Diversity analysis of blueberry clones and cultivars using EST-PCR and microsatellite markers

by

© Sapan Pravinbhai Tailor

A thesis submitted to the

School of Graduate Studies

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN BIOLOGY

DEPARTMENT OF BIOLOGY

Memorial University of Newfoundland

Date: March 3rd, 2016

St. John's Newfoundland and Labrador

Abstract

The blueberry (*Vaccinium* spp. L.) is a health promoting and economically important fruit crop. Genetic diversity and relatedness were studied in 63 blueberry genotypes, using eleven EST–PCR and nine microsatellite markers. Markers found to be polymorphic and detected 436 and 272 alleles, respectively. Other parameters such as PIC, expected and observed heterozygosity were also high for both the markers.

Cluster analyses such UPGMA, NJ and PCO identified clustering of genotypes according to their geographic collection site. Fingerprinting data was used for the STRUCTURE analysis which also separated clones and cultivars based on their geographic collection site. Similarity was found among the results of UPGMA, NJ, PCoA and STRUCURE analyses for both the markers analysed individually and combined. AMOVA analysis revealed high genetic diversity among genotypes. Results from the present study will help in germplasm management and for evaluating and selecting promising blueberry genotypes in the current blueberry improvement program.

Acknowledgement

As I sit down to write these final pages, I realize that my time at the Memorial University of Newfoundland and at the St. John's Research and Development Centre, Agriculture and Agri-Food Canada, has come to an end. This research could not have been done without the help of many other people. Without them, this great experience of my life would definitely not have been as special and gratifying as it was. Therefore, it is my pleasure to have the opportunity to express my thanks to those who have ever helped me.

I would sincerely like to thank Dr. Samir Debnath for having faith in me and providing this opportunity to work under his guidance. I would also like to thank my co-supervisor Dr. Andrei Igamberdieve and a committee member Dr. Natalia Bykova. They have been great source of knowledge, guidance and inspiration to me. My heartfelt thanks are also due to St. John's Research and Development Centre, Agriculture and Agri-Food Canada (AAFC), for extending all the facilities for successful completion of this research and for providing me a platform and motivation through entire program.

I would like to express my deepest gratitude to Darryl Martin, Sarah Leonard and Glenn Chubs for their technical support. They have been supportive in day-to-day situations and making me adapt to the new environment. Their help and guidance not only in my research but also about different situations in life will always be appreciated.

I am really thankful to my friends and fellow lab mates Juran Goyali, Dong An, Poorva Vyas, Junjie Hu (Grace), Amrita Ghosh and Dhrumit Bhatt for providing me all kinds of help and suggestions in my research. It was a great time with all of them. I am really going to miss them.

To my family, I thank you all for always being there for me, enabling me to be who I am and where I am today. You have been the constant source of love, support and encouragement, which makes everything possible. There are not enough words to describe all the thanks you deserve, but perhaps these three come close: I love you!

Finally, I sincerely thank from the bottom of my kindness to all my well-wishers and helpers, who was important to the successful realization of this research. For those of you whose names I did not mention here, I offer my apologies but know that those that have made this experience so special will never be forgotten.

Lastly, but not least, I thank GOD for providing me all the strength to fight the odds, stick to right and sustain my life in the intricate epoch.

Table of contents

List of tables

List of figures

Vaccinium

1 Introduction

1.1 Genus *Vaccinium*

Blueberries (*Vaccinium* spp L.) belong to the genus *Vaccinium* L., which contains approximately 400 species of shrubs or small trees worldwide (Vander Kloet 1988). *Vaccinium* belongs to the Vaccinieae tribe in the subfamily Vaccinioideae of the family Ericaceae. The genus features prominently in ecosystems throughout North and South America, Europe and Eurasia and the majority of species are also found in Malaysia and Southeast Asia (Vander Kloet 1988). The plants are generally found in acidic, sandy, peaty, or organic soils and readily invade disturbed or exposed areas.

The *Vaccinium* genus is traditionally divided into sub-generic groups called sections. From 30 sections of genus *Vaccinium* (*Aethopus, Baccula-Nigra, Barandanum, Batodendron, Brachyceratium, Bracteata, Ciliata, Cinctosandra, Conchophyllum, Cyanococcus, Cyanophtalmos, Eococcus, Epigynium, Euepigynium, Galeopetalum, Hemimyrtillus, Herpothamnus, Myrtillus, Neojunghuhnia, Nesococcus, Neurodesia, Oarianthe, Oxycoccoides, Oxycoccus, Polycodium, Praestantia, Pseudocephalanthos, Pyxothamnus, Rigiolepis, Vaccinium* and *Vitis-Idaea*), blueberries belong to section *Cyanococcus (*GRIN 2015*).* It was proposed that the origin of the section was in South America (Vander Kloet 1988). Genus *Vaccinium* has a very highly conserved genome in North America (Qu et al. 1998), in which polyploidy and interspecific hybridization are common (Lyrene et al. 2003). Therefore, it becomes difficult to infer taxonomic relationships between species within sections.

Berries produced by all the *Vaccinium* species, other than *V. arboreum* Marshall, have fruit that is edible by both birds and mammals, and their flavor varies from insipid to tart to sweet (Vander Kloet 1988). Blueberries (*Vaccinium* spp.), cranberries (*V. macrocarpon* Ait.) and lingonberries *(V. vitis-idaea* L.*)*, are commercially important *Vaccinium* species that have been domesticated in the twentieth century (Vander Kloet 1988).

1.2 Blueberries and their improvement

Blueberries are historically grown in Canada and the United States of America (USA). They grow well in sandy and acidic soil. Blueberries were introduced in some countries of South America, such as Chile and Argentina in the early 20th century. Few countries in Europe, such as the Netherlands, Germany and Sweden started production in mid-19th century (Naumann 1993). Blueberries were also introduced in East Asia and China in the middle of the 20th century (Celik and Islam 2014).

There are five major classes of blueberries grown commercially for fruit today. These classes include, 1) northern highbush, 2) southern highbush, 3) lowbush, 4) half-high, and 5) rabbiteye (Galletta 1996). Northern highbush blueberries (also referred to as highbush) are native to the North American regions. These blueberries grow up to 2 m in height. Most of these blueberries have parentage of *V. corymbosum* L., and some of them also have *V. angustifolium* Ait. as a parent. Northern highbush blueberries are the widely used cultivars in the world (Galletta 1996). Southern highbush blueberries have the same morphology and culture as northern highbush blueberries. These blueberries are mainly *V. corymbosum*, but

also contain the very low-chilling species *V. darrowii* in their parentage, and thus they are also called low-chill highbush blueberries. These blueberries are native to southeastern North America and grow in the same mild climate regions as rabbiteye blueberries, such as southern Virginia to northern Florida, along the Gulf Coast to east Texas and central Arkansas in the USA, and in subtropical regions in Australia (Galletta 1996). The cultivated highbush blueberry is primarily derived from a few wild clones of *V. corymbosum* L., but also includes genetic material from *V. angustifolium* (4X) and sometimes *V. darrowii* Camp (2X) (Qu et al. 1998; Qu and Hancock 1997).

Half-high blueberries are the cultivars derived from the crossing of lowbush and highbush blueberries (Galletta 1996). They attain the height of 0.5 to 1.0 m. These half – high blueberries are important in extending the range of blueberry production in northern areas, as their height is suitable to insulate the snow cover, and at the same time they give increased fruit size. These hybrids or backcross derivatives of lowbush-highbush hybrids involve parentage from *V. angustifolium* and *V. corymbosum.*

Lowbush blueberries are native to eastern Canada and the north-east USA, and they attain a height of 0.5 m. These plants are typically rhizomatous and include predominantly the tetraploid "sweet lowbush blueberry," *V. angustifolium*, but also include Canada blueberry, *V. myrtilloides* Michaux. Commercial production is confined mostly to Maine, Quebec and the Atlantic Canadian provinces (Luby et al. 1991). Rabbiteye blueberry is native to the southeastern North America, encompassing northern Florida, southern Georgia, and southern Alabama (Brightwell 1955).

The lowbush blueberry is a complex tetraploid (2n = 4x = 48) and is widely found in Atlantic Canadian provinces and in Maine. Native people harvested wild blueberries long before the European settlers arrived in North America. The lowbush blueberries were the first to be cultivated commercially by aboriginal people. They practiced blueberry growth by periodically burning blueberry fields, which would quickly grow again with new plants. They apparently knew from the experience that the bushes do not bear when overgrown with brush and weeds (Burgher et al. 2002). Harvested wild lowbush blueberry species include *Vaccinium angustifolium, V. angustifolium f. nigrum* and *V. myrtilloides*. The wild clones of lowbush blueberries are commercially managed in Atlantic Canada. These clones are morphologically and genetically more polymorphic (Burgher et al. 2002; Debnath 2009). Because of this polymorphic nature, the yield and antioxidant contents of berries vary among the blueberry clones.

Before 1900, the lowbush and highbush blueberries were abundant in the wild, which made commercial cultivation of the blueberries unnecessary. But with the increased population, the demand of blueberry increased; at the same time the suitable habitat for blueberry plantation decreased. This forced the plant into commercial cultivation and thus the blueberry breeding programs started. Dr. Frederick Vernon Coville recognized the potential of widespread commercial value of blueberry and started extensive research in various fields of North East USA, such as New Jersey, North Hampshire etc. He found out that bushes grow well in infertile, acidic, natural and non-limed soils, instead of rich, well matured and limed garden soils (Galletta 1996). He selected wild highbush blueberry plant for breeding purposes in 1908, which he named ‘Brooks’. After several failed attempts in

self-pollination, it was concluded that cross-pollination can be resulted in highest fruit production. Thus the second plant he selected for the breeding purpose was the wild lowbush blueberry plant, which he named 'Russell'. The first successful cross were made between them, and after that he selected several plants from the wild for the breeding purposes. By the time he died in 1937, he had propagated over 68,000 seedlings and introduced 15 cultivars for the commercial blueberry production. 'Dixi' is the last cultivar he released, which is Latin for "I am done" (Gough, Robert E. (Robert Edward) 1994.).

The early objectives in blueberry breeding included large berry size, light blue color, small scars, firm fruit, good flavour and high productivity (Ballington 2001). As these objectives were realized, breeders began selecting for extended harvest season and geographical range (Michigan, South Carolina, and Florida), adaptation to mechanical harvest, and disease resistance (e.g. mummy berry *(Monilinia)* (Ballington 2001). The future objective of the breeding experiments are to grow high quality blueberries with broader soil adaptation (less dependent on acidic soil), broader climatic adaptation, improved precocity, mechanical handling tolerance and superb fruit flavours (Galletta 1996).

1.3 Health benefits and medical functions of blueberries

Blueberries are known to possess great health benefits because of their high antioxidant capacity and total phenolics compounds, such as anthocyanin, flavonols and phenolic acids (Kalt et al. 1999). Today the blueberry's reputation as a "healthy" food makes it one of the most popular small fruits on the market. Much research suggests the

association of blueberry antioxidants with numerous health benefits (Halliwell 2007; Joseph et al. 2003; Willis et al. 2005). Blueberries have shown to have one of the highest antioxidant activities compared to the other fruits tested, including cranberries, apples, and red and green grapes (Prior et al. 1998; Wolfe and Liu 2007). Blueberry extract is believed to enhance signaling and prevent behavioral deficits in an Alzheimer disease model (Joseph et al. 2003), memory enhancement in a rat model (Andres-Lacueva et al. 2005), cancer prevention properties (Block et al. 1992) and have positive effect on night vision (Canter and Ernst 2004). The consumption of blueberries has also proved to reduce the size and growth of tumors in mice and rats (Kanaya et al. 2014; Jeyabalan et al. 2014).

1.4 Industry and production of blueberries

Early settlers in the Atlantic provinces of Canada first harvested the fruit for their own use or for local distribution (PRRP 2012). Canada's blueberries are commercially grown in both wild and cultivated varieties, which make them unique and no other Canadian fruit shares this distinction (AAFC 2011). Improvements in marketing and shipping and the establishment of canneries in Maine and along the Canada–USA border in the mid-1800 expanded the markets. Improved harvesting methods and management resulted in an expansion of production throughout the 20th century. Since the 1980's, production has increased dramatically because of advancements in management, including improved weed control and the increased use of bees for pollination (PRRP 2012).

Canada is the second largest blueberry producer after the USA and world's largest producer of "lowbush blueberries" (AAFC 2011). Most lowbush blueberries are grown

commercially in Quebec and the Atlantic provinces. Canada is the largest exporter of blueberries with its largest export to the USA and followed by Japan (Scrivener 2008; Ogg 2014). In 2011, the blueberry production in the USA was 188,926 metric tons (NASS 2015) while the Canadian blueberry production was 112,363 metric tons in 38,413 hectares of land (PRRP 2012). In 2012, Canada exported 88,477 metric tons of blueberries and in 2011, 54,486 metric tons, which shows 40% increase in export. Each year, as the export is increasing, the land used for the production is also increasing (Ogg 2014). Blueberries are produced largely, among all the fruits, in Canada and they consumes more than half the country's fruit-growing area (AAFC 2011; Scrivener 2008). This dictates the importance and public interest in blueberry. British Columbia is the highest blueberry growing region in Canada (AAFC 2011; Scrivener 2008).

Blueberries are used widely and have many usages in different industries, such as baking, wine, cosmetics, desserts, snack, pet foods, nutraceuticals and dairy (Villata 2015). Blueberries are delicious in pancakes, pies, tarts, muffins, sauces and cakes, but they are amazingly versatile in other dishes as well. Their fresh, natural flavour enhances the taste of pork, chicken and game, and chefs can combine them with almost any other fruit or berry to make a dessert (AAFC 2011). Health professionals are openly promoting the consumption of blueberry for healthy diet and high amounts of antioxidants. Because of their taste and health advantages, it is the number one favourite berry fruit (Villata 2015). Statistics detected that in 2012, blueberries were used in ~1000 different products in different industries (Brazelton 2013) compared to ~450 different products in year 1998. Baked good and dairy products such as: muffins, bagels, yogurt, juice etc. are the

most common uses. Thus, it can be determined that baking and snack industries are the largest buyer of the freshly grown blueberries. However, new categories such as pet foods continue to rise (Brazelton 2013).

1.5 Wild germplasm and importance of genetic diversity

Plant germplasm is the living tissue from which new plants can be grown. Germplasm is usually seed, or it can be another plant part, for example, a stem, a leaf, or pollen, or even just a few cells that can be cultured into a whole plant (AAFC 2014). Plant germplasm carries genetic information for the plant's hereditary makeup. For most crops, there is a very wide gene pool in the wild species in comparison with the limited range of genetic variability in collections in the existing gene centers (Hawkes 1977). Hawkes (1977) gave a general review of the importance of wild germplasm and evaluated its potential in plant breeding research. According to him, knowledge on the phylogeny, taxonomy and geographical distribution of the wild species is required for their best use in a breeding program. For instance, wild species may possess a whole set of favourable (for example pest resistance) or unfavourable agronomic features (like low yield and poor flavour) (Hawkes 1977). Thus, wild clones are very important for any breeding program as they are very diverse and give more options to the breeders.

Genetic diversity means genetic variation in a sum of individuals (Templeton 1991). Even though genetic diversity is at the lowest hierarchy of diversity analysis, it has an impact on the higher levels of biodiversity such as morphological and biochemical (Templeton 1991; Templeton 1993). Population needs genetic diversity among themselves

to evolve and adapt to environmental changes. Generally the wild species are more diverse than the cultivated varieties. However, the genepool for today's industry is expanding rapidly as new methods of interspecific hybridization are developed.

Canadian agriculture is based on crops that originated from areas outside of Canada. For example, wheat originated in the near east (in such countries as Iran), corn in Mexico and Guatemala, alfalfa in Turkey, and soybean in China. Crops of economic importance that are native to Canada are limited and include sunflowers, strawberries, raspberries, Saskatoon berries, blueberries, currants, and cranberries (AAFC 2014). Thus, Canada does not have enough germplasm to avoid the uniformity in the crop plants. The flood of genetic uniformity increases the potential for crop vulnerability to new pests and stresses (Lenne and Wood 1991; Hawkes 1977). Atlantic Provinces are rich with the wild germplasm for blueberries (Debnath 2007). Genetic diversity gives the sustained ability to develop new plant cultivars that can resist new pests, diseases and environmental stresses. Wild ancestors and relatives are the keys to genetic diversity (Hawkes 1977). Thus it is very important to have a knowledge on genetic diversity among the plants.

1.6 Measurement of genetic diversity

Until now, morphological, biochemical and molecular markers have been used for blueberry breeding programs. However, morphological and biochemical variabilities are often restricted, as these characters may not be obvious at all stages of the plant development and appearance may be affected by environmental changes. On the contrary, molecular markers are robust and do not get affected by the environmental changes. They

also gives similar results at all stages of the plant development. These benefits of molecular markers make them the first choice for breeding programs (Debnath et al. 2012).

1.6.1 Molecular tools for diversity analysis

Blueberry cultivars are vegetatively propagated and can be mis-identified based on phenotype. Keys for identifying blueberry cultivars based on leaf and plant characters are available for cultivars released before 1945 (Darrow 1966). A variety of different genetic markers have been proposed to assess genetic variability as complementary strategies to more traditional approaches in genetic resources management (Debnath et al. 2012). Molecular tools provide valuable data on diversity through their ability to detect variation at the protein or DNA level. DNA-based markers reveal a high degree of polymorphism and can accelerate genotype fingerprinting and studies of genetic relationships among accessions.

Numbers of different methods are available to show genetic diversity between organisms, and these methods are mainly based on protein and DNA molecules. The choice of technique for any specific use will depend upon the material being studied and the nature of the questions being addressed (Brown 2014; Debnath et al. 2012).

1.6.2 Diversity parameters

It is very important to determine, describe and quantify the amount and patterns of genetic variation within and among the populations, to understand the influence of selection, inbreeding and breeding interventions (Ojango 2011). Several methods have

been used for the genetic diversity analysis in *Vaccinium* species. In this and several other studies, various parameters were used, which showed variability and strengths of markers. These parameters include allele number, major allele frequency, expected and observed heterozygosity, inbreeding coefficient and polymorphic information content (PIC).

Allele number represents the number of alleles (N_A) provided by a marker. Total number of alleles observed for different populations is considered to be a reasonable indicator of genetic variation. Markers with a low total N_A have low genetic variation and markers with a high N_A have high genetic variation (Ojango 2011). The frequencies of an allele at loci are calculated manually by direct counting.

The simple measure of genetic variation in a population is the amount of heterozygosity observed (Weir 1996). High heterozygosity values for genotypes may be due to long-term natural selection or due to historic mixing of strains of different populations. Recombination can potentially increase variation and expected heterozygosity (H_E), either directly or indirectly (through the effects of selection). Thus, a higher recombination rate should translate in higher genetic diversity within a given genomic region, population or even species (Jaramillo-Correa et al. 2010). A low level of heterozygosity may be due to isolation with the subsequent loss of unexploited genetic potential. Marker heterozygosity is related to the polymorphic nature of each locus. A high level of average heterozygosity at a locus could be expected to correlate with high levels of genetic variation at loci. This parameter has critical importance for adaptive response to environmental changes (Ojango 2011).

The observed heterozygosity is defined as the percentage of loci heterozygous per individual, or the number of individuals heterozygous per locus (Ojango 2011). Marker heterozygosity is estimated by summing the heterozygosity at all loci for each genotype and averaging this quantity over all genotype. The alternative measure of variation, often referred to loosely as average heterozygosity, but more precisely known as expected heterozygosity (also referred to as gene diversity), is formed from the sum of squares of allele frequencies. It is more appropriate measure of variability for inbred population where there are very few heterozygotes, but there may be several different homozygous types (Nei 1987).

The inbreeding coefficient *F* (also called the fixation index) exhibits values ranging from -1 to +1. Values close to zero are expected under random mating, while substantial positive values indicate inbreeding or undetected null alleles. Negative values indicate excess of heterozygosity, due to negative assortative mating, or selection for heterozygotes (Peakall and Smouse 2012).

PIC refers to the value of a marker for detecting polymorphism within a population, depending on the number of detectable alleles and the distribution of their frequency; thus, it provides an estimate of the discriminating power of the marker. PIC is a modification of the heterozygosity measure that subtracts, from the heterozygosity value, an additional probability that an individual in a linkage analysis does not contribute information to the study (Nagy et al. 2012, Shete et al. 2000). Elston (2005) explained that PIC is a measure of a marker's usefulness for linkage analysis and gave a definition that the polymorphism information content of a marker is the probability that the marker genotype

of the offspring of a heterozygous parent affected with a dominant disease allows one to deduce which marker allele the offspring inherited from the parent. It is calculated using the below mentioned formula:

$$\mathbf{PIC} = \sum_{i=1}^{n} \boldsymbol{Pi^2} - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \boldsymbol{2Pi^2\, Pj^2}$$

Where *Pi* and *Pj* are the allele frequency for the *i*th and *j*th alleles respectively.

It is necessary to know, in wild populations, the relationship among the genotypes, and thus it is required to do cluster analysis. Cluster analysis is a technique for grouping individuals into unknown groups and to assess the relationship between the groups. In plant diversity studies, cluster analysis has been used to classify different cultivars and clones into groups on the basis of their genetic characteristics. In cluster analysis, distance measures are recommended before clustering, which define closeness or similarity of two observations or genotypes. In plant breeding, distance measures are called genetic distance (Weir 1996).

The commonly used methods of clustering fall into two general categories: hierarchical and non-hierarchical. Hierarchical procedures are the most commonly used in plant diversity studies. When the number of variables is more than two and the data set is large, dendrograms have been used. In a dendrogram, the horizontal axis lists the observations in a particular order. The vertical axis shows the successive steps or cluster numbers (Ojango 2011).

In plant diversity studies, genetic distance measures are used to construct the dendrograms. The two most commonly used analysis methods, for constructing the dendograms are, unweighted pair group method with arithmetic mean (UPGMA) and the neighbour-joining (NJ) method (Nei 1987). In this research the dendograms consist of branches and tips. The tips are the genotypes and the branch lengths between genotypes are graphical estimates of the genetic distance between the genotypes and give an indication of genetic relationships between genotypes. UPGMA dendograms give an indication of the time of separation (divergence) of genotypes. The higher the branch length the longer is the separation period between breeds (Ojango 2011). The difference between UPGMA and NJ analysis is that NJ allows unequal rates of evolution along the branches, and both methods use different statistical approach (Weir 1996). UPGMA method defines the intercluster distance as the average of all the pairwise distances for members of two clusters. NJ method is used for identifying closest pair, or neighbours of taxonomic units in a way to minimize the total length of a tree (Weir 1996).

Bootstrapping is usually done to provide confidence statements about the groupings of the genotypes as revealed by the dendrograms and hence test the validity of the clusters obtained. In bootstrapping new dendogram can be reconstructed by replacing node(s). The topology of this new dendogram is then compared to that of the original dendogram. Each interior branch of the original dendogram that is different from the bootstrap dendogram is given a score of 0; all other interior branches are given the value 1. This procedure of resampling the sites and the subsequent tree reconstruction is repeated several hundred or thousand times, and new dendograms are created. In this case, it is difficult

to present all the trees for publication. One way to solve this problem is to make a composite tree that includes all the trees. Such a composite tree is called a "consensus tree". There are several different types of consensus (Nei and Kumar 2000), but the most commonly used are the "strict consensus" and "majority-rule consensus trees". The strict consensus tree of a collection of trees contains only those clusters found in all the trees in the profile. Among the majority-rule consensus trees, the most commonly used is the 50% majority-rule consensus tree. In this tree a branching pattern that occurs with a frequency of >50% is adopted (RW.ERROR - Unable to find reference:431).

Multivariate analysis, such as principal coordinates analysis (PCoA) is a method to explore and to visualize similarities or dissimilarities of the data. It starts with a similarity matrix or dissimilarity matrix (distance matrix) and assigns for each item a location in a low-dimensional space. Interpretation of a PCoA plot is straightforward: objects ordinated closer to one another are more similar than those ordinated further away (Legendre and Legendre 1998).

The inherent genetic structure of populations can be assessed directly using a method developed by Pritchard et al. (2000) and implemented in the program STRUCTURE. The program implements a model-based clustering method to infer population structure, assign individuals to populations and identify migrants and admixed individuals using multilocus genotype data, independent of prior population information. The approach implemented in STRUCTURE assumes a model in which there

are *K* populations (where *K* may be unknown), each of which is characterized by a set of allele frequencies at each locus. Individuals in the sample are assigned probabilistically to populations or jointly to two or more populations if their genotypes indicate them to be admixed.

To see how the populations and groups are different from each other, generally analysis of molecular variance (AMOVA) method is used. AMOVA is a method of estimating population differentiation directly from molecular data and testing hypotheses about such differentiation (Excoffier et al. 1992). AMOVAs can be used to: (1) describe the partitioning of genetic variation among and within groups; and (2) test user-defined groupings of populations. AMOVA differs from a simple analysis of variance (ANOVA), in that data are arranged hierarchically and mean squares are computed for groupings at all levels of the hierarchy. This allows for hypothesis tests of between-group and within-group differences at several hierarchical levels (Ojango 2011).

All these methods have been used in the genetic diversity analysis of various plant species. The following sections illustrate the use of these methods.

1.6.3 Genetic diversity within *Vaccinium* species using allozyme markers

Protein based markers were the first markers used for genetic studies in plant diversity analysis (Krebs and Hancock 1989). For the first time in blueberry genetic analysis, segregation was observed with the help of six isoenzymes, encoded by four dimeric enzymes in highbush blueberry, *Vaccinium corymbosum* L. (Krebs and Hancock 1989). In 2000, 19 isoenzyme loci were studied in diploid and tetraploid populations of *V.*

oxycoccos (Mahy et al. 2000). It was reported that autotetraploid populations had more polymorphic loci, mean number of alleles and heterozygosity than diploids. Previous study by Hokanson and Hancock (1998), proved that number of allele per locus and the level of heterozygosity were lower in diploid *V. myrtilloides* than the tetraploids, *V. corymbosum* and *V. angustifolium*. The average level of heterozygosity, in this study, was almost 20% higher in *V. corymbosum* than *V. angustifolium* (Hokanson and Hancock 1998).

However, the number of polymorphic loci that can be assayed, and the level of polymorphisms observed at the loci are often low, which greatly limits their application in genetic diversity studies. With the development of new technologies, DNA polymorphisms have become the markers of choice for molecular-based surveys of genetic variation (Debnath et al. 2012; Die and Rowland 2013; Bell et al. 2008; Boches et al. 2006; Dhanaraj et al. 2004).

1.6.4 Genetic variation within *Vaccinium* species using DNA - based markers

The introduction of DNA markers for the investigation and exploitation of DNA polymorphism is one of the most critical developments in biology. DNA markers are increasingly applied in crop breeding programs around the world to investigate genetic diversity in and among the *Vaccinium* species (Debnath et al. 2012). Genotype identification and documentation of genetic variation are essential in blueberries for property rights protection, practical breeding program and germplasm characterization.

There are numerous DNA-based markers, which include restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNA (RAPD), inter simple sequence repeats (ISSRs), amplified fragment length polymorphisms (AFLPs), simple sequence repeats (SSR) and EST-PCR and EST-SSR.

1.6.4.1 Restriction fragment length polymorphism (RFLP)

In RFLP analysis, the DNA sample is digested into fragments, using specific restriction enzymes, and those fragments can be detected by southern hybridization after running agarose gel electrophoresis (Old 1980). Although RFLP is a good tool, it requires elaborate laboratory techniques, including development of specific probes, use of radioisotopes, southern blot hybridization procedure and autoradiography. These limitations make this technique time consuming and labour intensive (Debnath et al. 2012; Whisson et al. 1995; Kesseli et al. 1994).

1.6.4.2 Random amplified polymorphic DNA (RAPD)

The invention of PCR (polymerase chain reaction) is a milestone in the development of molecular techniques. RAPD and other methods explained below are based on PCR. In RAPD, the segments of DNA that are amplified are random. Random amplification of portions of DNA, with the primer, by PCR, is a useful technique in genetic diversity analysis (Qu and Hancock 1997; Nadarajan et al. 1999). RAPD primers are arbitrary, short primers (8–12 nucleotides). In RAPD markers, no knowledge of the DNA sequence of the targeted genome is required, as due to their short length the primers have high probability of binding somewhere in the sequence. The RAPD procedure does not

require the use of radioactive probes (Debnath et al. 2012; Nadarajan et al. 1999). However, RAPD markers are less informative than co-dominant markers (Nybom 2004). 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).

RAPD markers were developed for linkage mapping in *Vaccinium* (Levi and Rowland 1997; Levi et al. 1993; Rowland and Levi 1994; Qu and Hancock 1997; Rowland et al. 1999). This procedure resulted in reliable RAPD patterns for all organisms tested. Clonal structure in wild populations was also tested for wild populations of *V. stamineum* L. (Kreher et al. 2000), *V. vitis-idaea* (Persson and Gustavsson, 2001) and *V. myrtillus* L. (Albert et al. 2003). In these tests, RAPD markers gave mixed results. RAPDs have also been used to distinguish 26 genotypes in *V. angustifolium* (Burgher et al. 2002) and it was concluded that RAPD markers were useful tool for identification of lowbush blueberry genotypes. RAPD analysis was also used for the cultivar identification in rabbiteye and highbush blueberry (Aruna et al. 1995; Aruna et al. 1993; Levi and Rowland 1997), where cluster analysis of genetic distance assessment grouped siblings with each other and with one or both parents. This analysis also helped in identifying two wild selections of rabbiteye blueberry, 'Ethel' and 'Satilla'.

Debnath (2005) used RAPD markers to differentiate 13 cranberry genotypes. It was concluded that RAPDs were useful for the diversity analysis and germplasm maintenance (Debnath, 2005). RAPD markers have been widely used for cranberry genetic analysis (Novy et al. 1994; Novy and Vorsa 1996; Novy and Vorsa 1995; Novy et al. 1996). Twenty two cranberry varieties were assessed, where on the basis of 66 silver stained RAPD

(ssRAPD) markers, 17 unique ssRAPD-DNA profiles were identified rather than the expected 22. Fourteen varieties had unique profile, while the other 8 were represented by 3 ssRAPD profiles.

Debnath (2007a) assessed genetic variability in 43 wild cranberry clones collected from four Canadian provinces and five cranberry cultivars, using RAPD - PCR. A significant degree of genetic diversity were revealed. AMOVA indicated 10% of total variation was effected by geographical distribution and 90% of total variation was found among the clones (Debnath 2007a). Three ISSR markers and 15 RAPD primers were used to differentiate 15 high bush blueberry, 2 rabbiteye cultivars and one southern lowbush wild selection (Levi and Rowland 1997). RAPDs were suggested as a replacement for isozymes (Novy and Vorsa, 1995).

1.6.4.3 Amplified fragment length polymorphism (AFLP)

AFLP technology was introduced by Vos et al. (1995). In this technique, total genomic DNA is digested with two restriction enzymes. Then, double stranded nucleotide primers are ligated to the DNA fragments to serve as primers. These primers are binding sites for the PCR amplification. This technique has been extensively used with plant DNA for the development of high-resolution genetic maps and for the positional cloning of genes of interest (Debnath et al. 2012). A major drawback for AFLP is that it is dominant marker. Thus, heterozygotes cannot be distinguished from homozygotes with AFLP analysis (Nybom 2004).

1.6.4.4 Inter simple sequence repeat (ISSR)

ISSRs are DNA fragments of about 100–3000 bp, located between adjacent and oppositely oriented microsatellite regions. This technique was reported by Zietkiewicz et al. (1994). The primers based on microsatellites are utilized to amplify inter-SSR DNA sequences (Wang et al. 1994).

