I. EXECUTIVE SUMMARY

PROJECT TITLE

Generating Genomic Tools for Blueberry Improvement

PROJECT TYPE

Standard Research and Extension Project, Renewal Application

FOCUS AREA

Focus Area 1. Research in plant breeding, genetics, and genomics to improve crop characteristics (100%)

PROGRAM STAFF

Project Director:
Lisa J. Rowland, Research Geneticist, USDA/ARS, Genetic Improvement of Fruits and Vegetables Laboratory, BARC-West, Beltsville, MD 20705, Jeannine.Rowland@ars.usda.gov

Co-PDs:
Nahla V. Bassil, Research Geneticist, USDA/ARS, National Clonal Germplasm Repository, Corvallis, OR 97333, Nahla.Bassil@ars.usda.gov

Allan F. Brown, Assistant Professor of Applied Molecular Genetics, North Carolina State University, Dept. of Horticultural Science, Plants for Human Health Institute, Kannapolis, NC 28081, allan_brown@ncsu.edu

Frank A. Drummond, Professor of Biology and Cooperative Extension, University of Maine, School of Biology and Ecology, Orono, ME 04469, Frank.Drummond@umit.maine.edu

James F. Hancock, Professor of Genetics and Plant Breeding, Michigan State University, Dept. of Horticulture, East Lansing, MI 48824, hancock@msu.edu

James W. Olmstead, Assistant Professor of Horticulture, University of Florida, Horticultural Sciences Dept., Gainesville, FL 32611, jwolmstead@ufl.edu

James J. Polashock, Plant Pathologist, USDA/ARS, Genetic Improvement of Fruits and Vegetables Laboratory, Blueberry and Cranberry Research Center, Chatsworth, NJ 08019, James.Polashock@ars.usda.gov

Nadim Alkharouf, Associate Professor of Computer and Information Sciences (Bioinformatics), Towson University, 8000 York Road, Towson, MD 21252, nalkharouf@towson.edu

Chad Finn, Geneticist and Plant Breeder, USDA/ARS, Horticultural Crops Research Lab, Corvallis, OR 97333, finnc@hort.oregonstate.edu

Samuel Hanes, Assistant Professor of Agricultural Anthropology, University of Maine, Social/Cultural Anthropology, Orono, ME 04469, Samuel.Hanes@umit.maine.edu

Todd Mockler, Associate Member, Donald Danforth Plant Science Center, 975 N. Warson Rd., St. Louis, MO 63132, tmockler@danforthcenter.org

Collaborators:
Malcolm Campbell, Professor of Biology, Davidson College, Biology Dept., Davidson, NC 28036-7118, macampbell@davidson.edu

International collaborators:
Julie Graham and Susan McCallum, Genetics Programme, Scottish Crop Research Institute, The James Hutton Institute, Dundee, Scotland DD2 5DA, Julie.Graham@hutton.ac.uk and Susan.McCallum@hutton.ac.uk

Emily J. Buck and Roy Storey, The New Zealand Institute for Plant & Food Research Ltd (PFR), Palmerston North, New Zealand, Emily.Buck@plantandfood.co.nz

STAKEHOLDER NEED ADDRESSED

There has been increased demand for blueberries in recent years in part because of their recognized health benefits. Development of new blueberry cultivars with high fruit quality combined with early and late ripening and appropriate climatic adaptation is needed. With the availability of more genomic resources, marker-assisted breeding could be used in cultivar development to more efficiently combine traits for fruit quality with traits for climatic adaptation and early and late ripening to extend the season and increase the availability of fresh fruit. (p. 6, 17-19, in Project Narrative)

OUTREACH PLAN

Our project plan incorporates a multifaceted Outreach Program, which includes delivery of research results and training in marker-assisted breeding to the industry, breeders, and high school and college students. Our research plan itself involves working directly with industry groups and their breeding populations to identify markers linked to their traits of interest. High school interns and undergraduate college students will be integrally involved in carrying out many aspects of the proposed research. The social science component of our project has as its major goals helping molecular geneticists better target their science and outreach to address breeders' needs and concerns, clarifying areas of confusion, and identifying obstacles limiting adoption of molecular-assisted breeding. (p. 19, 28-32, in Project Narrative)

POTENTIAL ECONOMIC, SOCIAL, AND ENVIRONMENTAL BENEFITS

The incorporation of marker–assisted breeding should result in more cost-effective cultivar development strategies, principally by allowing breeders to better select parent plants for crosses and eliminate offspring that do not harbor the gene(s) for the desired traits before they are field planted for long-term evaluation. Socially, this project should result in more productive collaborations and working relationships between molecular geneticists and breeders. Environmental benefits include the development of new blueberry cultivars adapted to a wide range of climatic conditions and a better understanding of how cultivation practices may or may not have affected genetic diversity of wild, lowbush blueberry.

