

# Genetic Diversity in the Highbush Blueberry Evaluated with Microsatellite Markers

Peter Boches, Nahla V. Bassil, and Lisa Rowland

USDA–ARS, National Clonal Germplasm Repository, 33447 Peoria Road, Corvallis, OR 97333

ADDITIONAL INDEX WORDS. *Vaccinium corymbosum*, simple sequence repeats, SSRs, DNA fingerprinting

**ABSTRACT.** Sixty-nine accessions representing wild and domesticated highbush blueberry (*Vaccinium corymbosum* L.) germplasm were genotyped using 28 simple sequence repeats (SSRs). A total of 627 alleles was detected and unique fingerprints were generated for all accessions. Suspected duplicate accessions of ‘Coville’ and ‘Ivanhoe’ had DNA fingerprints that were identical to ‘Coville’ and ‘Ivanhoe’, respectively. Genetic similarity measures placed wild and cultivated blueberries in separate groups. Northern highbush blueberries grouped among ancestral clones that were used extensively in blueberry breeding such as ‘Rubel’ and ‘Stanley’. Southern highbush blueberries formed a separate group from northern highbush blueberries. The microsatellite markers used here show excellent promise for further use in germplasm identification, in genetic studies of wild *Vaccinium* L. populations, and for constructing linkage maps.

The highbush blueberry, *Vaccinium corymbosum*, is a perennial shrub cultivated for its edible fruit. Highbush blueberry is a high-value crop produced in the United States, Canada, Europe, Australia, New Zealand, Chile, and Argentina (Galletta and Ballington, 1996). The wild highbush blueberry is indigenous to subtropical and temperate climates in the eastern third of North America. Highbush is also the major cultivated blueberry type grown in temperate North America and the world.

*Vaccinium corymbosum* is believed to occur at three ploidy levels,  $2n = 2X = 24$ ,  $4X = 48$ , and  $6X = 72$  (Vander Kloet, 1988) but its taxonomy remains unsettled. The cultivated hexaploid was considered as a form of *V. corymbosum* by Vander Kloet. However, it is considered ecologically and morphologically distinct by some breeders (Qu et al., 1998; Ritzinger and Lyrene, 1998) and is referred to in the literature as *V. ashei* Reade or *V. virgatum* Aiton. In this article we refer to it as *V. virgatum*.

The highbush blueberry was domesticated in the 20th century. The first breeding program for the predominant type, known as “northern highbush,” was initiated by F. Coville of the U.S. Department of Agriculture (USDA) (Coville, 1937). Ehlenfeldt (1994) summarized the contributions of wild germplasm to the highbush blueberry genepool. Founder selections ‘Brooks’, ‘Sooy’, and ‘Rubel’ have contributed the majority of genes to the highbush blueberry genepool at overall levels of 23.5%, 12.5%, and 28.6%, respectively, for the period from 1908 to 1994. The lowbush blueberry, *V. angustifolium* Aiton, also contributed to the initial genepool through ‘Russell’ and ‘North Sedgewick’. There has also been a consistent increase in the average inbreeding coefficients of new highbush blueberry cultivars.

Southern highbush blueberry cultivars were recently developed from northern highbush cultivars through the introgression of genes from *V. darrowii* Camp, and to a lesser extent, *V. virgatum* and *V. tenellum* Aiton. Southern highbush blueberries are better adapted to warmer climates than northern highbush blueberries. Among southern highbush blueberry cultivars, *V. darrowii* is the largest contributor of genetic material after *V. corymbosum*. Despite concerns that inbreeding depression may be a serious problem in future blueberry breeding efforts (Haghighi and

Hancock, 1992; Hancock and Siefker, 1982), some of the most successful cultivars have high inbreeding coefficients (Ehlenfeldt, 1994).

Traditional methods of characterizing and identifying cultivars in fruit crops are based on phenotypic observations, a slow approach that is subject to environmental influences. DNA-based markers reveal a high degree of polymorphism and can accelerate genotype fingerprinting and studies of genetic relationships among cultivars. Randomly amplified polymorphic DNA (RAPD) markers were used to distinguish 26 genotypes of lowbush blueberry (*V. angustifolium* Aiton) (Burgher et al., 2002) and for cultivar identification of rabbiteye blueberry (*V. virgatum*) and highbush blueberry (Aruna et al., 1993, 1995; Levi and Rowland, 1997) as well as closely related cranberry (*V. macrocarpon* Aiton) (Novy and Vorsa, 1995, 1996a, 1996b; Novy et al., 1994). Three inter-simple sequence repeat (ISSR) and 15 RAPD primer pairs differentiated 15 highbush blueberry cultivars and two rabbiteye blueberry cultivars (Levi and Rowland, 1997). ISSR techniques are similar to RAPD techniques in their dominant nature and use of random primers. However, ISSR primer sequences are designed from microsatellite regions and the annealing temperatures used are higher than those used for RAPD markers. Clustering of *V. corymbosum* accessions based on molecular data did not agree with pedigree information (Levi and Rowland, 1997). Furthermore, RAPD markers are dominant and limited by lack of reproducibility between labs due to slight differences in thermocycling conditions or buffer composition (Powell et al., 1996). Therefore, more reproducible marker systems are needed in blueberry. Rowland et al. (2003) recently developed expressed sequence Tag-PCR (EST-PCR) and cleaved amplified polymorphism (CAPS) markers for blueberry (Rowland et al., 2003), which present a significant improvement over random primed DNA fingerprinting methods. However, when compared to microsatellite or simple sequence repeat (SSR) markers, EST-PCR markers are not as polymorphic and CAPS markers are not amenable to high throughput. SSR markers constitute one of the most reproducible and polymorphic types of markers and have not been previously reported in blueberry. SSRs are short (1–6 bp) tandemly repeated DNA sequences flanked by unique, conserved DNA sequences (Tautz and Rentz, 1984; Weber and May, 1989). Microsatellites were discovered in humans and have been found

in the genomes of most organisms (Li et al., 2002). They have become the preferred codominant marker for forensics and DNA fingerprinting (Wünsch and Hormaza, 2002). SSRs were used to examine crop origins and the domestication process in corn (*Zea mays* L.) (Matsuoka et al., 2002b), barley (*Hordeum vulgare* L.) (Ivancic et al., 2004), potato (*Solanum tuberosum* L.) (Raker and Spooner, 2002), almond (*Prunus dulcis* D.A. Webb) (Xu et al., 2004), and grape (*Vitis vinifera* L.) (Aradhya et al., 2003). SSR markers are frequently used to estimate genetic diversity in fruit crops such as apricot (*Prunus armeniaca* L.) (Romero et al., 2003), kiwifruit (*Actinidia deliciosa* A. Chev) (Zhen et al., 2004), peach (*Prunus persica* L.) (Aranzana et al., 2002), and apple (*Malus domestica* Borkh.) (Hokanson et al., 2001). DNA fingerprints created with SSR markers are assisting in the management of fruit germplasm collections in apple (Hokanson et al., 1998), cherry (*Prunus* L.) (Cantini et al., 2001), peach (Testolin et al., 2000), and grape (Dangl et al., 2001; Lamboy and Alpha, 1998; Martin et al., 2003). However, the main limitation of microsatellite markers is that they have to be isolated de novo for unrelated species.

The objective of this study was to use allelic variation at 28 microsatellite loci to assess the level of genetic diversity in 69 *Vaccinium* genotypes. These accessions include wild *V. corymbosum* from New York, North Carolina, and Maine, southern highbush

blueberry cultivars, historical and modern cultivars of northern highbush blueberry. The resulting SSR fingerprints should provide a reproducible means of identifying these cultivars for managing this germplasm in gene banks and nurseries.

## Materials and Methods

**PLANT MATERIAL AND DNA PREPARATION.** Plant materials sampled included 58 *Vaccinium* accessions held at the USDA–Agricultural Research Service (ARS)–National Clonal Germplasm Repository (NCGR) and 13 *Vaccinium* cultivars obtained from Fall Creek Farm and Nursery in Lowell, Ore. (Table 1). In total, 69 genotypes were sampled: 12 accessions of wild *V. corymbosum* (New York, North Carolina, Maine), one accession listed as a “*V. myrtilloides* Michx. hybrid”; 21 historical northern highbush blueberry cultivars (decade of hybridization prior to 1950) including two founding selections (‘Rubel’ and ‘Grover’) and 15 modern northern highbush blueberry cultivars (decade of hybridization after 1950); 14 southern highbush blueberry cultivars; one pentaploid (2n = 5x = 60) hybrid cultivar (‘Pearl River’); one half-high blueberry cultivar (‘Northsky’, a tetraploid hybrid of *V. angustifolium* and *V. corymbosum*) (Table 1). Also included were four cultivars (‘Draper’, ‘Legacy’, ‘Ozarkblue’, and ‘Sierra’) with chilling requirements and winter hardiness

Table 1. List of 69 blueberry accessions and two suspected duplicates analyzed with 28 SSR loci. USDA Plant introduction numbers (accession no.), name or taxon for species, pedigree or origin, and cultivar type for cultivars or wild for species are included. NHB(M) = modern northern highbush, NHB(H) = historical northern highbush, PHB = pentaploid highbush blueberry, SHB = southern highbush, I = intermediate to southern and northern highbush blueberries in their chilling requirements and winter hardiness with genetic contributions from southern highbush, HH = half-high.