This marker has proved to be more reproducible than RAPD marker, and generally indicates higher levels of polymorphism due to the longer sequences of ISSR primers and the higher annealing temperatures used (Gupta et al. 1994; Debnath, 2005). However, ISSR marker is a dominant marker. Because ISSR is a multilocus technique, disadvantages include the possible non-homology of similar sized fragments. Moreover, ISSRs, like RAPDs, can have reproducibility problems. Compared with RAPD and AFLP, ISSR overemphasizes differences between closely related populations and attribute less variation to differences over large geographical distances (Reddy et al. 2002).

Debnath (2009) reported a study on 43 wild lowbush blueberry (*V. angustifolium*) clones from four Canadian provinces and the cultivar 'Fundy', where he detected genetic similarity and variation using ISSR marker. Forty one genotypes were clustered into two main clusters by UPGMA and three genotypes were found to be outliers (Debnath 2009). There was also ISSR study on lingonberry, performed by Debanth (2007c), where 43 wild lingonberry clones were assessed. In total 10% of variation was found using AMOVA and explained by geographical distribution. ISSR markers have also been used for the genetic fingerprinting of cloudberry (*Rubus chamaemorus* L.) (Debnath 2007b).

1.6.4.5 Simple sequence repeats (SSR)

Microsatellite (also called simple sequence repeat (SSR)) markers are short (1-6 bp) tandemly repeated DNA sequences, flanked by unique, conserved DNA sequences that can be determined using PCR (Spritz 1981; Miesfeld et al. 1981; Hamada and Kakunaga 1982; Tautz and Renz 1984; Weber and May 1989). Microsatellites were discovered in humans but have been found in the genomes of most organisms. Microsatellite are also codominant markers (heterozygotes can be detected) and relatively easy to score (Li et al. 2002). Microsatellite containing genomic fragment have to be cloned and sequenced in order to design primers for specific PCR amplification. This approach was called sequence-tagged microsatellite site (STMS) (Beckmann and Soller 1990)

When designed and tested correctly, these markers are highly reproducible (Jones et al. 1997). The strengths of microsatellites include the codominance of alleles, their high genomic abundance in eukaryotes and their random distribution throughout the genome, with preferential association in low-copy regions (Morgante et al. 2002, Squirrell et al. 2003, Zane et al. 2002). Due to the use of long PCR primers, the reproducibility of microsatellites is high and analyses do not require high quality DNA. In the longer term, development of allele-specific markers for the genes controlling agronomic traits will be important for advancing the science of plant breeding (Beckmann and Soller 1990).

The main disadvantage to microsatellites is that they have to be isolated *de novo* for each species of interest (Li et al. 2002).. One of the main drawbacks of microsatellites is that high development costs are involved if adequate primer sequences for the species of

interest are unavailable, making them difficult to apply to unstudied groups. Although microsatellites are in principle co-dominant markers, mutations in the primer annealing sites may result in the occurrence of null alleles (no amplification of the intended PCR product), which may lead to errors in genotype scoring. A very common observation in microsatellite analysis is the appearance of stutter bands that are artifacts in the technique that occur by DNA slippage during PCR amplification. These can complicate the interpretation of the band profiles because size determination of the fragments is more difficult and heterozygotes may be confused with homozygotes (Beckmann and Soller 1990; Kumar et al. 2009).

Pyrosequencing technology has been used to generate SSR markers in cranberry (Zhu et al. 2012; Zalapa et al. 2012). Forty-eight polymorphic SSR loci with 2-15 alleles per locus, for a total of 323 allele were detected within 25 cranberry genotypes (Zhu et al. 2012). Fajardo et al.(2013) applied 12 SSR markers to run genetic cluster analysis to assess the genetic diversity of 21 cranberry cultivars, 11 experimental hybrid and 6 representative accessions of wild species. The most prospective clonal representative of some essential cranberry cultivars were identified by consensus genetic profiles via SSR alleles (Fajardo et al. 2013).

Schlautman et al. (2015) designed, synthesized, and tested new primers for 979 SSR loci of cranberries; 697 of the markers amplified allele patterns consistent with single locus segregation in a diploid organism and were considered polymorphic. Of the 697 polymorphic loci, 507 were tested on the panel of 13 cranberry cultivars which yielded 2278 alleles with an average of 4.49 alleles and a range of 2 to 11 alleles (N_A) per locus.

More than 80% of the 507 polymorphic SSRs tested had greater average number polymorphic information content (PIC) (0.59), observed heterozygosity (H_O) (0.72), and expected heterozygosity (H_E) (0.63) than the other motif classes. This comprehensive collection of developed and validated microsatellite loci represents a substantial addition to the molecular tools available for geneticists, genomicists, and breeders in cranberry and *Vaccinium.*

1.6.4.6 Expressed sequence tag (EST)-PCR and EST- simple sequence repeats (SSR)

Expressed sequence tags (ESTs) are sequenced portions of complementary DNA copies of mRNA—they represent part of the transcribed portion of the genome in given conditions. As expected, they mainly correspond to relatively conserved sequences. They are generated by largescale single-pass sequencing of randomly picked cDNA clones and have proven to be efficient and rapid means to identify novel genes. Techniques have been developed to reveal polymorphism associated with such sequences (Cato et al. 2001). Characterization of ESTs is a convenient and rapid way to identify new genes in various organisms (Liu et al. 1999; MacIntosh et al. 2001). There are numerous ongoing EST projects that provide expressed sequences by single-pass sequencing both upstream and downstream of cDNAs using arbitrarily selected cDNA libraries. This means that discovery of novel genes takes as little time as it takes to perform a computer search of a sequence database rather than the months or even years that were required to work at the bench (Rowland et al. 2003a; Rowland et al. 2003b; Bell et al. 2008; Yadav et al. 2011).

ESTs were originally intended as a way to identify gene transcripts, but have since been instrumental in gene discovery, for obtaining data on gene expression and regulation, sequence determination, and for developing highly valuable molecular markers, such as EST-based PCR and SSR (Debnath et al. 2012; Beckmann and Soller 1990). EST-SSR markers, unlike genomic-derived SSRs, are located in gene rich regions of the genome, making them more likely to be linked to useful traits (Weber and May 1989; Beckmann and Soller 1990). Since expressed sequences are highly conserved, EST-SSR markers are expected to result in better cross-species amplification than genomic SSRs (Wang et al. 1994).

Rowland et al. (2003b) developed PCR markers designed from expressed sequence tags (EST-PCR markers) for use in linkage mapping, cultivar identification, and genetic relatedness studies. These markers displayed polymorphism due to size differences in the amplicons or when the amplicons were cut with a restriction enzyme. The cDNA clones, from which the ESTs were derived, encoded a wide range of proteins including temperature stress-related protein, proteins involved in signal transduction and basic metabolic proteins (Rowland et al. 2003b). These EST-PCR markers are also conserved in related Ericaceae species (Rowland et al. 2003a). EST-PCR primers, which were derived for blueberry were used with cranberry genotypes (two wild selections of *V. oxycoccos* L. and two cultivars of *V. macrocarpon* Ait.). It was reported that 23 out of 26 primers produced amplification successfully and eight out of the 26 available primers produced polymorphic fragments among the cranberry genotypes. Rowland et al. (2010) tested 44 EST-PCR for genetic fingerprinting and relationship studies in rabbiteye blueberry. Of 44 EST-PCR primer pairs,

40 (91%) resulted in successful amplification, and 33 of those (83%) amplified polymorphic fragments among the rabbiteye genotypes. A dendrogram constructed from genetic similarity values, tended to group siblings and parent/progeny together, generally agreeing with pedigree information.

The EST-PCR markers developed for highbush blueberries were used to study genetic relationship and finger printing in lowbush blueberry (Bell et al. 2008; Debnath 2014). In one interspecific genetic relationship study, 14 genotypes, including more than two specimens of each of four closely related *Vaccinium* L. species and only four pedigreed cultivars of *V. angustifolium*, grouped separately as expected in a genetic similarity dendrogram (Bell et al. 2008). In another study, 21 wild lowbush blueberry clones from four Canadian provinces were compared with 6 cultivars using EST-PCR and EST-SSR markers. The structure analysis grouped all the wild clones according to their geographic location and grouped all the cultivars together as expected (Debnath 2014).

Several plant species in genus *Vaccinium* have been genetically profiled by EST-SSR markers (Liu et al. 2014; Boches et al. 2006; Boches et al. 2005; Yang et al. 2008). Boches et al.(2006) detected 627 alleles via 28 EST-SSR primers. Unique fingerprints were accessed for all 69 accessions representing wild and domesticated highbush blueberry germplasm. New EST-SSR primers were designed by Liu et al. (2014) from the EST database of blueberry. They designed 54 EST-SSR markers from 120 primers by assaying 12 samples of the *Vaccinium* genus section *Cyanococcus*. Results showed that those EST-SSR markers displayed considerable polymorphism and high transferability in both cultivated and wild species of *Vaccinium*. It was suggested that the primers will contribute

to the studies of genetic diversity, linkage mapping, QTL and molecular assisted breeding of *Vaccinium* crops (Liu et al. 2014).

1.7 Objectives of this project

In Canada, especially in Atlantic provinces and Quebec, lowbush blueberries are harvested either from commercially managed fields or from natural stands. It is undisputed that blueberries promote good health and carry very important phytochemicals. Due to its popularity and health advantages, the demand of high quality blueberries have increased in different industries, in recent years. To meet these demands, it is essential to develop cultivars suitable for cultivation in cool climates. The development of new cultivars requires selection of superior wild growing plants and crossing them with half-high or highbush blueberries. To address this issue, a program was developed to create improved blueberry cultivars at the St. John's Research and Development centre of Agriculture and Agri-Food Canada in Newfoundland and Labrador (Debnath 2011).

In order to accelerate genetic improvement, it is desirable to understand the inheritance pattern of complex quantitative traits like yield and adaptability. This can be facilitated by developing DNA-based markers and their applications to estimate the level of genetic diversity. The development and use of molecular markers to detect and exploit DNA polymorphisms is one of the most significant developments in the field of molecular genetics (Debnath et al. 2012). Among available molecular markers, simple sequence repeats (SSRs) or microsatellites have gained considerable importance in increasing plant genetic diversity, genetic mapping, population genetic analyses and molecular assisted

breeding techniques owing to their highly polymorphic and codominant nature (Chapman et al 2009). With the availability of large number of ESTs in public domain, the development of EST-derived markers is an efficient and cost-effective option. Since EST markers identify variability in the transcribed regions of the genome, it may lead to rapid identification of functional candidate genes and increase the efficiency of marker-assisted selection (MacIntosh et al. 2001). Debnath (2014) used EST-PCR and EST-SSR markers, which have been proved useful previously, to assess the genetic diversity among wild lowbush blueberry clones of four Canadian provinces. The number of genotypes and markers used in that study was low compared the present study. In this study, various statistical analyses were also performed to determine the genetic diversity among genotypes. The present study was conducted with 56 blueberry wild blueberry clones collected from four Atlantic Canadian provinces, six cultivars and one selection with the following objectives:

1. To identify EST-PCR and EST-SSR primers suitable for fingerprinting the present set of blueberry genotypes,
2. To assess the properties of markers and their importance in assessing the diversity in the present set of blueberry genotypes.
3. To evaluate genetic diversity and population structure amomg genotypes.

2 Materials and methods

2.1 Plant material

In this study, 63 blueberry genotypes were selected for the analysis. These genotypes include 56 wild lowbush blueberry clones collected from four Canadian provinces: 37 clones from Newfoundland and Labrador (designated as NL1 to NL37), 11 clones from Prince Edward Island (designated as PE1-PE11), six clones from Quebec (designated as QC1, QC2, QC4, QC5, QC7 and QC9) and two clones from New Brunswick (designated as NB1 and NB2). The other genotypes include three half-high blueberry cultivars Patriot, Chippewa, and St. Cloud (designated as PT, CH and SC, respectively); one highbush blueberry cultivars Polaris (designated as PO); two lowbush blueberry cultivars Fundy and Brunswick (designated as FU and BR, respectively); and one lowbush blueberry selection (designated as FO), obtained from the open-pollinated seedlings of lowbush blueberry cultivar Fundy.

All the blueberry wild clones were collected in August 2001.The information regarding location and collection sites for the lowbush blueberry clones is given in Table 1 and Figure 1, and the parentage information of blueberry cultivars is given in Table 2.

Table 1: Locations of wild lowbush blueberry clones collected from Canadian provinces: Newfoundland and Labrador (NL1-NL37), Prince Edward Island (PE1-PE11), Quebec (QC1, QC2, QC4, QC5, QC7 and QC9) and New Brunswick (NB1 and NB2).

Clones	Province	Community	Total genotypes	Longitude	Latitude
NL1 - 15	NL	Shearstown	15	47°35'	53°17'
NL16-36	NL	North river	21	47°32'	53°18'
NL37	NL	Logy bay	1	47°37'	52°40'
PE 1 - 11	PE	Blooming point	11	46°23'	62°58'
NB1	NB	Clifton	1	46°06'	64°01'
NB2	NB	Little Shemogue	1	47°43'	65°22'
QC1	QC	Baie-Comaeu	1	49° 13'	68° 08'
QC9	QC	Longue-Rive	1	48° 33'	69° 14'
QC2	QC	Pointe-Lebel	1	49° 09'	68° 13'
QC4, QC5, QC7	QC	Baie-Trinité	3	49° 25'	67° 18'

Clones were collected from three communities of Newfoundland and Labrador province: Shearstown, North River and Logy Bay. The clones belonging to particular communities are given in Table 2. For the Prince Edward Island province, clones were collected from only one community, Blooming-point. For the province of Quebec, there were four communities i.e. Baie-Comaeu, Longue-Rive, Pointe-Lebel and Baie-Trinité, from where the clones were collected. These were considered for AMOVA analysis. One clone each was collected from Clifton and Little Shemogue community of the province of New Brunswick.

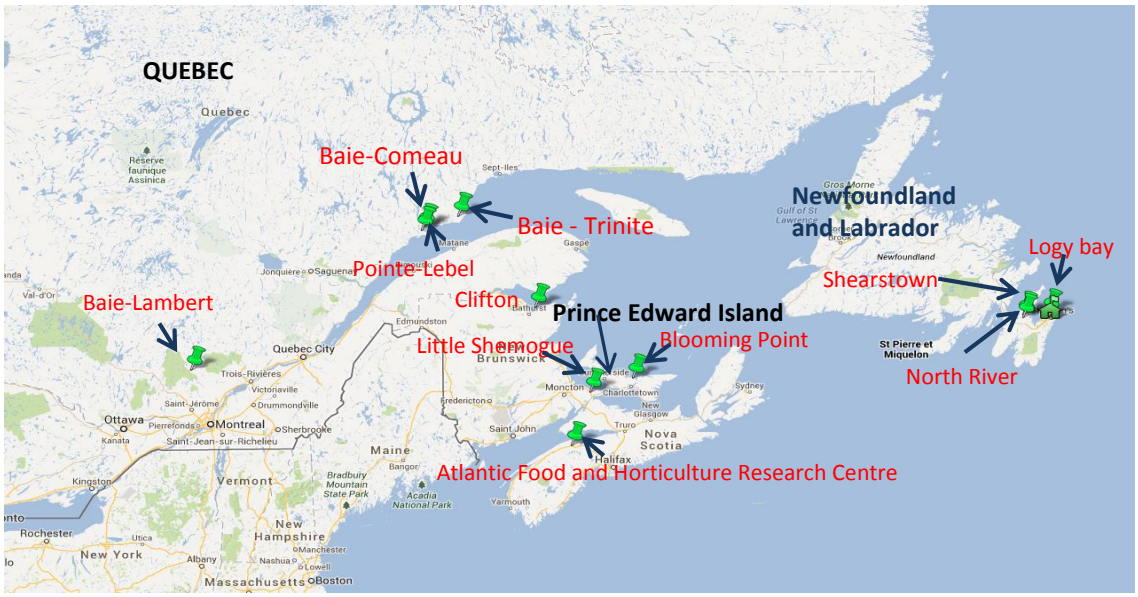

Figure 1: Collection sites for wild blueberry clones from four Atlantic Provinces. The genotype information belonging the locations is given in Table 1.

Two lowbush blueberry cultivars, Fundy and Brunswick, are vigorous in nature and contain large fruits. These two cultivars were developed at the Atlantic Food and Horticulture Research Centre, Kentville, Nova Scotia, Canada. Fundy was selected from open pollinated seedling of Augusta lowbush blueberry cultivar (Lyrene, 2002), whereas Brunswick was selected from local native lowbush blueberry clones (Aalders et al. 1977).

Table 2: Parentage of lowbush blueberry cultivars: Fundy (FU) and Brunswick (BR), half-high blueberry cultivars: St. Cloud (SC), Chippewa (CH) and Patriot (PT), highbush blueberry cultivar: Polaris (PO); and a selection (FO)

Cultivar/ Selection	Type	Parentage
FO*	Selection	Open pollinated seedling of cultivar Fundy
Fundy (FU)	Lowbush	Developed at AAFC centre Kentville, NS, Canada (Lyrene, 2002)
Brunswick (BR)	Lowbush	Fairfied, NB (natural selection) (Aalders et al. 1977)
Polaris (PO)	Highbush	B15 x Bluetta (Okie, 1997)
Patriot (PT)	Half-high	US3 (Dixi x Michigan lowbush No. 1) x Earliblue (Hepler and Draper, 1976; Finn et al. 1990)
Chippewa (CH)	Half-high	B18A (G65 x Ashworth) x US3 (Dixi x Michigan lowbush No.1) (Okie, 1997)
St. Cloud (SC)	Half-high	B19 (G65 x Ashworth) x US3 (Dixi x Michigan lowbush No.1) (Finn et al. 1990)

*FO = Fundy open pollinated seedling

Highbush blueberry cultivar Polaris is a cross between B15 (G65 x Ashworth) and Bluetta (Okie, 1997). This cultivar was developed in Minnesota and released by the University of Minnesota in 1996. Polaris is an upright variety which grow the height of 1.2 meters, which is slightly smaller than Bluetta (Okie, 1997). Half-high blueberry cultivar Patriot was developed by crossing (Dixi x Michigan lowbush No. 1) and cultivar Earliblue, by the cooperative efforts between the University of Maine and the U.S.D.A. Plants are upright, relatively open and vigorous. This variety attain the height of 1.5 m in Maine,

whereas in Minnesota it is slightly smaller (Hepler and Draper, 1976; Finn et al. 1990). There are a few breeders who believe that the Patriot is a highbush blueberry cultivar because of its height (1.5 m) (Mackenzie, 1997; Luby, 1991), whereas others believe that it is half-high blueberry cultivar because of the lowbush blueberry as one of the parents (Dixi x Michigan lowbush No. 1). In this study, it has been considered as half-high blueberry cultivar, because if its lowbush blueberry parent. Chippewa half-high blueberry cultivar is a cross between B18A (G65 X Ashworth) and US3 (Dixi x Michigan lowbush #1). This cultivar was developed in Minnesota and was released by the University of Minnesota in 1996 (Okie, 1997). It is also an upright variety which is self-pollinated but benefits from cross pollination. The fruits of this cultivar are firm and bigger in size. Another half-high cultivar St. Cloud is a cross between B19 (G65 x Ashworth) and US3 (Dixi x Michigan lowbush No.1) and was developed and released in Minnesota, United States of America in 1990. They attain the height of 1.3 m and the diameter of 1.7 m (Finn et al. 1990).

In the present study, blueberry clones were selected carefully from the wild, based on their optimal berry colour, optimal plant vigour, berry size, berry yield, and apparently free from insect and disease (Debnath 2009; Debnath 2014). During the collection process, it was taken into consideration that the distance between the two selected plants within a same community was more than 10 m (Debnath, 2009; Debnath, 2014).

All genotypes were grown in the greenhouse of the St. John's Research and Development Centre, Agriculture and Agri-Food Canada, St. John's, Newfoundland and Labrador; in plastic pots (25 cm in diameter and 18 cm deep, equivalent 6 L; EastChem

nc., Mount Pearl, NL, Canada) containing 2 peat: 1 perlite (v/v), under natural light conditions at a maximum PPF of 90 µmol m^{-2} s^{-1} at 20 ± 2ºC, 85% relative humidity. Irrigation was provided regularly and whenever needed. Peters Azalea Neutral Fertilizer 20N-20P-20K (Plant Products Co. Ltd., Brampton, ON, Canada) with concentration of 80g/100L applied three times a year to approximately 200 pots (i.e. 500ml/ 4 inch pot) . It was necessary to maintain chilling requirements, which were met by maintaining the plants at or below 6ºC for 12 weeks.

2.2 DNA extraction

A modified DNA extraction method was used based on Kim et al. (1997). For DNA extraction, young leaves were sampled from plants maintained in the greenhouse. Just after cutting leaves from branches, they were rinsed in a soapy water for few seconds, then dipped in a 70% alcohol for 3-4 seconds and then rinsed with the distilled water. These leaves were combined to make up 200 mg of weight and transferred to vials containing 300µl lysis buffer [100 mM EDTA (pH 8), 200 mM Tris-HCl (pH 8), 1M NaCl and 2% PVP-40 Sigma], and two big (6.35 mm) and few small beads (1.6 – 2.0 mm) (Glemnills Inc., Clifton, NJ, USA) were added.

The vials were then placed in the homogenizer (FastPrep®-24, MP Biomedicals) and homogenizer was allowed to run for 3 times at the speed of 5400 rpm for 45 seconds with the 1-minute interval in between. After crushing the leaves, 18 µl of 1M Dithiothreitol (DTT) (Sigma-Aldrich, USA) and 100 µl of 20% Sodium dodecyl sulphate (SDS) (Sigma-Aldrich, USA) were added to the vials. Then the vials were shaken and kept in the water

bath at 65ºC for 60 minutes. After incubation the vials were centrifuged at 5000 rpm for 1 minute and the lysates were transferred to another empty vials respectively. Here, 500 µl of 7.5 M ammonium acetate (Sigma-Aldrich, USA) solution was added to precipitate the protein and vials were kept on ice for 60 minutes. The vials were centrifuged at 9000 rpm for 9 minutes at 4ºC. Supernatants were transferred to new vials and equal volume of 99% isopropanol was added to precipitate the DNA.

The vials were then placed in -20ºC for 30 minutes and then centrifuged at 9,000 rpm for 9 minutes at 4ºC. Supernatant was discarded and pellets were dried in vacuum dryer for 10 minutes. Pellets were then dissolved by adding 500 µl of 1M NaCl-TE buffer and incubating the vials for 10 - 20 minutes at 65ºC. The vials were cooled down at the room temperature and then 3 µl of RNase [RNase Bovine Pancrease 5mg/ml concentration (Sigma-Aldrich, USA)] was added and incubated for 30 minutes at 37ºC. This step was performed for further purification if DNA from protein. To do that 500 µl of chloroform: isoamyl alcohol (24:1) (VWR, Canada) was added into the vial and mixed slowly. Vials were then centrifuged for 10 minutes at 9,000 rpm and 4ºC. Supernatants were transferred to another vials and Chloroform: isoamyl (VWR, Canada) alcohol step was repeated twice to remove protein.

Finally, 99% isopropanol (VWR, Canada) was added and vials were incubated for 10 minutes at the room temperature. Vials were centrifuged at 9000 rpm for 10 minutes at 18ºC. Supernatants were discarded and the vials were vacuum dried and the DNA pellets were dissolved in 100 µl of TE buffer (Sigma – Aldrich, USA) and incubated at 65ºC for 1

hour. DNA was then stored at 4^{o}C. DNA concentration and purity was measured using JENWAY GenovaNano spectrophotometer (Bibby scientific ltd., UK).

2.3 Primers, PCR and gel electrophoresis

Twelve EST-PCR (Table 3) and twelve microsatellite (ten EST-SSR and two genomic SSR) primer pairs (Table 4) were used for the study. These primer pairs worked well with *Vaccinium* species in other studies. (Boches et al. 2005; Boches et al. 2006; Bell et al. 2008; Dhanaraj et al. 2004). The primer pairs were procured form Integrated DNA Technologies, Inc. (IDT), Coralville, Iowa, USA. PCR reactions were performed using TopTaq DNA polymerase kit from QIAGEN. For PCR, the reaction mixture volume of 25µl was designed previously in the lab (Debnath 2014), which contained 2.5 µl of TopTaq 10X PCR buffer [Qiagen TopTaq contains Tris.Cl, KCl, $(NH_4)_2SO_4$, 1.5 mM $MgCl_2$, stabilizers; pH 8.7 at 20^{o}C], 0.2 µM of dNTP mix (Sigma Chemical Co.), 200µM of each forward and reverse primer, 25ng of DNA, 0.7 units of TopTaq DNA polymerase (Sigma-Aldrich, USA) and PCR grade distilled water (dH_2O) (Sigma-Aldrich, USA).

Table 3: Eleven EST-PCR primer pairs used to analyze relationships among blueberries. Primer names have prefixes which reflect origin of GenBank source sequence (CA= cold acclimated EST library, NA= non-acclimated EST library). For each primer its forward and reverse primers, putative function and annealing temperature (*Ta*) are given.

Primer	Forward/ Reverse primer	Putative function	*Ta* (ºC)
CA16	CCA ATG CCA CAA ACG AGA TT AGC CCC CAA CTT TCG TTC T	Nonidentified*	44
CA54	CCG GTG AAC TTC CAC TTG TT AGA TAC TAC TGG GGG TGG GG	Peroxiredoxin*	56
CA175	GAC AGA TTG CGT AAC CCG TAA CCA ATC CGC TTT GTC TGT TT	Nonidentified	53
CA227	TGG AGA CTG GAG TGA TGC AA TTT GCA AGA ACC ATG CTG AG	Nonidentified	52
CA287	AGG GCT TTC CCT CAA TCA CT CCT TGT TGT TCC TTC CTT CG	Late embryogenesis abundant*	58
CA1029	GAA GTT TTC CGT TCT CTG CAA CTG CAG CTA GGA CCG AAG AG	Lipid transferase protein*	52
CA1105	TGG TGC TTT CAT CCT GCT AA GCT TGC TTC TTG GGT GAC TC	Hypothetical protein*	50
CA1423	TCA TAG CCA ATA CAC TCG AAC C GCC CCA CCT TTA GCA AAC TC	Putative early light induced protein*	46
CA1590	AAC CCA GCA CCT CCT TTC TT CTC TGT TGC TGG CGT TGT GT	Hypothetical protein*	54
CA1785	CAC CAC CAC TGT CGT ACA CC GCA TGA GCC GAA CAT AAT CA	Argonaute/ Zwille-like protein 1*	60
NA27	CGC TCG CTC CAT TGT TTC TAT GCA TGA AGC TTG CCG TA	Actin-binding protein*	60
NA1068	CCG GAA GGA ATG GTG ACT AA ATC CCC ACA CAA ACA AAA GC	Nonidentified	60

*****Boches et al. (2006)

Table 4: Seven EST-SSR and two genomic SSR primer pairs used to analyze relationships among blueberries. Primer names have prefixes which reflect origin of GenBank source sequence (CA= cold acclimated EST library, NA= non-acclimated EST library). For each primer its forward and reverse primers, repeat motif (RM) and annealing temperature (*Ta*) are given.

Primer	Forward/ Reverse primer	RM	*Ta* (°C)
EST-PCR primer pairs			
CA23f	GAG AGG GTT TCG AGG AGG AG GTT TAG AAA CGG GAC TGT GAG ACG	(AGA)6	62
CA112	TCC ACC CAC TTC ACA GTT CA GTT TAT TGG GAG GGA ATT GGA AAC	(AG)7	56
CA169	TAG TGG AGG GTT TTG CTT GG GTT TAT CGA AGC GAA GGT CAA AGA	(GAT)4	62
CA236	GTT AAG CTT TTA GAT GAG TTG ATG G GTT TAA CCA GTC CCA GAC CCA AAT	(TG)17	48
CA421	TCA AAT TCA AAG CTC AAA ATC AA GTT TAA GGA TGA TCC CGA AGC TCT	(CT)25	60
CA483	GTC TTC CTC AGG TTC GGT TG GAA CCG CTC CGA AGA CAG	(TC)8	61
NA741	GCC GTC GCC TAG TTG TTG GTT TGA TTT TGG GGG TTA AGT TTG C	(TC)9	58
NA800	CAA TCC ATT CCA AGC ATG TG GTT TCC CTA GAC CAG TGC CAC TTA	(TC)13	54
NA961	TCA GAC ATG ATT GGG GAG GT GTT TGG AAT AAT AGA GGC GGT GGA	(TAC)5	53
NA1040	GCA ACT CCC AGA CTT TCT CC GTT TAG TCA GCA GGG TGC ACA A	(TC)11	47
Genomic SSR primer pairs			
VCC_I2	AGG CGT TTT TGA GGC TAA CA TAA AAG TTC GGC TCG TTT GC	(CT)14	62
VCC_S10	ATT TGG TGT GAA ACC CCT GA GTT TGC GGC TAT ATC CGT GTT TGT	(CT)22	61

Reaction mixtures were then amplified in PTC-100 Programmable Thermal Controller (MJ Research Inc., Watertown, MA, USA) using an initial "hot start" of 94^{o}C for 10 minutes denaturation step, followed by 40 cycles of 40s denaturation step at 94^{o}C, 70s annealing step at the approximate annealing temperature (given in the Table 3 and 4), and 120s extension step at 72^{o}C. The reaction was terminated with the final extension of 10 minutes at 72^{o}C before holding the samples at 4^{o}C for analysis.

The annealing temperature of each primer pair were calculated as (Old 1980)

Ta= 4^{o}C (#G's and C's in the primer) + 2^{o}C (#A's and T's in the primer)

However, the gradient PCR technique was used to check the actual annealing temperature, which can give good clear band.

Amplified fragments along with low ranger (100bp) and mid ranger (1kb) DNA Ladders (Norgen Biotek, Thorold, Canada), were separated using electrophoresis on 1.4% agarose gel. The gels were stained with the GelRed (Biotium, 3159 Corporate Place Hayward, CA, USA) dye solution following the manufacturer instruction. Images were captured using UV gel imaging system (INGENIUS-3, Syngene, Beacon House, Cambridge, UK) by using the settings for GelRed dye (Biotium, Inc., Hayward, CA). The size of the fragments was calculated by GeneTools software (Syngene) on the basis of comparison with the standard size marker. The experiment with each marker pair was repeated three times for each sample. The only loci amplified in each instance (i.e. all three trials) with reproducibility were scored and included in the analysis. The loci showing monomorphic band pattern were not included in the analysis. Bands of similar molecular

weight and migration distance across individuals were assumed to be homologous (Adams and Rieseberg 1998).

2.4 Data analysis

Considering the polyploid nature of the blueberries, their co-dominant scoring was complicated for the heterozygous samples (Esselink et al. 2003). In polyploid species, multiple genomes can amplify the same product, which makes it difficult to differentiate the genotypes based on the intensity of the band. Therefore, it is advisable to score polyploidy banding pattern as presence (1) or absence (0) of the band in a particular genotype (Gil-Ariza et al. 2009; Esselink et al. 2003; Horvath et al. 2011).

2.4.1 Banding pattern analysis

The ability of the most informative primer pair to distinguish between clones was assessed by calculating their resolving power (Rp), which has shown to be strongly correlated with their ability to distinguish between genotypes (Prevost and Wilkinson 1999). The Rp value was calculated using the following equation (Gilbert et al. 1999):

$$Rp = \sum [1 - (2 \times |0.5-p|)] \quad \text{(Equation 1)}$$

Where, p is the proportion of the 63 genotypes (56 wild blueberry clones from four Canadian provinces, six cultivars and one selection).