STAKEHOLDER ENGAGEMENT THROUGHOUT THE PROJECT

Many of our stakeholders are breeders that are integrally involved in the project as co-PDs. In addition, our research plan involves working directly throughout the project with industry groups and their breeding populations to identify markers linked to their traits of interest. Two major meetings were held in 2010 and 2011, as part of the initial funding of our "Generating Genomic Tools for Blueberry Improvement" project, where we presented our results, provided training in marker-assisted breeding, developed research plans, and received input from breeders, industry representatives, advisory board members, students, and others attending. Another such meeting is planned for the second year of this renewal. (p. 8, 15-16, 19, 30-31, in Project Narrative)

LOGIC MODEL

Our logic model supports the SCRI programmatic logic model in many ways. First, we have assembled a diverse international team of PI's and collaborators including scientists from the USDA/ARS, several universities throughout the U.S. (Michigan State University, University of Maine, North Carolina State University, University of Florida, Towson University, Davidson College), the Scottish Crop Research Institute/James Hutton Institute, and the New Zealand Institute for Plant & Food Research Ltd. Our team is multi-disciplinary and trans-disciplinary and includes geneticists and breeders, molecular geneticists, genomicists, bioinformaticists, an entomologist and ecologist, a plant pathologist, and a sociologist/anthropologist. Second, our stakeholders are actively engaged in our project--some are breeders that are integrally involved in the project as co-PDs, others are collaborating with us to validate and identify new markers linked to QTL in their breeding populations, others are members of our advisory board, and others are providing matching support for our project. Third, our project involves an Outreach Program, which includes delivery of research results and training in marker-assisted breeding to the industry, breeders, and high school and college students. Our research plan itself involves working directly with industry groups and their breeding populations to identify markers linked to their traits of interest. High school interns and undergraduate college students will be involved in carrying out many aspects of the proposed research and, in so doing, will receive training in mapping, marker-assisted selection, and bioinformatics. Our Outreach Program also involves a social science component that aims to help molecular geneticists better target their science, identify obstacles that limit the adoption of marker-assisted breeding, and ensure breeders' needs and concerns are met. Fourth, in terms of outcomes, in the short term we will generate new knowledge and products for blueberry scientists and stakeholders such as an expanded Blueberry Genomics Database website, assembled transcriptome sequences, a SNP chip, further saturated genetic linkage maps, markers linked to traits of interest, and identification of ways to better

target molecular research to meet breeders' needs, etc. More long term, we will actually identify genes responsible for QTL, implement marker-assisted breeding (MAB), release new varieties through its implementation, and help to train a new generation of students in MAB. Our logic model is shown below:

Inputs	Outputs		Outcomes		
	Activities	Outreach	Short	Medium	Long-term
Funding of project Expertise from diverse international team of breeders, molecular geneticists, bioinformaticists, horticulturalists, plant patho-logists, entomo-logists, and sociologists Stakeholder involvement Industry breeding populations, mapping populations Involvement of high school intern programs and college undergrad programs Advisory board	Complete assembly and annotation of all sequences from expressed genes Add SSR, HRM, SNP, and EST-PCR markers to maps to align the two maps and anchor the maps to the whole genome sequence Complete QTL analyses of populations Genotype new industry breeding populations and identify QTL for traits of interest Identify, propagate, and cross high yielding, good selfing lowbush clones	Use Blueberry Genomic Databases as tools to train undergrad bioinformatics students Have high school interns participate in the mapping projects Survey breeders and molecular geneticists on opinions of marker-assisted breeding (MAB) Hold meeting to describe results to and get input from breeders, industry representatives, advisory board members, etc. Grower meeting to describe potential for utilizing high yielding selfing clones via seed families	Expanded BBGD website Assembled sequences for use by scientists for building custom microarrays, etc. SNP chip Further saturated maps Mapped QTL Markers linked to traits of interest Identification of candidate genes near QTL Survey results Students trained in MAB and bioinformatics	Implementation of MAB Release of advanced selections/ cultivars from phenotyping studies More collaborations and better communication between breeders and molecular geneticists Use of maps to identify markers for MAB in more populations Identification of actual genes responsible for QTL Use of SNP chip Use of seedlings from selected high yielding, good selfing lowbush clones in non-productive areas in fields	Release of new cultivars from MAB New generation of breeders trained and experienced in using MAB