Accession no.	Name	Pedigree or origin	Type
PI 554833	1613A	Pioneer x Rubel	NHB(M)
PI 554798	Atlantic	Jersey x Pioneer	NHB(H)
None	Aurora <sup>w</sup>	Brigitta Blue x Elliott	NHB(M)
PI 554883	Berkeley	Stanley x (Jersey x Pioneer)	NHB(H)
PI 618193	Biloxi	Sharpblue x US 329	SHB
PI 554827	Bluecrop	(Jersey x Pioneer) x (Stanley x June)	NHB(H)
PI 554846	Bluejay	Berkeley x (Pioneer x Taylor)	NHB(M)
PI 554799	Blueray	(Jersey x Pioneer) x (Stanley x June)	NHB(H)
PI 554837	Bluetta	(North Sedgewick lowbush blueberry x Coville) x Earliblue	NHB(M)
PI 554826	Cabot	Brooks x Chatsworth	NHB(H)
None	Chandler <sup>w</sup>	Darrow x M-23	NHB(M)
PI 554829	Coville <sup>y</sup>	(Jersey x Pioneer) x Stanley	NHB(H)
PI 618035	Darrow	F-72 x Bluecrop	NHB(M)
None	Draper <sup>w</sup>	Duke x (290-2 x MU 652)	I
PI 554872	Duke	(Ivanhoe x Earliblue) x (E 30 x E 11)	NHB(M)
PI 554893	Earliblue	Stanley x Weymouth	NHB(H)
PI 554894	Elliott	Burlington x US 1	NHB(M)
None	Emerald <sup>w</sup>	FL 91-69 x NC 1528	SHB
PI 554957	Flordablue	FL 63-29 x FL 63-12	SHB
PI 554873	Georgiagem	(Bluecrop x E-118) x (Fla. 4B x Bluecrop)	SHB
PI 554804	Grover	Wild founder selection, New Jersey	NHB(H)
PI 554805	Herbert	Stanley x GS-149	NHB(H)
PI 554807	Ivanhoe <sup>y</sup>	(Rancocas x Carter) x Stanley	NHB(H)
PI 554808	Jersey	Rubel x Grover	NHB(H)
None	Jewel <sup>w</sup>	Unknown backcross selection, Univ. of Florida	SHB
PI 554810	June	(Brooks x Russell) x Rubel	NHB(H)
PI 618164	Legacy	Elizabeth x (Fla. 4B x Bluecrop)	I
None	Liberty <sup>w</sup>	Brigitta Blue x Elliott	NHB(M)
PI 618194	Magnolia	(Harrison x Avonblue) x FL 2-5	SHB
PI 554832	Meador	Earliblue x Bluecrop	NHB(M)

Table 1 continued on next page.

intermediate between southern and northern highbush cultivars. For ‘Coville’ and ‘Ivanhoe’, two plants of each cultivar that were donated to the NCGR by different cooperators and are maintained as separate inventory items were sampled.

DNA was extracted in duplicate from young leaves of plants grown at the NCGR in Corvallis with the PUREGENE kit (Gentra Systems Inc., Minneapolis) using the optional RNase A treatment.

**SIMPLE SEQUENCE REPEAT GENOTYPING.** The 28 primer pairs and PCR reaction conditions were previously described (Boches et al., 2005). SSR loci with sequence similarity to known genes were determined as described previously (Boches et al., 2005). The locations of SSRs relative to gene features such as the 5′ and 3′ untranslated region (5′UTR and 3′UTR), exons, and introns were predicted by aligning the SSR containing DNA sequence to the most similar genomic sequences in GenBank using ClustalX

(Chenna et al., 2003). The DNA duplicate samples were tested with primers CA112F and CA23F for identical profiles before one sample was selected for the final analysis. The success of the PCR reaction was verified by agarose (2%) gel electrophoresis. Failure to amplify was scored as a null genotype after three failed attempts on a DNA template that amplified using other primers.

PCR reactions were carried out separately for each primer pair and up to three PCR products (one per SSR primer set) were pooled in a multiplex and separated using capillary electrophoresis. PCR reactions for primers CA23F, CA112F, CA169F, NA398, and NA741 were carried out in 10-μL volumes using fluorescently labeled forward primers (FAM, HEX, or NED) and unlabeled reverse primers. The PCR reactions were diluted with water to a factor ranging from 1:80 to 1:320 and 0.5 μL was injected into an ABI 3100 capillary sequencer (Applied Biosystems, Foster City, Calif.) at the Central Services Laboratory (Oregon State

Table 1. Continued from previous page.

Accession no.	Name	Pedigree or origin	Type
None	Millennia <sup>w</sup>	FL 85-69 x O’Neal	SHB
PI 555317	Misty	FL 67-1 x Avonblue	SHB
PI 554952	Northland	Berkeley x 19-H	NHB(M)
PI 554943	Northsky	B-6 x R2P4	HH
None	Nui	(Ashworth x Earliblue) x Bluecrop	NHB(M)
PI 554812	Olympia	Pioneer x Harding	NHB(H)
PI 554944	O’Neal <sup>w</sup>	Wolcott x FL 64-15	SHB
None	Ozarkblue	G 144 x FL 4-76	I
PI 554843	Patriot	US3 x Earliblue	NHB(M)
PI 618192	Pearl River	G 136 x Beckyblue	PHB
PI 554815	Pioneer	Brooks x Sooy	NHB(H)
PI 554816	Rancocas	(Brooks x Russell) x Rubel	NHB(H)
PI 554817	Rubel	Wild founder selection, New Jersey	NHB(H)
None	Santa Fe <sup>w</sup>	Avonblue OP seedling	SHB
None	Sapphire <sup>w</sup>	Unknown backcross selection, Univ. of Florida	SHB
PI 554948	Sharpblue <sup>w</sup>	FL 61-5 x FL 62-4	SHB
PI 618099	Sierra	US 169 x G-156	I
None	Southmoon <sup>w</sup>	FL 80-46 x unknown northern highbush	SHB
PI 554845	Spartan	Earliblue x US 11-93	NHB(M)
PI 554820	Stanley	Katharine x Rubel	NHB(H)
None	Star <sup>w</sup>	FL 80-31 x O’Neal	SHB
PI 618023	Toro	Earliblue x Ivanhoe	NHB(M)
PI 554841	11-104	(Jersey x Pioneer) x (Stanley x June)	NHB(H)
PI 554825	F-72	Wareham x Pioneer	NHB(H)
PI 554821	Wareham	Rubel x Harding	NHB(H)
PI 554823	Weymouth	June x Cabot	NHB(H)
PI 267844 <sup>z</sup>	<i>Vaccinium corymbosum</i>	Maine	Wild
PI 267847 <sup>z</sup>	<i>V. corymbosum</i>	New York	Wild
PI 267848 <sup>z</sup>	<i>V. corymbosum</i>	New York	Wild
PI 296395 <sup>z</sup>	<i>V. corymbosum</i>	New York	Wild
PI 296396 <sup>z</sup>	<i>V. corymbosum</i>	Maine	Wild
PI 296400 <sup>z</sup>	<i>V. corymbosum</i>	Maine	Wild
PI 296403 <sup>z</sup>	<i>V. corymbosum</i>	Maine	Wild
PI 296414 <sup>z</sup>	<i>V. myrtilloides</i> hybrid	New York	Wild
PI 554792 <sup>z</sup>	<i>V. corymbosum</i>	Maine	Wild
PI 554793 <sup>z</sup>	<i>V. corymbosum</i>	Maine	Wild
PI 554795 <sup>z</sup>	<i>V. corymbosum</i>	Maine	Wild
PI 554880 <sup>x</sup>	<i>V. corymbosum</i>	North Carolina	Wild
PI 554881 <sup>x</sup>	<i>V. corymbosum</i>	North Carolina	Wild

<sup>z</sup>Unknown ploidy level.

<sup>y</sup>Two plants of ‘Ivanhoe’ (CVAC 60.001, CVAC 90.001) and two plants of ‘Coville’ (CVAC 53.001, CVAC 95.001) were sampled.

<sup>x</sup>Identifies diploid wild germplasm.

<sup>w</sup>Material obtained from Fall Creek Farm and Nursery (Lowell, Ore.).

University, Corvallis). GeneScan (version 2.1; Applied Biosystems) and Genotyper (version 2.0; Applied Biosystems) were used for automated data collection and computation of allele size and accurate visualization of the alleles, respectively. Alleles were scored by fitting peaks into bins encompassing less than one nucleotide, and peaks significantly smaller than the tallest peak (e.g., <20% of the height of the tallest peak in the lane) were not scored.

The amplification products of the remaining primers were analyzed using a CEQ 8000 genetic analyzer (Beckman Coulter Inc., Fullerton, Calif.). PCR reactions were carried out in 15- $\mu$ L volumes for each primer pair using fluorescently labeled forward primers [using either WellRED (Beckman Coulter Inc.) or Light-Sabre (Synthegen, LLC., Houston)] and unlabeled reverse primers. The optimal amount of PCR product (ranging from 0.05 to 1.5  $\mu$ L) to inject into the CEQ 8000 genetic analyzer was determined experimentally for each primer pair. Allele sizing and visualization were performed using the fragment analysis module of the CEQ 8000 software (Beckman Coulter Inc.). Alleles were scored by fitting peaks into bins less than one nucleotide wide. Unique alleles were verified by comparison to neighboring alleles. Stutter peaks were not scored. For primers that displayed non-template A addition, only the plus A peak was scored. Primers that do not produce stutter peaks or non-template A addition are indicated (Table 2). True peaks from a single primer pair were treated as belonging to the same locus for the purpose of computing genetic distances.

**DATA ANALYSIS.** Genetic distance matrices were computed using PowerMarker (version 3.0; Liu and Muse, 2004) with the proportion of shared alleles distance ( $D_{sa}$ ):

$$D_{sa} = \frac{1}{m} \sum_{j=1}^m \sum_{i=1}^{a_j} \min(p_{ij}, q_{ij})$$

where  $p_{ij}$  and  $q_{ij}$  are the frequencies of the  $i$ th allele at the  $j$ th locus,  $m$  is the number of loci examined, and  $a_j$  is the number of alleles at the  $j$ th locus. The exact copy number of each allele could not be inferred in the tetraploid blueberry individuals. Therefore, individuals were scored for the presence or absence of each allele and each allele was treated as a separate locus. The bootstrap option of Powermarker was used to create 1000 dendrograms using both the neighbor joining (NJ) and unweighted pair group method with arithmetic mean (UPGMA). The MEGA3 software (Kumar et al., 2004) was used to generate and edit consensus dendrograms. Microsat (Minch et al., 1997) was used to calculate the number of unique alleles for each locus. Genetic diversity for each locus and subpopulation was calculated using Shannon's index ( $H$ ) (Shannon and Weaver, 1949):

$$H = -\sum p_i \log_2 p_i$$

where  $p_i$  is the frequency of the presence or absence of the  $i$ th allele. Shannon's index (Table 3) was calculated using the Perl scripting language (ActiveState Programmer Network, 2005). The number of unique genotypes per locus was also calculated using the Perl scripting language, where a genotype is defined as a unique SSR banding pattern. SAS (version 9.1; SAS Institute Inc., Cary, N.C.) was used to analyze differences in genetic diversity between subpopulations.