2.4.2 Genetic diversity analysis

PowerMarker v3.25 (Liu and Muse 2005) software was used to calculate the diversity measurements at each locus, including major allele frequency (Major AF), the number of alleles per locus (N_A), expected heterozygosity (H_E), observed heterozygosity (H_O) and polymorphic information content (PIC) of both type of markers.

For each marker type, the total number of alleles (N_A) was calculated using equation 2 (Pejic et al. 1998)

$$\mathbf{N_A = \sum Ne} \qquad \text{(Equation 2)}$$

Where, Ne is effective number of alleles per locus and was calculated using equation 3 (Morgante et al. 1994).

$$\mathbf{Ne = 1/\sum P_i^2} \qquad \text{(Equation 3)}$$

Where *Pi* is the allele frequency for the *i*th allele.

The expected heterozygosity (H_E) of the polymorphic locus for each genetic marker was computed with the following equation 4 (Belaj et al. 2003):

$$\mathbf{H_E = 1 - \sum Pi^2} \qquad \text{(Equation 4)}$$

Where *Pi* is the allele frequency for the *i*th allele.

The measure of the amount of heterozygosity across loci can be applied as a general indicator of the amount of genetic variability (Lynch 1990). The observed heterozygosity

was obtained through GenAlEx 6.5 (Peakall and Smouse 2012) software which divides number of heterozygotes with total number of samples. Heterozygosity levels are linked directly to reduced population fitness via inbreeding depression. This leads to the expectation that levels of heterozygosity and fitness, at the population level, will be correlate (Reed and Frankham 2003). Finite populations lose genetic variation as a consequence of genetic drift and at the same time become inbred. This loss of heterozygosity can be described by the inbreeding coefficient, F, which is related to the amount of genetic variation present, in the absence of mutation and selection, and the Inbreeding coefficient (*F*) was calculated as

$$\boldsymbol{F = 1 - (H_O / H_E)} \qquad \text{(Equation 5)}$$

The PIC was calculated with the following equation 6 (Botstein et al. 1980):

$$\mathbf{PIC} = \sum_{i=1}^{n} \boldsymbol{Pi^2} - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \boldsymbol{2Pi^2\, Pj^2} \qquad \text{(Equation 6)}$$

Where *Pi* and *Pj* are the allele frequency for the *i*th and *j*th alleles respectively.

Genetic diversity for each locus was calculated using Shannon's index (I) equation 7 (Shannon and Weaver 1999). Shannon's index is different than before mentioned parameters as it does not require knowledge of allele frequencies. This makes it an accurate measure of diversity in polyploidy species, due to the difficulty in distinguishing the copy number of individual alleles (Boches et al. 2006)

$$\boldsymbol{I = \sum Pi\, log_2 Pi} \qquad \text{(Equation 7)}$$

Where, *Pi* is the frequency of the presence or absence of the *ith* allele.

2.4.2.1 Comparison of genetic diversity among groups based on their geographic locations using EST-PCR and microsatellite markers

Parameters such as number of alleles (N_A), expected heterozygosity (H_E) and polymorphic information content (PIC) are reported separately for the different grouping made according to the geographic collection site/ origin of the genotypes where they collected from. The genotypes were divided into five groups. Group 1, 2 and 3 included wild clones collected from Newfoundland and Labrador (designated as NL), Prince Edward Island (designated as PE) and Quebec (designated as QC), respectively. Group 4, included half-high and highbush blueberry cultivars (designated as HHCU) and a group 5 included two wild clones from New Brunswick (designated as NB), two lowbush blueberry cultivars (Fundy (FU) and Brunswick (BR)) and a selection. This group was designated as "NB and LBCU". Parameters were calculated separately for EST-PCR and microsatellite data.

2.4.3 Cluster analysis

Genetic distance matrix was computed using PowerMarker (Version 3.25) (Liu and Muse 2005) with the equation 8 proportion of shared alleles distance (D_{sa}):

$$D_{sa} = \frac{1}{m} \sum_{j=1}^{m} \sum_{i=1}^{aj} \min\left(p_{ij}, q_{ij}\right) \qquad \text{(Equation 8)}$$

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. Only very clear and distinct bands were considered as 'true' bands/ picks. Presence (1) or absence (0) of each allele which was treated as a separated locus and the matrices were created. Separate matrices

were created for both type of markers and for both the matrices combined. The bootstrap option of PowerMarker was used to create 1000 dendrograms using neighbor joining (NJ) and unweighted pair group methods (UPGMA). These 1000 dendograms were used to create a strict consensus tree, which gives an idea about common pattern. In these analyses both dendograms were rooted at midpoint (Boches et al. 2006).

2.4.4 Principle coordinate analysis (PCoA)

The binary data were also subjected to principle coordinate analysis (PCoA) to partition the variance using GenAlEx 6.501 (Peakall and Smouse, 2012) and the first two components were plotted into two-dimensional plots. PCoA was done using marker data of EST-PCR, SSR and combined data.

2.4.5 Structure analysis

In order to infer the population structure of the entire set of genotypes without regard to the pre-existing subspecies classification or geographical information, the model-based program, STRUCTURE version 2.3.4 (Pritchard et al. 2000) was used. The software provides Bayesian approach to infer population structure by using our markers EST-PCR, EST-SSR and combined marker datasets to identify the number of clusters (*K)* to which the program then assigns each individual genotype (Falush et al. 2007; Falush et al. 2003; Pritchard et al. 2000).

In this model, a number of populations (*K*) are assumed to be present, each of which characterized by a set of allele frequencies at each locus. Individuals in the sample are

assigned to population (clusters), or jointly to more populations if their genotypes indicate that they are admixed. All loci are assumed to be independent, and each *K* population is assumed to follow Hardy - Weinberg equilibrium. Each run was performed with 100,000 burn in iterations and 100,000 subsequent Monte Carlo Markov Chain (MCMC) runs with *K* ranging from 1 to 9. To check the consistency of the results between the runs with the same *K*, five replicates were run for each assumed *K* value.

Optimal value of *K* was identified using two methods, one as described by Pritchard (Pritchard et al. 2000) using L(*K*) (Equation 10) and the other developed to determine delta K (Δ*K*) (Equation 9) (Evanno et al. 2005). In case of first method, when posterior probabilities of *K* is approaching to plateaus (or continues increasing slightly) and has high variance between runs (Rosenberg et al. 2001) then that *K* is considered true value of *K*. In second method, data calculated based on the second order rate of change of the likelihood the Δ*K* (Evanno et al. 2005) shows a clear peak at the true value of *K* in graph. Evanno et al. (2005) showed that while *L*(*K*), the (ad hoc) estimate for the number of groups, given by structure, often does not correspond to the real number. Δ*K*, another ad hoc quantity based on the second order rate of change of the likelihood function with respect to *K*, has a mode at the true *K* for most of the situations investigated. When the mode of Δ*K* at the true *K* was absent, it was either because sample size and marker number was small, leading to an absence of signal, or visual inspection of the values of *L* (*K*) would have identified runs of the MCMC with outlying values for *L* (*K*).

$$\Delta K = \text{mean}(|\text{L}''(\text{K})|) \,/\, \text{sd}(\text{L}(\text{K})) \quad \text{(Equation 9)}$$

Where, $L''(K) = L'(K)_n - L'(K)_{n-1}$

Where, $L'(K) = L(K)_n - L(K)_{n-1}$

Where, L (K) = an average of 5 values of Ln P (D). (Equation 10)

The results obtained from the STUCTURE software package were placed in a website (http://taylor0.biology.ucla.edu/structureHarvester/#) based STRUCTURE HARVESTER program (Earl and Vonholdt 2012). This program, when provided with the result file, runs a script and after calculations presents with the results and graphs of Ln (*K*) and ΔK.

2.4.6 Analysis of molecular variance (AMOVA)

Arlequin version 3.5.1.2 (Excoffier et al. 2005) was used to conduct the analysis of molecular variance (AMOVA) among and within populations and subpopulations. Pairwise difference method was used to calculate the difference with 1000 permutations. The p-value was set at 0.05.

AMOVA analysis was performed among five groups based on geographic locations/ place of origin of the genotypes; where groups 1, 2 and 3 are wild clones from Newfoundland and Labrador, Prince Edward Island and Quebec, respectively. Group 4 represented two New Brunswick clones, two lowbush cultivars Fundy and Brunswick and a selection Fundy open pollinated seedling, because all of these genotypes originated from the same province. Group 5, represented all the half-high and highbush blueberry cultivars. Pairwise difference method was used to calculate the genetic distance and 1000 permutations were performed. The p-value was set at 0.05 significance level.

3 Results

3.1 DNA extraction

DNA purity was good and 260 nm: 280 nm ratios were within 1.8 to 1.9, with the average ratio of 1.87; 260 nm: 230 nm ratios were within 2 to 2.2, with the average ratio of 2.09. Some amounts of the DNA samples were diluted to the working concentration of 12.5 ng/ µl and stored at -20^{o}C for further use.

3.2 Gel electrophoresis

Out of 10 EST-SSR primer pairs, CA23f was found to be monomorphic and CA112 and CA169 did not give reproducible results. Out of 12 EST-PCR primer pairs, CA227 did not give reproducible results. EST-PCR primer CA227 and EST-SSR primer pairs CA23f, CA112 and CA169 were not considered in the further calculations and analysis. Seven EST-SSR, two genomic SSR and 11 EST-PCR primer pairs showed polymorphic nature and reproducibility, and were considered for further analysis.

3.3 Banding pattern analysis

Tables 5 and 6 show the number of polymorphic bands, size range of the bands and the resolving power (Rp) of EST-PCR and SSR markers, respectively. The EST-PCR markers showed polymorphic bands ranging from two for NA27 and CA1590 to 16 for CA16, with 87 total bands and an average of 7.9 bands per primer. The number of polymorphic bands, in EST-SSR markers, ranged from two for CA236 to 16 for NA961,

with 39 total bands and 5.6 bands per primer. Two genomic SSRs, VCC_I2 and VCC_S10 contributed 14 bands in total with an average of 7 bands per primer

Table 5: Identity, polymorphism and resolving power for EST-PCR primer pairs tested on 63 blueberry genotypes.

Primer	Polymorphic bands (no.)	Size range (bp)	Resolving power (Rp)
CA16	16	350 bp - 2600 bp	12.1
CA54	14	250 bp - 3250 bp	7.6
CA175	4	850 bp - 1950 bp	1.7
CA287	6	550 bp - 1650 bp	2.2
CA1029	3	150 bp - 100 bp	1.7
CA1105	6	250 bp - 1600 bp	3.2
CA1423	8	450 bp - 850 bp	5.9
CA1590	2	1500 bp - 2000 bp	1.7
CA1785	14	450 bp - 2950 bp	7.5
NA27	2	700 bp - 1100 bp	1.2
NA1068	12	250 bp - 2650 bp	6.0

Table 6: Identity, polymorphism and resolving power for EST-SSR primer pairs tested on 63 blueberry genotypes.

Primer	Polymorphic bands (no.)	Size range (bp)	Resolving power (Rp)
CA236	2	300 bp – 500 bp	0.3
CA421	6	400 bp - 1100 bp	2.6
CA483	5	150 bp - 650 bp	2.7
NA741	3	200 bp - 2500 bp	2.1
NA800	3	200 bp - 1300 bp	1.4
NA961	16	200 bp - 3250 bp	10.1
NA1040	4	200 bp - 1750 bp	1.8
VCC_I2	10	100 bp - 900 bp	2.8
VCC_S10	4	150 bp - 350 bp	3.2

The Rp values for the EST-PCR markers ranged from 1.2 for NA27 to 12.1 for CA16, while for the EST-SSR markers these values ranged from 0.3 for CA236 to 10.1 for NA961. The Rp value of VCC_I2 was 2.8 and of VCC_S10 was 3.2.

Figures 2, 3 and 4 represent the gel pictures of banding pattern obtained by CA16 EST-PCR marker, genomic SSR marker VCC_I2 and EST-SSR marker CA421, respectively. It can be seen that each marker is polymorphic in nature.

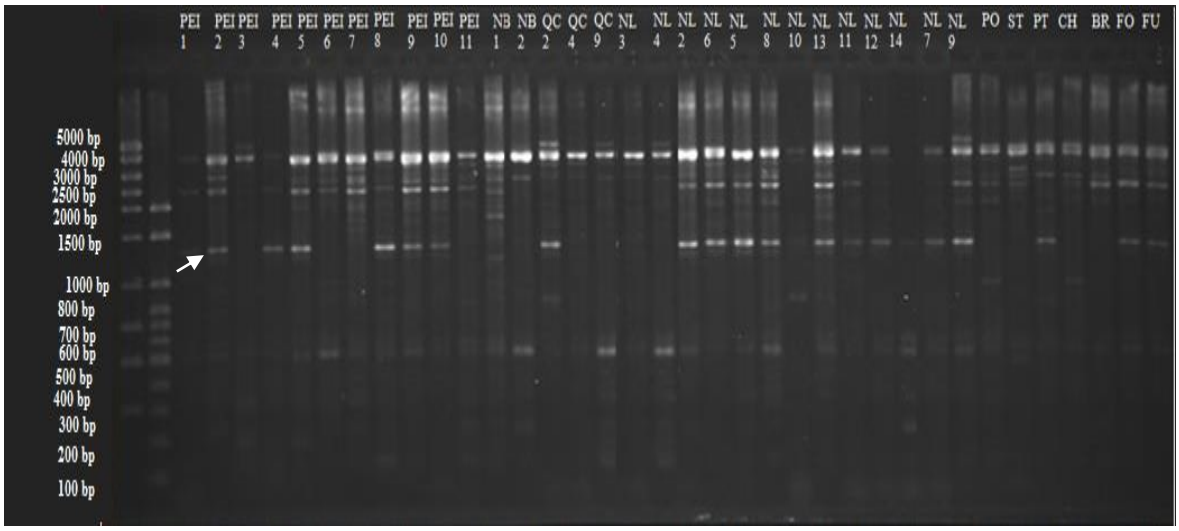

Figure 2: EST-PCR primer (CA16) banding pattern for 36 blueberry genotypes including 13 Newfoundland and Labrador (NL2-NL14), two New Brunswick (NB1-NB2), three Quebec (QC2, QC4 and QC9) and 11 Prince Edward Island (PE1-11) lowbush blueberry wild clones; one highbush blueberry cultivar Polaris (PO); three half-high blueberry cultivars Chippewa (CH), St. Cloud (SC) and Patriot (PT); two lowbush blueberry cultivars Brunswick (BR) and Fundy (FU); and one Fundy open pollinated seedling selection (FO). Standard molecular ladders 1 kb ladder (left) and 100 bp ladder (right).

In Figure 2, pointed band is considered to be a polymorphic band as it is not present in all the genotypes and present in some. For example it is present in PE2, PE4, PE5, PE8-PE10, QC2, NL2, NL6, NL5, NL8, NL13, NL11, NL12, NL7, NL9, PT, FO and FU, whereas it is not present in other genotypes. This band was scored present (1) in the analysis for the genotypes it is present in.

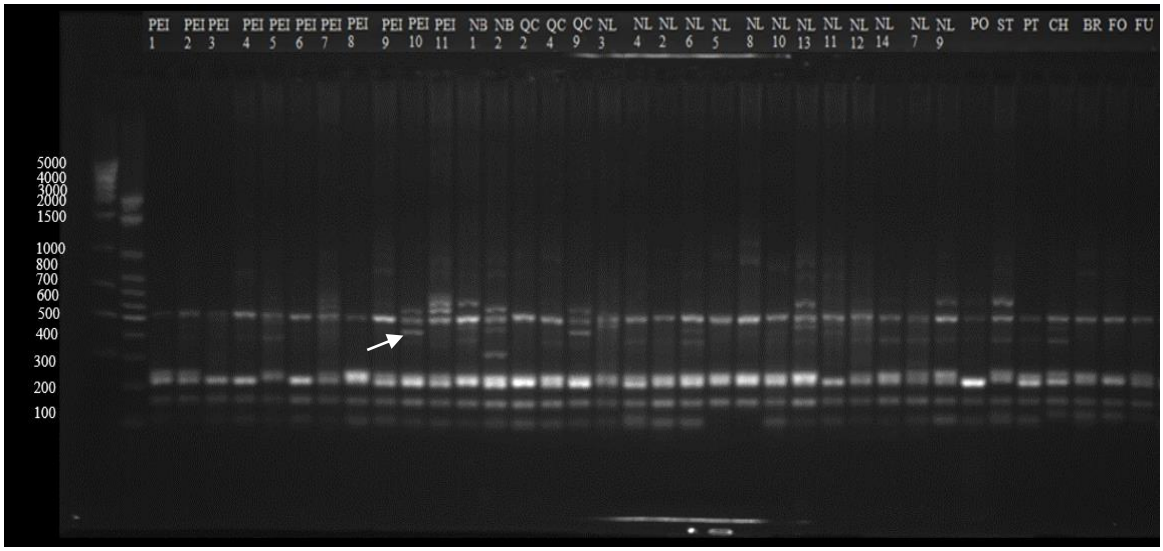

Figure 3: Genomic SSR primer (VCC_I2) banding pattern for 36 blueberry genotypes including 13 Newfoundland and Labrador (NL2-NL14), two New Brunswick (NB1-NB2), three Quebec (QC2, QC4 and QC9) and 11 Prince Edward Island (PE1-11) lowbush blueberry wild clones; one highbush blueberry cultivar Polaris (PO); three half-high blueberry cultivars Chippewa (CH), St. Cloud (SC) and Patriot (PT); two lowbush blueberry cultivars Brunswick (BR) and Fundy (FU); and one Fundy open pollinated seedling selection (FO). Standard molecular ladders 1 kb ladder (left) and 100 bp ladder (right).

In Figure 3, pointed band is a polymorphic band and is not present in many genotypes except few. As seen in the figure it is present in PE9, QC9 and NL6, whereas other genotypes did not have this band.

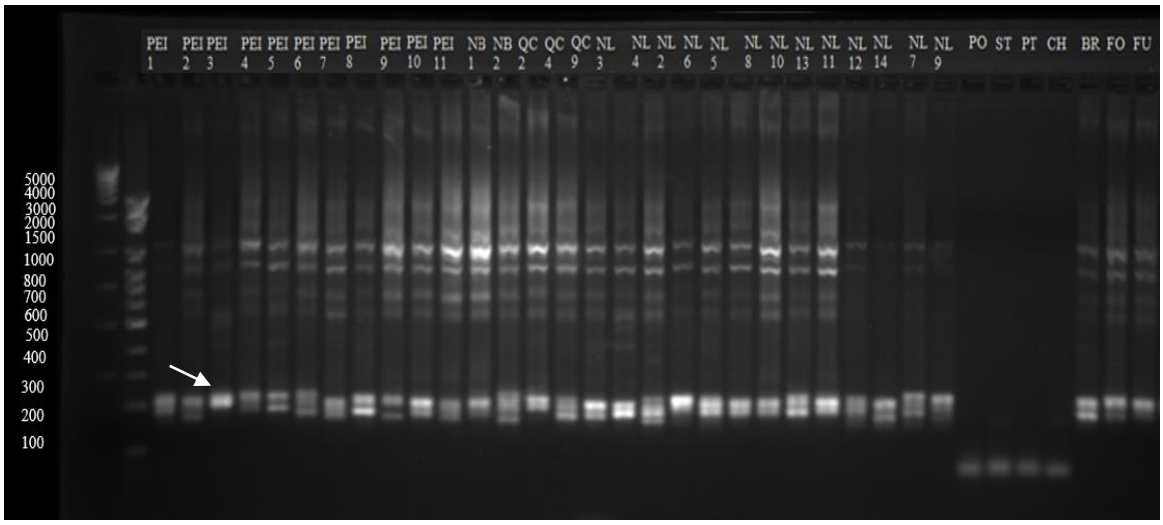

Figure 4: EST-SSR primer (CA421) banding pattern for 36 blueberry genotypes including 13 Newfoundland and Labrador (NL2-NL14), two New Brunswick (NB1-NB2), three Quebec (QC2, QC4 and QC9) and 11 Prince Edward Island (PE1-11) lowbush blueberry wild clones; one highbush blueberry cultivar Polaris (PO); three half-high blueberry cultivars Chippewa (CH), St. Cloud (SC) and Patriot (PT); two lowbush blueberry cultivars Brunswick (BR) and Fundy (FU); and one Fundy open pollinated seedling selection (FO). Standard molecular ladders 1 kb ladder (left) and 100 bp ladder (right).

Figure 4 shows few stutter bands, but it still shows some polymorphic bands. Pointed band is very dark and distinct. It is not present in all the genotypes, however none of the bands are present in highbush and half-high blueberries. Thus the pointed band is considered polymorphic and its presence in the genotypes will be notes as 1 in the matrix.

3.4 Diversity parameters

3.4.1 Diversity parameters based on EST-PCR markers

For EST-PCR markers, the total number of alleles (N_A) was 272, with the range of $N_A = 4$ to $N_A = 52$ alleles, for the markers CA1029 and CA16, respectively (Table 7). The average allele number for 11 EST-PCR markers was $N_A = 23$. The major allele frequency for the EST-PCR markers ranged from 0.06 for CA1785 to 0.64 for NA27.

The values of expected heterozygosity (H_E) for EST-PCR markers ranged from $H_E = 0.51$ for NA27 primer to $H_E = 0.98$ for CA1785, while the values of observed heterozygosity (*Ho*) ranged from *Ho* = 0.17 for CA1029 to *Ho* = 0.54 for CA175. The inbreeding coefficient (*F*) value for the EST-PCR markers ranged from *F* = 0.33 for CA175 to *F* = 0.74 for CA1029, with the average *F* of *F* = 0.60.

Another important parameter, Shannon index (*I*), produced wide range of values for EST-PCR markers, which ranged from *I* = 0.88 for NA27 to *I* = 4.01 for CA16, with the average *I* of *I* = 2.17. Polymorphic information content (PIC) values, which are very important for polymorphism information, ranged from 0.44 for NA27 to 0.98 for CA1785, with the average PIC value of 0.80.

Table 7: Allelic diversity among blueberry genotypes for 11 EST-PCR primer pairs. Major allele frequency (Major.AF), allele number (N_A), expected (H_E) and observed (H_O) heterozygosity, inbreeding coefficient (F), shannon's index (I) and polymorphic information content (PIC) are given for all 63 genotypes.

Marker	**Major.AF**	**N_A**	***H_E***	***H_O***	***F***	***I***	**PIC**
CA16	0.063	52	0.976	0.221	0.774	4.099	0.976
CA54	0.127	44	0.962	0.330	0.657	3.783	0.961
CA175	0.365	10	0.799	0.536	0.329	1.835	0.778
CA287	0.460	14	0.756	0.266	0.648	2.029	0.741
CA1029	0.444	4	0.649	0.171	0.737	1.850	0.583
CA1105	0.175	14	0.883	0.252	0.715	2.500	0.872
CA1423	0.190	19	0.906	0.419	0.538	2.486	0.900
CA1590	0.333	4	0.707	0.358	0.494	1.081	0.650
CA1785	0.063	51	0.977	0.326	0.666	4.033	0.976
NA27	0.635	4	0.511	0.279	0.454	0.879	0.441
NA1068	0.190	33	0.932	0.398	0.573	4.099	0.929
Total	-	272	-	-	-	-	-
Mean	-	23	0.824	0.337	0.599	2.619	0.801

3.4.2 Diversity parameters based on microsatellite markers

SSR markers produced 164 alleles. The allele numbers (N_A) varied from $N_A = 3$ to $N_A = 54$, for the markers CA236 and NA1040, respectively (Table 8). The average allele number was $N_A = 18.22$. The major allele frequency for the SSR markers ranged from 0.09 to 0.83 for the markers VCC_S10 and CA236, respectively.

The values of H_E for SSR ranged from $H_E = 0.30$ for CA236 to $H_E = 0.98$ for NA1040, while the values of *Ho* ranged from *Ho* = 0.14 for CA483 to *Ho* = 0.44 for NA1040. The *F* values, which were determined from the values of H_E and H_O, ranged from $F = 0.08$ for CA236 to $F = 0.84$ for CA438, with the average value of $F = 0.55$.

I values for SSR markers ranged from $I = 0.58$ for CA236 to $I = 3.97$ for NA1040, with the average value of $I = 2.02$. PIC value is essential for detecting marker's polymorphism, and its values for SSR markers ranged from 0.28 for CA236 to 0.98 for NA1040, with the average value of 0.77.

Table 8: Allelic diversity among blueberry genotypes for seven EST-SSR and two genomic SSR primer pairs. Major allele frequency (Major.AF), allele number (N_A), expected (H_E) and observed (*Ho*) heterozygosity, inbreeding coefficient (*F*), shannon's index (*I*) and polymorphic information content (PIC) are given for all 63 genotypes.

Marker	**Major.AF**	**N_A**	***H_E***	***H_O***	***F***	***I***	**PIC**
CA236	0.825	3	0.303	0.280	0.076	0.580	0.282
CA421	0.302	10	0.811	0.344	0.576	1.768	0.788
CA483	0.190	16	0.894	0.144	0.839	2.263	0.886
CA741	0.381	6	0.735	0.322	0.562	2.082	0.692
NA800	0.238	15	0.865	0.442	0.489	2.545	0.851
NA961	0.413	7	0.723	0.320	0.557	1.487	0.682
NA1040	0.063	54	0.977	0.444	0.546	3.972	0.977
VCC_I2	0.365	13	0.786	0.395	0.497	1.976	0.762
VCC_S10	0.095	40	0.964	0.196	0.797	3.615	0.963
Total	-	164	-	-	-	-	-
Mean	-	18.22	0.784	0.321	0.549	2.027	0.765

3.4.3 Comparison of genetic diversity parameters between EST-PCR and microsatellite (SSR) markers

Table 9 illustrates the differences between EST-PCR and SSR markers by taking into consideration some of the most important diversity parameters. As seen in the Table, 11 EST-PCR markers produced 87 polymorphic bands while nine microsatellite markers produced only 53 polymorphic bands.

Table 9: Comparison of information on genetic diversity parameters, obtained with EST-PCR and SSR markers in 56 wild lowbush blueberry clones, six cultivars and one blueberry selection.

	Marker system	
Property	**EST-PCR**	**SSR**
Number of assay units	11	9
Number of polymorphic bands	87	53
Average number of polymorphic bands	9.75	7.18
Average of resolving power	4.6	2.6
Average of polymorphism information content	0.801	0.765
Average expected heterozygosity	0.824	0.784
Average observed heterozygosity	0.337	0.321
Average value of Shannon's index	2.619	2.027

EST-PCR markers had higher values for most of the parameters. For instance, the number of average polymorphic bands was higher in the EST-PCR markers, 9.75, than in the microsatellite markers, 7.81. Also, the mean resolving power (Rp) of the EST-PCR markers was higher, 4.6, compared to, 2.6, in the microsatellite markers.

PIC provides an estimate of the discriminatory power of a locus by taking into account the number of alleles that are expressed and the relative frequencies of those alleles. This was used to assess the informativeness level of each marker (high, PIC > 0.5; moderate, 0.5 > PIC > 0.25; low, PIC < 0.25; Botstein et al. 1980). Average PIC is higher for the EST-PCR markers, 0.80, than for the microsatellite markers, 0.77.

The same was noticed for average H_E and *Ho*, where average H_E for EST-PCR markers was found to be $H_E = 0.82$, while for the microsatellites it was $H_E = 0.78$, which was lower than the former. Average *Ho* for EST-PCR markers was found to be *Ho* = 0.34, which is higher than *Ho* of SSR markers, *Ho* = 0.32. Shannon index (*I*) was also higher for EST-PCR markers, $I = 2.62$, than for the microsatellite markers, $I = 2.03$.

3.4.4 Comparison of genetic diversity among groups based on their geographic locations using EST-PCR and microsatellite markers

While studying diversity, it becomes important to do the diversity analysis within the groups, and check which groups are more diverse than the others. The groups here were created based on geographic collection sites of the genotypes. The result of the differences between these groups is given in Tables 10 and 11, for the EST-PCR and the microsatellite markers, respectively.

In Table 10 CA16 showed highest N_A for NL, PE, QC and HHCU groups: NL group had highest allele number, N_A = 36, followed by PE group, N_A = 11; QC group, N_A = 5; and HHCU group, N_A = 5. CA1785 and NA1068, produced the highest allele number, N_A = 5, for the NB and LBCU group.

CA16 showed the highest expected heterozygosity (H_E) for NL group, H_E = 0.97; followed by QC group, H_E = 0.78; and HHCU group, H_E = 0.63, while CA54 and CA1785 produced the highest H_E for PE group, H_E = 0.89. For the NB and LBCU group, CA1785 and NA1068 produced the highest H_E = 0.80.

The marker with the highest PIC values also varied among groups, where CA16 produced the highest PIC values for the NL group, 0.97; the QC group, 0.74; and the HHCU group, 0.55. At the same instance, CA54 and CA1785 produced the highest PIC value, 0.88, for the PE group, and markers CA1785 and NA1068 produced highest PIC value, 0.77, for the NB and LBCU group.

Table 10: Allelic diversity among blueberry genotypes for 11 expressed sequence tag – polymerase chain reaction (EST-PCR primer pairs). Number of alleles (N_A), expected heterozygosity (H_E) and polymorphic information content (PIC) are reported separately for the wild clones of Newfoundland and Labrador (NL), Prince Edward Island (PE) and Quebec (QC). These parameters were also reported for half-high and highbush cultivars (HHCU) group and New Brunswick clones and Lowbush cultivars (NB and LBCU) group, which contained two NB clones, two lowbush blueberry cultivars and a selection.

Marker	NL			PE			QC			HHCU			NB and LBCU		
	N_A	H_E	PIC	N_A	H_E	PIC	N_A	H_E	PIC	N_A	H_E	PIC	N_A	H_E	PIC
CA16	36	0.972	0.971	11	0.793	0.763	5	0.778	0.744	3	0.625	0.555	2	0.320	0.269
CA54	26	0.926	0.923	10	0.893	0.883	4	0.722	0.671	2	0.375	0.305	3	0.560	0.499
CA175	7	0.786	0.755	4	0.727	0.678	1	0.000	0.000	1	0.000	0.000	3	0.560	0.499
CA287	8	0.687	0.659	6	0.645	0.579	1	0.000	0.000	2	0.375	0.305	2	0.320	0.269
CA1029	3	0.536	0.474	6	0.645	0.572	2	0.444	0.346	1	0.000	0.000	1	0.000	0.000
CA1105	10	0.804	0.778	5	0.694	0.637	3	0.611	0.535	1	0.000	0.000	2	0.480	0.365
CA1423	11	0.837	0.818	5	0.694	0.649	3	0.500	0.449	2	0.375	0.305	2	0.480	0.365
CA1590	3	0.631	0.560	3	0.595	0.526	2	0.278	0.239	1	0.000	0.000	1	0.000	0.000
CA1785	32	0.966	0.965	10	0.893	0.883	4	0.722	0.671	1	0.000	0.000	5	0.800	0.768
NA27	4	0.577	0.510	2	0.000	0.000	2	0.444	0.346	1	0.000	0.000	1	0.000	0.000
NA1068	18	0.859	0.850	11	0.843	0.825	1	0.000	0.000	1	0.000	0.000	5	0.800	0.768
Total	158	-	-	58	-	-	28	-	-	16			27		
Mean	14	0.780	0.751	5	0.765	0.636	3	0.409	0.364	1.4	0.159	0.134	2	0.393	0.346

Table 11: Allelic diversity among blueberry genotypes for seven expressed sequence tag – simple sequence repeats (EST-SSR) and two genomic SSR primer pairs. Number of alleles (N_A), expected heterozygosity (H_E) and polymorphic information content (PIC) are reported separately for the wild clones of Newfoundland and Labrador (NL), Prince Edward Island (PE) and Quebec (QC). These parameters were also reported separately for half-high and highbush cultivars (HHCU) group and New Brunswick clones and Lowbush cultivars (NB and LBCU) group which contained two NB clones, two lowbush blueberry cultivars and a selection.

