymorphic fragments among the genotypes tested.

Clone	Forward and reverse	Annealing	Genetic variation	Type of
name	primer sequences	temp (°C)	among genotypes	inheritance
4	GCC GCT CTT CTC TTC CTA GC	58	Polymorphic	Dominant
	CCA TCA AAC ACC ACC TAT GC			
15	CTA GAG GCT GCA GTG GAA GC	58	Polymorphic	Multigene family
	TTG CTC GTG TCG TCC TTA TG			
21	TCC GAT AAC CGT TAC CAA GC	56	Polymorphic	nv ^z
	TAT ACA GCG ACA CGC CAA AA			
22	CTG GTC ATA CCG GGC AAC	58	Polymorphic	Multigene family
	GAT CAA CGG GGT TCA TGG T			
23	GTT AGA GAG GGT TTC GAG GA	54	Polymorphic	Multigene family
	AGC AAA AAC TTC ACG CCA AT			
28	AAT GGA AAG AAG CTC TGA CA	54	Polymorphic	Multigene family
	ATC AGC CTC AGA TCC ACC AC			
31	AGC ATT TGA CAC CAG TCA CG	52	Polymorphic	Multigene family
	TTA CAG GAG GGG GAT TTT			
34	CTA AAG ACG GGC CTG AAG TG	60	Polymorphic	Dominant
	TCT GGT GAG AAC TGG TCG TG			
39	TAA TGA GTC TGT GGC GAA CG	56	Polymorphic	Multigene family
	AAC AAG ACC AAA CCC CAC AT			
43	GGC ACG AGG TCA CAG AAT GT	54	Polymorphic	nv
	ACA ATC CCA CCC AAA AAC AA			
44	AGC AGC GGT ATT CTC CTC AA	58	Polymorphic after digestion with AluI	Codominant
	CAC AGA ATC CAT TGA CAG CG			
45	CAA TGT TGG GGA GAA TGC TT	56	Polymorphic before and after digestion	nv before digestion
	AAG AGC AAA ATA CAC GCA CG		with AluI	codominant after digestion
47	CTG CTG ATC CTA GCC ACC TC	56	Polymorphic	nv
	AAA GGT TGC CCA AAA GTT CC			
49	GGG CAA AAC ATT TGA TTG GT	54	Polymorphic	Dominant
	CTT TCC GGC TTC TCA CAA AA			
51	GCT GCT CTT GTA CAG GGC TC	55	Polymorphic	Dominant
	TTG CGC ACA CAT AAA CCT AAA			
52	GGC AAG TGG GCA ATG ATA GT	58	Polymorphic after digestion with AluI	Codominant
	TCA CAA ACT GAC CCC GTA CA			
148	CAA GGG TGC ACG TGA ACT TA	56	Polymorphic	Multigene family
	AAA AGC ATT GCA GTC ACA CG			

 2 nv = not verifiable. The type of inheritance was not verifiable in the mapping population either because the amplified fragments were not polymorphic between the parent plants of the population or because the segregation pattern could fit either a dominant or a codominant model of inheritance. In this case, there were generally two fragments present in the parent plants, one that was present in both parents and one that was present in only one parent. This second fragment could have been an allele of the first fragment segregating in a codominant fashion or a separate gene from the first, segregating in a dominant fashion.

of EST-PCR markers included amplification products from four primer pairs , 21, 22, 34, and 49.

CLONING AND SEQUENCE ANALYSIS OF EST-PCR PRODUCTS. The homology of 20 EST-PCR markers was investigated by DNA sequencing. In making our selection of PCR products to sequence, we chose a sufficient number to give a good indication of whether the fragments were generally homologous to the cDNAs from which the primer sequences were derived. We sequenced approximately equal numbers of monomorphic and polymorphic fragments, including several fragments that would be useful for mapping in our populations (polymorphic between the original parent plants of the mapping populations). Specifically, eight monomorphic and 12 polymorphic fragments were cloned and sequenced and the

Α.



В.