TABLE OF CONTENTS

II. INTRODUCTION

A. The Objectives, the Need for the Research, and Stakeholder Involvement

Blueberry (*Vaccinium*, section *Cyanococcus*) is the most economically important fruit crop within the *Ericaceae* family. Consumer demand for blueberries is at an all time high because of its many recognized health benefits. Blueberry fruit is rich in anthocyanins, which have been linked to improved night vision, prevention of macular degeneration, anti-cancer activity, and reduced risk of heart disease (Cho et al. 2004; Kalt et al. 2007). The compound resveratrol, found in blueberries, has also been linked to reduced risk of heart disease and cancer, and another compound, pterostilbene, has been shown to lower cholesterol (Rimando et al. 2004). The high consumer demand for blueberries, combined with the loss of many plant breeding positions from government and academic institutions, has resulted in a real need by the industry for the rapid development of new blueberry cultivars with high fruit quality, expanded fruiting season, and adapted to a wide range of climatic conditions. Up until recently, there have been very few genomic resources available for blueberry, or for the *Ericaceae* family in general. Now, with the availability of more genomic resources, largely from the funding of our previous SCRI grant and partly from the whole genome sequencing effort currently underway, we should be able to use marker-assisted breeding of blueberry to more efficiently combine traits for fruit quality with traits for climatic adaptation, and early and late ripening.

This renewal proposal requests two more years of funding of our "Generating Genomic Tools for Blueberry Improvement" project. Our original objectives were:
1) deeper sequencing of the highbush blueberry transcriptome,
2) developing molecular markers from the available expressed sequence tags (ESTs),
3) further saturating the diploid and tetraploid genetic linkage maps, mapping quantitative trait loci (QTL) for chilling requirement, cold tolerance, and fruit quality traits in highbush blueberry, carrying out studies on the population genetic structure of wild lowbush blueberry, and
4) delivering research results and training in molecular breeding through our Outreach Program.
We have completed or nearly completed all of these objectives. A detailed Progress Report is provided in the following section (see p. 10-16).

Briefly, regarding the first objective (transcriptome sequencing), we have generated over 600,000 new ESTs from blueberry by next generation 454 sequencing, which have been assembled into ~15,000 contigs and ~124,000 singletons (Rowland et al. submitted). Sequences were generated from fruit at different stages of development, flower buds at different stages of cold acclimation, and leaves. The sequences are publicly available in the short read archive of GenBank and are also available on our website (http://bioinformatics.towson.edu/BBGD454), along with their annotations and various assemblies. Regarding the second objective (marker development), we have mined the newly generated sequences for SSRs and carried out sequencing for SNP development. We have developed more EST-PCR and high resolution melting (HRM) markers. From the 454 sequences, 15,886 blueberry EST-SSR loci were identified. Regarding the third objective (mapping of highbush blueberry and population genetic studies on lowbush blueberry), we have added many SSR, HRM, and EST-PCR markers to our diploid and tetraploid maps. Before this renewal would begin, we expect our maps to be comprised of well over 300 markers each. The diploid mapping population was phenotyped for

cold hardiness over two years and for chilling requirement for one year. Chilling requirement is being determined again this year. We have identified QTL for cold hardiness that are consistent between the two years. The tetraploid mapping population (an actual breeding population) was established at five locations and phenotyped for 19 traits. QTL for all the traits have been identified at at least some of the locations, and analyses are still ongoing. The population was also evaluated for chilling requirement in a greenhouse study and an evaluation of cold tolerance is underway. Studies have been carried out to investigate the spatial genetic structure (patterns of genetic relationship) of lowbush blueberry on both a local scale (within a field) and on a wide scale (across several fields in Maine and the eastern U.S.). A gene flow study on lowbush blueberry is near completion. Regarding our fourth objective (Outreach), 11 high school interns received training while participating in this project. Twelve undergraduate and seven graduate students also received training while contributing to this project. One of the high school interns and one college student traveled to New Zealand and worked on the project there as well. Two major meetings were held where we presented our results and provided training in marker-assisted breeding to breeders, industry representatives, advisory board members, and students.