Marker	NL			PE			QC			HHCU			NB and LBCU		
	N_A	H_E	PIC	N_A	H_E	PIC	N_A	H_E	PIC	N_A	H_E	PIC	N_A	H_E	PIC
CA236	3	0.279	0.252	1	0.000	0.000	1	0.000	0.000	1	0.000	0.000	1	0.000	0.000
CA421	7	0.720	0.681	2	0.165	0.152	1	0.000	0.000	1	0.000	0.000	1	0.000	0.000
CA483	11	0.811	0.788	5	0.793	0.761	1	0.000	0.000	1	0.000	0.000	3	0.560	0.499
NA741	3	0.659	0.585	3	0.628	0.551	1	0.000	0.000	2	0.375	0.305	4	0.720	0.672
NA800	7	0.698	0.665	1	0.000	0.000	1	0.000	0.000	1	0.000	0.000	1	0.000	0.000
NA961	32	0.960	0.958	9	0.876	0.863	6	0.833	0.810	2	0.375	0.305	5	0.800	0.768
NA1040	8	0.711	0.685	3	0.579	0.486	3	0.611	0.535	3	0.625	0.555	3	0.640	0.563
VCC_I2	21	0.938	0.934	10	0.893	0.883	5	0.778	0.744	3	0.625	0.555	5	0.800	0.768
VCC_S10	11	0.855	0.839	4	0.612	0.555	3	0.611	0.535	2	0.375	0.305	3	0.560	0.499
Total	103	-	-	38	-	-	22	-	-	16	-	-	26	-	-
Mean	11.4	0.737	0.710	4	0.505	0.472	2.4	0.315	0.292	1.8	0.264	0.225	2.9	0.453	0.419

Among the five groups, NL group produced the highest values for all the parameters followed by the PE, QC, HHCU and NB and LBCU groups. Three markers (CA175, CA287 and NA1068), in the QC group, produced only one allele and null H_E and PIC values. Similarly, in the HHCU and the NB and LBCU groups, respectively, six (CA175, CA1029, CA1105, CA1590, CA1785, NA27 and NA1068) and three (CA1029, CA1590 and NA27) markers produced only one allele and null H_E and PIC.

In the SSR marker analysis (Table 11), NA961 produced the highest allele number, N_A = 32, for the NL group, followed by the QC group, N_A = 6; and the NB and LBCU group, N_A = 5. NA961 did not produce the highest number of alleles in the other remaining two groups. For example, only VCC_I2 produced the highest alleles, N_A = 10, for the PE group, and both NA1040 and VCC_I2 produced the highest alleles, N_A = 3, for HHCU group. Marker VCC_I2 with the NA961 produced the highest alleles, N_A = 5, for the group the NB and LBCU.

For H_E parameter, markerVCC_I2 produced highest value, H_E = 0.94, in the NL group; and H_E = 0.89, in the PE group. NA961 produced the highest value of H_E = 0.83 in the QC group, whereas markers NA1040 and VCC_I2 produced the highest value of H_E = 0.63 for the HHCU group. Markers NA961 and VCC_I2 produced the value of H_E = 0.80 for the NB and LBCU group.

PIC values also varied among the groups. SSR marker analysis showed that, NA961 produced the highest PIC value, 0.96, for the NL group and, 0.81, for the QC group. VCC_I2 produced the highest PIC value of 0.88 for the PE group, while markers NA1040

and VCC_I2 produced the highest PIC value of 0.56 for the HHCU group, and both NA961 and VCC_I2 produced the highest PIC value of 0.77 for the NB and LBCU group.

NL group produced the highest values for all the parameters followed by the PE, QC, HHCU and NB and LBCU groups. Two markers (CA236 and NA800) produced only one allele and null H_E and PIC values in PE group. Five markers (CA236, CA421, CA483, NA741 and NA800), four markers (CA236, CA421, CA483 and NA800) and three markers (CA236, CA421 and NA800), respectively, produced only one allele and null H_E and PIC values for the QC, HHCU and NB and LBCU groups.

3.5 Cluster analysis

3.5.1 Cluster analysis using neighbour joining (NJ) dendogram

3.5.1.1 NJ analysis using EST-PCR markers

In the NJ analysis for EST-PCR data (Figure 5), wild blueberry clones and cultivars formed distinct groups and the genetic distances between the genotypes ranged from 0.18 to 1.00.

The NJ dendogram here is a strict consensus dendogram created from 1000 dendograms. In this dendogram, a branching pattern that occurs in 1000 dendogram with a frequency of 100% is adopted. There were five major clusters, three of which (clusers I, II and V) were comprised of NL clones only. The biggest cluster (cluster IV) contained blueberry clones from all other provinces and included lowbush blueberry cultivars and a

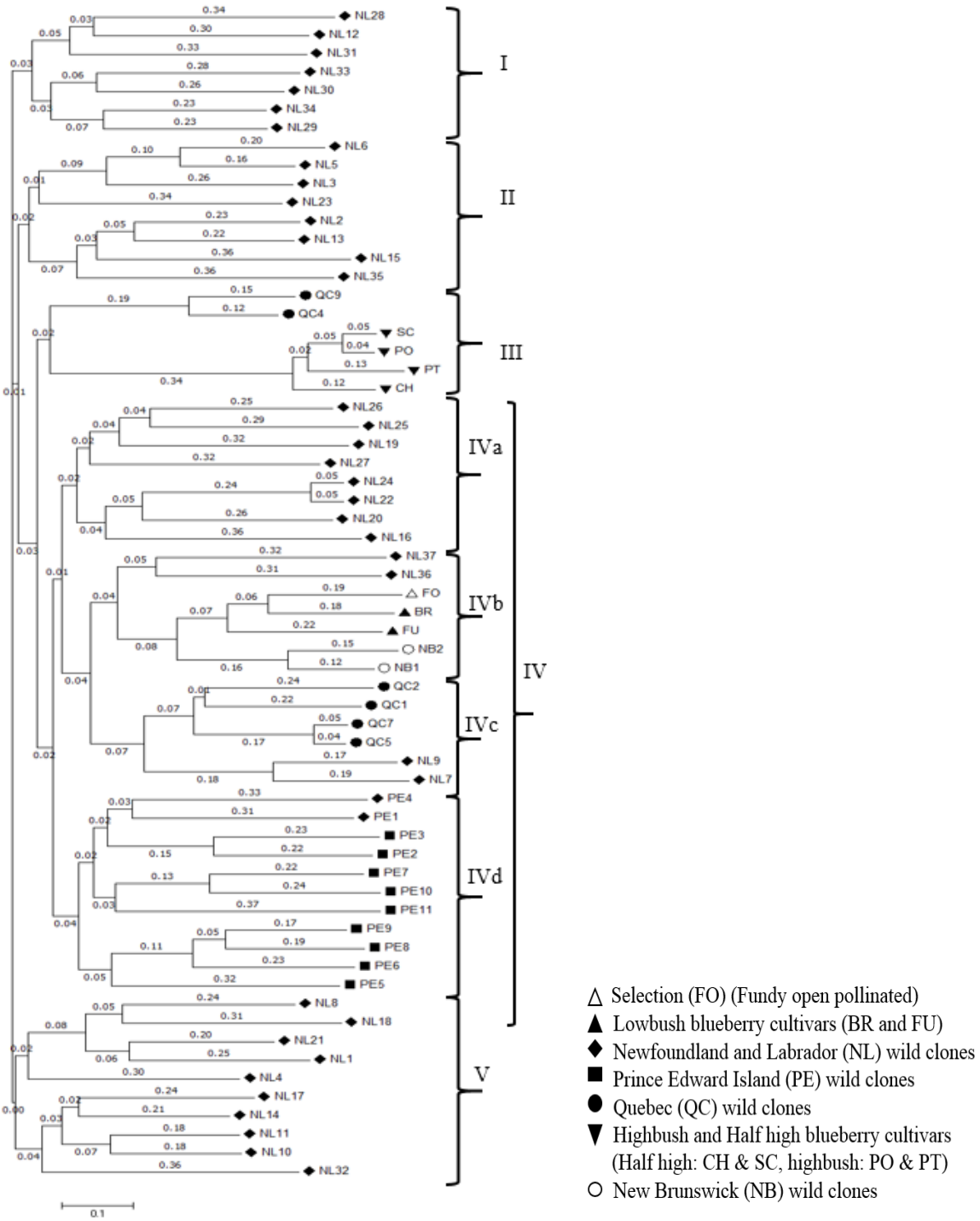

Figure 5: Neighbor joining (NJ) dendrogram of 63 blueberry genotypes based on the proportion of shared alleles distance for EST-PCR markers. Numbers refer to branch lengths.

selection (FO). Cluster I grouped seven NL clones and it was further divided into two subclusters. The first conatined three clones NL12, NL28 and NL31, and the second contained NL29, NL30, NL33 and NL34.

Cluster II was comprised of eight NL blueberry clones and was further broken up into two subclusters. The first of which included NL2, NL5, NL6 and NL23, and the second included NL2, NL13, NL15 and NL36.

Cluster III included the highbush blueberry cultiva Polaris and the three half-high blueberry cultivars Patriot, Cheppawa and St. Cloud, with two QC clones QC4 and QC9. QC clones made a very distinct subcluster from a subcluster of halfhigh and highbush blueberry cultivars.

Cluster IV was divided in two main clusters. The first cluster was further divide into three subclusters IVa, IVb and IVc. Subcluster IVa contained eight NL clones (NL16, NL19, NL20, NL22, NL24, NL25, NL26 and NL27), which were divided into two sub - subclusters. One of the sub - subclusters contained NL19, NL25, NL26 and NL27. The second sub - subcluser of subcluster IVa conatained NL16, NL20, NL22 and NL24. Subcluster IVb contained two NL clones, two NB clones, two lowbush blueberry cultivars (BR and FU) and a selection (FO). Subcluster IVb was divided into two small sub - subclusters. This sub - subcluster was further divided, where two NB clones NB1 and NB2 contained in one sub-sub-subcluser, while the second sub-sub-subcluster contained lowbush blueberry cultivars BR and FU and a selection FO. Subcluster IVc contained two

NL clones and four QC clones. Subcluster IVc was further divided into two sub-subclusters, where both NL clones NL7 and NL9 contained in the first sub-subcluster and four QC clones QC1, QC2, QC5 and QC7 contianed in the second sub-subcluster. The another main cluster of cluster IV was denoted as IVd. It contained all PE clones. Subcluster IVd was divided into two sub-subclusters. The first sub-subcluster contained PE1, PE2, PE3, PE4, PE7, PE10 and PE11, and the second sub-subcluster contained four PE clones, PE5, PE6, PE8 and PE9.

Cluster V consisted of 10 NL clones and was splitted in two subclusters of five clones each. The first of which contained NL1, NL5, NL8, NL18 and NL21 and the second contained NL10, NL11, NL14, NL17 and NL32.

3.5.1.2 NJ analysis using EST-SSR and genomic SSR markers

In the NJ analysis of EST-SSR and genomic SSR marker data (Figure 6), blueberry wild clones and cultivars formed distict groups. Strict consensus NJ dendogram showed five major clusters. The genetic distnace between genotypes ranged from 0.11 to 1.00.

In the dendogram, clusters I was divided into subclusters Ia and Ib. Subcluster Ia conatined all the genotypes belonging to NB provinnce, which included two NB clones NB1 and NB2, two lowbush blueberry cultivars FU and BR and one selection FO. Subcluster Ib contained six NL clones; NL18, NL23, NL29, NL31, NL36 and NL37.

Cluster II is divided into two subcluster IIa and IIb. IIa contained four PE clones; PE1, PE2, PE3 and PE4, and one NL clone, NL28, whereas IIb contained 11 NL clones;

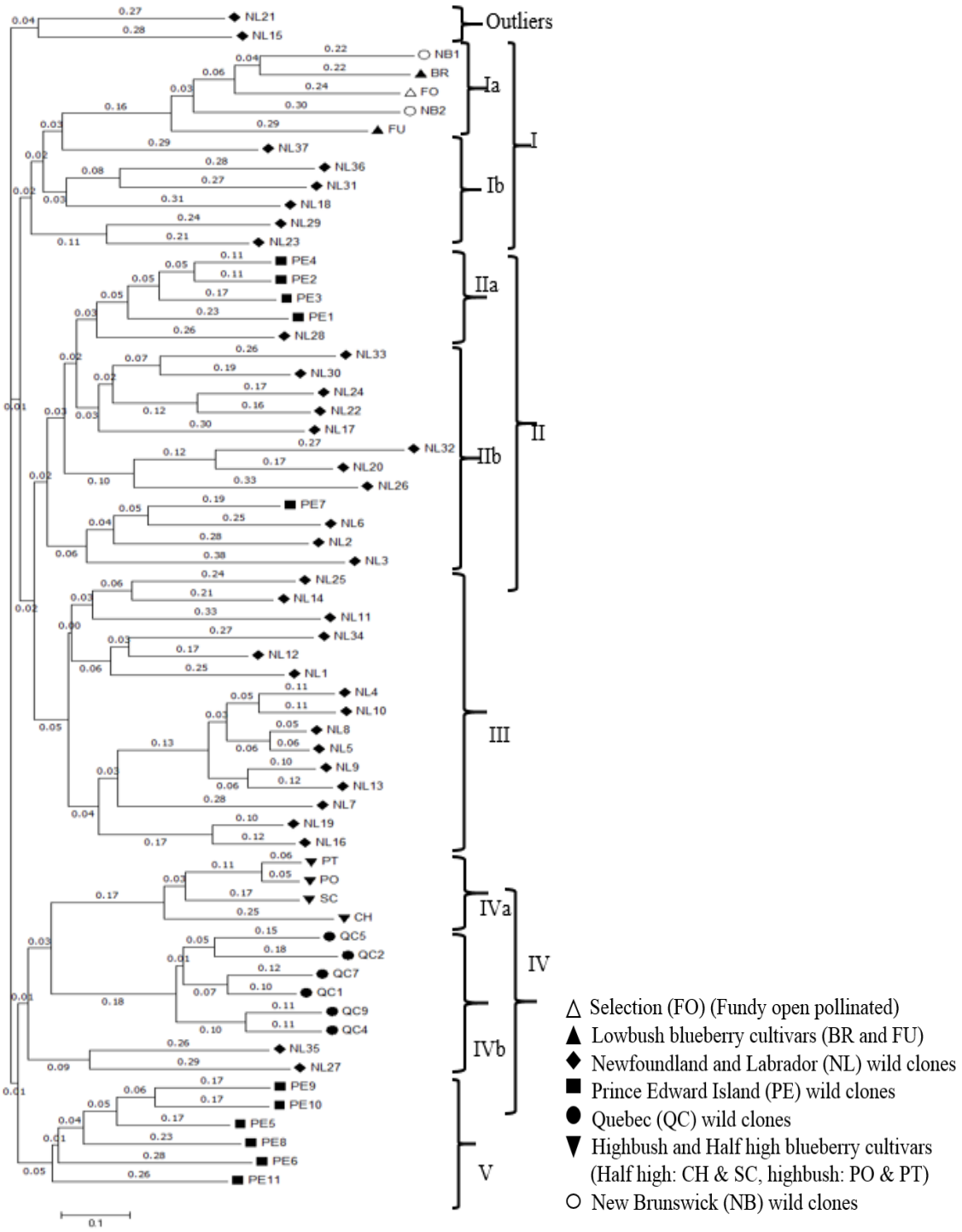

Figure 6: Neighbor joining (NJ) dendrogram of 63 blueberry genotypes based on the proportion of shared allele distance for EST-SSR and genomic SSR markers. Numbers refer to branch lengths.

NL2, NL3, NL6, NL17, NL20, NL22, NL24, NL26, NL30, NL32 and NL33 and one PE clone, PE7.

Cluster III, which contained 15 NL clones, was divided into two subclusters. The first subcluster included NL1, NL11, NL12, NL14, NL25 and NL34 and the second subcluster included NL4, NL5, NL7, NL8, NL9, NL10, NL13, NL16 and NL19.

Cluster IV divided into two subclusters, IVa and IVb. Subcluster IVa was comprised of one highbush blueberry cultivars PO with three half-high blueberry cultivars PT, CH and SC. Subcluster IVb contained all QC clones; QC1, QC2, QC4, QC5, QC7 and QC9. Cluster V contained six PE clones; PE5, PE6, PE8, PE9, PE10 and PE11.

3.5.1.3 Combined NJ analysis using EST-PCR, EST-SSR and genomic SSR markers

In a combined analysis for EST-PCR, EST-SSR and genomic SSR data, 63 genotypes were divided into five distinct clusters (Figure 7). The genetic distnace between genotypes ranged from 0.10 to 1.00.

Cluster I contained five NL clones NL23, NL27, NL28, NL32 and NL33, which were divided into two subclusters. The first subcluster contained NL23, NL27 and NL28, and the second subcluster contained NL32 and NL33.

Cluster II was further divided into two subclusters IIa and IIb. IIa contained all PE clones, from PE1 to 11. It was further divided into two sub-subclusters. The first sub-subcluster contained PE1 to 4, while the second sub-subcluster contained PE5 to 11. Subcluster IIb contained all six QC clones; QC1, QC2, QC4, QC5, QC7 and QC9. IIb was

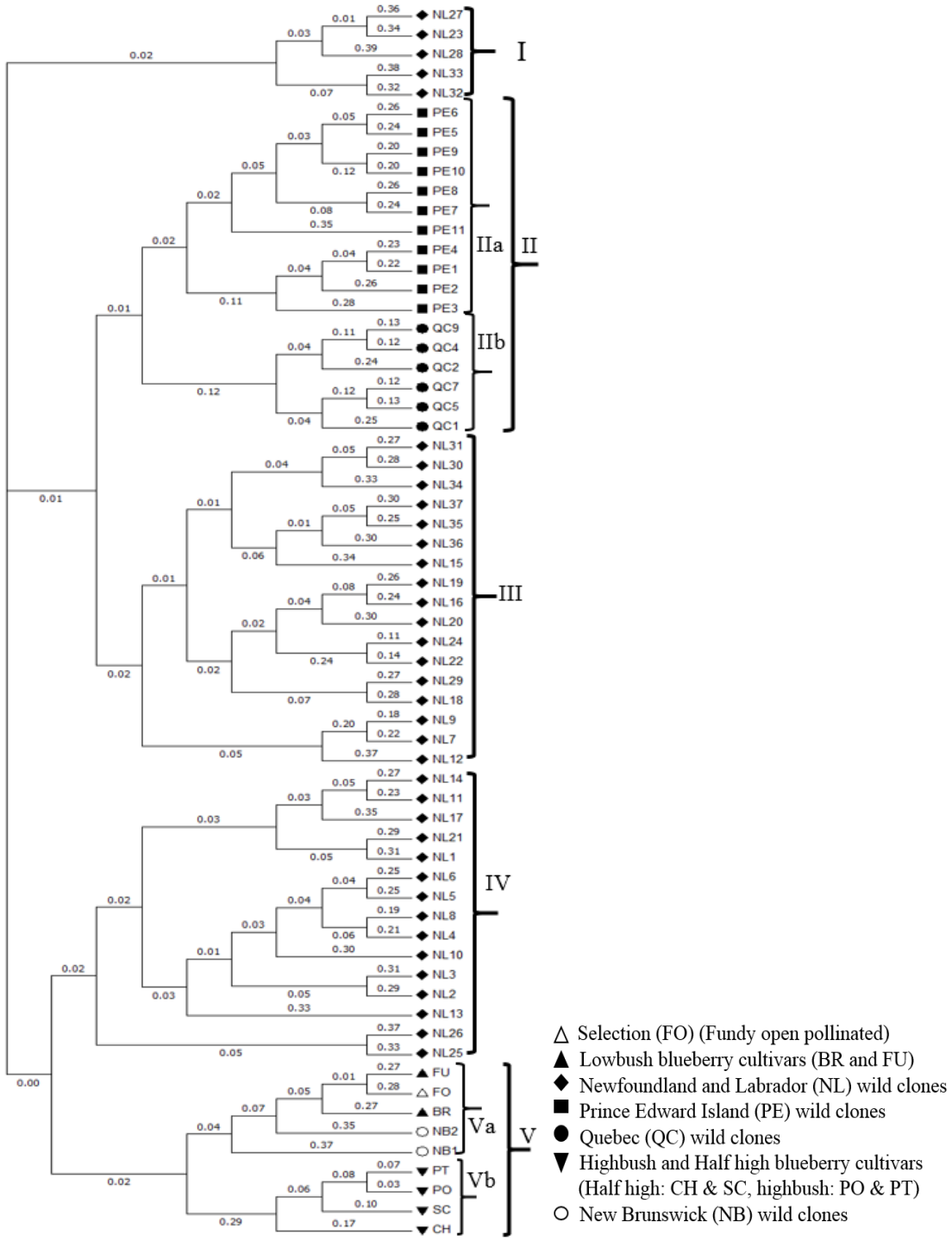

Figure 7: Neighbor joining (NJ) dendrogram of 63 blueberry genotypes based on the proportion of shared allele distance for EST-PCR, EST-SSR and genomic SSR markers. Numbers refer to branch lengths.

divided into two small sub-subclusters. The first sub-subcluster contained QC2, QC4 and QC9 and the second sub-subcluster contained QC1, QC5 and QC7.

Cluster III contained 17 NL clones which were further divided into two subclusters. From these two subclusters, the first subcluster was separated into two sub-subclusters. The first sub-subcluster contained seven NL clones (NL15, NL30, NL31, NL34, NL35, NL36 and NL37) and the second contained seven NL clones (NL16, NL18, NL19, NL20, NL22, NL24 and NL29). The second subcluster of the cluster III contained NL7, NL9 and NL12.

Cluster IV contained 15 NL clones, which included NL1, NL2, NL3, NL4, NL5, NL6, NL8, NL10, NL11, NL13, NL14, NL17, NL21, NL25 and NL26. Cluter IV was divided into two subclusters, the first of which contained five NL clones (NL1, NL11, NL14, NL17 and NL21), and the second contained eight NL clones (NL2, NL3, NL4, NL5, NL6, NL8, NL10 and NL13).

Cluster V was divided into two distinct subclusters i.e. subcluster Va and Vb. Subcluster Va contained all the genotypes related to NB province, which included two NB clones NB1 and NB2; two lowbush blueberry cultivas FU and BR and one selection FO. Vb contained four blueberry cultivars which include one highbush blueberry cultivar PO, and three halfhigh blueberry cultivars PT, CH and SC.

3.5.2 Cluster analysis using unweighted pair group method analysis (UPGMA) dendogram

3.5.2.1 UPGMA analysis using EST-PCR markers

In a strict consensus UPGMA dendogram (Figure 8) for EST-PCR marker data, blueberry wild clones and cultivars formed five distinct groups. The genetic distance between genotypes ranged from 0.10 to 1.00.

Cluster I contained four blueberry cultivars, which included one higbush blueberry cultivar PO and three halfhigh blueberry cultivars PT, CH and SC.

Cluster II contained eight genotype of which six were QC clones and two were NL clones. This cluster was divided into two subcluster, where the first subcluster contained two NL clones, NL7 and NL9 and the second subcluster contained QC1, QC2, QC4, QC5, QC7 and QC9.

Cluster III contained 22 NL clones: NL1, NL2, NL3, NL4, NL5, NL6, NL8, NL10, NL11, NL13, NL14, NL15, NL17, NL18, NL21, NL29, NL30, NL31, NL32, NL33, NL34, and NL35. This cluster was divided into two subclusters. The first subcluster contained NL2, NL3, NL5, NL6, NL13, NL15 and NL35 and the second subcluster contained NL1, NL4, NL8, NL10, NL11, NL14, NL17, NL18, NL21, NL29, NL30, NL31, NL32, NL33 and NL34.

Cluster IV contained 16 genotypes, which included five genotypes related to NB province (two NB clones NB1 and NB2, two lowbush blueberry cultivars (FU and BR) and a selection, FO) and 11 NL clones. These genotypes were divided into two subclusters IVa

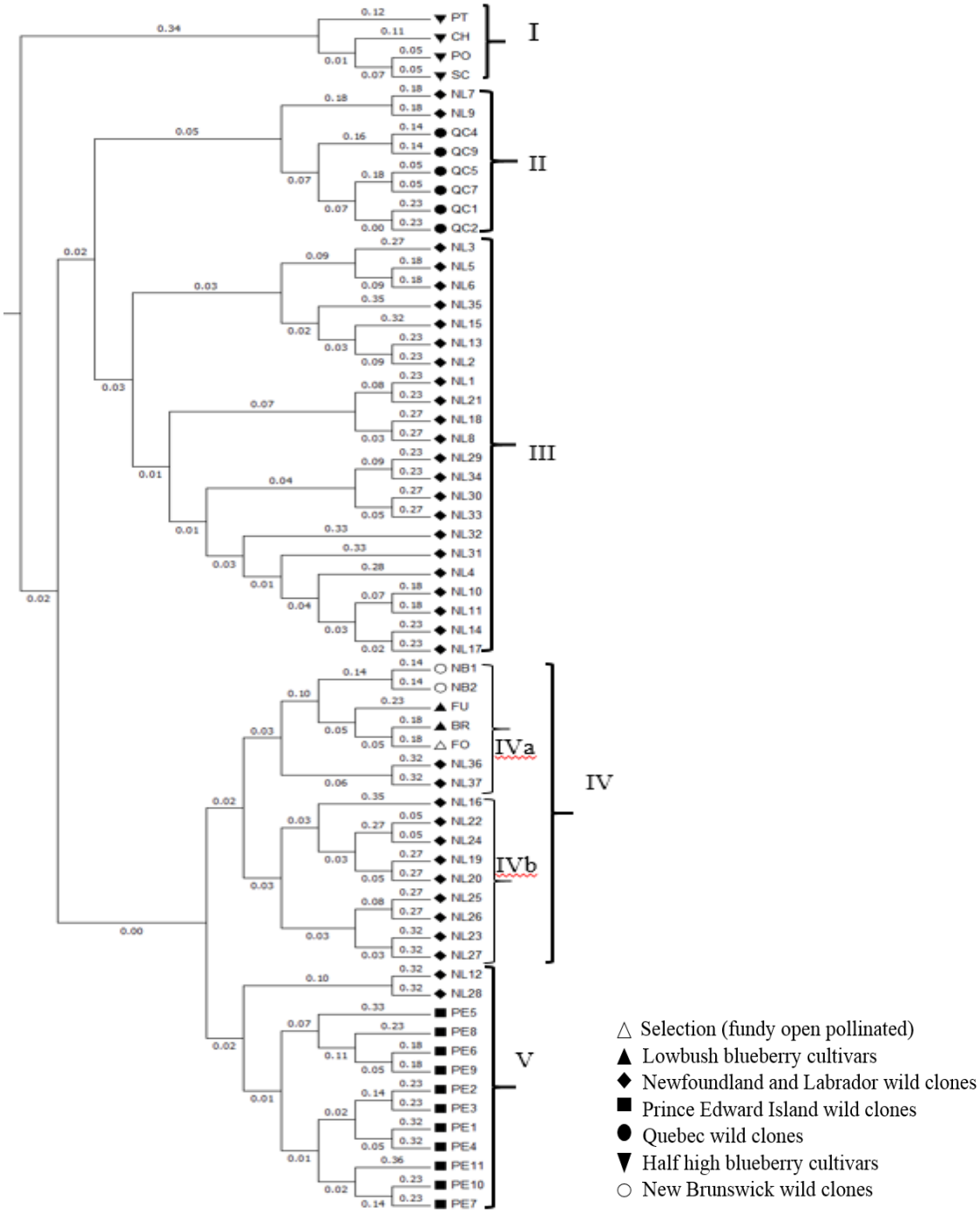

Figure 8: Unweighted pair group method analysis (UPGMA) dendrogram of 63 blueberry genotypes based on the proportion of shared allele distance for EST-PCR markers. Numbers refer to branch lengths.

and IVb. Subcluster IVa contained seven genotypes, which divided into two sub-subclusters. The first of which contained two lowbush blueberry cultivars, FU and BR; two NB clones NB1 and NB2; and one selection FO. The second sub-subcluster was comprised of two NL clones, NL36 and NL37. Subcluster IVb contained nine NL clones, which contained NL16, NL19, NL20, NL22, NL23, NL24, NL25, NL26 and NL27. These genotypes were divided into two sub-subclusters. The first was comprised of five NL clones: NL16, NL19, NL20, NL22 and NL24, and the second contained four NL clones: NL23, NL25, NL26 and NL27.

Cluster V contained 11 PE clones and two NL clones. Cluster V was splited into two subclusters. The first contained two NL clones: NL12 and NL28, and the second contained all PE clones, PE1 to 11. The second was further broken into two sub-subclusters. The first of which contained PE5, PE6, PE8 and PE9, and the second contained PE1, PE2, PE3, PE4, PE7, PE10 and PE11.

3.5.2.2 UPGMA analysis using EST-SSR and genomic SSR markers

In a strict consensus UPGMA dendogram (Figure 9) for EST-SSR and genomic SSR markers, wild clones and cultivars formed distinct groups, and total five major clusters were formed.

Cluster I contained five genotypes, which contained two lowbush blueberry cultivars FU and BR, two wild clones NB1 and NB2 and a selection FO. All these genotypes originated in the province of NB. Cluster II contained only three NL clones NL20, NL26 and NL32.