To determine the level of correlation between the genetic distances and pedigree information, coancestry coefficients for a subset of the data were obtained and Pearson correlation coefficients were calculated using the SAS Procedure CORR. The proportion of shared alleles distances were first converted into

Table 2. Twenty-eight SSR primer pairs used to analyze relationships among blueberries. Locus name prefixes reflect origin of GenBank source sequence (CA= cold acclimated EST library, NA= non-acclimated EST library, VCC= enriched genomic library). GenBank accession numbers and repeat motifs are listed. Primer pairs with the same multiplex group number were sized simultaneously. Allele scoring quality refers to the presence or absence of PCR artifacts and factors that complicate allele scoring (G = good, S = stutter, P = split peaks, M = multiple loci, and N = null genotypes; see discussion). For each primer pair, the allele size range is shown.

Primer pair	GenBank accession no.	Repeat motif	Multiplex group	Allele scoring quality	Allele size range (bp)
CA23F	CF810543	(AGA) <sub>6</sub>	9	G	155–164
CA94F	CF811011	(AG) <sub>7</sub>	6	G	368–420
CA112F	CF810443	(AG) <sub>7</sub>	10	S	142–184
CA169F	CF811071	(GAT) <sub>4</sub>	9	G	109–136
CA190R	CF811085	(TGC) <sub>5</sub>	8	G	233–251
CA236F	CF810540	(TG) <sub>17</sub>	3	S,N	214–241
CA344F	CF810639	(GCG) <sub>6</sub>	6	P	134–179
CA421F	CF810704	(CT) <sub>25</sub>	none	S, P	153–230
CA483F	CF810754	(TC) <sub>8</sub>	5	N	292–347
CA642F	CF810880	(CT) <sub>17</sub>	4	S, P, M	302–387
CA787F	CF810934	(GAA) <sub>7</sub>	1	G	278–376
CA794F	CF810941	(GA) <sub>12</sub>	7	S	141–226
CA855F	CF811000	(GA) <sub>14</sub> (CGA) <sub>5</sub>	1	G	225–258
NA41	CF811380	(CT) <sub>10</sub> (CT) <sub>7</sub>	8	S	186–242
NA398	CF811369	(AAAT) <sub>5</sub>	9	G	211–232
NA741	CF811540	(TC) <sub>9</sub>	10	S	247–300
NA800	CF811589	(TC) <sub>13</sub>	none	S, M	180–287
NA824	CF811613	(AG) <sub>17</sub>	1	S, M	158–190
NA961	CF811674	(TAC) <sub>5</sub>	none	G	176–201
NA1040	CF811165	(TC) <sub>11</sub>	3	S	173–263
VCC_H9	AY762677	(CT) <sub>13</sub>	6	S	152–279
VCC_I2	AY762678	(CT) <sub>14</sub>	5	S, P	205–245
VCC_I8	AY762679	(TG) <sub>8</sub>	none	G	97–149
VCC_J1	AY762680	(CA) <sub>6</sub>	2	S, M	213–315
VCC_J3	AY762681	(AAG) <sub>15</sub>	none	P	115–166
VCC_J5	AY762682	(TC) <sub>17</sub>	2	S	250–303
VCC_J9	AY762683	(TG) <sub>9</sub> (GA) <sub>23</sub>	7	S	215–250
VCC_K4	AY762684	(TC) <sub>16</sub> ...(TC) <sub>12</sub>	4	S	169–300

similarity measures by subtracting them from 1. Pairwise coefficients of coancestry for the 10 accessions described in Rowland et al. (2003) were provided by M. Ehlenfeldt (USDA-ARS, Chatsworth, N.J.). Coefficients of coancestry were calculated based on assumptions of tetrasomic inheritance (Ehlenfeldt, 1994). Genetic similarity coefficients based on EST-PCR markers from Rowland et al. (2003) were also included in the correlation analysis for comparison.

The microsatellite data were transformed into a binary data matrix, with "1" indicating the presence and "0" the absence of an allele. The distance matrix was constructed based on Jaccard's coefficient using the NTSYS-pc (version 2.1; Exeter Software, Setauket, N.Y.) software package (Rohlf, 2000). The SIMQUAL program was used to calculate the Jaccard's coefficient, a common estimator of genetic identity and was calculated as follows:

$$\text{Jaccard's coefficient} = N_{AB} / (N_{AB} + N_A + N_B)$$

where  $N_{AB}$  is the number of bands shared by samples,  $N_A$  represents amplified fragments in sample A, and  $N_B$  represents fragments

Table 3. Allelic diversity among highbush blueberries for 28 SSR primer pairs. The number of unique alleles ( $A_u$ ) and genotypes were reported in all accessions while the total number of alleles (A) and Shannon's index (H) were estimated separately for all accessions, wild populations, historical as well as modern northern highbush cultivars (NHB), and southern highbush cultivars (SHB).

Primer pair	All accessions				Wild		Historical NHB		Modern NHB		SHB	
	$A_u$	Genotypes (no.)	A	H	A	H	A	H	A	H	A	H
CA23F	0	9	4	1.66	4	2.02	3	1.32	2	0.65	4	1.92
CA94F	7	30	20	6.36	14	7.39	5	3.91	7	3.87	9	5.75
CA112F	5	26	13	4.65	8	5.41	5	3.09	7	3.60	7	3.53
CA169F	1	26	8	4.98	5	3.80	5	3.61	6	4.57	6	4.68
CA190R	1	8	4	2.53	3	2.68	3	1.58	3	2.50	3	2.39
CA236F	1	25	11	5.49	6	4.63	9	4.84	6	3.89	8	5.24
CA344F	2	46	16	8.08	12	8.73	11	6.41	9	5.38	12	7.59
CA421F	7	66	30	12.64	20	11.49	14	10.51	13	9.33	17	10.58
CA483F	6	39	19	6.88	9	5.48	8	6.05	11	5.46	10	5.17
CA642F	9	68	48	22.71	34	21.98	32	18.61	26	17.17	36	21.30
CA787F	1	15	6	3.61	4	3.27	5	4.11	5	2.68	6	2.34
CA794F	3	58	17	9.03	14	8.46	11	8.73	10	7.87	12	7.40
CA855F	5	62	25	11.07	17	10.00	16	10.08	14	9.06	12	7.30
NA41	5	46	21	8.61	16	9.21	11	7.33	9	5.73	12	7.60
NA398	1	20	10	4.36	6	3.52	5	2.28	7	3.62	6	4.38
NA741	11	65	35	13.12	22	11.73	15	11.17	18	10.62	14	9.59
NA800	20	68	78	35.69	49	31.38	44	29.92	49	31.18	54	31.90
NA824	7	51	19	8.88	15	8.66	12	8.30	10	7.01	9	8.01
NA961	2	15	10	4.81	4	2.81	9	5.78	8	5.42	4	2.53
NA1040	14	65	31	11.40	17	9.33	15	9.89	13	8.79	20	11.36
VCC_H9	10	44	26	10.19	20	11.19	14	8.87	12	7.63	9	6.59
VCC_I2	3	48	15	8.20	13	9.48	10	7.33	10	6.02	10	5.90
VCC_I8	2	19	9	4.02	7	4.58	3	2.32	6	3.86	7	4.39
VCC_J1	13	68	58	25.79	46	29.69	37	20.68	33	20.78	30	20.30
VCC_J3	3	34	13	5.65	10	7.09	6	3.94	7	4.15	8	5.00
VCC_J5	6	65	30	12.36	22	12.45	18	10.48	16	9.57	15	9.65
VCC_J9	4	55	22	9.23	19	10.24	10	7.50	11	7.00	13	7.77
VCC_K4	8	63	29	11.42	18	10.14	16	9.96	17	9.17	13	8.64
<b>Total</b>	157	1204	627	273.43	434	266.84	345	228.59	354	216.60	366	228.81
<b>Average</b>	5.6	43	22.4	9.77	15.5	9.53	12.3	8.16	12.6	7.74	13.1	8.17

in sample B. Similarity matrices based on these indices were calculated. The principal coordinate analysis was performed with the DCENTER and EIGEN modules and illustrated by the 3D PLOT module.

## Results

**PRIMER PAIR PRODUCT QUALITY.** Nine out of 28 blueberry SSR loci did not exhibit any PCR artifacts (loci with allele scoring quality = "G" in Table 2). One of two PCR artifacts was observed in the remaining SSRs: stutter caused by production of additional peaks differing from the true peak by a multiple of the repeat unit; and split peaks that result from unequal non-templated addition of adenosine by *Taq* polymerase. Stutter bands were more common (57% of loci) than split peaks (18% of loci). Stuttering was observed in 16 loci: CA112F, CA236F, CA421F, CA642F, CA794F, NA41, NA741, NA800, NA824, NA1040, VCC\_H9, VCC\_I2, VCC\_J1, VCC\_J5, VCC\_J9, and VCC\_K4.

Null alleles are caused by mutations leading to the loss of a primer binding site, and result in the loss of codominance. The presence of null alleles can be ascertained by segregation analysis. Null genotypes, as indicated by lack of amplification at a locus were observed in two di-nucleotide SSRs (7%), CA483F and CA236F.