Fig. 1. Amplification products resulting from PCRs using forward and reverse primer pairs derived from (A) EST 34 and (B) EST 52 and DNA from the 19 blueberry genotypes. Lanes 1-20 were loaded in the following order: (1) 1-kb ladder, (2) 'Berkeley', (3) 'Bluegold', (4) 'Bluecrop', (5) 'Bluejay', (6) 'Blueray', (7) 'Cooper', (8) 'Duke', (9) 'Georgiagem', (10) 'Gulfcoast', (11) 'Nelson', (12) 'Patriot', (13) 'Sierra', (14) 'Sunrise', (15) 'Toro', (16) 'Weymouth', (17) 'Climax', (18) 'Tifblue', (19) Fla4B, and (20) W85-20.

sequences were analyzed (Table 4). Of the 20 EST-PCR products examined, all but three were homologous to the cDNAs from which the primer sequences were designed. The three that were not homologous were generated from primers that amplified some of the highest numbers of fragments (>5) per genotype of all the primers. On the other hand, three different cloned fragments amplified using primer pair 28, which yielded numerous PCR fragments, all showed homology to the original cDNA clone 28. These results show that, if an EST-PCR or CAPS marker has been mapped, one should not assume that the particular gene from which the PCR primer sequences were derived has been mapped until the EST-PCR product has been cloned and sequenced and its homology has been established. However, even the EST-PCR markers that lack homology to the original cDNA clones may still be used for DNA fingerprinting and genetic studies if they are reproducible, as are all the markers described in this study.

GENETIC SIMILARITY COEFFICIENTS AND COEFFICIENTS OF COAN-CESTRY. Similarity coefficients based on the molecular marker data were calculated for each pair of genotypes using the simple matching (symmetrical, including double-zeros) and Jaccard (asymmetrical, excluding double-zeroes) coefficient functions of the NTSYS-pc program. The dendrograms derived from the two matrices were nearly identical; therefore, only the similarity matrix based on the simple matching coefficient function (data not shown) is discussed. The simple matching similarity values ranged from 0.432 to 0.932, and the average similarity value was 0.740. The average similarity value among just the V. corymbosum cultivars was 0.789. The most divergent V. corymbosum cultivar was 'Sierra', with an average pairwise similarity value of 0.728 (based on just the V. corymbosum cultivar/V. corymbosum cultivar pairwise comparisons). 'Sierra' is a hybrid of four different species (V. corymbosum, V. darrowi, V. constablaei Gray, and V. ashei); thus, this divergence is consistent with its pedigree (Table 1). The most divergent genotypes overall were the *V.ashei* cultivar Tifblue (average similarity value of 0.605), followed by the V. darrowi selection Fla4B (average similarity value of 0.645), the diploid V. corymbosum selection W85-20 (average similarity value of 0.683), and the V. corymbosum cultivar Sierra (average similarity value of 0.685). The genotypes 'Bluecrop' and 'Nelson' had the highest similarity value (0.932) of all the pairwise comparisons. This also is in agreement with their pedigrees, since 'Bluecrop' is a parent of 'Nelson'.

Coefficients of coancestry were calculated for all possible pairs of the 19 genotypes based on complete pedigree information (data not shown). Polysomic inheritance was assumed for the tetraploids based on the findings of Krebs and Hancock (1989) and Qu and Hancock (1995), as was done by Ehlenfeldt (1994), for calculations of inbreeding coefficients. Coefficients of coancestry ranged from 0, for those genotypes that shared no parents in common, to 0.044, for the two pairs of genotypes that are parent and offspring ('Bluecrop'/'Nelson' and 'Berkeley'/'Bluejay'), to even 0.45 for 'Weymouth' and 'Toro', which shared several ancestors in common. The average coefficient of coancestry value was 0.006. The average coefficient of coancestry value among the *V. corymbosum* cultivars was 0.008.

GENETIC RELATEDNESS TREES. The dendrogram based on a cluster analysis of the simple matching similarity matrix is shown in Fig. 2A. A cophenetic value matrix was produced from the tree matrix to measure the goodness of fit to the similarity matrix. An r value of 0.775 indicated a very good fit (data not shown).

A dendrogram based on cluster analysis of the coefficients of coancestry matrix is shown in Fig. 2B. This tree is similar to the tree based on similarity coefficients in some general ways but differs

Га	ble 4. Summary	of EST-PCR	products that w	ere cloned and	sequenced,	including the	primers and	d genotypes	(abbreviation	s given in '	Table 1)
	that were used,	, sizes of fragr	nents in bp, gene	etic variation of	oserved amo	ong genotypes	, and whethe	er the PCR p	roducts were	homologo	us to the
	original cDNA	clones from	which the prime	r sequences we	re derived.						