Our renewal would build and expand upon our previous work and includes objectives for further development of genomic tools and utilization of tools by molecular geneticists and breeders for blueberry improvement. At the conclusion of this work, blueberry breeders will be able to use marker-assisted selection in their breeding programs. The specific objectives are:

For highbush blueberry

(1) complete the assembly and annotation of all available transcriptome sequences of blueberry,
(2) add simple sequence repeat (SSR) markers, developed from the scaffolds of the whole genome sequence assembly, and other markers (developed from candidate genes of traits of interest) to the current diploid and tetraploid genetic linkage maps of blueberry,
(3) anchor the maps to the whole genome sequence and align the two maps to each other,
(4) develop a single nucleotide polymorphism (SNP) chip and add SNPs to the genetic linkage maps,
(5) increase marker coverage in the vicinity of QTL (identified in the previous project) associated with chilling requirement, harvest date, and fruit quality (soluble solids, pH, titratable acidity, and picking scar), and validate and/or identify new QTL linked to these traits in additional actual breeding populations,
(6) identify and test candidate genes near QTL for controlling traits of interest,

For lowbush blueberry

(7) identify and propagate high yielding, good selfing clones (genotypes) with good fruit quality
(8) carry out a diallel crossing design of clones to generate seed families for field testing
(9) assess tetraploid highbush genetic linkage map for use with lowbush blueberry

For social science/outreach component

(10) survey breeders, growers, and molecular geneticists with the goal of helping molecular geneticists better target their science and outreach to address breeders' and growers' needs and identify obstacles limiting adoption of molecular-assisted breeding

(11) deliver research results and training to our blueberry stakeholders and provide training in mapping and bioinformatics to high school and college students.

Our objectives were arrived at after the second of our two major meetings, held in East Lansing, MI, in 2011. Blueberry breeders and industry representatives in the U.S. were invited to attend and participate, including members of our advisory board. A social scientist from the University of Maine was also invited to attend. All the co-PDs and collaborators involved in our project presented their research results. In addition, talks were given to provide background and training on construction of genetic linkage maps, identifying QTL, etc. Informal open discussion sessions were held to get input from breeders on marker-assisted breeding—when they would use it, when they would not use it, traits they were most interested in, traits that are most difficult to evaluate, etc. After this meeting and some follow-up discussions with industry groups, co-PDs, and collaborators, the objectives listed above were decided upon. Support of the proposed objectives is evidenced by pledged matching funds from industry groups (Berry Blue LLC, Driscolls, and Fall Creek Nursery) and from 'in kind' matching funds from co-PIs and international collaborators with the Scottish Crop Research Institute/James Hutton Institute and Plant & Food Research in New Zealand.

B. Background

Commercial importance. Blueberry is an economically important small fruit crop adapted to acidic, sandy soils that otherwise might be considered worthless from an agronomic standpoint (Galleta and Ballington 1996). Blueberry is a member of the *Ericaceae* family, which includes many acid-loving species, such as the commercially important berry crops, blueberry, cranberry, and lingonberry, and the floral and nursery crops, rhododendron, azalea, and mountain laurel. North America is the major producer of blueberries, although production of blueberries is on the rise worldwide. About 2/3 of blueberry production in the U.S. is from improved cultivars of the tetraploid highbush blueberry (*Vaccinium corymbosum*), and to a lesser extent hexaploid rabbiteye blueberry (*V. virgatum*), and about 1/3 is from wild, managed fields of tetraploid lowbush blueberry (*V. angustifolium*) (United States Department of Agriculture National Agricultural Statistics Service). The commonly used highbush/lowbush nomenclature in blueberry refers to stature. Lowbush are rhizomatous with stems from 0.30 to 0.60 m; highbush and rabbiteye are crown forming and generally maintained between 1.8 and 2.5 m and between 2.0 and 4.0 m, respectively (Galletta and Ballington 1996).