Twenty-four EST-SSR primer pairs appeared to amplify a single

locus, based on the presence of two or fewer alleles in diploid individuals and four or fewer alleles in tetraploid individuals. Mapping of these loci is necessary to verify their presence in one location in the blueberry genome. For nine of these primer pairs, the location of the SSR relative to gene features (5' and 3'UTR, coding, intron) could be predicted by aligning them to similar sequences in GenBank. Measures of allelic diversity (number of alleles, number of unique alleles, unique number of genotypes, and Shannon's index) ranged from slightly less than average to

Table 4. Allelic diversity for single copy EST-SSRs in highbush blueberry with predicted locations. Mean alleles (no.), unique alleles (no.), and Shannon's index (H) for SSRs with predicted locations in the 5' UTR, intron, coding (CDS), or unknown location due to lack of BLAST hit (NA for not available).

SSR location	Loci (no.)	Mean alleles (no.)	Unique alleles	Genotypes (no.)	Shannon's index (H)
5'UTR	3	21	4.3	55.3*	9.6
Intron	1	30	7	66*	12.6
CDS	4	12.75	2.75	28.75*	5.8
3'UTR	1	31	14	65*	11.4
NA	8	13.13	3.38	26.125*	5.4

\*Significant differences between values in column, single factor ANOVA ( $P < 0.05$ ).

above average for SSRs with predicted gene locations (Table 4). The least polymorphic class of SSR loci was located in coding regions or present at undetermined locations. The differences in unique genotype number per locus were statistically significant ( $\alpha = 0.05$ , single factor analysis of variance) depending on the predicted location of the SSR. The number of unique genotypes per locus was highest in SSRs located in introns, 3'UTR and 5'UTR.

**ALLELIC DIVERSITY.** Sixty-nine blueberry germplasm accessions comprising wild populations (13), northern highbush blueberry (36), southern highbush blueberry (14), pentaploid blueberry (1), half-high blueberry (1), and cultivars that exhibit an intermediate northern/southern phenotype (4) were genotyped using 28 unmapped SSR markers. Four markers amplified more than one locus indicated by the presence of more than four alleles in tetraploids. The total number of alleles (A), number of unique alleles ( $A_u$ ), number of unique genotypes, and Shannon's index (H) for each locus were estimated (Table 3). The 28 SSR primer pairs amplified a total of 627 alleles ( $\bar{x} = 22.4$  alleles per primer pair). Of these alleles, 157 (25%) were unique. For 24 apparently single copy SSRs, the number of alleles and of unique alleles ranged from 4 to 35 ( $\bar{x} = 17.7$ ) and 0 to 14 ( $\bar{x} = 4.5$ ), respectively. Four SSR primer pairs appeared to amplify multiple loci, the number of alleles and of unique alleles ranged from 19 to 78 ( $\bar{x} = 50.8$ ) and 7 to 20 ( $\bar{x} = 12.3$ ), respectively. Shannon's index (H) ranged from 1.66 to 13.12, ( $\bar{x} = 7.51$ ) for single copy SSRs and from 8.88 to 35.69 ( $\bar{x} = 23.27$ ) for multiple copy SSRs.

The average number of alleles and the average Shannon's index per locus were compared among groups of blueberry accessions (wild, historical northern highbush, modern northern highbush, and southern highbush) (Table 3). For the purposes of this analysis, blueberry accessions with an intermediate chilling phenotype ('Draper', 'Legacy', 'Ozarkblue', and 'Sierra') and the pentaploid cultivar ('Pearl River') were included in the southern highbush group and the half-high blueberry cultivar ('Northsky') was included in the modern northern highbush group. The average number of alleles, the average Shannon's index per locus, and the number of unique alleles ( $A_u$ ) were greater in wild accessions (15.5, 9.53, and 7.41, respectively) than in domesticated accessions (12.6, 8.04, and 2.45, respectively). Analysis of variance was conducted to compare differences in H, A, and ( $A_u$ ) between groups, using SAS proc GLM. Differences between groups of blueberries were highly significant ( $P < 0.01$ ) for A, H, and  $A_u$ . The Tukey Studentized range test (Tukey's honestly significant difference) was used to test differences between group means. All differences between wild blueberries and domesticated blueberries were significant ( $P < 0.01$ ) for A, H, and  $A_u$ . Differences in A, H, and  $A_u$  among domesticated blueberries were not significant. Northern highbush blueberry cultivars which had four or more unique alleles included 'Bluejay' (6), 'Cabot' (7), and 'Olympia' (5) while southern highbush blueberry cultivars with four or more  $A_u$  consisted of 'Flordablue' (4), and 'Misty' (4). 'Northsky', the half-high blueberry, had a high  $A_u$  (6). The cultivar with the highest number of unique alleles was the pentaploid blueberry 'Pearl River' (11).

Several diversity parameters were compared among single copy SSRs of different classes (dinucleotide versus trinucleotide SSRs and genomic vs. EST-SSRs). The average number of alleles, number of unique alleles, number of genotypes, and Shannon's index (per locus) were significantly higher in dinucleotide SSRs

(22.06, 6.06, 48.5, and 9.04 respectively) than in trinucleotide SSRs (8.71, 1.43, 21.86, and 4.47 respectively) (*t* test assuming unequal variances,  $P < 0.002$ ). A slight increase in diversity parameters was obtained in genomic SSRs as compared to EST-SSRs but the difference was not statistically significant (*t* test assuming unequal variances,  $P < 0.21$ ). The average number of alleles (16.47), unique alleles (4.24), genotypes (36.5), and Shannon's index (7.02) per EST-SSR locus increased to 20.57, 5.24, 46.8, and 8.72, respectively, in genomic SSRs.

Suspected duplicate accessions of 'Ivanhoe' and 'Coville' had DNA fingerprints identical to 'Ivanhoe' and 'Coville', respectively. Elimination of duplicate accessions is one way that DNA fingerprinting can be used to manage germplasm collections. All other accessions were easily distinguished from one another. Full DNA fingerprints for all cultivars are available through the USDA-ARS Germplasm Resources Information Network (GRIN) (USDA-ARS, 2005).

Wild and domesticated blueberries formed distinct groups in the clustering analysis. Genetic distances (Dsa) among the 69 unique germplasm accessions ranged from 0.008 (PI 554792 - PI 554795) to 0.247 ('Pearl River' - 'Misty'). In the neighbor joining dendrogram (Fig. 1), distances between individual accessions were large, and distances between major clades were small. There were three major clades: a large mixed clade of wild blueberry accessions, northern highbush blueberry, and some southern highbush blueberry accessions; a clade of southern highbush blueberries; and a clade of northern highbush blueberries. Accessions hybridized before and after 1950 did not cluster separately.

The wild accessions formed a single cluster and grouped according to their geographical origins. Wild blueberry accessions from New York formed a single cluster, sister to wild blueberry accessions from Maine and North Carolina. Diploid wild blueberry accessions from North Carolina clustered with the wild blueberry accessions from Maine. The blueberry accessions from New York came from a single, inland collection site. In contrast, the blueberry accessions from Maine and North Carolina came from at least five separate collection sites but all were from coastal swamps or hills. The blueberry cultivars Northsky and Pearl River formed monotypic clusters sister to the wild blueberry accessions.

The cluster of wild blueberry accessions was sister to one of the two northern highbush blueberry clusters. This cluster of northern highbush blueberry had two major clades. One of these clades contained 'Blucrop' and its full-sibs or progeny ('Blueray', 'USDA 11-104', 'Georgiagem', 'Meader', and 'Nui'), a northern adapted southern highbush blueberry accession ('Ozarkblue'), the wild blueberry selection 'Grover' and its progeny ('Jersey'), and the northern highbush blueberry cultivar Olympia. The other major clade contained 'Rubel' and its descendants ('June', 'Rancocas', 'Wareham', '1613A'), 'Darrow', and 'Legacy'), 'Pioneer' and its descendants ('USDA F-72', 'Atlantic', and 'Chandler'), and 'Elliott' and its progeny ('Aurora' and 'Liberty').

The other major clade of northern highbush blueberry cultivars in the neighbor joining dendrogram was composed of three major subclusters. The first subcluster contained 'Stanley' and its direct descendants ('Coville' and 'Herbert') as well as indirect descendants through 'Earliblue' ('Toro', 'Spartan', and 'Bluetta'). The second subcluster contained 'Cabot' and its descendants ('Weymouth', 'Earliblue', and 'Patriot') as well as 'Sierra' (which has an interspecific background similar to southern