EST-PCR		Fragment size	Monomorphic (M) or	Homologous
primer pair	Genotype	(bp)	polymorphic (P)	to cDNA
4	FL	1200	Р	Yes
15	GG	2100	Р	Yes
		2000	М	Yes
22	FL	1300	Р	Yes
28	FL	1650	Р	Yes
		1500	М	Yes
		300	М	Yes
31	BR	1650	Р	Yes
39	FL	450	Р	No
43	W2	1200	М	Yes
44	FL	1500	М	Yes
45	FL	1900	Р	Yes
		1600	М	Yes
47	ТВ	2000	Р	Yes
		1650	М	Yes
49	W2	1000	Р	Yes
		400	Р	No
51	W2	1400	Р	Yes
52	FL	700	М	Yes
148	BJ	2100	Р	No

in much of the detail. In the tree based on similarity coefficients, the first division, at 65% similarity, separated the V. ashei cultivars, Tifblue and Climax, from the remaining genotypes. Next, at about 66% similarity, Fla4B and W85-20, the only diploid V. darrowi genotype and the only diploid V. corymbosum genotype, respectively, were separated from the remaining genotypes. The next division, at about 73% similarity, separated 'Sierra', the cultivar with multiple species in its background from the other V. corymbosum cultivars. In the tree based on coefficients of coancestry, the V. ashei cultivars, Tifblue and Climax, grouped separately from the V. corymbosum cultivars, as did the only diploid V. corymbosum genotype, W85-20. In the coefficients of coancestry tree, however, the diploid V. darrowi genotype, Fla4B, clustered together with the southern V. corymbosum x V. darrowi hybrid cultivars, Cooper, Georgiagem, and Gulfcoast, for which it is a parent, rather than remaining separate. In the similarity coefficient tree, 'Cooper' and 'Gulfcoast', full sibs, clustered together with 'Sunrise', their half sib, rather than with 'Georgiagem' and Fla4B. In the tree based on similarity coefficients, 'Bluecrop', 'Nelson', 'Blueray', and 'Georgiagem' clustered together. 'Bluecrop' is a parent of 'Nelson', a full sib of 'Blueray', and a grandparent of 'Georgiagem' on both sides. Likewise, in the coefficients of coancestry tree, 'Bluecrop', 'Nelson', and 'Blueray' grouped together, but this cluster included 'Bluegold' rather than 'Georgiagem'. Clustering of most of the other V. corymbosum cultivars was quite different between the two trees.

COMPARISON OF RELATIONSHIP MEASURES. A correlation test was performed on the genetic similarity matrix based on the molecular marker data and the coefficient of coancestry matrix based on pedigree information. A fair positive correlation (r = 0.317) was found, which was highly significant (p < 0.0001). This was not surprising since an empirical examination of the two dendrograms suggested that they agreed in some ways but differed in others. The similarity matrix and resulting dendrogram is based on only 37 EST-PCR markers and seven digestion profiles; thus, a better correlation with the coefficients of coancestry matrix might be found with the development and analysis of additional markers. However, in our

previous studies in blueberry using hundreds of RAPD markers (Levi and Rowland, 1997) and in strawberry using hundreds of AFLP markers (Degani et al., 2001), we still found only weak to moderate correlations between genetic similarity values based on molecular marker data and coefficients of coancestry. A cause for these fair, at best, correlations may be inaccuracies in both estimates. Coefficients of coancestry are based on probabilities of inheriting the same genes; therefore, they are indirect estimates of similarity and may not always reflect the true relationships between accessions (Graner et al., 1994; Russell et al., 1997). The assumptions underlying the coefficients of coancestry calculations also may not be completely fulfilled. Coefficients of coancestry assume no natural or breeder's selection and assume that all original ancestors are equally unrelated, which is probably not true (Schut et al., 1997). The accuracy of genetic similarity estimates based on molecular markers depends on several variables as well, such as the number of markers analyzed, their distribution over the genome if linkage disequilibrium is involved, and correct scoring of the markers (Schut et al., 1997).