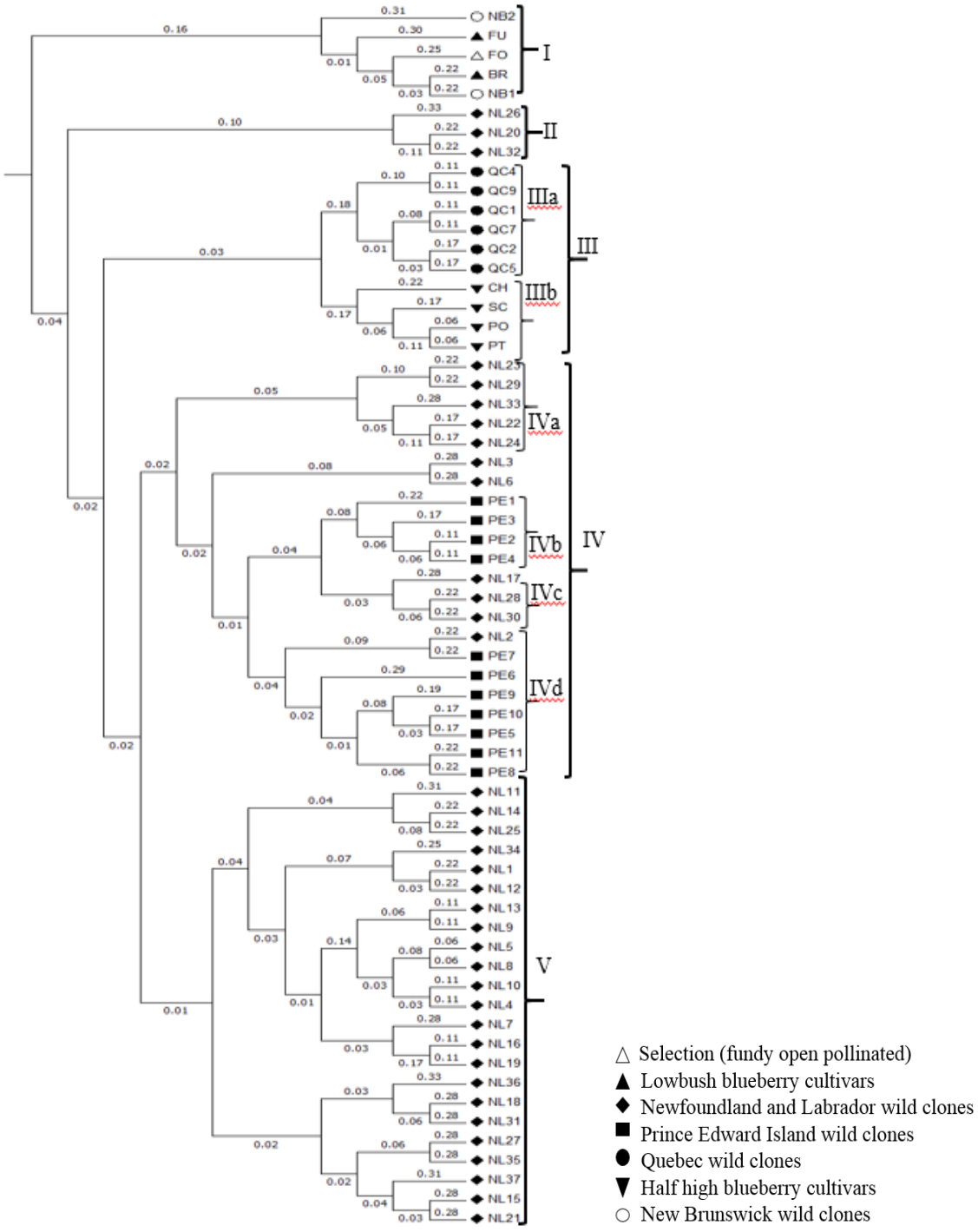

Figure 9: Unweighted pair group method analysis (UPGMA) dendrogram of 63 blueberry genotypes based on proportion of shared alleles distance for EST-SSR and genomic SSR markers. Numbers refer to branch lengths.

Cluster III is divided into two sub clusters, IIIa and IIIb. IIIa contained six QC clones; QC1, QC2, QC4, QC5, QC7 and QC9, and subcluster IIIb contained four cultivars, which included one highbush blueberry cultivar PO and three half-high blueberry cultivars PT, CH and SC.

Cluster IV contained 22 genotypes which were broken into four sub clusters: IVa, IVb, IVc and IVd. IVa contained five NL clones: NL22, NL23, NL24, NL29 and NL33. Subcluster IVb contained four PE clones: PE1, PE2, PE3 and PE4. The remaining seven PE clones PE5, PE6, PE7, PE8, PE9, PE10 and PE11 were grouped into IVd. However, IVc contained only three NL clones NL17, NL28 and NL30.

Cluster V contained 23 NL clones, which were NL1, NL4, NL5, NL7, NL8, NL9, NL10, NL11, NL12, NL13, NL14, NL15, NL16, NL18, NL19, NL21, NL25, NL27, NL31, NL34, NL35, NL36 and NL37. These clones were further divided into two subclusters, where the first one was comprised of NL1, NL4, NL5, NL7, NL8, NL9, NL10, NL11, NL13, NL14, NL16, NL19 and NL25 and the second contained the rest of the 10 clones.

3.5.2.3 Combined UPGMA analysis using EST-PCR, EST-SSR and genomic SSR markers

In a strict consesnus UPGMA dendogram (Figure 10), for combined EST-PCR, EST-SSR and genomic SSR markers, wild clones and cultivars formed distinct groups and separated into total five major clusters.

Cluster I contained four cultivars which included one highbush blueberry cultivar PO and three half-high blueberry cultivars PT, CH and SC.

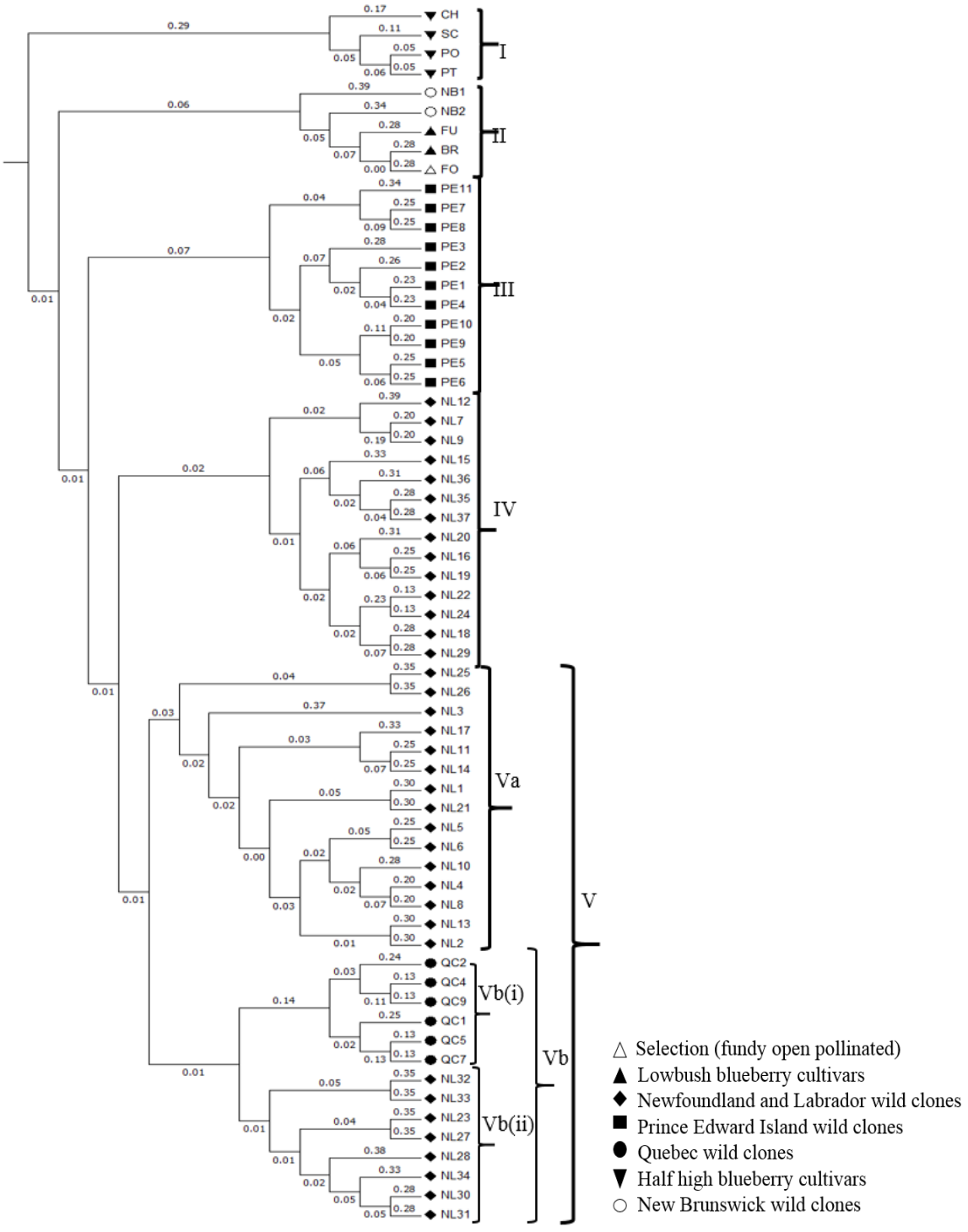

Figure 10: Unweighted pair group method analysis (UPGMA) dendrogram of 63 blueberry genotypes based on proportion of shared alleles distance for the combined EST-PCR, EST-SSR and genomic SSR markers. Numbers refer to branch lengths.

Cluster II contained five genotypes which included two NB clones NB1 and NB2, two lowbush cultivars FU and BR and one selection FO. All these five genotypes were originated from NB province.

Cluster III contained 11 PE clones from PE1 to 11. This cluster was further splitted into two subclusters. The first contained three PE clones: PE7, PE8 and PE11, and the second contained the rest of the eight PE clones: PE1, PE2, PE3, PE4, PE5, PE6, PE9 and PE10.

Cluster IV contained 14 NL clones: NL7, NL9, NL12, NL15, NL16, NL18, NL19, NL20, NL22, NL24, NL29, NL35, NL36 and NL37. This clones were further divided into two subclusters. The first of which contained three NL clones: NL7, NL9 and NL12, and the second contained other remaining clones: NL15, NL16, NL18, NL19, NL20, NL22, NL24, NL29, NL35, NL36 and NL37.

Cluster V was divided into two subclusters Va and Vb. Va contained 15 NL clones: NL1, NL2, NL3, NL4, NL5, NL6, NL8, NL10, NL11, NL13, NL14, NL17, NL21, NL25, and NL26. However, Vb was divided into two sub-subclusters Vb(i) and Vb(ii). Vb(i) conatined all six QC clones, which were QC1, QC2, QC4, QC5, QC7 and QC9, and the second sub-subcluster Vb(ii) contained eight NL clones, which were NL23, NL27, NL28, NL30, NL31, NL32, NL33 and NL34.

3.6 Principle coordinate analysis (PCoA)

3.6.1 PCoA using EST-PCR markers

In the principle coordinate analysis (PCoA) (Figure 11) of EST-PCR markers, five major clusters were formed (designated as I, II, III, IV and V), which distributed genotypes according to their geographic collection sites. The clusters were made among closely grouped genotypes. The first three axes together represent 34.17% of total genetic variation (axis 1 = 15.09%, axis 2 = 10.44% and axis 3 = 8.63%).

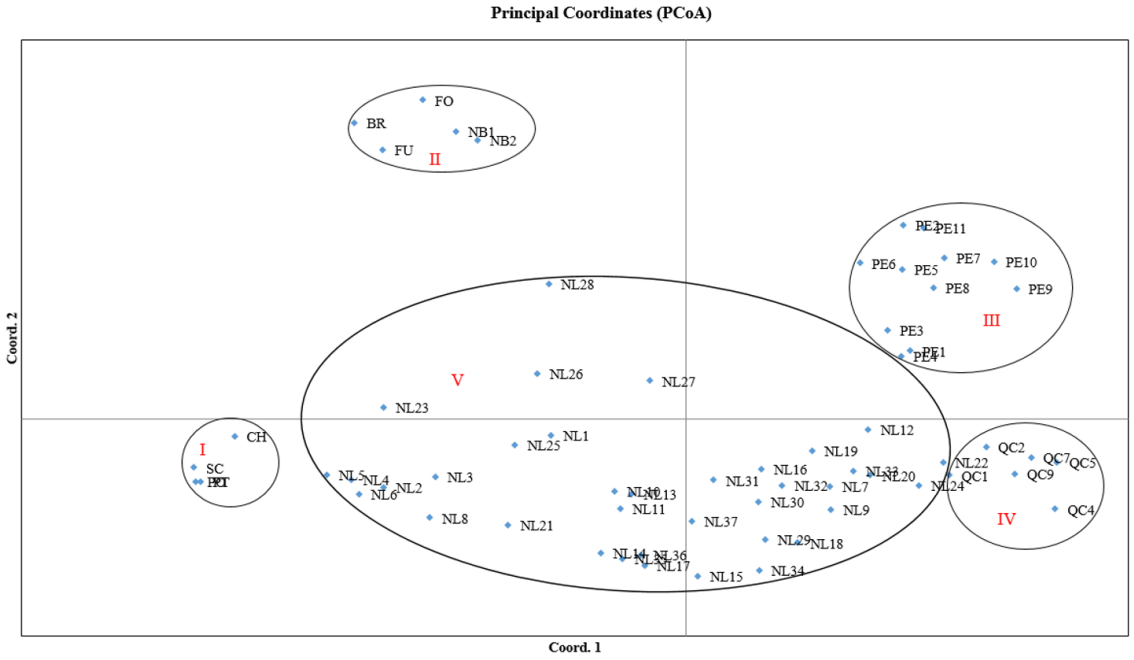

Figure 11: Two dimensional Principle co-ordinate analysis of 63 blueberry genotypes using EST-PCR markers. The circles indicate groupings of closely related genotypes. PCoA revealed that 15.09%, 10.44% and 8.63% (total = 34.17%) of the genetic variability can be explained by 1st, 2nd and 3rd axes.

Cluster I contained four genotypes, which included one highbush blueberry cultivar PO and three half-high blueberry cultivars PT, CH and SC. Cluster II contained two lowbush cultivars FU and BR, two NB clones NB1 and NB2 and a selection FO.

Cluster III contained all 11 PE clones: PE1, PE2, PE3, PE4, PE5, PE6, PE7, PE8, PE9, PE10 and PE11, and cluster IV contained all QC clones: QC1, QC2, QC4, QC5, QC7 and QC9. Cluster V contained all 37 NL clones from NL1 to NL37.

3.6.2 PCoA using EST-SSR and genomic SSR markers

In the PCoA (Figure 12) of seven EST-SSR and two genomic SSR markers, five major clusters were created (designated as I, II, III, IV and V). The clusters were made among closely grouped genotypes. This analysis distributed genotypes according to their geographic collection sites, but combined QC and PE clones, at the same time divided NL clones into two separate groups. The first three axes together represent 36.21% of total genetic variation (axis 1 = 14.40%, axis 2 = 12.31% and axis 3 = 9.50%).

Cluster I contained a highbush blueberry cultivar PO, and three half-high blueberry cultivars PT, CH and SC. Cluster II contained five genotypes which include two NB clones NB1 and NB2, two lowbush cultivars FU and BR and a selection FO.

Cluster III contained 18 NL clones: NL15, NL17, NL18, NL19, NL21, NL22, NL23, NL24, NL26, NL27, NL28, NL30, NL31, NL32, NL33, NL35, NL36 and NL37 and cluster IV contained 19 NL clones: NL1, NL3, NL4, NL5, NL7, NL8, NL9, NL10, NL11, NL12, NL13, NL14, NL16, NL20, NL25, NL29 and NL34.

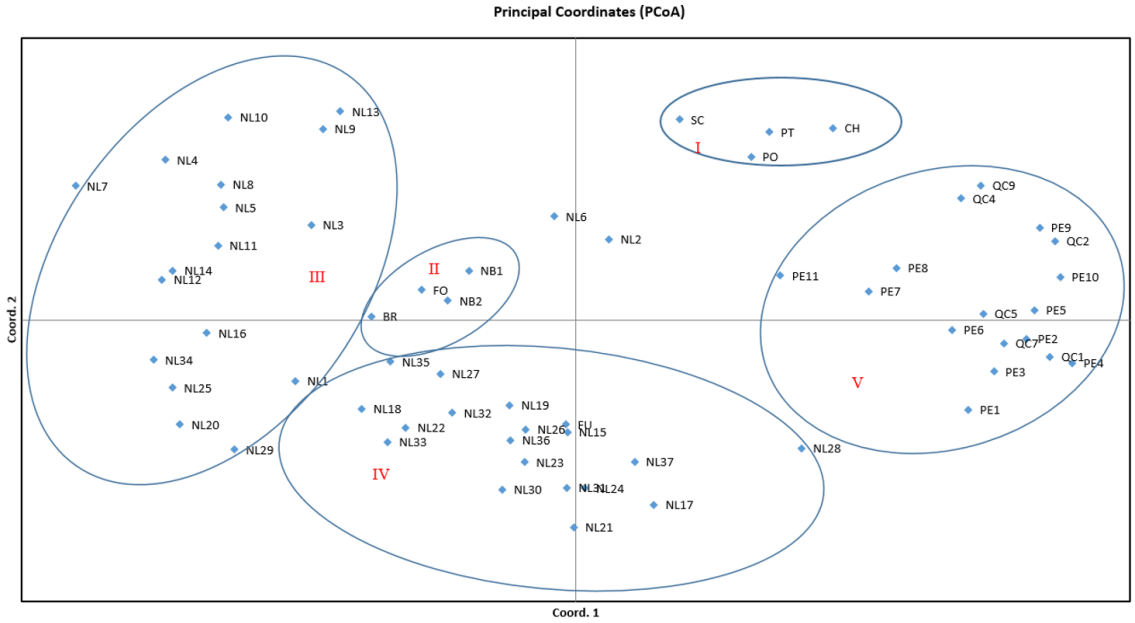

Figure 12: Two dimensional Principle co-ordinate analysis of 63 blueberry genotypes using microsatellite markers. The circles indicate groupings of closely related genotypes. PCoA revealed that 14.40%, 12.31% and 9.50% (total = 36.21%%) of the genetic variability can be explained by 1st, 2nd and 3rd axes.

Cluster V contained all 11 PE clones from PE1 to PE11, and all six QC clones: QC1, QC2, QC4, QC5, QC7 and QC9.

3.6.3 Combined PCoA analysis using EST-PCR, EST-SSR and genomic SSR markers

Five major clusters (designated as I, II, III, IV and V) were created in the principle coordinate analysis (PCoA) (Figure 13) of EST-PCR, EST-SSR and genomic SSR markers. This analysis distributed genotypes according to their geographic collection sites. The clusters were made among closely grouped genotypes. First three axes together represent 29.35% of total genetic variation (axis 1 = 12.11%, axis 2 = 9.37% and axis 3 = 7.88%).

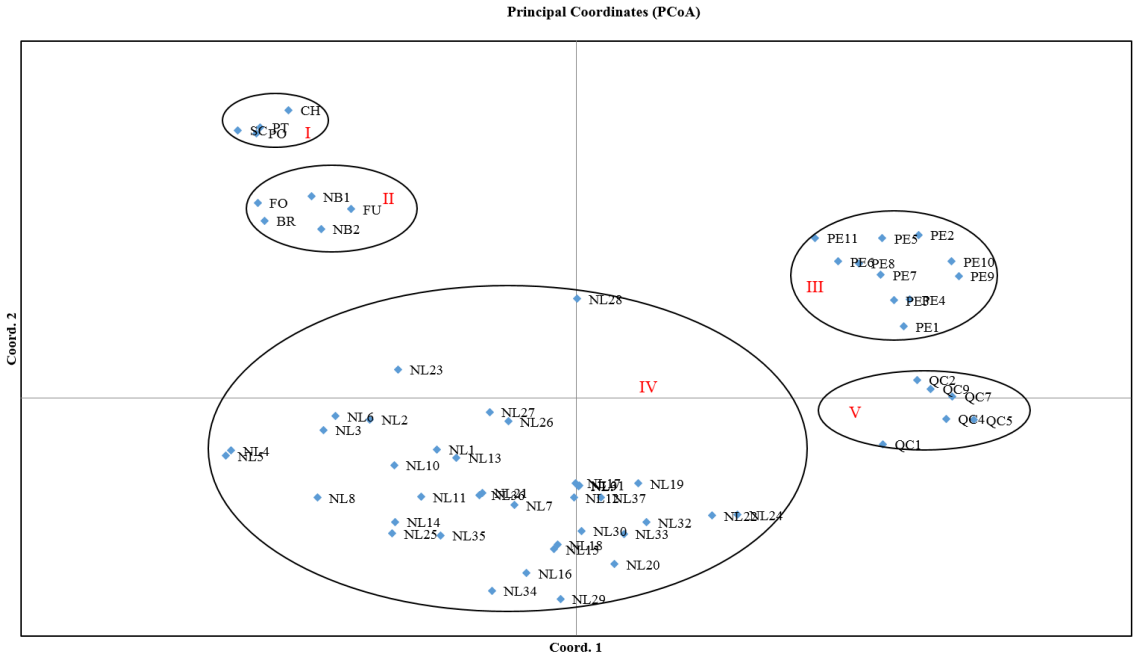

Figure 13: Two dimensional Principle co-ordinate analysis of 63 blueberry genotypes using EST-PCR and microsatellite markers. The circles indicate groupings of closely related genotypes. PCoA revealed that 12.11%, 9.37% and 7.88% (total = 29.35%) of the genetic variability can be explained by 1st, 2nd and 3rd axes.

Cluster I contained four blueberry cultivars, two of which contained highbush blueberry cultivars PO and PT, and the other two contained CH and SC. Cluster II contained five genotypes, which contained two NB clones NB1 and NB2, two lowbush blueberry cultivars BR and FU and a selection FO. These five genotypes originated from NB province.

Cluster III contained all PE clones from PE1 to PE11 and cluster IV contained six QC clones QC1, QC2, QC4, QC5, QC7 and QC9. Cluster V contained all 37 NL clones. This PCoA analysis shows that the genotypes are clustered according to their geographic origin.

3.7 Structure analysis

3.7.1 Structure analysis using EST-PCR marker data

Total 11 EST-PCR primer pairs were used to group 63 genotypes using STRUCURE program. From the plot of Ln (*K*) (Figure 14) and Δ*K* (Figure 15) it was determined that the genotypes were divided into two (*K* = 2) and six (*K* = 6) clusters. In Figure 14, graph plateaus at *K* = 6. So according to the Pritchard et al. (2000), genotypes is divided into 6 clusters.

However, in Δ*K* graph (Figure 15), generally the *K* showing highest peak is considered for clustering. In this case that peak is at *K*= 2. So the question arises at the peak of *K* = 6 in Figure 15, which has lower value than the peak at *K* = 2. In this case we can consider both the *K* values 2 and 6, considering *K* = 2 for broader structure and *K* = 6 for finer. Evanno et al. (2005) has explained that, as Δ*K* helps in identifying the correct number of clusters in most situations, it should not be used exclusively and should be used together with the other information provided by structure, such as *L*(*K*) itself. So, as seen in Figure 14, in L(*K*) graph, *K* plateaus at *K* = 6 and the there is also a good peak at *K* = 2 and 6 in Δ*K* graph (Figure 15). It makes *K* = 2 and 6, good choices. In addition, when bar plot was created using *K*=6, it showed clustering of genotypes according to their geographic collection site as seen in Figure 17. Thus, clustering at *K* = 6 is also considered for this data even though it did not give highest peak, but Pritchard et al. (2000) gives rationale to select *K*=6 for L(*K*) graph.

The bar plot (Figure 16) was sorted by *Q value* and plotted in multiple lines. Genotype clustering / distribution can be seen with the colour code in the bar plot.

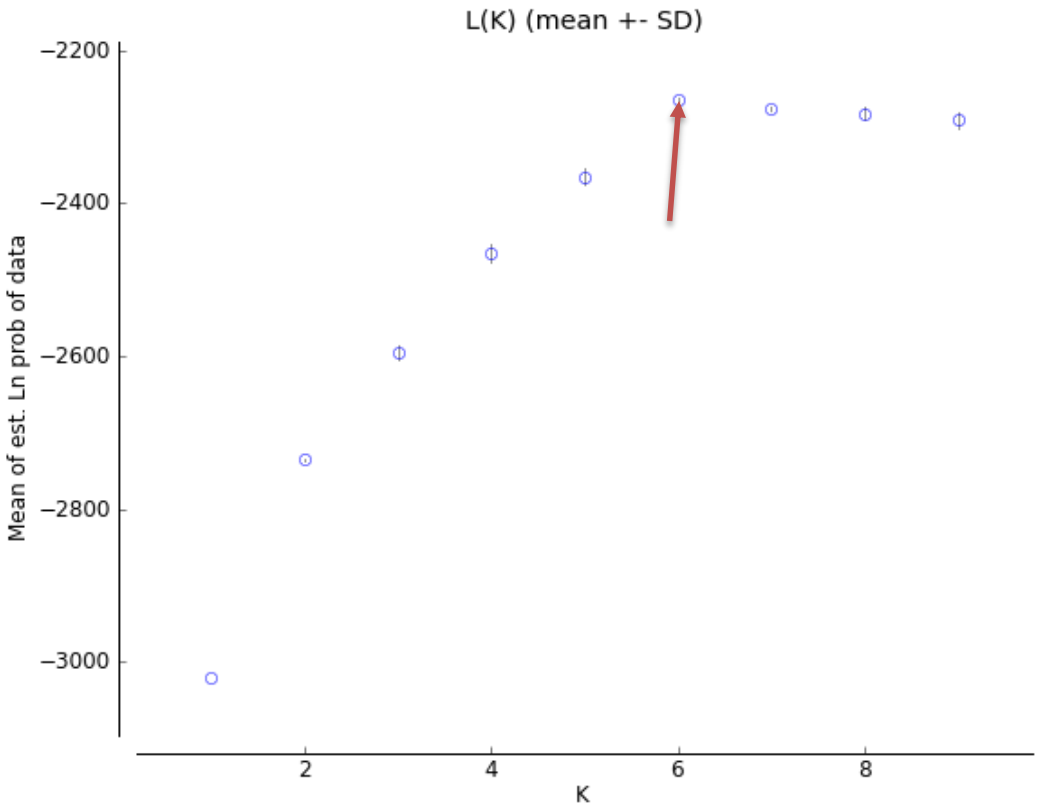

Figure 14: Graph of Ln (probability of data) for *K* ranging from 1 to 9 for EST-PCR marker analysis obtained from STRUCTURE HARVESTER software (Earl and Vonholdt, 2012; Pritchard et al. 2000).

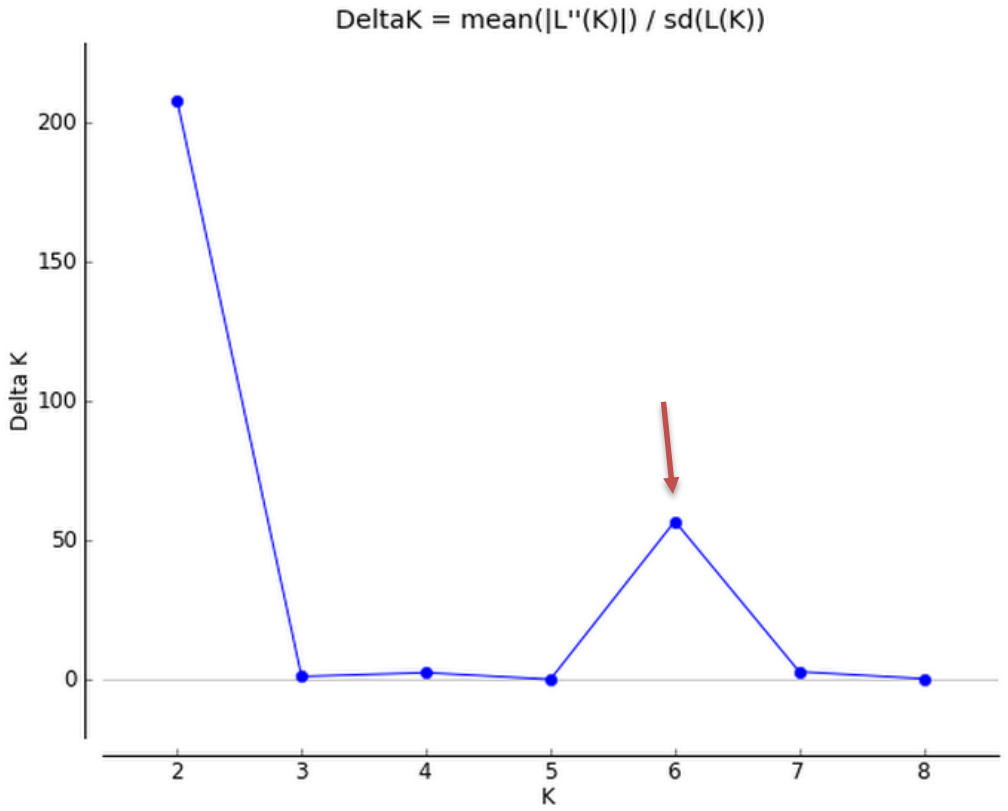

Figure 15: The graph obtained from STRUCTURE HARVESTER software (Earl and Vonholdt, 2012) estimating the number of clusters, using Δ*K* values for *K* ranging from 1 to 9 for EST-PCR marker data, using a method proposed by Evanno et al. (2005) .

The clustering in Figure 16 shows that all 63 genotypes were divided into two clusters. Cluster I was comprised of highbush blueberry cultivar PO and three half-high blueberry cultivars PT, CH and SC, two NB wild clones NB1 and NB2, two lowbush blueberry cultivars FU and BR and a selection FO, and 12 NL clones: NL7, NL9, NL11, NL12, NL14, NL15, NL16, NL17, NL18, NL19, NL20, NL21 and NL28. The cluster was coded by the red colour. NL14, NL21 and NL28 showed admixture of ~40% to 45% from cluster II; clones NL7, NL11 and NL12 had admixture of ~20%, and NL15, NL20 and NB1

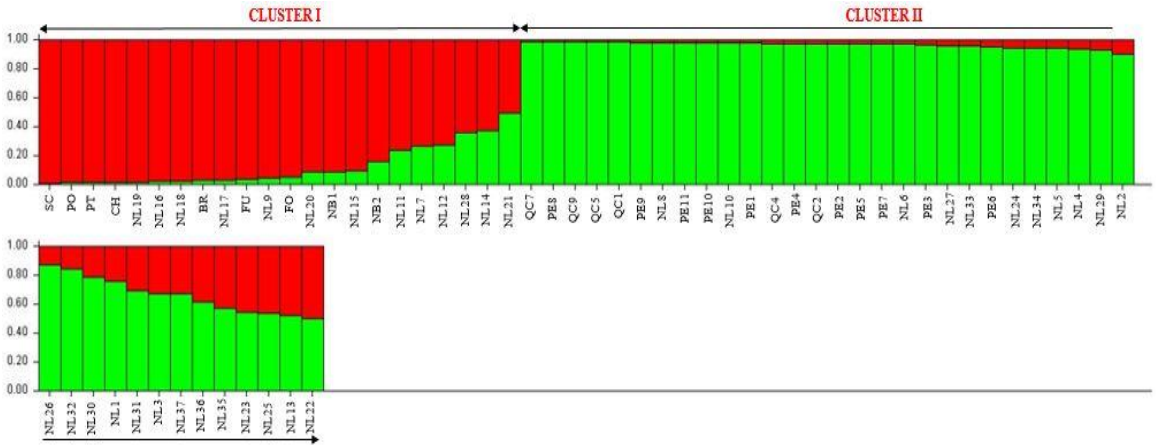

Figure 16: Q-plot showing clustering of 63 genotypes, for *K*=2, based on analysis of EST-PCR genotypic data using STRUCTURE. Each genotype is represented by a vertical bar. The colored subsections within each vertical bar indicate membership coefficient (Q) of the accession to different clusters. Identified clusters are I (red colour) and II (green colour).

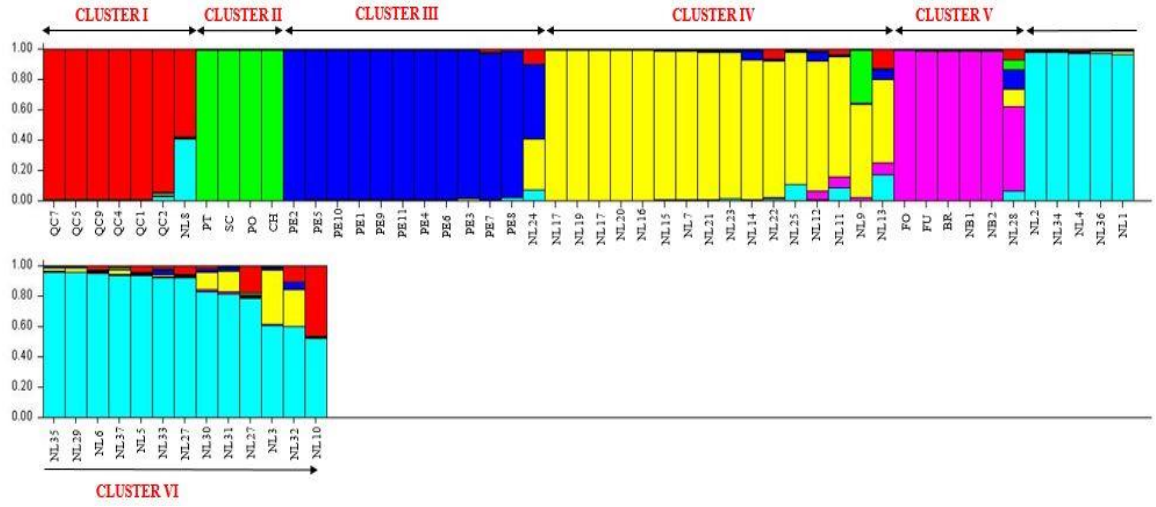

Figure 17: Q-plot showing clustering of 63 genotypes, for *K*=6, based on analysis of EST-PCR genotypic data using STRUCTURE. Each genotype is represented by a vertical bar. The colored subsections within each vertical bar indicate membership coefficient (Q) of the accession to different clusters. Identified clusters are I (red colour), II (green colour), III (blue colour), IV (yellow colour), V (pink colour) and VI (cyan colour).

showed admixture of ~9%. Cluster II contained all other genotypes. This cluster was coded by green colour.