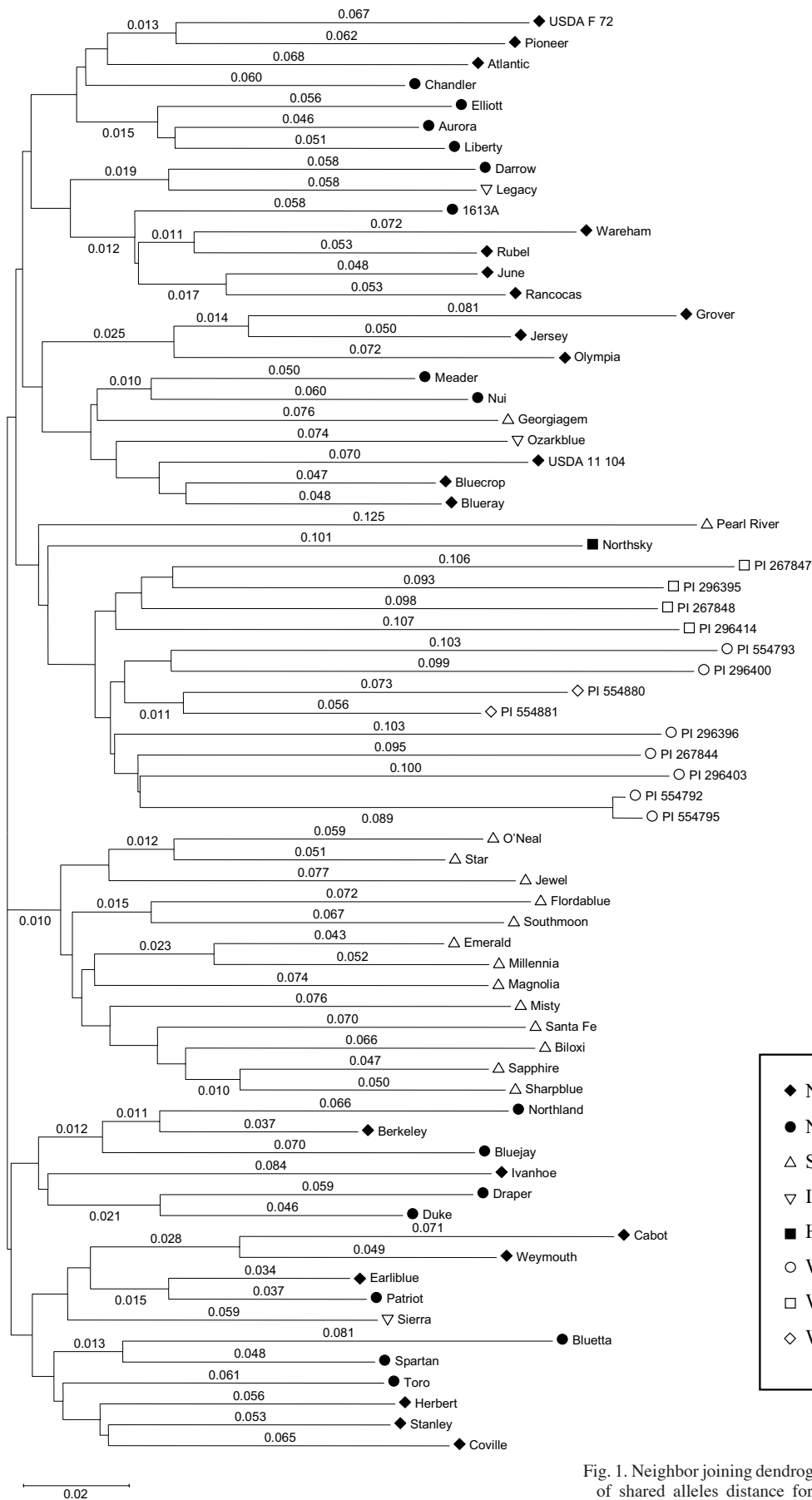


Fig. 1. Neighbor joining dendrogram of *Vaccinium* cultivars based on proportion of shared alleles distance for 627 SSR polymorphisms. Numbers refer to branch lengths.

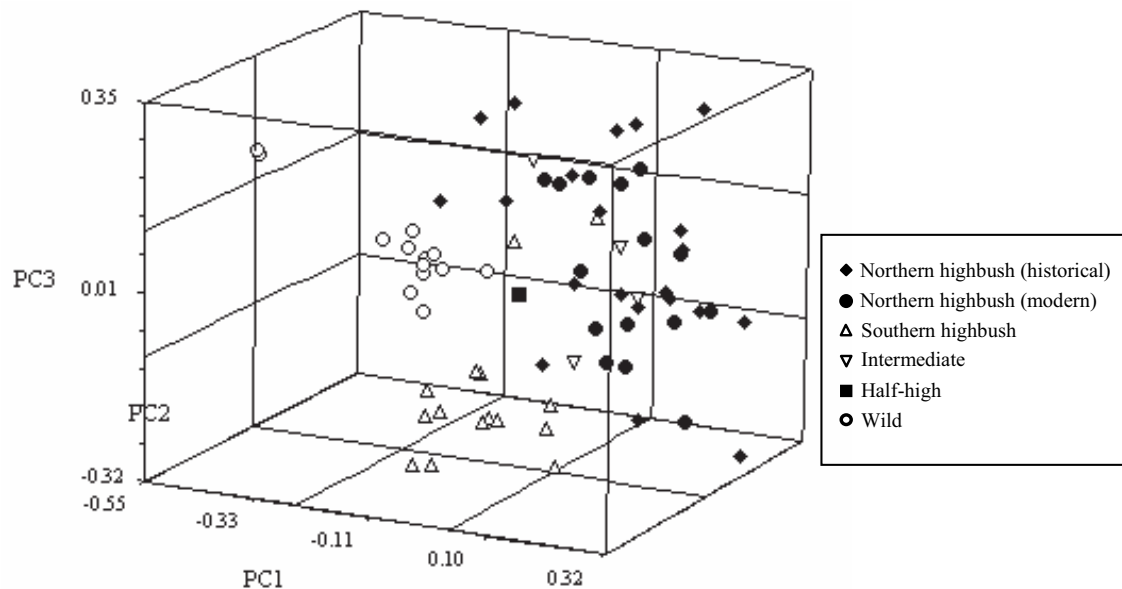


Fig. 2. Principal components analysis of highbush blueberry cultivars based on Jaccard's coefficient for 627 SSR polymorphisms.

highbush blueberry cultivars, but is northern adapted). The third subcluster contained 'Duke', 'Draper', 'Ivanhoe', 'Bluejay', and 'Berkeley'. Each of the accessions in the third subcluster are related to 'Stanley' through either 'Ivanhoe' or 'Berkeley'.

In the southern highbush blueberry cultivars cluster, subclusters were poorly resolved. Still, some pedigree associations were evident. 'O' Neal' clustered closely with its progeny 'Star' but not with 'Millennia'. The three cultivars Magnolia, Misty, and Biloxi share 'Avonblue' in their pedigree and grouped together.

In the principal component analysis (PCA) (Fig. 2), the first axis separated the wild blueberry accessions from the domesticated blueberry accessions. Horticulturally desirable northern highbush blueberry accessions such as 'Bluecrop', 'Stanley', and 'Earliblue' fell at the extreme opposite of wild accessions from Maine along the first principal component axis. In the middle of the first component axis were southern highbush blueberry accessions and historical northern highbush blueberry accessions (i.e., 'Grover', 'Cabot', 'Pioneer', and USDA F-72). The second axis did not correspond to any obvious groupings among blueberry genotypes. The third axis separated accessions with a low-chilling requirement from accessions with a high-chilling requirement with the exception of 'Georgiagem' which grouped with northern accessions that require more chilling. Along the third axis, wild accessions grouped more closely to northern highbush blueberry accessions than to southern highbush blueberry accessions. Some of the early-maturing blueberry accessions ('Earliblue', 'Weymouth', and 'Duke') had principal components scores closer to the low-chill blueberry accessions along the third axis.

Many of the same clades seen in the dendrogram were present as groups in the PCA. Northern highbush blueberry accessions from the upper clade in the neighbor joining dendrogram had more positive scores along the first and third axes than accessions from the lower clade of northern highbush blueberry accessions. In the PCA, southern highbush blueberry cultivars did not break into distinct groups. Among southern highbush blueberry cultivars there was not always correspondence between the associations

seen in the PCA and the dendrogram. 'Jewel' and 'Sapphire', for example, were very closely associated in the PCA but they appeared on different subclusters in the dendrogram.

**COMPARISON TO PEDIGREE INFORMATION.** A correlation test was performed on the genetic similarity values from the 28 SSR loci, tetrasomic coancestry coefficients (Rowland et al., 2003), and genetic similarity measures based on 17 EST-PCR primer pairs (Rowland et al., 2003) for a subset of 10 cultivars ('Berkeley', 'Bluecrop', 'Bluejay', 'Blueray', 'Duke', 'Georgiagem', 'Patriot', 'Sierra', 'Toro', and 'Weymouth'). Moderate correlation was found between the SSR data and the coancestry coefficients ( $r = 0.352$ ,  $P = 0.017$ ). The level of correlation between the SSR data and the coancestry values exceeded the correlation between EST-PCR marker data and the coancestry values ( $r = 0.194$ ,  $P = 0.2009$ ). The correlation between the EST-PCR markers and coancestry values was greater ( $r = 0.317$ ,  $P \leq 0.0001$ ) in the original dataset (Rowland et al., 2003) that included nine additional genotypes not used in this study.

## Discussion

Nine out of 28 blueberry SSR loci did not exhibit any PCR artifacts and are recommended for exchanging DNA fingerprints between labs. These markers were able to distinguish each unique accession tested and gave identical fingerprints for suspected duplicate clones. Stutter was observed only in loci with dinucleotide repeats as previously reported (Holton, 2001) and is thought to result from replication slippage in vitro (Ellegren, 2004).

More than four alleles were amplified by four out of 28 (14%) primer pairs. The amplification of SSR alleles from multiple loci in the genome has been reported in other plant species (Fisher et al., 1998). Due to the production of allelic "ladders" by closely spaced alleles, stutter bands, and non-template adenylation, true alleles and PCR artifacts were difficult to distinguish at these loci. However, the fingerprints produced by these loci were highly distinctive and had the same reproducibility as fingerprints from other loci. In some cases, the majority of the individuals



examined were distinguished using only one locus (e.g., NA800 which amplified a total of 78 alleles). Thus, these loci appear to have advantages if the only objective of the analysis is to verify the identity of the individual. Still, the high mutation rate that is implied at these loci could lead to incorrect genotyping of two otherwise identical clones (Crespan, 2004). Additional loci and multiple plants from the same cultivar should be tested to verify a mismatch.