In conclusion, the EST-PCR and CAPS markers proved very effective at distinguishing the cultivars in this study and many were polymorphic between the parent plants of the mapping population. A selection of 20 EST-PCR products were cloned and sequenced and homology to the original cDNA clones confirmed for all but three. A dendrogram constructed from genetic similarity values, based on molecular marker data, agreed fairly well with pedigree information. Thus, these markers should have general utility for DNA fingerprinting and for estimates of genetic relatedness in blueberry. In addition, these markers are being added to our current genetic linkage maps (data not shown). Because the markers are derived from ESTs produced from a cDNA library made from RNA from cold-acclimated plants, they should be particularly useful for mapping QTLs controlling cold hardiness.

In studies with other plants, amplification using EST-PCR primers generally had to be followed by either digestion with restriction enzymes to generate CAPS markers, heteroduplex analysis,



B.



Fig. 2. Dendrograms of 19 blueberry genotypes generated by UPGMA cluster analysis of the (A) simple matching similarity matrix based on molecular markers and (B) coefficients of coancestry matrix based on pedigree information. Genotypes are designated by abbreviations given in Table 1.

or SSCP analysis to detect polymorphisms (Cato et al., 2001). In the case of blueberry, 15 out of the 30 tested EST-PCR primer pairs resulted in amplification of polymorphic fragments that were detectable directly after ethidium bromide staining of agarose gels, and amplification reactions with as few as four of the primer pairs could distinguish all the genotypes in this study. Blueberry species are primarily outcrossing and exhibit low to moderate levels of self-fertility. Self-pollinations are not often used in blueberry cultivar development due to inbreeding depression as evidenced by reduced seed set and poor germination (Galletta and Ballington, 1996). The inherent breeding behavior of blueberry together with past breeding techniques used for cultivar development may explain the high level of polymorphic EST-PCR markers observed without the need for additional modifications or analyses.

Literature Cited

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410.

Aruna, M., P. Ozias-Akins, M.E. Austin, and G. Kochert. 1993. Genetic relatedness among rabbiteye blueberry (Vaccinium ashei) cultivars determined by DNA amplification using single primers of arbitrary sequence. Genome 36:971-977.

Aruna, M., M.E. Austin, and P. Ozias-Akins. 1995. Randomly amplified polymorphic DNA fingerprinting for identifying rabbiteye blueberry (Vaccinium ashei Reade) cultivars. J. Amer. Soc. Hort. Sci. 120:710-713.

Burgher, K.L., A.R. Jamieson, and X. Lu. 2002. Genetic relationships among lowbush blueberry genotypes as determined by randomly amplified polymorphic DNA analysis. J. Amer. Soc. Hort. Sci. 127:98-103.

Cato, S.A., R.C. Gardner, J. Kent, and T.E. Richardson. 2001. Arapid PCR-based method for genetically mapping ESTs. Theor. Appl. Genet. 102:296-306. Degani, C., L.J. Rowland, J.A. Saunders, S.C. Hokanson, E.L. Ogden, A. Golan-Goldhirsh, and G.J. Galletta. 2001. A comparison of genetic relationship measures in strawberry (Fragaria x ananassa Duch.) based on AFLPs, RAPDs, and pedigree data. Euphytica 117:1-12.

Doyle, J.J. and J.L. Doyle. 1990. Isolation of plant DNA from fresh tissue. Focus 12:13-15.

Ehlenfeldt, M.K. 1994. The genetic composition and tetrasomic inbreeding coefficients of highbush blueberry cultivars. HortScience 29:1342-1345.

Galletta, G. J. and J.R. Ballington. 1996. Blueberries, cranberries, and lingonberries, p. 1-107. In: J. Janick and J.N. Moore (eds.). Fruit breeding, Vol. II: Vine and small fruits crops. Wiley, New York.

Graner, A., W.F. Ludwig, and A.E. Melchinger. 1994. Relationships among European barley germplasm. II. Comparison of RFLP and pedigree data. Crop Sci. 34:1199-1205.

Jaccard, P. 1908. Nouvelles reseherches sur la distribution florale. Bul. Soc. Vaud. Sci. Nat. 44: 223-270.

Jones, C.J., K.J. Edwards, S. Castaglione, M.O. Winfield, F. Sala, C. van de Wiel, G. Bredemeijer,

B. Vosman, M. Matthes, A. Daly, R. Brettschneider, P. Bettini, M. Buiatti, E. Maestri, A. Malcevschi, N. Marmiroli, R. Aert, G. Volckaert, J. Rueda, R. Linacero, A. Vazquez, and A. Karp. 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Mol. Breeding 3:381-390.