The clustering in Figure 17 shows that all 63 genotypes were divided into six clusters. Cluster I was comprised of six QC clones (coded with pink colour) and NL8. NL8 showed high admixture of ~40% from cluster VI.

Cluster II contained highbush blueberry cultivar PO and three half-high blueberry cultivars PT, CH and SC (coded with green colour). Cluster III was comprised of 11 PE clones. All the genotypes in this cluster were coded by blue colour.

Cluster IV contained 17 clones: NL7, NL9, NL11, NL12, NL13, NL14, NL15, NL16, NL17, NL18, NL19, NL20, NL21, NL22, NL23, NL24 and NL25. NL9, NL13 and NL24 showed highest admixture of ~40% from multiple clusters. Clones NL11, NL12, NL14, NL22 and NL25 showed admixtures of ~5-9% from multiple clusters.

Cluster V contained two NB wild clones NB1 and NB2, two lowbush blueberry cultivars FU and BR, selection FO and NL28. NL28 contained ~40-45% admixture from multiple clusters. This cluster is coded with cyan colour in the plot.

Cluster VI included 18 NL clones: NL1, NL2, NL3, NL4, NL5, NL6, NL10, NL26, NL27, NL29, NL30, NL31, NL32, NL33, NL34, NL35, NL36 and NL37. NL3, NL10 and NL32 contained admixture of ~40-50% from different clusters, and clones NL27, NL30 and NL31 contained ~ 20% admixture from different clusters. This cluster was coded with cyan colour.

3.7.2 Structure analysis using EST-SSR and genomic SSR markers

Total nine microsatellite primer pairs were used to group 63 genotypes using STRUCURE software. From the plot of Ln (*K*) (Figure 18) and Δ*K* (Figure 19) it can be determined that the genotypes are divided into six (*K* = 6) clusters.

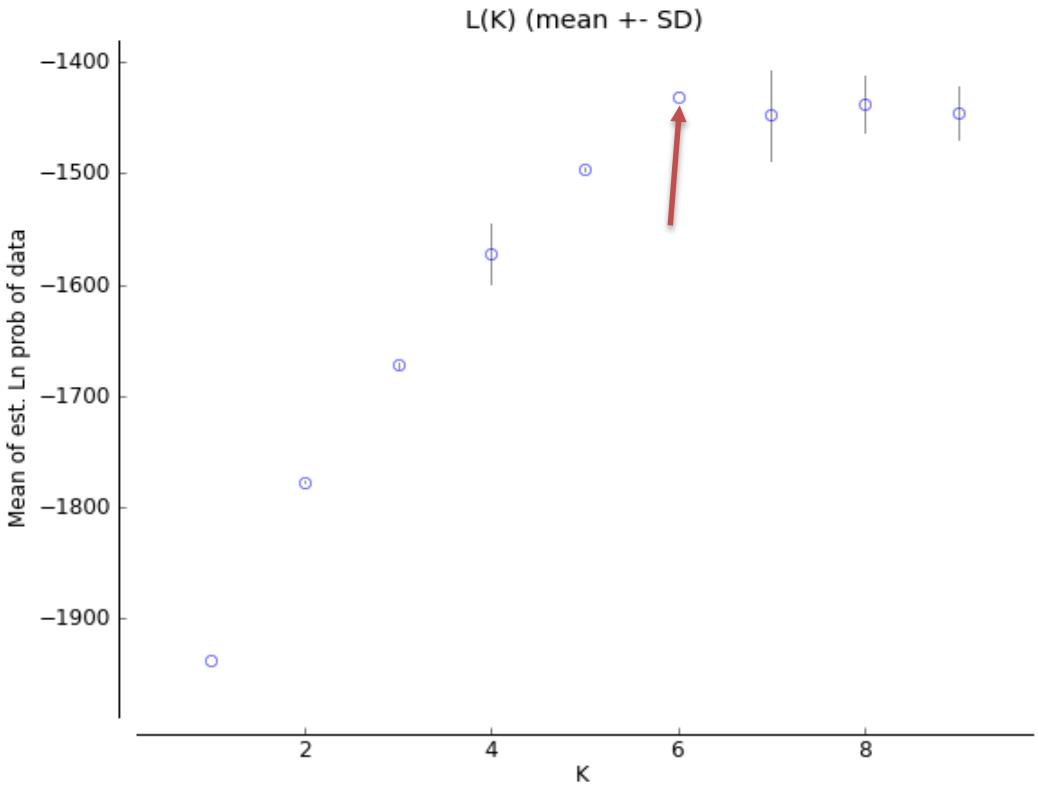

Figure 18: Graph of Ln (probability of data) for *K* ranging from 1 to 9 for microsatellite marker analysis obtained from STRUCTURE HARVESTER software (Earl and Vonholdt, 2012; Pritchard et al. 2000).

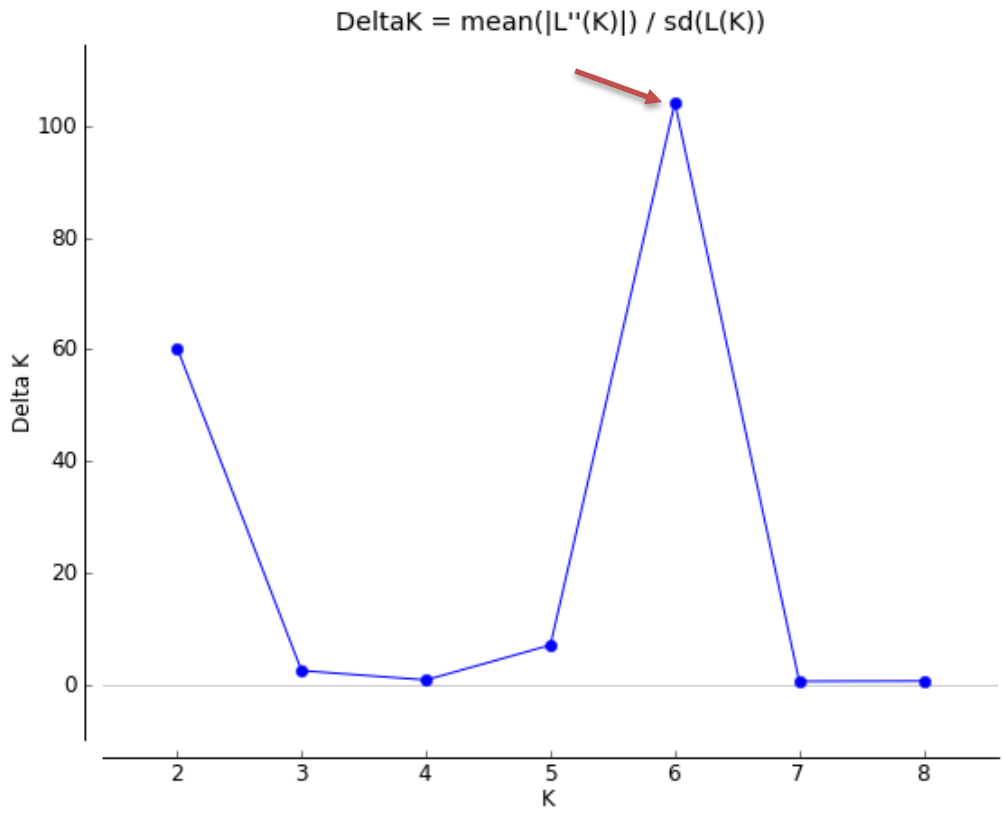

Figure 19: The graph obtained from STRUCTURE HARVESTER software (Earl and Vonholdt, 2012) estimating the number of clusters, using Δ*K* values for *K* ranging from 1 to 9 for microsatellite marker data, using a method proposed by Evanno et al. (2005) .

Here Δ*K* graph (Figure 19) shows highest peak at *K*=6 and in L (*K*) graph, *K* plateaus at *K* = 6. In addition, when bar plot was created, it also showed clustering of genotypes according to their geographic sampling site as seen in Figure 20. Thus, clustering at *K* = 6 is a good choice and is considered for this data.

The bar plot (Figure 20) created by STRUCTURE software was sorted by *Q value* and plotted in multiple lines. Genotype clustering / distribution can be seen with the colour code in the bar plot.

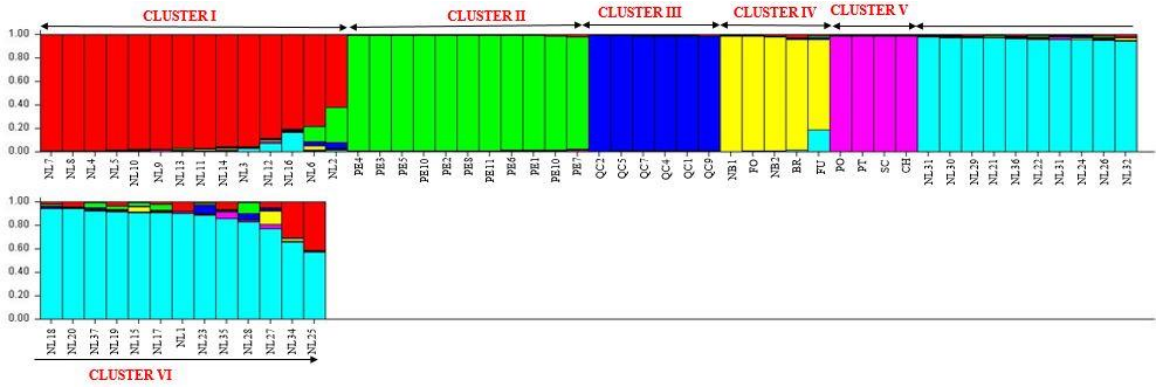

Figure 20: Q-plot showing clustering of 63 genotypes, for *K*=6, based on analysis of microsatellite genotypic data using STRUCTURE software. Each genotype is represented by a vertical bar. The colored subsections within each vertical bar indicate membership coefficient (Q) of the accession to different clusters. Identified clusters are I (red colour), II (green colour), III (blue colour), IV (yellow colour), V (pink colour) and VI (cyan colour).

The 63 genotypes were divided into six clusters (Figure 20). Cluster I contained 14 NL clones, which included NL2, NL3, NL4, NL5, NL6, NL7, NL8, NL9, NL10, NL11, NL12, NL13, NL14 and NL16. All the genotypes of this cluster were coded by red colour. NL2 showed ~30% admixture from other clusters, NL6 and NL16 showed ~20% admixture from different clusters. NL12 showed ~10% admixture from cluster VI.

Cluster II included all 11 PE clones, which were coded by green colour, Cluster III was comprised of six QC clones (coded with blue colour). Cluster IV contained two NB wild clones NB1 and NB2, two lowbush blueberry cultivars FU and BR and a selection

FO. FU had admixture of ~20% from cluster VI. This cluster was coded with yellow colour in the plot.

Cluster V included four cultivars, one of which was highbush blueberry cultivar PO, and the other three were half-high blueberry cultivars PT, CH and SC. All four cultivars were coded by pink colour

Cluster VI included 22 NL clones, which included NL1, NL15, NL17, NL18, NL19, NL20, NL21, NL22, NL23, NL24, NL25, NL26, NL27, NL28, NL29, NL30, NL31, NL32, NL33, NL34, NL35, NL36 and NL37. All the genotypes of this cluster were coded by cyan colour. Wild clones from this cluster showed high level of admixture from different clusters. Clones NL27, NL34 and NL35 showed admixture of ~35% to ~42%. Clones such as NL1, NL15, NL17, NL18, NL19, NL20, NL23, NL28, NL35 and NL37 showed ~10% to 20% admixture.

3.7.3 Structure analysis using combination of EST-PCR and microsatellite markers

In this analysis of combined markers 63 genotypes were divided into six clusters. As seen in Ln (*K*) (Figure 21) graph, the graph plateaus at $K = 6$. So according to Pritchard et al. (2000), genotypes is divided into 6 clusters. However, as seen in section 3.7.1, combined analysis also showed two peaks in ΔK graph (Figure 22). In common practice *K* showing highest peak is considered for clustering. In this case that peak is at $K= 2$. Here we can consider both the *K* values 2 and 6, considering $K = 2$ for broader structure and $K = 6$ for finer. The rationale behind selecting two *K* values is explained in section 3.7.1.

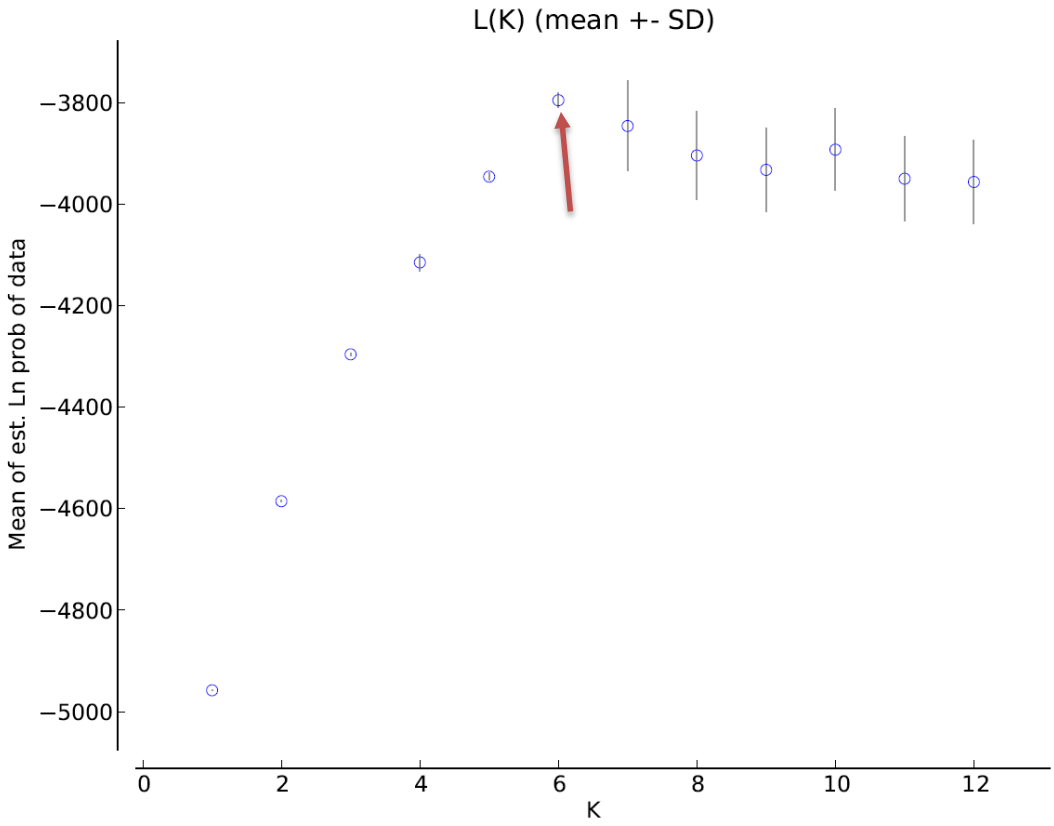

Figure 21: Graph of Ln (probability of data) for *K* ranging from 1 to 9 for EST-PCR and microsatellite markers' combined analysis obtained from STRUCTURE HARVESTER software (Earl and Vonholdt, 2012; Pritchard et al. 2000).

The bar plots (Figure 23 and Figure 24) created by STRUCTURE software were sorted by *Q value* and plotted in multiple lines. Genotype clustering / distribution can be seen with the colour code in the bar plot.

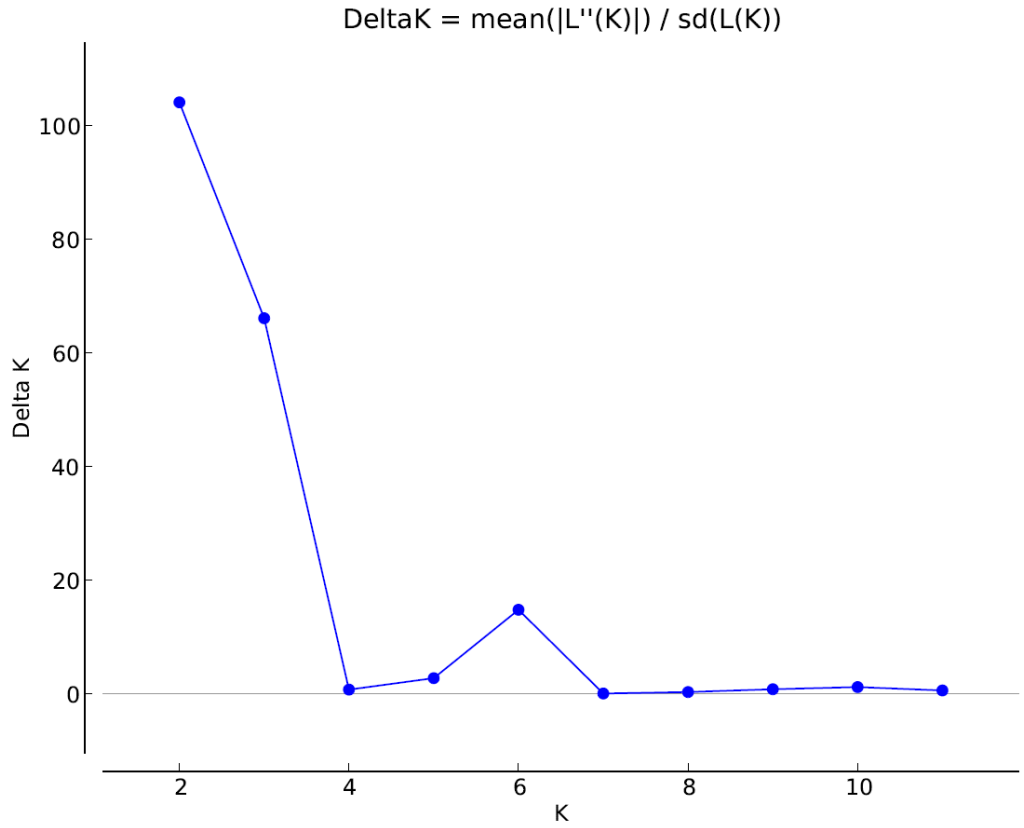

Figure 22: The graph obtained from STRUCTURE HARVESTER software (Earl and Vonholdt, 2012) estimating the number of clusters, using Δ*K* values for *K* ranging from 1 to 9 for EST-PCR and microsatellite marker data, using a method proposed by Evanno et al. (2005).

STRUCTURE plot, Figure 23, divided 63 genotypes into two clusters. Cluster I was comprised of all eleven PE wild clones and all six QC wild clones, and three NL clones NL22, NL24 and NL28. The cluster was coded by the red colour. QC2 showed admixture of ~6% from cluster II, and clones NL22, NL24 and NL28 had admixture of ~40% - 50% admixture with cluster II. Cluster II contained all other genotypes. This cluster was comprised of green colour.

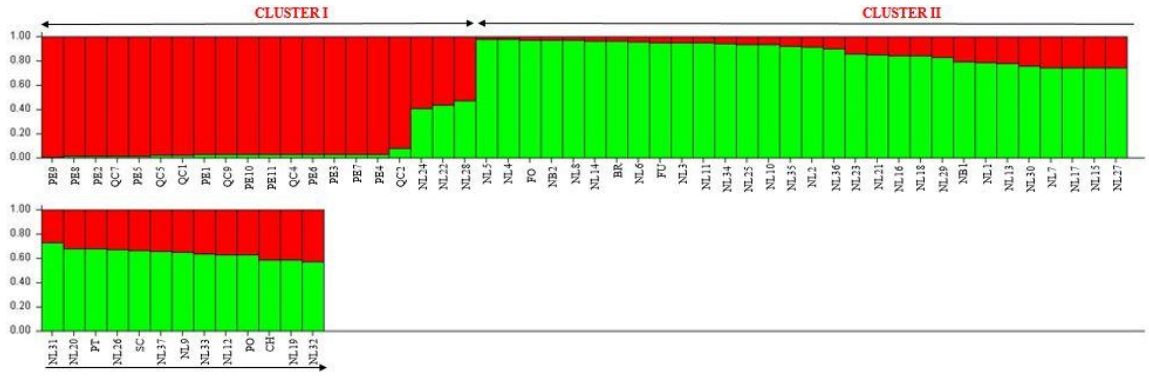

Figure 23: Q-plot showing clustering of 63 genotypes, for *K*=2, based on analysis of combined genotypic data of EST-PCR and microsatellite using STRUCTURE software. Each genotype is represented by a vertical bar. The colored subsections within each vertical bar indicate membership coefficient (Q) of the accession to different clusters. Identified clusters are I (red colour) and II (green colour).

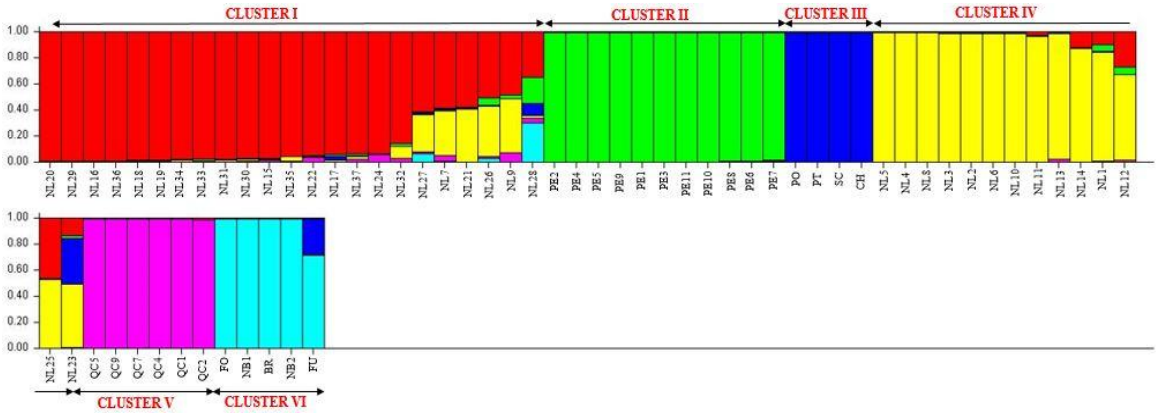

Figure 24: Q-plot showing clustering of 63 genotypes, for *K*=6, based on analysis of combined data of EST-PCR and microsatellite genotypic data using STRUCTURE software. Each genotype is represented by a vertical bar. The colored subsections within each vertical bar indicate membership coefficient (Q) of the accession to different clusters. Identified clusters are I (red colour), II (green colour), III (blue colour), IV (yellow colour), V (pink colour) and VI (cyan colour).

The clustering in Figure 24 shows that all 63 genotypes were divided into six clusters. Cluster I contained 23 clones: NL7, NL9, NL15, NL16, NL17, NL18, NL19,

NL20, NL21, NL22, NL24, NL26, NL27, NL28, NL29, NL30, NL31, NL32, NL33, NL34, NL35, NL36 and NL37. From these clones, clones NL17, NL22, NL24, NL35 and NL37 have ~5% admixture from other clusters. NL32 showed admixture of ~19% and clones NL7, NL9, NL21, NL26, NL27 and NL28 showed admixture of ~39% to ~70%.

Cluster II was comprised of 11 PE clones (coded by green colour), and cluster III contained highbush blueberry cultivar PO and three half-high blueberry cultivars PT, CH and SC (coded with blue colour).

Cluster IV included 14 NL clones: NL1, NL2, NL3, NL4, NL5, NL6, NL8, NL10, NL11, NL12, NL13, NL14, NL23 and NL25. NL1 and NL14 contained admixture of ~18% from different clusters, and clones NL12, NL23 and NL25 contained ~ 40% - 70% admixture from different clusters. This cluster was coded with yellow colour.

Cluster V was comprised of six QC clones (coded with pink colour) and cluster VI contained two NB wild clones NB1 and NB2, two lowbush blueberry cultivars FU and BR and a selection FO. This cluster is coded with cyan colour in the plot.

3.8 Analysis of molecular variance (AMOVA) based on geographic location

AMOVA analysis was performed separately EST-PCR and microsatellite marker data and for combined data.

3.8.1 AMOVA using EST-PCR marker data

The results for AMOVA analysis are shown in the Table 12, where eleven EST-PCR markers found the variation of 36.54% among the groups, 17.42% variation among the communities within the groups and 46.04% variation among genotypes within communities.

Table 12: Analysis of molecular variance based on geographic collection site/ the type of genotypes (half-high/lowbush cultivars) with EST-PCR marker data in 63 genotypes of blueberries.

Source of variation	def.	Sum of squares	Variance components	% of variation
Among groups	4	46,881	7.00	36.54
Among communities within groups	12	14,126	3.34	17.42
Among genotypes within communities	8803	77,713	8.83	46.04
Total	8819	1,38,720	19.17	100

3.8.2 AMOVA using data of EST-SSR and genomic SSR markers

Nine SSR markers, as seen from Table 13, found the variation of 26.50% among the groups, 32.01% variation among the communities within the groups and 41.49% variation among the genotypes within the communities.

Table 13: Analysis of molecular variance based on geographic collection site/ the type of genotypes (half-high/lowbush cultivars) with EST-SSR and genomic SSR marker data in 63 genotypes of blueberries.

Source of variation	def.	Sum of squares	Variance components	% of variation
Among groups	4	28,793	3.29	26.50
Among communities within groups	12	16,729	3.97	32.01
Among genotypes within communities	8803	45,314	5.15	41.49
Total	8819	90,836	12.41	100

3.8.3 AMOVA analysis using combined data of EST-PCR, EST-SSR and genomic SSR markers

When the AMOVA analysis done with the combined data of 11 EST-PCR and nine microsatellite markers (Table 14), the variation of 32.60% was found among the groups, 23.15% variation was found among the communities within the groups and 44.25% variation was found among genotypes within the communities.

Table 14: Genetic differentiation among blueberry genotypes by analysis of molecular variance based on geographic collection site/ the type of genotypes (half-high/lowbush cultivars) with EST-PCR and microsatellite markers' combined data.

Source of variation	d.f.	Sum of squares	Variance components	% of variation
Among groups	4	76,673	10.29	32.60
Among communities within groups	12	30,854	7.31	23.15
Among genotypes within communities	8803	123,028	13.98	44.25
Total	8819	229,555	31.58	100

4 Discussion

Blueberries are an economically very important fruit crop in North America and some European countries. They also provide several health benefits and are favourite berries for people across the world. With increasing demand in quality and quantity of blueberries, their development was undertaken by several research groups in the USA. Plant breeding is an essential technique to develop new varieties of crop plants. It can help develop crops which grow in previously unexplored regions, and can also help improve the quality and yield of crops including blueberries. It is pertinent to have knowledge about genetic diversity among wild and cultivated blueberry genotypes, as it plays an important role in crop improvement and plant breeding programs.

Molecular markers serve to achieve this knowledge on genetic diversity. Several molecular markers have been used in distinguishing wild and cultivated genotypes of blueberries (Bell et al. 2008; Boches et al. 2006; Debnath 2014; Rowland et al. 1994; Rowland et al. 2003a; Rowland et al. 2010). EST-PCR (Bell et al. 2008; Rowland et al. 2010; Debnath 2014), EST-SSR (Boches et al. 2008; Debnath 2014) and genomic SSR markers (Boches et al. 2006) have been found effective in diversity analysis of different types of blueberry genotypes. SSR markers are excellent genetic markers because they are co-dominant, multi-allelic and reproducible, but at the same time they are expensive and can be laborious to develop. Primer pairs developed for highbush blueberry cultivar 'Bluecorp' (Rowland et al. 2003b) were utilized in the present study to estimate genetic diversity in blueberry clones, cultivars and a selection and were found effective. This

mitigates the cost of cDNA preparation and development of SSR markers for the present material.

The goal of this study was to identify EST-PCR and EST-SSR primer pairs suitable for fingerprinting the present set of blueberry genotypes, to assess the properties of markers and their importance in assessing the diversity in the present set of blueberry genotypes, and to evaluate genetic diversity and population structure among the genotypes.

4.1 Identification of suitable primer pairs

In this study, 12 EST-PCR, 10 EST-SSR and two genomic SSR primer pairs were included to check for their suitability and effectiveness for diversity analysis in lowbush, highbush and half-high blueberry cultivars, wild clones and a selection. Among them, 11 EST-PCR, seven EST-SSR and two genomic SSR primer pairs gave clear and polymorphic bands and were included in this study. EST-SSR primer pairs CA23, CA112 and CA169, and EST-PCR primer CA227 were not considered for further analysis. The reason could lie in their inefficiency to amplify the genes in the current pool of wild lowbush blueberry clones and blueberry cultivars. Other possible reasons which could lead to non-reproducibility of these markers can be gene mutation or the tetraploid nature of the lowbush blueberry (Boches et al. 2006).

In the present study, as shown in Table 5 and 6, 11 EST-PCR primer pairs produced 87 polymorphic bands, with the average of 7.9 bands per primer, and nine microsatellite primer pairs produced 53 polymorphic bands, with the average of 5.9 bands per primer, indicating that considerably more polymorphism has been detected in the present study

than those observed by Bell et al. (2008), who detected a total of 81 polymorphic bands in 14 genotypes, using 17 primer pairs with 4.8 bands per primer in diploid and tetraploid *Vaccinium* species. This was expected as the present material was collected from a wide range of geographical regions of Canadian provinces.

Due to the production of allelic 'ladders' by closely spaced alleles and stutter bands, it was extremely difficult to distinguish true alleles and PCR artifacts at these loci. However, the fingerprints produced by these loci were highly polymorphic and had the same reproducibility as fingerprints from other loci. In some cases, it was possible to distinguish the majority of the individuals examined using only one locus (e.g. CA16 for EST-PCR and NA1040 for SSR, which amplified a total of 52 and 54 alleles, respectively). 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 the incorrect genotyping of two otherwise identical clones (Crespan 2004). Additionally, loci and multiple accessions from the same cultivar should be tested to verify a mismatch.

The ability of the most informative primer pair to differentiate between genotypes was assessed by estimating their resolving power (Rp) (Prevost and Wilkinson, 1999). The Rp value compares the diagnostic effectiveness of primers (Gilbert et al. 1999). The higher Rp value suggests increased ability of a primer to distinguish genotypes (Gilbert et al. 1999). The collective Rp value of 11 EST-PCR markers was 50.8 with an average of 4.61, while the collective Rp value of nine SSR markers was 27 with an average of 3. This result suggests that 11 EST-PCR markers are superior to nine microsatellite markers in

distinguishing various genotypes in the present material. These Rp values are comparable with other studies from other species too (Hameed et al. 2012). The Rp value suggests that the marker has more power to detect polymorphism because of its ability to produce more bands. In this study, EST-PCR markers produced more Rp values than microsatellite markers. Even the average values were higher for EST-PCR markers, which indicates that EST-PCR markers are better than microsatellite markers for current set of genotypes. However, the resolving power provides no information on the ability of a primer to reflect the genetic or taxonomic relationships of a group of genotypes under study (Prevost and Wilkinson 1999).