Microsatellite analysis reflected a high level of heterozygosity in 69 blueberry accessions as indicated by an average of 17.7 alleles per single locus and a Shannon's index (H) of 9.77. Low numbers of SSR alleles per locus have been reported for self-pollinated crops like rice (*Oryza sativa* L.) (5.13 alleles/locus for genomic SSR, 2.78 alleles/locus for EST-SSR; Cho et al., 2000) in comparison to cross-pollinated crops like maize (*Zea mays* L.) (12.6 alleles per locus, Matsuoka et al., 2002a; 8.02 alleles per locus, Reif et al., 2004) and sunflower (*Helianthus annuus* L.) (12.0 alleles/locus, Tang and Knapp, 2003). Nonetheless, the number of alleles per locus in cross-pollinated crops ranges from a low of 2.8 in diploid self-incompatible sweet cherry (*Prunus avium* L.) (Dirlewanger et al., 2002) to 26.4 in a large sample of diverse apple germplasm (Hokanson et al., 2001). The 17.7 alleles per locus found in blueberry is similar to the 23.7 alleles per locus found in kiwifruit, another genus that is cross-pollinated, has polyploid species, and is clonally propagated (Zhen et al., 2004). A high number of alleles per locus was also observed in other plant species that are cross-pollinated, perennial, and vegetatively propagated, such as hops (*Humulus lupulus* L.) (10.6; Stajner et al., 2005), grape [(8; Dangl et al., 2001); (11; Martín et al., 2003); (24.4; Lamboy and Alpha, 1998)], apple (26.4; Hokanson et al., 2001) and tetraploid sour cherry (*Prunus cerasus* L.) (10.7; Cantini et al., 2001). In other studies, a large number of alleles per locus was not always associated with an increase in ploidy level. In allotetraploid cotton (*Gossypium hirsutum* L.) that exhibits disomic inheritance, five alleles per locus were found (Liu et al., 2000) while tetraploid fescue (*Festuca arundinacea* Schreb.) had 2.71 alleles per locus (Saha et al., 2004).

The average Shannon's index (H) per locus in blueberry germplasm was high at 9.77 as compared to 3.21 in globe artichoke (*Cynara cardunculus* L. var. *scolymus* L.) based on 519 AFLP polymorphisms (Lanteri et al., 2004). Shannon's index is a measure of diversity that takes into account both the number of classes (allelic richness) and the distribution of individuals among classes (allelic evenness). Unlike other commonly used measures of diversity such as expected heterozygosity ( $H_e$ ), Shannon's index does not require knowledge of allele frequencies. Therefore, Shannon's index is an accurate measure of diversity in polyploid species where allele frequencies cannot be determined with certainty due to the difficulty in distinguishing the copy number of individual alleles.

Wild blueberry accessions had significantly larger numbers of alleles and a higher average Shannon's index per locus than modern northern highbush accessions. These two diversity parameters did not differ significantly between the southern and northern highbush blueberry groups. Wild blueberry accessions also possessed significantly greater numbers of unique alleles than cultivated types. Surprisingly, the numbers of unique alleles present in southern and northern highbush blueberry accessions were not significantly different.

Many of the unique alleles in the northern highbush blueberry group were found in older cultivars such as 'Cabot', and 'Olympia'. These unique alleles may represent regions of the

genome that contain unfavorable alleles and are of little use to breeders. However, some recently released highbush cultivars such as 'Bluejay', 'Flordablue', and 'Misty' also had unique alleles. 'Bluejay' probably gets its unique alleles from 'Taylor' and 'Flordablue' from *V. darrowii* and/or *V. virgatum*. 'Misty' may receive its unique alleles from *V. darrowii*.

Differences in ploidy level between accessions used in this analysis could potentially interfere with the use of genetic distances and measures of genetic diversity. However, separate cluster analysis of diploid and tetraploid blueberry taxa using the proportion of shared alleles distance gave very similar results to a combined analysis of diploid and polyploid taxa (Boches, 2005). Thus, the proportion of shared alleles distance appears to be robust to differences in ploidy level.

Patterns of clustering based on the neighbor joining dendrogram (Fig. 1) depicted the association of blueberry cultivars with one of their parents or ancestors and the genetic distinctness of southern highbush blueberry cultivars (middle clade). Wild accessions were a sister group to one of two clades of northern highbush selections: that containing highbush selections from the wild like 'Rubel' and 'Grover'. The southern highbush blueberry cluster that grouped separately from northern highbush blueberry clades probably did so due to their combined *V. darrowii* and *V. virgatum* ancestry, despite the phenotypic similarity of these two groups. Southern highbush blueberry cultivars have been backcrossed to northern highbush blueberry cultivars and are selected for predominantly northern highbush blueberry phenotype. The PCA revealed the same basic groups as the dendrogram. In addition to the distinct group of wild blueberry accessions and that of the southern highbush blueberry accessions, an overlapping set of northern highbush blueberry groups corresponded to the two northern highbush blueberry clades seen in the neighbor joining diagram.

In the neighbor joining dendrogram, wild blueberry accessions were separated into groups based on their geographical origins. The placement of the wild blueberry accessions from North Carolina with the wild blueberry accessions from Maine probably reflects their common proximity (150 km) to the coast. Camp (1945) noted that one of the major factors associated with morphological variation in the highbush blueberry was the coastal vs. inland origin of the race. In the PCA, two wild blueberry accessions from Maine formed an extreme group, and wild blueberry accessions from North Carolina grouped near each other.

Among northern highbush blueberry cultivars, blueberry cultivars hybridized before and after 1950 were intermixed in various clades. In the dendrograms, one major cluster of northern highbush blueberry contained 'Stanley' ('Katharine' x 'Rubel') and the other contained one of its parents, 'Rubel'. In the PCA, 'Rubel' and 'Stanley' were near the intersection of the two major subsets of northern highbush blueberry cultivars. Clustering of blueberry cultivars near 'Stanley' could represent the influence of 'Katharine' ('Brooks' x 'Sooy') and future studies should include 'Katharine'.

Frequently, northern highbush cultivars clustered near one of their parents or siblings. 'Bluecrop' for example grouped with its progeny 'Nui', 'Meader', and 'Georgiagem' and its full-sibs USDA 11-04 and 'Blueray' in both the dendrograms and the PCA. Often, the parents of a given cultivar were genetically distant from each other, and the cultivar would cluster closely to only one of its parents in the dendrogram but not in the PCA. For example, 'Weymouth' ('June' x 'Cabot') grouped with its parent 'Cabot' and not 'June' ['Rubel' x ('Brooks' x 'Russell')] in the neighbor

joining dendrogram. In PCA however, 'Weymouth' did not group closely to either parent and was equidistant from both.

The northern and southern highbush blueberry clusters overlapped. 'Georgiagem' is the only southern highbush blueberry cultivar that grouped among the northern highbush blueberries. 'Bluecrop' was used twice as this cultivar's grandparent which could explain its clustering close to 'Bluecrop' in the dendrogram and between 'Bluecrop' and the southern highbush blueberry types in the PCA. 'Draper', 'Legacy', 'Ozarkblue', and 'Sierra' are intermediate in their chilling requirement and winter hardiness, in agreement with their positions in the dendrogram and PCA.

The lower resolution in clustering of southern highbush cultivars could have resulted from incomplete pedigrees due to the use of bulked pollen and open pollination in their breeding.

The pentaploid cultivar Pearl River was genetically very distinct and had the greatest number of unique alleles (12). This cultivar is unique in containing a high proportion of *V. virgatum* (probably 60%) from its parent 'Beckyblue'. Other accessions used in this study with contributions from *V. virgatum* include the cultivars Draper, Flordablue, O'Neal (and its seedlings Millenia and Star), Ozarkblue, Magnolia, Misty, Santa Fe, Sharpblue, Sierra, and Southmoon. These cultivars clustered near other southern highbush blueberry with the exception of 'Sierra', which clustered near 'Earliblue'.

Another genetically distinct blueberry cultivar was 'Northsky', the only blueberry with half-high stature included in this analysis (although many cultivars contain up to 25% *V. angustifolium* in their parentage, such as 'Bluetta', 'Northland', and 'Patriot'). 'Northsky' is estimated to be 25% *V. angustifolium* and 75% *V. corymbosum*. Among its *V. corymbosum* parents, 25% of contributions are from the hardy wild clone 'Ashworth', 25% from miscellaneous *V. corymbosum*, and 25% from founding clones such as 'Brooks', and 'Rubel' (Ehlenfeldt, 1994). In the neighbor joining dendrogram, 'Northsky' clustered with wild blueberry highbush accessions. 'Northsky' also clustered near the wild blueberry accessions in the PCA, probably due to the presence of miscellaneous wild *V. corymbosum* ancestors and 'Ashworth' in its ancestry. In comparison to 'Northsky', 'Northland' is 25% *V. angustifolium* but groups into the 'Stanley' cluster because the remainder of its parentage comes mainly from founding clones (18.8% each from 'Brooks', 'Sooy', and 'Rubel').

Rowland et al. (2003) studied 19 blueberry genotypes using 30 EST-PCR markers. In agreement with clustering based on coefficients of coancestry, 'Nelson', 'Bluecrop', and 'Blueray' clustered together according to DNA markers. However, 'Cooper', 'Georgiagem', and 'Nelson' would be predicted to cluster based on coancestry but did not based on markers. 'Georgiagem' clustered with its twice grandparent 'Bluecrop' in Rowland et al. (2003) and the present study as well. Clustering by DNA markers was significantly different from that based on coefficients of coancestry for the other genotypes studied by Rowland et al. (2003). Tetraploid coefficients of coancestry were not calculated for accessions used in this study. However, for 10 accessions common to both studies, the correlation between genetic distances and coancestry coefficients in our study exceeded that in Rowland et al. (2003) due to the larger number of polymorphisms used (627 vs. 37). The accuracy of genetic similarity estimates depends on several variables including the number of markers analyzed, their distribution in the genome and correct scoring (Schut et al., 1997). The greater correlation between the EST-PCR markers and coancestry values reported in the full dataset (Rowland

et al., 2003) is due to the larger number of genotypes used (19 vs. 10). Because the resolution provided by SSR analysis in this study is at least as good as that obtained from EST-PCR analysis (Rowland et al., 2003), it appears that the association between 'Bluecrop' and 'Georgiagem' may not be an anomaly.