- Krebs, S.L. and J.F. Hancock. 1989. Tetrasomic inheritance of isozyme markers in the highbush blueberry, Vaccinium corymbosum L. Heredity 63:11-18.
- Levi, A., G.R. Panta, C.M. Parmentier, M.M. Muthalif, R. Arora, S. Shanker, and L.J. Rowland. 1999. Complementary DNA cloning, sequencing, and expression of a unique dehydrin from blueberry floral buds. Physiol. Plant. 107:98-109.
- Levi, A. and L.J. Rowland. 1997. Identifying blueberry cultivars and evaluating their genetic relationships using randomly amplified polymorphic DNA (RAPD) and simple sequence repeat- (SSR-) anchored primers. J. Amer. Soc. Hort. Sci. 122:74-78.

Levi, A., L.J. Rowland, and J.S. Hartung. 1993. Production of reliable randomly amplified polymorphic DNA (RAPD) markers from DNA of woody plants. HortScience 28:1188-1190.

Moore, J.N. 1993. The blueberry industry of North America. Acta Hort. 346:15-26.

- Panta, G.R., L.J. Rowland, R. Arora, E.L. Ogden, and C.C. Lim. Inheritance of cold hardiness and dehydrin genes in diploid mapping populations of blueberry. J. Crop Prod. (in press).
- Perry, D.J. and J. Bousquet. 1998. Sequence-tagged-site (STS) markers of arbitrary genes: the utility of black spruce-derived STS primers in other conifers. Theor. Appl. Genet. 97:735–743.
- Qu, L. and J.F. Hancock. 1995. Nature of 2n gamete formation and mode of inheritance in interspecific hybrids of diploid *Vaccinium darrowi* and tetraploid *V. corymbosum*. Theor. Appl. Genet. 91:1309–1315.
- Qu, L. and J.F. Hancock. 1997. Randomly amplified polymorphic DNA-(RAPD-) based genetic linkage map of blueberry derived from an interspecific cross between diploid *Vaccinium darrowi* and tetraploid *V. corymbosum.* J. Amer. Soc. Hort. Sci. 122:69–73.
- Rowland, L.J. and A. Levi. 1994. RAPD-based genetic linkage map of blueberry derived from a cross between diploid species (*Vaccinium darrowi* and *V. elliottii*). Theor. Appl. Genet. 87:863–868.
- Rowland, L.J., E.L. Ogden, R. Arora, C.C. Lim, J.S. Lehman, A. Levi, and G.R. Panta. 1999. Use of blueberry to study genetic control of chilling requirement and cold hardiness in woody perennials. HortScience 34: 1185–1191.
- Russell, J.R., J.D. Fuller, M. Macaulay, B.G. Hatz, A. Jahoor, W. Powell, and R. Waugh. 1997. Direct comparisons of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs.

Theor. Appl. Genet. 95:714-722.

- Schubert, R., G. Mueller-Starck, and R. Riegel. 2001. Development of EST-PCR markers and monitoring their intrapopulational genetic variation in *Picea abies* (L.) Karst. Theor. Appl. Genet. 103:1223–1231.
- Schut, J.W., X. Qi, and P. Stam. 1997. Association between relationship measures based on AFLP markers, pedigree data and morphological traits in barley. Theor. Appl. Genet. 95:1161–1168.
- Stommel, J.R., G.R. Panta, A. Levi, and L.J. Rowland. 1997. Effects of gelatin and BSA on the amplification reaction for generating RAPD. BioTechniques 22:1064–1066.
- Sunnucks, P. 2000. Efficient genetic markers for population biology. Tree 15:199–203.
- Temesgen, B., G.R. Brown, D.E. Harry, C.S. Kinlaw, M.M. Sewell, and D.B. Neale. 2001. Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda* L.). Theor. Appl. Genet. 102:664–675.
- Tsumura, Y., Y. Suyama, K. Yoshimura, N. Shirato, and Y. Mukai. 1997. Sequence-tagged-sites (STSs) of cDNA clones in *Cryptomeria japonica* and their evaluation as molecular markers in conifers. Theor. Appl. Genet. 94:764–772.
- Williams, J.G.K., A.R. Kubelik, K.J. Livak, J.A. Rafalski, and S.V. Tingey. 1990. DNA polymorphism amplified by arbitrary primers are useful as genetic markers. Nucl. Acids Res. 18:6531–6535.