4.2 Polymorphic nature of markers and genetic diversity

EST-PCR and SSR markers showed different discriminating capacities in genetic diversity analysis. The capacity of a marker's potential in detecting genetic polymorphism and variation was shown with the parameters such as allele number (N_A), polymorphic information content (PIC), expected (*He*) and observed heterozygosity (*Ho*), inbreeding coefficient (F) and Shannon's index (I). Generally, the higher the values of any parameter for a marker system, the greater its capacity in detecting polymorphism and variation between genotypes. The higher value also suggests that the marker is more informative (Ojango 2011).

In our study, we found that EST-PCR and SSR analyses yielded a high mean allele number of 23 and 18.2 per locus (primer) in 63 blueberry genotypes, respectively, which is comparable with other cross-pollinated polyploidy species: Boches et al. (2006) reported

18 alleles for the highbush blueberries, Bian et al. (2014) reported 14.24 alleles per locus for 150 cultivated blueberry accessions, 24 alleles were reported in kiwifruit (*Actinidia* L.) by Zhen et al. (2004) and Liu et al. (2014) reported 10 alleles per locus for 54 EST-SSR primer pairs in 12 *Vaccinium* spp. Low numbers of SSR alleles per locus have been reported for self-pollinated crops like rice (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 [12.6 alleles per locus (Matsuoka et al. 2002); 8.02 alleles per locus (Reif et al. 2004)] and sunflower [12.0 alleles/locus (Tang and Knapp 2003)]. 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 [10.6 (Stajner et al. 2005), grapes [8 (Dangl et al. 2001); 11 (Martin et al. 2003); 24.4 (Lamboy and Alpha 1998)], apples [26.4 (Hokanson et al. 2001)] and tetraploid sour cherries [10.7 (Cantini et al. 2001)]. However, large number of alleles per locus is not always associated with an increase in ploidy level. In allotetraploid cotton *(Gossypium)*, which exhibits disomic inheritance, five alleles per locus were found (Liu et al. 2000), while tetraploid fescue *(Festuca)* had 2.71 alleles per locus (Saha et al. 2004). Eleven EST-PCR primer pairs and eight SSR primer pairs (89%) amplified more than four alleles. The amplification of PCR and SSR alleles from multiple loci in the genome has been reported in other plant species also (Fisher et al. 1998).

Polymorphic information content (PIC) is also one of the parameters that helps to identify the power of markers based on the marker's ability to provide polymorphic information for the genotypes. PIC provides an estimate of the discriminatory power of a locus by taking into account the number of alleles that are expressed and the relative

frequencies of those alleles (Ojango 2011). In this study, the average PIC value is higher for EST-PCR markers, 0.80, than for SSR markers, 0.77. However, the PIC values of the EST-PCR markers ranged from 0.44 for NA27 to 0.98 for CA16 and CA1785, and the PIC values for SSR markers ranged from 0.28 for CA236 to 0.98 for NA1040. These values can be comparable to other studies. Lamboy and Alpha (1998) observed the mean PIC value of 0.91 for 110 accessions of grape spices, where only five SSR markers were studied. Fifty four EST-SSR primer pairs exhibited polymorphism among 12 cultivars with an average PIC value of 0.77 (Liu et al. 2014). Cantini et al. (2001) used 10 SSR markers for 59 tetraploid cherry accessions, and discovered a mean PIC value of 0.81 (Cantini et al. 2001). Another study, using 50 EST-PCR primer pairs, found low mean PIC values of 0.19 in cultivated rubber trees and 0.22 in wild rubber tress (Li et al. 2012). In our study, we did not find much difference between the mean PIC values of EST-PCR and microsatellite markers, which suggests that both markers are highly polymorphic and able to find genetic diversity among blueberry genotypes. These results suggest high genetic diversity among blueberry genotypes in this study.

Marker heterozygosity gives an idea on marker's ability to provide heterozygous information. A high level of average heterozygosity at a locus could be expected to correlate with high levels of genetic variation at loci. Heterozygosity of genotype is affected by adaptive response of genotypes to environmental changes (Reed and Frankham 2003). There are two types of heterozygosity, observed (*Ho*) and expected (H_E). In this study, primer pairs showed very high values of observed as well as expected heterozygosity. EST-PCR showed average expected heterozygosity of 0.82 and observed heterozygosity of 0.34,

however SSR showed average expected heterozygosity of 0.78 and observed heterozygosity of 0.32. In other words, H_E and *Ho* values for EST-PCR are higher than those of SSR, indicating that EST-PCR markers are more powerful. These markers can show more heterozygosity among the genotypes from different populations and collection sites. These results are comparable with other observations done by other research groups in different species. Cantini et al. (2001) observed the mean observed and expected heterozygosity values of 0.95 and 0.81, respectively, using 10 SSR markers, where 59 tetraploid cherry accessions were analysed. In another study 12 blueberry cultivars were analysed using 54 EST-SSR markers, where mean H_E was 0.80 (Liu et al. 2014). Bian et al. (2014) analysed 150 blueberry accessions with 42 SSR markers, which gave an average H_E of 0.87. These results suggest that blueberry genotypes are very heterozygous in nature.

In the present study, the discrepancy between the values of expected and observed heterozygosity suggests some inbreeding among the blueberry genotypes. The inbreeding coefficients (*F*) are also given in the results, which come to an average of 0.60 for EST-PCR and 0.55 for SSR markers; this shows that the EST-PCR markers are more powerful in showing inbreeding behavior among genotypes. The higher the inbreeding value of marker from other markers, the greater its chance to detect inbreeding among the genotypes (Reed and Frankham 2003). As blueberries grow in wild, there are greater chances of inbreeding. The inbreeding coefficient is a very important parameter. Early acting inbreeding depression seems to be the primary factor in limiting self-fertility (Hokanson and Hancock 2001). Research shows that severe inbreeding depressions were encountered upon selfing early cultivars (Coville 1937). However, the blueberries have polysomic

inheritance (Ehlenfeldt 1994), which helps increase their heterozygosity level (Soltis and Soltis 2000).

Shannon's index (*I*), which is another diversity parameter, takes into account both the number of classes (allelic richness) and the distribution of individuals among classes (allelic evenness) (Boches et al. 2006). In contrast to other parameters such as expected heterozygosity, Shannon's index does not require knowledge of allele frequencies. Therefore, *I* is an accurate measure of diversity in polyploidy species where allele frequencies cannot be determined with certainty due to the difficulty in distinguishing the copy number of individual alleles (Boches et al. 2006). In present studies, the average Shannon's index of 2.61 for 11 EST-PCR, and 2.03 for nine SSR markers, were higher than those reported by Debnath (2014), who recorded a Shannon's index of 1.93 in mixed genotypes (including lowbush and half-high blueberry cultivars, and wild lowbush blueberry clones from different communities). Albert et al. (2005) recorded Shannon's index of 0.65 for *V. uliginosum* clones, and Albert et al. (2003) recorded 0.55 for 27 *V. myrtillus* clones. Pearson and Gustavsson (2001) reported Shannon's index of 0.57 for *V. vitis-idaea* and Bian et al. (2014) showed an average Shannon's index of 0.62 for 150 blueberry accessions. Only Boches et al. (2006), who used 28 SSRs to analyse 69 accessions, reported higher values of 9.77 for *V. corymbosum*. The results suggest that, the blueberries, because of their tetraploid genetic makeup are highly heterozygous and diverse in nature. In addition, wild clones among all blueberry genotypes are more diverse and heterozygous than highbush cultivars or clones (Boches et al. 2006), and as this study involves 56 lowbush wild clones, the markers provide higher Shannon's index values.

As seen from Table 10 and 11, it can be said that the genotypes from Newfoundland and Labrador (NL) group are more diverse than the genotypes from other groups, such as Prince Edward Island (PE) group, Quebec (QC) group, half-high and highbush cultivars (HHCU) group; and New Brunswick and lowbush cultivars (NB and LBCU) group . The NL group had significantly large number of alleles than any other groups. The same was true for other two parameters, H_E and PIC. These three diversity parameters did not differ significantly among PE, QC, HHCU, and LBCU and NB groups. The HHCU group detected a decreased number of alleles for both EST-PCR and SSR markers. Seven out of 11 EST-PCR markers and four out of nine SSR markers gave only one allele for this group. SSR markers detected more alleles for the HHCU group, 1.8, which is higher than EST-PCR markers, 1.4. These kind of comparative studies were shown by Boches et al. (2006) using 28 SSR markers in 69 highbush blueberry accessions, where primer pairs detected more number of average alleles (15.5) for wild clones, than the historical northern highbush, modern northern highbush and southern highbush cultivars (12.3, 12.6 and 13.1, respectively).

In this analysis, only three markers CA16, CA54 and CA1423, out of 11 EST-PCR markers detected more than one allele for all groups; other eight markers (CA175, CA287, CA1029, CA1105, CA1590, CA1785, NA27 and NA1068) which detected only one allele also detected null heterozygosity and null polymorphic information content, which give them less power to distinguish those particular genotypes from each other. The same can be interpreted for other markers which detected only one allele in those particular groups. For instance, CA175 detected one allele, null heterozygosity and PIC for QC group, which

include six QC clones. As the heterozygosity and PIC of this marker is null for these genotypes, this marker is not efficient to distinguish these genotypes from each other, but marker NA961, which detected 6 alleles, 0.83 H_E and 0.81 PIC, can distinguish these six genotypes with more power. The same can be explained for SSR markers (CA236, CA421, CA483, NA741 and NA800), which detected one allele for QC clones, and thus null H_E and PIC (Boches et al. 2006).

4.3 Cluster analysis and population structure

NJ, UPGMA, PCoA, STRUCTURE and AMOVA analyses were also used to assess the population structure and genetic diversity among genotypes. Three methods of multivariate molecular analysis: NJ, UPGMA and PCoA were employed to cluster the genotypes. Patterns of clustering based on NJ, UPGMA and PCoA were similar for most of the genotypes. Three NJ and UPGMA dendograms were created, each for EST-PCR, SSR and combined analysis. Similarly, three PCoA graphs were created. All six dendograms divided 63 accessions into five main clusters and all three PCoA graphs also divided genotypes into five groups. All six dendrograms depicted the genetic distinctiveness of half-high and highbush cultivars and the wild clones collected from different provinces.

NJ dendogram based on EST-PCR, SSR and combined markers' analysis separated half-high and highbush cultivars from lowbush blueberry clones and cultivars. In the EST-PCR dendogram, two QC (QC4 and QC9) clones were grouped with the half-high and highbush cultivars. Lowbush cultivars and a selection from open pollinated seedling of the

cultivar Fundy (FO), grouped together with NB clones. This makes it the geographic clustering, as both the lowbush blueberry cultivars included in a study are originated from the province of New Brunswick. Cultivar Fundy was selected from open pollinated seedling of Augusta lowbush blueberry cultivar (Lyrene 2002), and cultivar Brunswick was selected from local native lowbush blueberry clones of the province (Aalders et al. 1977). This grouping for NB clones and lowbush blueberry cultivars was also seen in NJ dendograms of SSR markers and combined marker data. Wild blueberry clones from the province of Quebec made a separate cluster from all other clones but the cluster was close to NL and NB clones and lowbush cultivars. NL clones were divided into four clusters. The relation among the NL clones was found to be almost the same in SSR and combined analysis dendogram. NL clones were collected from three different communities of NL province (two of them (North River and Shearstown) are 7 km apart from each other, and the other Logy Bay is 90 km apart from both of them) and their sample size was also large which gives more genetic pool to analyse and thus contributes to more genetic diversity. These clones have been not been domesticated by breeding, but are grown in wild, under natural conditions, to adapt them and optimize their growth, deterring competition from weeds, and minimizing the losses incurred by fungal diseases and insect damage (Yarborough, 1998).

In the NJ dendogram created using SSR marker data, half-high and highbush blueberry cultivars made a separated sub cluster (IVa) in cluster (IV) with other QC clones in another sub cluster (IVb). PE clones were divided into two separate clusters, but the genetic relatedness between PE clones was always seen with the other lowbush blueberry

clones from NL province (NL22, NL27 and NL35). A combined analysis also revealed five clusters for the NJ dendogram, and it clustered blueberry clones and cultivars the same way as seen in the NJ dendogram of EST-PCR and SSR markers. In the EST-PCR dendogram, two QC clones were seen with half-high and highbush blueberry cultivars, while they were not together in SSR and combined analysis. This shows the importance of using more than one type of marker for the genetic diversity analysis. If one marker is not powerful enough, then the other marker can compensate what the first one is lacking.

The UPGMA dendogram based on EST-PCR, SSR and combined marker data had five major clusters. These five clusters genetically separated half-high and highbush blueberry cultivars from lowbush blueberry clones and cultivars. Half-high and highbush blueberry cultivars formed a separate cluster. Lowbush blueberry cultivars Fundy and Brunswick, and a selection FO, grouped together with the clones from the province of New Brunswick, as they were grouped together in NJ analysis. Based on EST-PCR marker data, wild QC clones made a separate cluster, with two NL clones NL7 and NL9. NL clones were divided into three clusters. The same grouping and genotypes relatedness observed in this UPGMA dendogram also matched with the dendograms of SSR and combined analysis. In all three UPGMA dendograms, genetic relatedness between PE clones was always seen with the other lowbush blueberry clones (NL2, NL12, NL27 and NL28). In the UPGMA dendogram, created using combined marker data, PE clones made a separate cluster III and grouped together in it. Similarly, all six QC clones grouped in a separated sub-sub cluster (Vb(i)). The pattern of UPGMA combined analysis also matches the pattern of combined NJ analysis. In the EST-PCR UPGMA dendogram, two NL clones were seen

with QC clones, while they were not together in SSR and combined analysis. This shows the importance of using more than one type of marker for the genetic diversity analysis.

Debnath (2014), used 10 EST-PCR and two EST-SSR markers to differentiate 36 blueberry genotypes, which included 28 blueberry wild clones, six blueberry cultivars and two selections. The NJ dendogram created based on Jaccard's similarity matrix clustered all PE clones, NB clones and NL clones in separate clusters from each other. All half-high and highbush cultivars clustered separately in another cluster. This clustering was in agreement with the present study. This particular study was done in the same research group with same range of genotypes from different province of Atlantic Canada. But the number of genotypes included was low and the study used less number of primers for the study than what have been used in this study. Boches et al. (2006) used 69 accessions representing wild and domesticated highbush blueberry germplasm. The diversity analysis was performed among them using 28 SSR markers. NJ dendogram placed wild and cultivated blueberries in separate clusters. Southern highbush blueberries formed a separate cluster from northern highbush blueberries. In present study also half-high and highbush blueberry cultivars formed separate clusters from lowbush blueberry clones and cultivars. In addition, lowbush blueberry clones and cultivars also detected separate genotype clustering based on their location of origin, which as in agreement with Liu et al. (2014), who observed separate clustering of wild and cultivated blueberry species from a total of 30 blueberries genotypes, using 54 EST-SSR primer pairs.

Another multivariate molecular analysis, PCoA also formed five groups in EST-PCR, SSR and combined analyses. PCoA based on EST-PCR markers and the combined

markers created separate groups for half-high and highbush blueberry cultivars. Separate groups were also seen for six QC clones, 11 PE clones and 37 NL clones. Lowbush cultivars Fundy and Brunswick, which were obtained from the wild stands of the province of NB, grouped together with the clones from NB province. Thus it can be said that EST-PCR markers and combination of both markers distinguished all the wild clones according to their location of origin and separated half-high and highbush blueberry cultivars from wild clones. This kind of association and grouping was also seen in NJ and UPGMA dendograms. In PCoA of 36 blueberry genotypes, Debnath (2014) observed that all genotypes separated according to their geographic location and made four groups. In that study, similarities were seen among different lowbush blueberry clones, which was also seen in the present study, where SSR PCoA grouped QC and PE blueberry clones together. Liu et al. (2014) used 54 EST-SSR markers, where they were successful in grouping different accession of same species together in PCO analysis.

Model based Bayesian cluster analysis of EST-PCR, SSR and combined data set was implemented to determine whether any genetic structure could be detected without knowledge of sampling locations of individuals. In crop plants, STRUCTURE analysis has been employed to identify the presence of subgroups in core collections (Abdurakhmonov et al. 2008; Anderson et al. 2009; Debnath 2014; Kottapalli et al. 2011; Sakiroglu et al. 2010; Tyagi et al. 2014), while in two instances STRUCTURE analysis has been employed with blueberry genotypes (Bian et al. 2014; Debnath, 2014). For the STRUCTURE analysis, the number of clusters best fitting the EST-PCR, SSR and combined data were K = 2 and 6, K = 6 and K = 2 and 6, respectively, as indicated by the modal value of ΔK

(Figure 15, 19 and 22, respectively). Two studies with blueberry genotypes found various K values such as K = 4 for 36 genotypes using EST-PCR and EST-SSR markers (Debnath, 2014) and K = 3 for 89 highbush blueberry accessions using 54 EST-SSR markers (Bian et al. 2014). In this study we have structured groups according to both L(K) and ΔK graphs.

STRUCTURE analysis for EST-PCR data revealed two and six clusters. High level of admixture was seen among genotypes in both structure bar plots (Oliveria et al. 2012). Oliveira et al. (2012) observed that most tetraploid wheat accessions received alleles from more than a single group, and the "admixture" model of the STRUCTURE analysis produced results that seemed more informative about gene flow between the different groups. STRUCTURE result for K=6 showed finer structure of genotypes than K=2. It (K=6) shows the grouping of all the genotypes based on their respective geographic location of origin/ collection site. NL clones were divided mainly into two clusters, which suggests the diversity among NL clones in comparison to other genotypes. Eleven PE clones also created a single cluster with NL24. NL24 detected high ratio of admixture with cluster IV, which included all NL clones suggesting that NL24 shares genes from both PE and NL clones. All highbush and half-high blueberry cultivars made a separate cluster (cluster II) which included only four cultivars.

STRUCTURE for microsatellite markers revealed only six clusters. There was no broader structure detected at K=2, but it gave finer structure at K=6. All 11 PE clones made a separate cluster just like EST-PCR based STRUCTURE analysis. Two NB clones (NB1 and NB2); two lowbush blueberry cultivars (BR and FU); and one selection (FO), which originated from same geographic collection site, made a separate cluster V, as also seen in

EST-PCR analysis. Cultivar Fundy showed high level of admixture from cluster VI, which suggest that Fundy shares genetic material with NL clones. Six QC clones, made a separate cluster; the same clustering was observed in the STURCTURE analysis of EST-PCR.

STRUCTURE analysis of combined data revealed two structure, one broader structure at K=2 with two clusters and a finer structure at K=6 with six clusters. All half-high and high bush cultivars were in the same cluster as was seen in two individual EST-PCR and SSR based STRUCTURE analyses. All PE clones formed a separate cluster. One cluster only contained two lowbush cultivars (FU and BR), two NB clones (NB1 and NB2) and a selection (FO). All six QC clones grouped into one separate cluster, as seen in previous two individual STRUCTURE analyses. NL clones were divided into two clusters seen in previous two STRUCTURE analyses. STRUCTURE analysis depicts how these markers help differentiate clones and cultivars based on their respective geographic locations of origin or collection sites. Debnath (2014) used STRUCURE analysis to find genetic structure among 36 blueberry genotypes using 10 EST-PCR and two EST-SSR markers, where four clusters separated half-high blueberry cultivars and lowbush blueberry wild clones and cultivars into genetically distinct clusters, which was also seen in this study. In this study NL clones were also divided into two separate clusters which were seen together in one cluster by Debnath (2014). This can be due to the fact that more number of genotype and markers were studied into this project. Bian et al. (2014) showed the substantial interspecific differentiation between rabbiteye blueberry and other blueberry types, using 89 highbush blueberries, where these blueberries were divided into three clusters. This stratification method is particularly important to blueberry improvement

programs. The population structure assessment and the genetic diversity measurement together can help to better utilize and manage diversity within blueberries in future genetic studies in blueberry and other *Vaccinium* species.

As it is seen from the graphs of STRUCTURE, UPGMA, NJ and PCoA, few clones (NL8, NL27 and NL28) classified by the STRUCTURE analysis, fall into different groups of UPGMA, NJ or PCoA. The reason can be STRUCTURE software, where the loci within a population were assumed to be in Hardy Weinberg equilibrium (HWE) and linkage disequilibrium. However, non-random mating, mutations, selection, random genetic drift, gene flow and meiotic drive in fact might have disturbed HWE (Hardy, 1908). In all cluster analysis it was largely found that separate clusters were made according to the respective geographic origin of the genotypes.

The AMOVA analysis detected abundant variation among genotypes within communities, among communities with groups (provinces) and among groups (provinces). High degree of variation among genotypes within communities (46.04% for EST-PCR markers, 41.49% for SSR markers and 44.25% for combined markers) was found. These results were confirmed with the previous studies of 36 blueberry accessions using EST-PCR and EST-SSR markers (43.52%) (Debnath, 2014) and with wild lowbush blueberry clones using ISSR marker (73%) (Debnath, 2009). Other reports of AMOVA analysis in other *Vaccinium* species include: lingonberry using ISSR (90%) (Debnath 2007c) and RAPD markers (89%) (Pearson and Gustavsson 2001); *V. uliginosum* (96%) (Albert et al. 2005); and *V. myrtillus* using RAPD marker (86%) (Albert et al. 2004). In present study, high level of variation was also observed among communities within groups (17.42% for

EST-PCR markers, 32.01% for SSR markers and 23.15% for combined markers), which can be explained by geographic distribution among the communities within groups (province). In the present study, there was also lot of variation among groups (36.54% for EST-PCR markers, 26.50% for SSR markers and 32.60% for combined markers), which was less than reported by Debnath (2014), 7.46%. This can be explained by existence of diverse gene pool from province to province. This diversity was also seen in previous analysis where genotypes formed clusters according to their geographic collection site, and thus all genotypes from the same province made separate clusters from other provinces. This explains large percentage variation among groups. AMOVA analysis is helpful in identifying percentage variation at different levels of population.

From the results of STRUCTURE, UPGMA, NJ, PCoA and AMOVA analyses, it is evident that genotypes were very diverse and there was more variation among genotypes from the same communities. There was no report on variability study with EST-PCR and SSR markers with so many genotypes as used in this study. Debnath (2014) also observed that 12 EST primer pairs (10 EST-PCR and two EST- SSR) distinguished 36 genotypes according to their geographic location and half-high and highbush blueberries grouped together. In the present study, 11 EST-PCR, seven EST-SSR and two genomic SSR markers were helpful and sufficient to differentiate 63 blueberry genotypes.

5 Summary

Genetic variation was studied among 56 wild lowbush blueberry clones, collected from Newfoundland and Labrador (NL1 - NL37), Prince Edward Island (PE1 – PE11), Quebec (QC1, QC2, QC4, QC5, QC7 and QC9), and New Brunswick (NB1 and NB2); six blueberry cultivars including three half-high (Chippewa, St. Cloud and Patriot), one highbush (Polaris) and two lowbush cultivars (Brunswick and Fundy); and one lowbush blueberry selection, obtained from the lowbush blueberry cultivar Fundy, using 11 EST-PCR, seven EST-SSR and two genomic SSR markers.

The markers detected high polymorphism among 63 genotypes. A combined analysis of two types of markers produced 140 polymorphic bands and 436 alleles. The average resolving power of the EST-PCR and microsatellite markers were 4.6 and 2.6, respectively. Average PIC values for EST-PCR and microsatellite were 0.80 and 0.77, respectively. The other genetic diversity parameters were expected and observed heterozygosity, inbreeding coefficient and Shannon's index which had average value of 0.82, 0.34, 0.6 and 2.6 for the EST-PCR markers, respectively; and 0.78, 0.32, 0.55 and 2.0 for microsatellite markers, respectively.

The clustering analysis was done for two types of markers individually and for the combined data. Multivariate clustering analyses, UPGMA, NJ and PCoA, formed five groups in each analysis and clustered the genotypes according to their place of origin/ geographic origin.

STRUCTURE analysis divided 63 blueberry genotypes into two kinds of structure for EST-PCR and combined data analysis. One broader structure at K=2, which divided genotypes in two clusters, and a finer structure at K=6, which divided genotypes in six groups. Microsatellite markers divided genotypes only in six clusters. The Clones and cultivars were grouped according to their place of origin. The grouping was same for most of the genotypes. Substantial amount of differentiation was found among the blueberry genotypes from different provinces.

The results of NJ, UPGMA, PCO and STRUCTURE analyses were confirmed by AMOVA analysis. AMOVA detected a sufficient account of variation among genotypes within communities, among communities within groups/ provinces, and among groups/provinces with EST-PCR and microsatellite markers, and also with the combination of two markers.

In present study, the markers detected a high level of genetic variation among 63 blueberry genotypes, indicating that these markers may be suitable for further molecular analysis such as genetic diversity analysis with other blueberry genotypes. These kind of analyses will be helpful in DNA fingerprinting, for selecting useful clones as parents in a breeding program, management of blueberry germplasm and for conservation of intellectual property rights.

6 Literatures cited

AAFC. 2011. Canadian blueberries. The true taste of nature. [Canadian blueberries. Agriculture and Agri-Food Canada. Government of Canada. AAFC No. 11509E] Available: http://www.agr.gc.ca/eng/industry-markets-and-trade/exporting-importing-and-buying-assistance/buy-from-canada/canadian-blueberries/?id=1426167712421 .

AAFC. 2014. Plant gene resources of Canada. [Agriculture and Agri-Food Canada. Government of Canada.] Available: http://pgrc3.agr.gc.ca/index_e.html .

Aalders, L. E., Hall, I. V., Jackson, L. P. 1977. Brunswick lowbush blueberry. Canadian Journal of Plant Science. **57**: 301-301.

Abdurakhmonov, I. Y., Kohel, R. J., Yu, J. Z., Pepper, A. E., Abdullaev, A. A., Kushanov, F. N., Salakhutdinov, I. B., Buriev, Z. T., Saha, S., Scheffler, B. E., Jenkins, J. N., Abdukarimov, A. 2008. Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm. Genomics **92**(6): 478-487.

Adams, R. P. and Rieseberg, L. H. 1998. The effects of non-homology in RAPD bands on similarity and multivariate statistical ordination in *Brassica* and *Helianthus*. Theoretical and Applied Genetics. **97**: 323-326.

Albert, T., Raspe, O., Jacquemart, A. L. 2003. Clonal structure in *Vaccinium myrtillus* L. revealed by RAPD and AFLP markers. International Journal of Plant Sciences. **164**: 649-655.

Albert, T., Raspe, O., Jacquemart, A. 2004. Clonal diversity and genetic structure in *Vaccinium myrtillus* populations from different habitats. Belgian Journal of Botany **137**(2): 155-162.

Albert, T., Raspe, O., Jacquemart, A. L. 2005. Diversity and spatial structure of clones in *Vaccinium uliginosum* populations. Canadian Journal of Botany-Revue Canadienne De Botanique. **83**: 211-218.

Aldrich, P. R. and Doebley, J. 1992. Restriction fragment variation in the nuclear and chloroplast genomes of cultivated and wild *Sorghum bicolor*. Theoretical and Applied Genetics. **85**: 293-302.

Anderson, W. F., Maas, A., Ozias-Akins, P. 2009. Genetic variability of a forage bermudagrass core collection. Crop Science. **49**(4): 1347-1358.

Andres-Lacueva, C., Shukitt-Hale, B., Galli, R. L., Jauregui, O., Lamuela-Raventos, R. M., Joseph, J. A. 2005. Anthocyanins in aged blueberry-fed rats are found centrally and may enhance memory. Nutritional Neuroscience. **8**: 111-120.

Aruna, M., Austin, M. E., Oziasakins, P. 1995. Randomly amplified polymorphic DNA-fingerprinting for identifying rabbiteye blueberry (*Vaccinium-ashei* Reade) cultivars. Journal of the American Society for Horticultural Science. **120**: 710-713.

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

Ballington, J. R. 2001. Collection, utilization, and preservation of genetic resources in *Vaccinium*. HortScience. **36**: 213-220.

Beckmann, J. S. and Soller, M. 1990. Toward a unified approach to genetic-mapping of eukaryotes based on sequence tagged microsatellite sites. Bio-Technology. **8**: 930-932.

Belaj, A., Satovic, Z., Cipriani, G., Baldoni, L., Testolin, R., Rallo, L., Trujillo, I. 2003. Comparative study of the discriminating capacity of RAPD, AFLP and SSR markers and of their effectiveness in establishing genetic relationships in olive. Theoretical and Applied Genetics. **107**: 736-744.

Bell, D. J., Rowland, L. J., Polashock, J. J., Drummond, F. A. 2008. Suitability of EST-PCR markers developed in highbush blueberry for genetic fingerprinting and relationship studies in lowbush blueberry and related species. Journal of the American Society for Horticultural Science. **133**: 701-707.

Bian, Y., Ballington, J., Raja, A., Brouwer, C., Reid, R., Burke, M., Wang, X., Rowland, L. J., Bassil, N., Brown, A. 2014. Patterns of simple sequence repeats in cultivated blueberries (*Vaccinium* section *Cyanococcus* spp.) and their use in revealing genetic diversity and population structure. Molecular Breeding. **34(**2): 675-689.

Block, G., Patterson, B., Subar, A. 1992. Fruit, vegetables, and cancer prevention: A review of the epidemiological evidence. Nutrition and Cancer. **18**: 1-29.

Boches, P. S., Bassil, N. V., Rowland, L. J. 2005. Microsatellite markers for *Vaccinium* from EST and genomic libraries. Molecular Ecology Notes. **5**: 657-660.

Boches, P., Bassil, N. V., Rowland, L. 2006. Genetic diversity in the highbush blueberry evaluated with microsatellite markers. Journal of the American Society for Horticultural Science. **131**: 674-686.

Botstein, D., White, R. L., Skolnick, M., Davis, R. W. 1980. Construction of a genetic-linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics. **32**: 314-331.

Brazelton, C. 2013. World blueberry acreage and production. [North American Blueberry Council.]

Available: http://www.chilealimentos.com/2013/phocadownload/Aprocesados_congelados/nabc_2012-world-blueberry-acreage-production.pdf .

Brightwell, W.T., Woodard, O., Darrow, G.M., Scott, D.H. 1955. Observations on breeding blueberries for the southeast. Proceedings of American Society of Horticulture Science. **65**: 5.

Brown, J., Caligari, P., Campos, H. 2014. Plant breeding. John Wiley & Sons, Hoboken, NJ.

Brown, T. A. 2010. Gene cloning and DNA analysis an introduction. 6th ed. Hoboken: Wiley, Hoboken. .

Bruederle, L. P. and Vorsa, N. 1994. Genetic differentiation of diploid blueberry, *Vaccinium* sect *Cyanococcus* (Ericaceae). Systematic Botany. **19**: 337-349.

Burgher, K. L., Jamieson, A. R., Lu, X. W. 2002. Genetic relationships among lowbush blueberry genotypes as determined by randomly amplified polymorphic DNA analysis. Journal of the American Society for Horticultural Science. **127**: 98-103.

Canter, P. H. and Ernst, E. 2004. Anthocyanosides of *Vaccinium* myrtillus (bilberry) for night vision - A systematic review of placebo-controlled trials. Survey of Ophthalmology. **49**: 38-50.