This study demonstrates that microsatellites can resolve and assess relationships among tetraploid blueberries. A decrease in genetic diversity among cultivated blueberries as opposed to wild blueberries was observed, as measured by the total number of alleles, the number of unique alleles, and Shannon's index (Table 3). However, substantial genetic diversity is present in the cultivated blueberry gene pool as indicated by the lack of significant differences between the northern and southern highbush blueberry accessions for A, H, and  $A_u$  (Table 3, Fig. 3), as well as the genetic distances among accessions in the northern highbush blueberry cluster compared to those in the wild blueberry populations (Figs. 1, 2). Many of the genetic distances between accessions of northern highbush blueberries were similar to those observed among wild accessions. Indeed, the smallest genetic distance observed among all blueberry accessions in the neighbor joining dendrogram was between two wild accessions from Maine (PI 554792 and PI 554795). In the PCA, the wild accessions formed a much denser cluster than did the northern highbush accessions, which spanned a considerable distance across each axis. The founding clones we sampled were genetically very distinct from each other and from other wild blueberries. Substantial genetic diversity has been preserved by breeder selection and the reproductive habits of *Vaccinium* despite the narrow germplasm base of the domesticated blueberry. Blueberry has reproductive characteristics that are associated with high heterozygosity. Blueberry flowers have pollen shedding stigmas shaped like an inverted funnel that encourage cross-pollination by bees (Ehlenfeldt, 2001; Harrison et al., 1993; Shutak and Marucci, 1966). Blueberry cultivars also show varying degrees of self-fertility (Ehlenfeldt, 2001; Harrison et al., 1993; Krebs and Hancock, 1990) even though they lack a true system of self-incompatibility. Blueberry has polysomic inheritance (reviewed in Ehlenfeldt, 1994), which dramatically increases heterozygosity (reviewed in Soltis and Soltis, 2000). Clonal propagation, common in some wild blueberry species (Camp, 1945), also favors the evolution of a small number of highly heterozygous genotypes (Balloux et al., 2003). These characteristics of blueberry, along with breeder selection for high vigor, are likely reasons for the high levels of allelic diversity observed in blueberry cultivars relative to many cultivated crops.

#### Literature Cited

- ActiveState Programmer Network. 2005. ActivePerl. 4 Apr. 2005. <<http://aspn.activestate.com/ASPN/Downloads/ActivePerl/>>.
- Aradhya, M.K., G.S. Dangi, B.H. Prins, J.-M. Boursiquot, A.M. Walker, C.P. Meredith, and C.J. Simon. 2003. Genetic structure and differentiation in cultivated grape, *Vitis vinifera* L. Genet. Res. (Cambridge) 81:179-192.
- Aranzana, M.J., J. Carbó, and P. Arús. 2002. Microsatellite variability in peach [*Prunus persica* (L.) Batsch]: Cultivar identification, marker mutation, pedigree inferences and population structure. Theor. Appl. Genet. 106:1341-1352.
- 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 am-

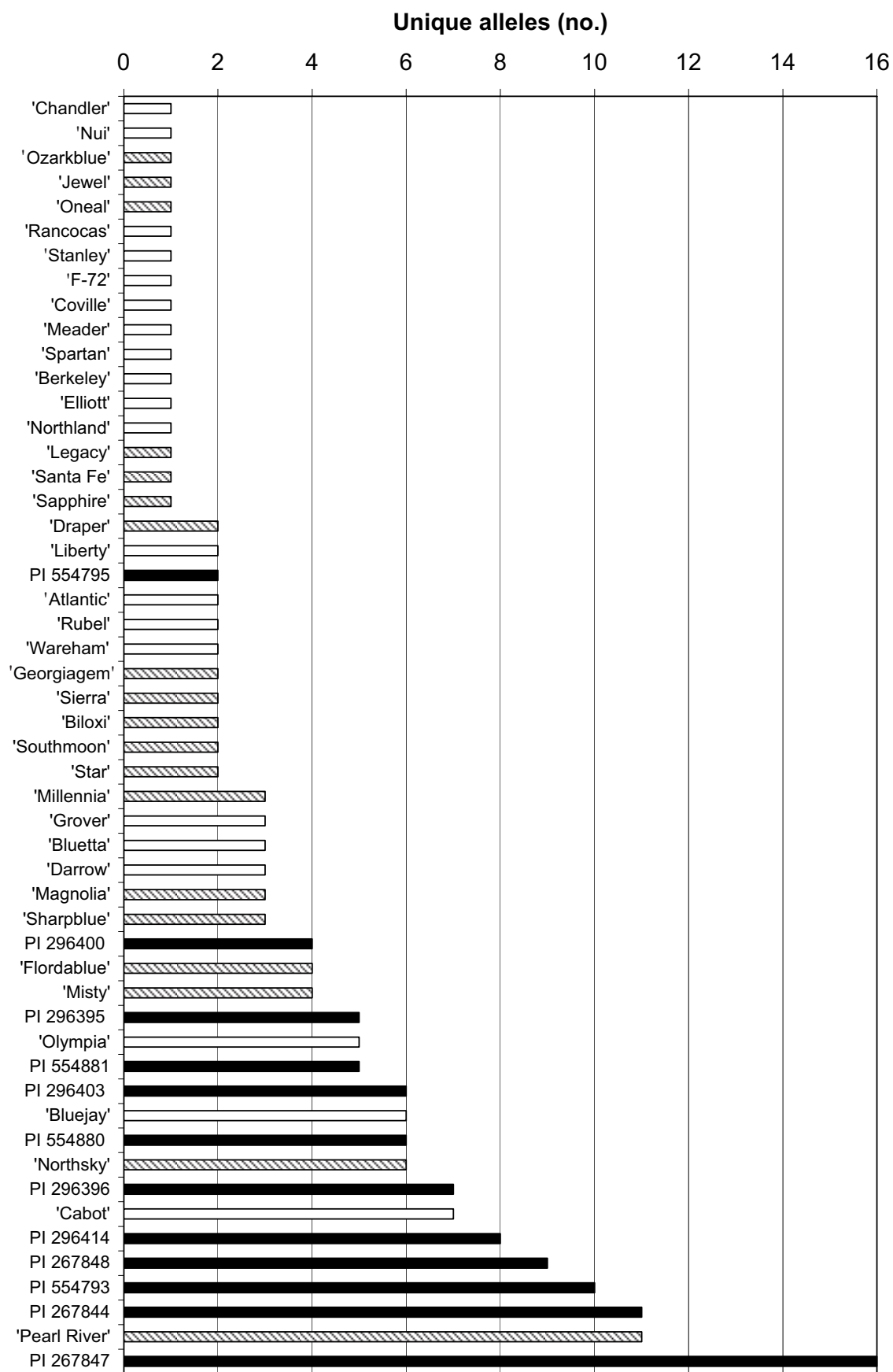


Fig. 3. The number of unique alleles per accession for 28 SSR primer pairs evaluated in northern highbush blueberry (*open bars*), southern highbush, pentaploid or half-high (*diagonally hatched bars*), and wild populations (*black filled bars*).