Cantini, C., Iezzoni, A., Lamboy, W., Boritzki, M., Struss, D. 2001. DNA fingerprinting of tetraploid cherry germplasm using simple sequence repeats. Journal of American Society of Horticultural Science. **126**(2): 205-209.

Cato, S. A., Gardner, R. C., Kent, J., Richardson, T. E. 2001. A rapid PCR-based method for genetically mapping ESTs. Theoretical and Applied Genetics. **102**: 296-306.

Celik, H. and Islam, A. 2014. Blueberry species introduction, selection and cultivation practice in northeastern part of Anatolia. X International Symposium on *Vaccinium* and Other Superfruits. **1017**: 441-446.

Chapman, M. A., Hvala, J., Strever, J., Matvienko, M., Kozik, A., Michelmore, R. W., Tang, S., Knapp, S. J., Burke, J. M. 2009. Development, polymorphism, and cross-

taxon utility of EST-SSR markers from safflower (carthamus tinctorius L.). Theor. Appl. Genet. **120**(1): 85-91.

Cho, Y. G., Ishii, T., Temnykh, S., Chen, X., Lipovich, L., McCouch, S. R., Park, W. D., Ayres, N., Cartinhour, S. 2000. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theoretical and Applied Genetics. **100** (5): 713-722.

Coville, F. V. (1937). Improving the wild blueberry. Pages 559-574 *in*: USDA Yearbook of Agricultire 1937. 75th congress, 1st session, house document No. 28.

Crespan, M. 2004. Evidence on the evolution of polymorphism of microsatellite markers in varieties of *Vitis vinifera* L. Theoretical and Applied Genetics. **108**(2): 231-237.

Dangl, G., Mendum, M., Prins, B., Walker, M., Meredith, C., Simon, C. 2001. Simple sequence repeat analysis of a clonally propagated species: A tool for managing a grape germplasm collection. Genome **44**(3): 432-438.

Darrow, G. and Scott DH. 1966. Varieties and their characteristics. Pages 95-95 *in:* P, E., ed. Blueberry culture. Quinn and Boden Company, New Jersey.

Debnath, S. C. 2005. Differentiation of *Vaccinium* cultivars and wild clones using RAPD markers. Journal of Plant Biochemistry and Biotechnology. **14**: 173-177.

Debnath, S. C. 2007a. An assessment of the genetic diversity within a collection of wild cranberry (*Vaccinium macrocarpon* ait.) clones with RAPD-PCR. Genetic Resources and Crop Evolution. **54**: 509-517.

Debnath, S. C. 2007b. Inter-simple sequence repeat (ISSR)-PCR analysis to assess genetic diversity in a collection of wild cloudberry (*Rubus chamaemorus* L.) clones. Journal of Horticultural Science & Biotechnology. **82**: 727-732.

Debnath, S. C. 2007c. Inter simple sequence repeat (ISSR) to assess genetic diversity within a collection of wild lingonberry (*Vaccinium vitis-idaea* L.) clones. Canadian Journal of Plant Science. **87**: 337-344.

Debnath, S. C. 2009. Development of ISSR markers for genetic diversity studies in *Vaccinium* angustifolium. Nordic Journal of Botany. **27**: 141-148.

Debnath, S. C. 2011. Conventional methods combined with biotechnology in blueberry improvement. Agriculture and Agri-Food Canada. Catalog. No. A22e528/2011E-PDF, AAFC No. 1142E, ISBN 978-1-100-18251-3.

Debnath, S. C. 2014. Structured diversity using EST-PCR and EST-SSR markers in a set of wild blueberry clones and cultivars. Biochemical Systematics and Ecology. **54**: 337-347.

Debnath, S. C., Siow, Y. L., Petkau, J., An, D., Bykova, N. V. 2012. Molecular markers and antioxidant activity in berry crops: Genetic diversity analysis. Canadian Journal of Plant Science. **92**: 1121-1133.

Dhanaraj, A. L., Slovin, J. P., Rowland, L. J. 2004. Analysis of gene expression associated with cold acclimation in blueberry floral buds using expressed sequence tags. Plant Science. **166**: 863-872.

Die, J. V. and Rowland, L. J. 2013. Advent of genomics in blueberry. Molecular Breeding. **32**: 493-504.

Duvick, D. N. 1984. Genetic diversity in major farm crops on the farm and in reserve. Economic Botany. **38**: 161-178.

Earl, D. A. and Vonholdt, B. M. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources. **4**: 359-361. http://taylor0.biology.ucla.edu/structureHarvester/#

Ehlenfeldt, M. 1994. The genetic composition and tetrasomic inbreeding coefficients of highbush blueberry cultivars. Hortscience **29**(11): 1342-1345.

Elston, R. C. 2005. Polymorphism information content. *in:* Encyclopedia of biostatistics. John Wiley & Sons, Ltd,

Esselink, G., Smulders, M., Vosman, B. 2003. Identification of cut rose (*Rosa hybrida*) and rootstock varieties using robust sequence tagged microsatellite site markers. Theoretical and Applied Genetics. **106**(2): 277-286.

Evanno, G., Regnaut, S., Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molecular Ecology. **14**: 2611-2620.

Excoffier, L., Laval, G., Schneider, S. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evolutionary Bioinformatics. **1**: 47-50.

Excoffier, L., Smouse, P. E., Quattro, J. M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics **131**(2): 479-491.

Fajardo, D., Morales, J., Zhu, H., Steffan, S., Harbut, R., Bassil, N., Hummer, K., Polashock, J., Vorsa, N., Zalapa, J. 2013. Discrimination of American cranberry cultivars and assessment of clonal heterogeneity using microsatellite markers. Plant Molecular Biology Reporter. **31**: 264-271.

Falush, D., Stephens, M., Pritchard, J. K. 2007. Inference of population structure using multilocus genotype data: Dominant markers and null alleles. Molecular Ecology Notes. **7**: 574-578.

Falush, D., Stephens, M., Pritchard, J. K. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics. **164**: 1567-1587.

Finn, C., Luny, J., Wildung, D. 1990. Half-high blueberry cultivars. Fruit Varieties Journal. **44**(2): 63-68.

Fisher, P., Richardson, T., Gardner, R. 1998. Characteristics of single- and multi-copy microsatellites from *Pinus radiata*. Theoretical and Applied Genetics. **96**(6-7): 969-979.

Galletta, G. J. and Ballington, J. R. 1996. Blueberries, cranberries and lingonberries. Pages 1-107 *in:* Janick, J. and Moore, J. N., eds. Fruit breeding, volume II: Vine and small fruit crops. John Wiley and Sons, Inc, Prentice Hall, New York, USA.

Gil-Ariza, D. J., Amaya, I., Lopez-Aranda, J. M., Sanchez-Sevilla, J. F., Botella, M. A., Valpuesta, V. 2009. Impact of plant breeding on the genetic diversity of cultivated strawberry as revealed by expressed sequence tag-derived simple sequence repeat markers. Journal of the American Society for Horticultural Science. **134**: 337-347.

Gilbert, J. E., Lewis, R. V., Wilkinson, M. J., Caligari, P. D. S. 1999. Developing an appropriate strategy to assess genetic variability in plant germplasm collections. Theoretical and Applied Genetics. **98**: 1125-1131.

Gough, Robert E. (Robert Edward). 1994. The highbush blueberry and its management. Food Products Press, New York. USA.

GRIN. 2015. USDA, ARS, national genetic resources program.*germplasm resources information network - (GRIN)* [online database]. National germplasm resources laboratory, beltsville, maryland. [Genus *Vaccinium* L.] Available: https://npgsweb.ars-grin.gov/gringlobal/taxonomygenus.aspx?id=12610.

Gupta, M., Chyi, Y. S., Romeroseverson, J., Owen, J. L. 1994. Amplification of DNA markers from evolutionarily diverse genomes using single primers of simple-sequence repeats. Theoretical and Applied Genetics. **89**: 998-1006.

Haghighi, K. and Hancock, J. F. 1992. DNA restriction fragment length variability in the genomes of highbush blueberry. HortScience. **27**: 44-47.

Halliwell, B. 2007. Dietary polyphenols: Good, bad, or indifferent for your health? Cardiovascular Research. **73**: 341-347.

Hamada, H. and Kakunaga, T. 1982. Potential z-DNA forming sequences are highly dispersed in the human genome. Nature. **298**: 396-398.

Hameed, U., Pan, Y. B., Muhammad, K., Afghan, S., Iqbal, J. 2012. Use of simple sequence repeat markers for DNA fingerprinting and diversity analysis of sugarcane (*Saccharum* spp) cultivars resistant and susceptible to red rot. Genetics and Molecular Research **11**(2): 1195-1204.

Hardy, G. H. 1908. Mendelian proportions in a mixed population. Science. **28**: 49-50.

Hawkes, J. G. 1977. Importance of wild germplasm in plant breeding. Euphytica. **26**: 615-621.

Hepler, P. and Draper, A. 1976. Patriot blueberry. Hortscience. **11**(3): 272-272.

Hokanson, K. and Hancock, J. 1998. Levels of allozymic diversity in diploid and tetraploid *Vaccinium* sect. *Cyanococcus* (blueberries). Canadian Journal of Plant Science. **78**: 327-332.

Hokanson, S., Lamboy, W., Szewc-McFadden, A., McFerson, J. 2001. Microsatellite (SSR) variation in a collection of malus (apple) species and hybrids. Euphytica **118**(3): 281-294.

Horvath, A., Sanchez-Sevilla, J. F., Punelli, F., Richard, L., Sesmero-Carrasco, R., Leone, A., Hoefer, M., Chartier, P., Balsemin, E., Barreneche, T., Denoyes, B. 2011. Structured diversity in octoploid strawberry cultivars: Importance of the old European germplasm. Annals of Applied Biology. **159**: 358-371.

Jaramillo-Correa, J. P., Verdu, M., Gonzalez-Martinez, S. C. 2010. The contribution of recombination to heterozygosity differs among plant evolutionary lineages and life-forms. Bmc Evolutionary Biology **10**: 22.

Jeyabalan, J., Aqil, F., Munagala, R., Annamalai, L., Vadhanam, M. V., Gupta, R. C. 2014. Chemopreventive and therapeutic activity of dietary blueberry against estrogen-mediated breast cancer. Journal of Agricultural and Food Chemistry. **62**: 3963-3971.

Jones, C. J., Edwards, K. J., Castaglione, S., Winfield, M. O., Sala, F., vandeWiel, C., Bredemeijer, G., Vosman, B., Matthes, M., Daly, A., Brettschneider, R., Bettini, P., Buiatti, M., Maestri, E., Malcevschi, A., Marmiroli, N., Aert, R., Volckaert, G., Rueda, J., Linacero, R., Vazquez, A., Karp, A.1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Molecular Breeding. **3**: 381-390.

Joseph, J. A., Denisova, N. A., Arendash, G., Gordon, M., Diamond, D., Shukitt-Hale, B., Morgan, D. 2003. Blueberry supplementation enhances signaling and prevents behavioral deficits in an Alzheimer disease model. Nutritional Neuroscience. **6**: 153-162.

Kalt, W., Forney, C. F., Martin, A., Prior, R. L. 1999. Antioxidant capacity, vitamin C, phenolics, and anthocyanins after fresh storage of small fruits. Journal of Agricultural and Food Chemistry. **47**: 4638-4644.

Kanaya, N., Adams, L., Takasaki, A., Chen, S. A. 2014. Whole blueberry powder inhibits metastasis of triple negative breast cancer in a xenograft mouse model through modulation of inflammatory cytokines. Nutrition and Cancer-an International Journal. **66**: 242-248.

Kesseli, R. V., Paran, I., Michelmore, R. W. 1994. Analysis of a detailed genetic-linkage map of *Lactuca sativa* (lettuce) constructed from RFLP and RAPD markers. Genetics. **136**: 1435-1446.

Kim, C. S., Lee, C. H., Shin, J. S., Chung, Y. S., Hyung, N. I. 1997. A simple and rapid method for isolation of high quality genomic DNA from fruit trees and conifers using PVP. Nucleic Acids Research. **25**: 1085-1086.

Kottapalli, P., Upadhyaya, H. D., Kottapalli, K. R., Payton, P., Dwivedi, S., Burow, M., David, K. O., Sanogo, S., Puppala, N. 2011. Population structure and diversity in Valencia peanut germplasm collection. Crop Sci. **51**(3): 1089-1100.

Krebs, S. L. and Hancock, J. F. 1989. Tetrasomic inheritance of isoenzyme markers in the highbush blueberry, *Vaccinium corymbosum* l. Heredity. **63**: 11-18.

Kreher, S. A., Fore, S. A., Collins, B. S. 2000. Genetic variation within and among patches of the clonal species, *Vaccinium stamineum* L. Molecular Ecology. **9**: 1247-1252.

Kumar, P., Gupta, V. K., Misra, A. K., Modi, D. R., Pandey, B. K. 2009. Potential of molecular markers in plant biotechnology. Plant Omics. **2**: 141-162.

Lamboy, W. and Alpha, C. 1998. Using simple sequence repeats (SSRs) for DNA fingerprinting germplasm accessions of grape (*Vitis* L.) species. Journal of the American Society for Horticultural Science. **123**(2): 182-188.

Legendre, P. and Legendre, L. 1998. Numerical ecology. 2nd ed. Elsevier, Amsterdam.

Lenne, J. M. and Wood, D. 1991. Plant-diseases and the use of wild germplasm. Annual Review of Phytopathology. **29**: 35-63.

Levi, A. and Rowland, L. J. 1997. Identifying blueberry cultivars and evaluating their genetic relationships using randomly amplified polymorphic DNA (RAPD) and simple sequence repeat-(SSR-) anchored primers. Journal of the American Society for Horticultural Science. **122**: 74-78.

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

Li, D., Xia, Z., Deng, Z., Liu, X., Dong, J., Feng, F. 2012. Development and characterization of intron-flanking EST-PCR markers in rubber tree (H*evea brasiliensis muell.* Arg.). Molecular Biotechnology. **51**(2): 148-159.

Li, Y. C., Korol, A. B., Fahima, T., Beiles, A., Nevo, E. 2002. Microsatellites: Genomic distribution, putative functions and mutational mechanisms: A review. Molecular Ecology. **11**: 2453-2465.

Liu, K. J. and Muse, S. V. 2005. PowerMarker: An integrated analysis environment for genetic marker analysis. Bioinformatics. **21**: 2128-2129.

Liu, S., Cantrell, R., McCarty, J., Stewart, J. 2000. Simple sequence repeat-based assessment of genetic diversity in cotton race stock accessions. Crop Science **40**(5): 1459-1469.

Liu, Y., Liu, S., Liu, D., Wei, Y., Liu, C., Yang, Y., Tao, C., Liu, W. 2014. Exploiting EST databases for the development and characterization of EST-SSR markers in blueberry (*Vaccinium*) and their cross-species transferability in *Vaccinium* spp. Scientia Horticulturae. **176**: 319-329.

Liu, Z. J., Karsi, A., Dunham, R. A. 1999. Development of polymorphic EST markers suitable for genetic linkage mapping of catfish. Marine Biotechnology. **1**: 437-447.

Luby, J. 1991. Breeding cold-hardy fruit crops in Minnesota. Hortscience **26**(5): 507-512.

Lynch, M. 1990. The similarity index and DNA fingerprinting. Molecular Biology and Evolution. **7**: 478-484.

Lyrene, P. 2002. Blueberry. HortScience. **37**: 252.

Lyrene, P. M., Vorsa, N., Ballington, J. R. 2003. Polyploidy and sexual polyploidization in the genus *Vaccinium*. Euphytica. **133**: 27-36.

MacIntosh, G. C., Wilkerson, C., Green, P. J. 2001. Identification and analysis of *Arabidopsis* expressed sequence tags characteristic of non-coding RNAs. Plant Physiology. **127**: 765-776.

MacKenzie, K. 1997. Pollination requirements of three highbush blueberry (*Vaccinium corymbosum* L.) cultivars. Journal of the American Society for Horticultural Science. **122**(6): 891-896.

Mahy, G., Bruederle, L. P., Connors, B., van Hofwegen, M., Vorsa, N. 2000. Allozyme evidence for genetic autopolyploidy and high genetic diversity in tetraploid cranberry, *Vaccinium oxycoccos* (Ericaceae). American Journal of Botany. **87**: 1882-1889.

Martin, J., Borrego, J., Cabello, F., Ortiz, J. 2003. Characterization of spanish grapevine cultivar diversity using sequence-tagged microsatellite site markers. Genome **46**(1): 10-18.

Matsuoka, Y., Mitchell, S. E., Kresovich, S., Goodman, M., Doebley, J. 2002. Microsatellites in Zea - variability, patterns of mutations, and use for evolutionary studies. Theoretical and Applied Genetics. **104**: 436-450.

Miesfeld, R., Krystal, M., Arnheim, N. 1981. A member of a new repeated sequence family which is conserved throughout eukaryotic evolution is found between the human delta-globin and beta-globin genes. Nucleic Acids Research. **9**: 5931-5947.

Morgante, M., Hanafey, M., Powell, W. 2002. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nature Genetics. **30**: 194-200.

Morgante, M., Rafalski, A., Biddle, P., Tingey, S., Olivieri, A. M. 1994. Genetic-mapping and variability of 7 soybean simple sequence repeat loci. Genome. **37**: 763-769.

Nadarajan, N., Vivekanandan, P., Henry, R. J. 1999. Use of molecular markers for diversity analysis in rice. Indian Journal of Genetics and Plant Breeding. **59**: 247-259.

Nagy, S., Poczai, P., Cernak, I., Gorji, A. M., Hegedus, G., Taller, J. 2012. PICcalc: An online program to calculate polymorphic information content for molecular genetic studies. Biochem. Genet. **50**(9-10): 670-672.

NASS. 2015. National statistics for blueberries. National agricultural statistics service. United States Department of Agriculture. .

Naumann, W. D. 1993. Overview of the *Vaccinium* industry in Western Europe. Acta Horticulture (ISHS). **346**: 53-58.

Nei, M. 1987. Molecular evolutionary genetics. Columbia university press, New York. USA.

Nei, M. and Kumar, S. 2000. Molecular evolution and phylogenetics. Oxford University Press, New York.

Novy, R. G. and Vorsa, N. 1996. Evidence for RAPD heteroduplex formation in cranberry: Implications for pedigree and genetic-relatedness studies and a source of co-dominant RAPD markers. Theoretical and Applied Genetics. **92**: 840-849.

Novy, R. G. and Vorsa, N. 1995. Identification of intracultivar genetic-heterogeneity in cranberry using silver-stained RAPDS. HortScience. **30**: 600-604.

Novy, R. G., Vorsa, N., Patten, K. 1996. Identifying genotypic heterogeneity in 'McFarlin' cranberry: A randomly-amplified polymorphic DNA (RAPD) and phenotypic analysis. Journal of the American Society for Horticultural Science. **121**: 210-215.

Novy, R. G., Kobak, C., Goffreda, J., Vorsa, N. 1994. RAPDs identify varietal misclassification and regional divergence in cranberry *Vaccinium macrocarpon* (Ait) pursh. Theoretical and Applied Genetics. **88**: 1004-1010.

Nybom, H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Molecular Ecology. **13**: 1143-1155.

Ogg, M. 2014. Blueberry industry in figures. International special edition: Blueberries of the world 2014. www.freshfruitportal.com. : 5-7.

Ojango, J.M., Malmfors, B. and Okeyo,A.M. 2011. Module 4: Quantitative methods to improve the understanding and utilisation of animal genetic resources. [Animal genetic training resources (AGTR). International Livestock Research Institute (ILRI)]

Available: http://agtr.ilri.cgiar.org/index.php?option=com_content&view=article&id=28&Itemid=268.

Okie, W. 1997. Register of new fruit and nut varieties. Brooks and olmo list 38. HortScience. **32**: 785-805.

Old, R. W. 1980. Principles of gene manipulation: An introduction to genetic engineering. Primrose, S. B., ed. Berkeley, CA: University of California Press, USA.

Oliveira, H. R., Campana, M. G., Jones, H., Hunt, H. V., Leigh, F., Redhouse, D. I., Lister, D. L., Jones, M. K. 2012. Tetraploid wheat landraces in the Mediterranean basin: Taxonomy, evolution and genetic diversity. Plos One **7**(5): e37063.

Peakall, R. and Smouse, P. E. 2012. GenAlEx 6.5: Genetic analysis in excel. Population genetic software for teaching and research - an update. Bioinformatics. **28**: 2537-2539.

Pejic, I., Ajmone-Marsan, P., Morgante, M., Kozumplick, V., Castiglioni, P., Taramino, G., Motto, M. 1998. Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs, and AFLPs. Theoretical and Applied Genetics. **97**: 1248-1255.

Persson, H. A. and Gustavsson, B. A. 2001. The extent of clonality and genetic diversity in lingonberry (*Vaccinium vitis-idaea* L.) revealed by RAPDs and leaf-shape analysis. Molecular Ecology. **10**: 1385-1397.

Powell, E. A. and Kron, K. A. 2002. Hawaiian blueberries and their relatives- A phylogenetic analysis of *Vaccinium* sections *Macropelma*, *Myrtillus*, and *Hemimyrtillus* (Ericaceae). Systematic Botany. **27**: 768-779.

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S., Rafalski, A. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding. **2**: 225-238.

Prevost, A. and Wilkinson, M. J. 1999. A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. Theoretical and Applied Genetics. **98**: 107-112.

Prior, R. L., Cao, G. H., Martin, A., Sofic, E., McEwen, J., O'Brien, C., Lischner, N., Ehlenfeldt, M., Kalt, W., Krewer, G., Mainland, C. M. 1998. Antioxidant capacity as influenced by total phenolic and anthocyanin content, maturity, and variety of *Vaccinium* species. Journal of Agricultural and Food Chemistry. **46**: 2686-2693.

Pritchard, J. K., Stephens, M., Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics. **155**: 945-959.

PRRP. 2012. Crop profile for highbush blueberry in Canada, 2011. Pesticide risk reduction program. Pest management Centre and Agriculture and Agri-Food Canada. Ottawa, Canada.

Qu, L. P. and Hancock, J. F. 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*. Journal of the American Society for Horticultural Science. **122**: 69-73.

Qu, L. P., Hancock, J. F., Whallon, J. H. 1998. Evolution in an autopolyploid group displaying predominantly bivalent pairing at meiosis: Genomic similarity of diploid *Vaccinium darrowi* and autoteraploid *V corymbosum* (Ericaceae). American Journal of Botany. **85**: 698-703.

Reddy, M. P., Sarla, N., Siddiq, E. A. 2002. Inter simple sequence repeat (ISSR) polymorphism and its application in plant breeding. Euphytica. **128**: 9-17.

Reed, D. H. and Frankham, R. 2003. Correlation between fitness and genetic diversity. Conservation Biology. **17**: 230-237.

Reif, J. C., Xia, X. C., Melchinger, A. E., Warburton, M. L., Hoisington, D. A., Beck, D., Bohn, M., Frisch, M. 2004. Genetic diversity determined within and among CIMMYT maize populations of tropical, subtropical, and temperate germplasm by SSR markers. Crop Science. **44**(1): 326-334.

Rosenberg, N. A., Burke, T., Elo, K., Feldmann, M. W., Freidlin, P. J., Groenen, M. A. M., Hillel, J., Maki-Tanila, A., Tixier-Boichard, M., Vignal, A., Wimmers, K., Weigend, S. 2001. Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. Genetics. **159**: 699-713.

Rowland, L. J. and Levi, A. 1994. RAPD-based genetic-linkage map of blueberry derived from a cross between diploid species (*Vaccinium darrowi* and *Vaccinium elliottii*). Theoretical and Applied Genetics. **87**: 863-868.

Rowland, L. J., Dhanaraj, A. L., Polashock, J. J., Arora, R. 2003a. Utility of blueberry-derived EST-PCR primers in related Ericaceae species. HortScience. **38**: 1428-1432.

Rowland, L. J., Mehra, S., Dhanaraj, A. L., Ogden, E. L., Slovin, J. P., Ehlenfeldt, M. K. 2003b. Development of EST-PCR markers for DNA fingerprinting and genetic relationship studies in blueberry (*Vaccinium*, section *Cyanococcus*). Journal of the American Society for Horticultural Science. **128**: 682-690.

Rowland, L. J., Ogden, E. L., Arora, R., Lim, C. C., Lehman, J. S., Levi, A., Panta, G. R. 1999. Use of blueberry to study genetic control of chilling requirement and cold hardiness in woody perennials. HortScience. **34**: 1185-1191.

Rowland, L. J., Ogden, E. L., Ehlenfeldt, M. K. 2010. EST-PCR markers developed for highbush blueberry are also useful for genetic fingerprinting and relationship studies in rabbiteye blueberry. Scientia Horticulturae. **125**: 779-784.

Saha, S., Tomaro-Duchesneau, C., Tabrizian, M., Prakash, S. 2012. Probiotics as oral health bio therapeutics. Expert Opinion on Biological Therapy. **12**: 1207-1220.

Sakiroglu, M., Doyle, J. J., Brummer, E. C. 2010. Inferring population structure and genetic diversity of broad range of wild diploid alfalfa (*Medicago sativa* L.) accessions using SSR markers. Theoretical and Applied Genetics. **121**(3): 403-415.

Schlautman, B., Fajardo, D., Bougie, T., Wiesman, E., Polashock, J., Vorsa, N., Steffan, S., Zalapa, J. 2015. Development and validation of 697 novel polymorphic genomic and EST-SSR markers in the American cranberry (*Vaccinium macrocarpon* Ait.). Molecules **20**(2): 2001-2013.

Scrivener, L. 2008. Economy singing the blues, but berries are booming. [www.thestar.com. Toronto] Available: http://www.thestar.com/news/2008/07/20/economy_singing_the_blues_but_berries_are_booming.html .

Shannon, C. E. and Weaver, W. 1999. The mathematical theory of communication. University of Illinois Press, Urbana. .

Shete, S., Tiwari, H., Elston, R. C. 2000. On estimating the heterozygosity and polymorphism information content value. Theoretical Population Biology. **57**(3): 265-271.

Soltis, P. S. and Soltis, D. E. 2000. The role of genetic and genomic attributes in the success of polyploids. Proc. Natl. Acad. Sci. U. S. A. **97**(13): 7051-7057.

Spritz, R. A. 1981. Duplication-deletion polymorphism 5'- to the human beta-globin gene. Nucleic Acids Research. **9**: 5037-5047.

Squirrell, J., Hollingsworth, P. M., Woodhead, M., Russell, J., Lowe, A. J., Gibby, M., Powell, W. 2003. How much effort is required to isolate nuclear microsatellites from plants? Molecular Ecology. **12**: 1339-1348.

Stajner, N., Jakse, J., Kozjak, P., Javornik, B. 2005. The isolation and characterisation of microsatellites in hop (*Humulus lupulus* L.). Plant Science **168**(1): 213-221.

Tang, S. X. and Knapp, S. J. 2003. Microsatellites uncover extraordinary diversity in Native American land races and wild populations of cultivated sunflower. Theoretical and Applied Genetics. **106**(6): 990-1003.

Tautz, D. and Renz, M. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Research. **12**: 4127-4138.

Templeton, A. R. 1991. Genetics and conservation biology. 15–29 *in*: A. seitz and V. loeschcke., eds. Species conservation: A population- biological approach: Birkhauser Verlag, Basel.

Templeton, A. R. 1993. Translocation as conservation tool. 315-325 *in:* Szaro, R. C. and Johnston, D. W**.**, eds. Biodiversity in managed landscapes: Theory and practice. Oxford university press.

Tyagi, P., Gore, M. A., Bowman, D. T., Campbell, B. T., Udall, J. A., Kuraparthy, V. 2014. Genetic diversity and population structure in the US upland cotton (gossypium hirsutum L.). Theoretical and Applied Genetics. **127**(2): 283-295.

USHBC. 2013. USHBC news: North American new products continue to roll. [U.S. highbush blueberry council] Available: http://www.blueberry.org/news.htm .

Van Heemstra, M. I., Bruederle, L. P., Vorsa, N. 1991. Inheritance and linkage of thirteen polymorphic isozyme loci in diploid blueberry. Journal of the American Society for Horticultural Science. **116**: 89-94.

Vander Kloet, S. P. 1988. The genus *Vaccinium* in North America. Canada. Agriculture Canada. Research Branch, ed. Ottawa: Agriculture Canada Research Branch, Ottawa.

Villata, M. 2015. Market research shows increasing demand and use. U.S. Highbush Blueberry Council annual report 2013-2014. .

Vorsa, N., Manos, P. S., Vanheemstra, M. I. 1988. Isozyme variation and inheritance in blueberry. Genome. **30**: 776-781.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Vandelee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., Zabeau, M. 1995. AFLP - a new technique for DNA-fingerprinting. Nucleic Acids Research. **23**: 4407-4414.

Wang, Z., Weber, J. L., Zhong, G., Tanksley, S. D. 1994. Survey of plant short tandem DNA repeats. Theoretical and Applied Genetics. **88**: 1-6.

Weber, J. L. and May, P. E. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain-reaction. American Journal of Human Genetics. **44**: 388-396.

Weir, B. S. 1996. Genetic data analysis II: Methods for discrete population genetic data. Sinauer associates, Inc., Sunderland. .

Whisson, S. C., Drenth, A., Maclean, D. J., Irwin, J. A. G. 1995. *Phytophthora sojae* Avirulence genes, RAPD, and RFLP markers used to construct a detailed genetic linkage map. Molecular Plant-Microbe Interactions. **8**: 988-995.

Willis, L., Bickford, P., Zaman, V., Moore, A., Granholm, A. C. 2005. Blueberry extract enhances survival of intraocular hippocampal transplants. Cell Transplantation. **14**: 213-223.

Wolfe, K. L. and Liu, R. H. 2007. Cellular antioxidant activity (CAA) assay for assessing antioxidants, foods, and dietary supplements. Journal of Agricultural and Food Chemistry. **55**: 8896-8907.

Yadav, H. K., Ranjan, A., Asif, M. H., Mantri, S., Sawant, S. V., Tuli, R. 2011. EST-derived SSR markers in *Jatropha curcas* L.: Development, characterization, polymorphism, and transferability across the species/genera. Tree Genetics & Genomes. **7**: 207-219.

Yang, S., Bishop, J. G., Webster, M. S. 2008. Colonization genetics of an animal-dispersed plant (*Vaccinium membranaceum*) at Mount St. Helens, Washington. Molecular Ecology. **17**: 731-740.

Yarborough, D. 1997. Production trends in the wild blueberry industry in North America. Acta Horticulturae. Sixth International Symposium on *Vaccinium* Culture (446): 33-35.

Zalapa, J. E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E., McCown, B., Harbut, R., Simon, P. 2012. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. American Journal of Botany. **99**: 193-208.

Zane, L., Bargelloni, L., Patarnello, T. 2002. Strategies for microsatellite isolation: A review. Molecular Ecology. **11**: 1-16.

Zhen, Y. Q., Li, Z. Z., Huang, H. W., Wang, Y. 2004. Molecular characterization of kiwifruit (*Actinidia*) cultivars and selections using SSR markers. Journal of the American Society for Horticultural Science. **129**: 374-382.

Zietkiewicz, E., Rafalski, A., Labuda, D. 1994. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain-reaction amplification. Genomics. **20**: 176-183.

Zhu, H., Senalik, D., McCown, B. H., Zeldin, E. L., Speers, J., Hyman, J., Bassil, N., Hummer, K., Simon, P. W., Zalapa, J. E. 2012. Mining and validation of pyrosequenced simple sequence repeats (SSRs) from American cranberry (*Vaccinium macrocarpon* Ait.). Theoretical and Applied Genetics. **124**: 87-96.