- plified polymorphic DNA fingerprinting for identifying rabbiteye blueberry (*Vaccinium ashei* Reade) cultivars. *J. Amer. Soc. Hort. Sci.* 120:710–713.
- Balloux, F., L. Lehmann, and T. de Meeûs. 2003. The population genetics of clonal and partially clonal diploids. *Genetics* 164:1635–1644.
- Boches, P. 2005. Microsatellite marker development and molecular characterization in highbush blueberry (*Vaccinium corymbosum* L.) and *Vaccinium* species. M.S. Diss., Oregon State Univ., Corvallis.
- Boches, P., L.J. Rowland, and N.V. Bassil. 2005. Microsatellite markers for *Vaccinium* from EST and genomic libraries. *Mol. Ecol. Notes* 5:657–660.
- 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.
- Brownstein, M.J., J.D. Carpten, and J.R. Smith. 1996. Modulation of non-template nucleotide addition by Taq DNA polymerase: Primer modifications that facilitate genotyping. *Biotechniques* 20:1004–1010.
- Camp, W.H. 1945. The North American blueberries with notes on other groups of Vacciniaceae. *Brittonia* 5:203–275.
- Cantini, C., A.F. Iezzoni, W.F. Lamboy, M. Boritzki, and D. Struss. 2001. DNA fingerprinting of tetraploid cherry germplasm using simple sequence repeats. *J. Amer. Soc. Hort. Sci.* 126:205–209.
- Chenna, R., H. Sugawara, T. Koike, R. Lopez, T.J. Gibson, D.G. Higgins, and J.D. Thompson. 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* 31:3497–3500.
- Cho, Y.G., T. Ishii, S. Temnykh, X. Chen, L. Lipovich, S.R. McCouch, W.D. Park, N. Ayres, and S. Cartinhour. 2000. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 100:713–722.
- Coville, F.V. 1937. Improving the wild blueberry, p. 559–574. In: G. Hambidge (ed.). *USDA yearbook of agriculture*. U.S. Govt. Printing Office, Washington, D.C.
- Crespan, M. 2004. Evidence on the evolution of polymorphism of microsatellite markers in cultivars of *Vitis vinifera* L.. *Theor. Appl. Genet.* 108:231–237.
- Dangl, G.S., M.L. Mendum, B.H. Prins, M.A. Walker, C.P. Meredith, and C.J. Simon. 2001. Simple sequence repeat analysis of a clonally propagated species: A tool for managing a grape germplasm collection. *Genome* 44:432–438.
- Dirlwanger, E., P. Cosson, M. Tavaud, M.J. Aranzana, C. Poizat, A. Zanetto, P. Arús, and F. Legret. 2002. Development of microsatellite markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). *Theor. Appl. Genet.* 105:127–138.
- Ehlenfeldt, M.K. 1994. The genetic composition and tetrasomic inbreeding coefficients of highbush blueberry cultivars. *HortScience* 29:1342–1345.
- Ehlenfeldt, M.K. 2001. Self- and cross-fertility in recently released highbush blueberry cultivars. *HortScience* 36:133–135.
- Ellegren, H. 2004. Microsatellites: Simple sequences with complex evolution. *Nature Rev. Genet.* 5:435–445.
- Fisher, P.J., T.E. Richardson, and R.C. Gardner. 1998. Characteristics of single- and multiple-copy microsatellites from *Pinus radiata*. *Theor. Appl. Genet.* 96:969–979.
- 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: Small fruit crops*. Prentice Hall, New York.
- Haghighi, K. and J.F. Hancock. 1992. DNA restriction fragment length variability in the genomes of highbush blueberry. *HortScience* 27:44–47.
- Hancock, J.F. and J.H. Siefker. 1982. Levels of inbreeding in highbush blueberry cultivars. *HortScience* 17:363–366.
- Harrison, R.E., J.J. Luby, and P.D. Ascher. 1993. Genetic characteristics of self-fertility in highbush and half-high blueberries. *Euphytica* 67:79–88.
- Hokanson, S.C., W.F. Lamboy, A.K. Szewc-McFadden, and J.R. McFerson. 2001. Microsatellite (SSR) variation in a collection of *Malus* (apple) species and hybrids. *Euphytica* 118:281–294.
- Hokanson, S.C., A.K. Szewc-McFaddon, W.F. Lamboy, and J.R. McFerson. 1998. Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus x domestica* Borkh. core subset collection. *Theor. Appl. Genet.* 97:671–683.
- Holton, T.A. 2001. Plant genotyping by analysis of microsatellites, p. 15–26. In: R.J. Henry (ed.). *Plant genotyping: The DNA fingerprinting of plants*. CAB Intl. Press, Oxon, U.K.
- Ivandic, V., C.A. Hackett, E. Nevo, R. Keith, W.T. Thomas, and B.P. Forster. 2004. Analysis of simple sequence repeats (SSRs) in wild barley from the Fertile Crescent: Association with ecology, geography, and flowering time. *Plant Mol. Biol.* 48:511–527.
- Krebs, S.L. and J.F. Hancock. 1990. Early acting inbreeding depression and reproductive success in the highbush blueberry, *Vaccinium corymbosum* L. *Theor. Appl. Genet.* 79:825–832.
- Kumar, S., K. Tamura, and M. Nei. 2004. MEGA 3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings Bioinformatics* 5:150–163.
- Lanteri, S., A. Acquadro, E. Saba, and E. Portis. 2004. Molecular fingerprinting and evaluation of genetic distances among selected clones of artichokes (*Cynara cardunculus* L. var. *scolymus* L.). *J. Hort. Sci. Biotechnol.* 79:863–870.
- Lamboy, W.F. and C.G. Alpha. 1998. Using simple sequence repeats (SSRs) for DNA fingerprinting germplasm accessions of grape (*Vitis* L.) species. *J. Amer. Soc. Hort. Sci.* 123:182–188.
- 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.
- Li, Y., A.B. Korol, T. Fahima, A. Beiles, and E. Nevo. 2002. Microsatellites: Genomic distribution, putative functions, and mutational mechanisms: A review. *Mol. Ecol.* 11:2453–2465.
- Liu, K. and S. Muse. 2004. PowerMarker: New genetic data analysis software. Version 3.0. 10 Oct. 2005. <<http://www.powermarker.net>>.
- Liu, S., R.G. Cantrell, J.C. McCarty Jr., and J. McD Stewart. 2000. Simple sequence repeat-based assessment of genetic diversity in cotton race stock accessions. *Crop Sci.* 40:1459–1469.
- Martín, J.P., J. Borrego, F. Cabello, and J.M. Ortiz. 2003. Characterization of Spanish grapevine cultivar diversity using sequence-tagged microsatellite site markers. *Genome* 46:10–18.
- Matsuoka, Y., S.E. Mitchell, S. Kresovich, M. Goodman, and J. Doebly. 2002a. Microsatellites in *Zea* variability, patterns of mutations, and use for evolutionary studies. *Theor. Appl. Genet.* 104:436–450.
- Matsuoka, Y., Y. Vigouroux, M.M. Goodman, G.J. Sanchez, E. Buckler, and J. Doebly. 2002b. A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl. Acad. Sci. USA* 99:6080–6084.
- Minch, E., A. Ruiz-Linares, D. Goldstein, M. Feldman, and L.L. Cavalli-Sforza. 1997. MICROSAT: A computer program for calculating various statistics on microsatellite allele data. 10 Oct. 2005. <<http://hpgl.stanford.edu/projects/microsat/>>.
- Novy, R.G. and N. Vorsa. 1996a. Identifying genotypic heterogeneity in 'McFarlin' cranberry: A randomly-amplified polymorphic DNA (RAPD) and phenotypic analysis. *J. Amer. Soc. Hort. Sci.* 121:210–215.
- Novy, R.G. and N. Vorsa. 1996b. Evidence for RAPD heteroduplex formation in cranberry: Implications for pedigree and genetic-relatedness studies and a source of co-dominant RAPD markers. *Theor. Appl. Genet.* 92:840–849.
- Novy, R.G. and N. Vorsa. 1995. Identification of intracultivar genetic heterogeneity in cranberry using silver-stained RAPDs. *HortScience* 30:600–604.
- Novy, R.G., C. Kokak, J. Goffreda, and N. Vorsa. 1994. RAPDs identify varietal misclassification and regional divergence in cranberry [*Vaccinium macrocarpon* (Ait.) Pursh.]. *Theor. Appl. Genet.* 88:1004–1010.
- Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, and S. Tingey. 1996. The comparison of RFLP, AFLP, and SSR (microsatellite) markers for germplasm analysis. *Mol. Breeding* 2:225–238.

- Qu, L., J.F. Hancock, and J.H. Whallon. 1998. Evolution in an autopolyploid group displaying predominantly bivalent pairing at meiosis: Genomic similarity of diploid *Vaccinium darrowii* and autotetraploid *V. corymbosum* (Ericaceae). *Amer. J. Bot.* 85(5):698–703.
- Raker, C.M. and D.M. Spooner. 2002. Chilean tetraploid cultivated potato, *Solanum tuberosum*, is distinct from the Andean populations: Microsatellite data. *Crop Sci.* 42:1451–1458.
- Reif, J.C., X.C. Xia, A.E. Melchinger, M.L. Warburton, D.A. Hoisington, D. Beck, M. Bohn, and M. Frisch. 2004. Genetic diversity within and among CIMMYT maize populations of tropical, subtropical, and temperate germplasm by SSR markers. *Crop Sci.* 44:326–334.
- Ritzinger, R. and P.M. Lyrene. 1998. Pollen fertility of hybrids between rabbiteye blueberry and *Vaccinium constablei*. *Fruit Var. J.* 52:96–100.
- Rohlf, F.J. 2000. NTSYS-pc numerical taxonomy and multivariate analysis system, version 2.1. Exeter Publ., Setauket, N.Y.
- Romero, C., A. Pedryc, V. Munoz, G. Llacer, and M.L. Badenes. 2003. Genetic diversity of different apricot geographical groups determined by SSR markers. *Genome* 46:244–252.
- Rowland, L.J., S. Mehra, A. Dhanaraj, E.L. Ogden, and J.P. Slovin. 2003. Development of EST-PCR markers for DNA fingerprinting and genetic relationship studies in blueberry (*Vaccinium*, section *Cyanococcus*). *J. Amer. Soc. Hort. Sci.* 128:682–690.
- Saha, M.C., M.A. Mian, I. Eujayl, J.C. Zwonitzer, L. Wang, and G.D. May. 2004. Tall fescue EST-SSR markers with transferability across several grass species. *Theor. Appl. Genet.* 109:783–791.
- 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.
- Shannon, C.E. and W. Weaver. 1949. The mathematical theory of communication. Univ. of Illinois Press, Urbana.
- Shutak, V.G. and P.E. Marucci. 1966. Plant and fruit development, p. 179–198. In: P. Eck (ed.). *Blueberry culture*. Rutgers Univ. Press, New Brunswick, N.J.
- Soltis, P.S. and D.E. Soltis. 2000. The role of genetic and genomic attributes in the success of polyploids. *Proc. Natl. Acad. Sci. USA* 97:7051–7057.
- Stajner, N., J. Jakse, P. Kozsjak, and B. Javornik. 2005. The isolation and characterisation of microsatellites in hops (*Humulus lupulus* L.). *Plant Sci.* 168:213–221.
- Tang, S. and S. Knapp. 2003. Microsatellites uncover extraordinary diversity in native American land races and wild populations of cultivated sunflower. *Theor. Appl. Genet.* 106:990–1003.
- Tautz, D. and M. Rentz. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res.* 12:4127–4137.
- Testolin, R., T. Marrazzo, G. Cipriani, R. Quarta, I. Verde, M.T. Dettori, M. Pancaldi, and S. Sansavini. 2000. Microsatellite DNA in peach (*Prunus persica* L. Batsch) and its use in fingerprinting and testing the genetic origin of cultivars. *Genome* 43:512–520.
- U.S. Department of Agriculture, Agricultural Research Service. 2005. Genetic Resources Information Network. 10 Oct. 2005. <<http://www.ars-grin.gov/cor/vaccin/blueberry-loci-05.xls>>.
- Vander Kloet, S.P. 1988. The genus *Vaccinium* in North America. Can. Govt. Publ. Ctr., Ottawa, Ont.
- Vorsa, N., G. Jelenkovic, A.D. Draper, and W.V. Welker. 1987. Fertility of 4x x 5x and 5x x 4x progenies derived from *Vaccinium ashei/corymbosum* pentaploid hybrids. *J. Amer. Soc. Hort. Sci.* 112:993–997.
- Weber, J.L. and P.E. May. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Amer. J. Human Genet.* 44:388–396.
- Wünsch, A. and J.J. Hormaza. 2002. Cultivar identification and genetic fingerprinting of temperate fruit tree species using DNA markers. *Euphytica* 125:59–67.
- Xu, Y., R.-C. Ma, H. Xie, J.-T. Liu, and M.Q. Cao. 2004. Development of SSR markers for the phylogenetic analysis of almond trees from China and the Mediterranean region. *Genome* 47:1091–1104.
- Zhen, Y., Z. Li, and H. Huang. 2004. Molecular characterization of kiwifruit (*Actinidia*) cultivars and selections using SSR markers. *J. Amer. Soc. Hort. Sci.* 129:374–382.