Traditional breeding of blueberry and its limitations. Much progress has been made in traditional breeding of highbush and rabbiteye cultivars since their domestication in the early twentieth century. In general, breeding efforts have focused on development of cultivars with broader climatic adaptation (increased freezing tolerance for northern regions and reduced chilling requirements for southern regions), season extension, disease and pest resistance, mechanical handling tolerance, and high fruit quality (Galletta and Ballington 1996).

In an effort to expand the growing areas of highbush blueberry, breeders have employed interspecific hybridization of species within the *Cyanococcus* section in the development of both northern and southern highbush types, blending species with very different chilling requirements and cold tolerances. Cold tolerance is the ability of a plant to survive winter

temperatures without damage principally, in blueberry, to flower buds, as this will result in losses in fruit yield. Flower buds are formed in the late summer and fall, and chilling requirement is the amount of low temperature exposure that is required for bud break the following spring. Developing low-chill blueberry cultivars has been critical for the expansion of highbush blueberry to the southern U.S., where rabbiteye blueberry is traditionally grown. Degree of mid-winter cold tolerance is an important consideration in developing cultivars for the northern and midwestern U.S., and spring frost tolerance is important in nearly all growing areas. Evaluating the climatic adaptations of hybrid breeding populations in the field is quite difficult, however, as chilling hours and temperatures vary greatly from year to year and site to site, necessitating long trialing periods to adequately evaluate a genotype's adaptive zone. The identification of QTL regulating chilling hour requirements and cold tolerance would allow for the more precise prediction of a plant's adaptive range and more rapid genetic improvement through marker-assisted breeding.

The identification and mapping of genes associated with high fruit quality is also highly desirable. The fruit characteristics most sought after in blueberries are flavor, large size, light blue color (a heavy coating of wax), a small scar where the pedicel detaches, easy fruit detachment for hand or machine harvest, firmness, and a long storage life. High antioxidant capacity has become an important nutritional quality parameter in blueberries also, although specific breeding for this characteristic has not yet been undertaken. Genetic improvement could be rapid, as considerable amounts of quantitative variability have been observed in this characteristic (Ehlenfeldt and Prior 2001; Connor et al. 2002a and b). Marker-assisted breeding could aid highbush breeders in combining the genes associated with high fruit quality with other traits, such as those determining regionally appropriate chilling hour and cold requirements.

Expanding the fruiting season of highbush blueberry is also highly desirable. One of the major obstacles that have been faced by both the northern and southern highbush breeders has been generating very early and late harvested genotypes that have a pleasant sugar to acid ratio and small scar. Some early and late harvested genotypes do emerge with sweet flavors and small scars, but a high percentage of them are very tart with a large scar. An understanding of the linkage relationships between harvest date and fruit quality QTL could aid significantly in developing breeding strategies and improve efficiency through marker assisted breeding. Useful parents could be more efficiently identified using molecular markers and progeny with undesirable traits could be selected out in the greenhouse before field planting.

Management of wild, lowbush blueberry. The commercial lowbush blueberry is a unique agricultural crop of northeastern North America and is Maine's most important fruit crop. Plants are wild, having reclaimed burned and cleared stretches of land over several centuries. Phenotypic variation among lowbush genotypes, referred to as clones, is high with adjacent clones showing as much as 12-15 fold differences in yield, for example. Lowbush is predominantly outcrossing (Aalders and Hall 1961; Hall et al. 1979) and self-fertility is poor due to early-acting inbreeding depression (Hokanson and Hancock 2000). There has been little traditional breeding of lowbush blueberry, but improved cultural practices have resulted in a 4-fold increase in yield over the last few decades (Yarborough 2004).

Genomic resources. Before funding of our SCRI proposal "Generating Genomic Tools for Blueberry Improvement" at the end of 2008, few genomic resources were available for blueberry, or for the *Ericaceae* family in general. Only about 5,000 ESTs were available from blueberry (Dhanaraj et al. 2004, 2007) and about 1,200 from rhododendron (Wei et al. 2005). A limited number of robust molecular markers, like SSRs (Boches et al. 2005, 2006) and EST-PCRs (Rowland et al. 2003a, b, c), had been developed from some of the blueberry ESTs and were being used in limited genetic diversity and mapping studies. Efforts to construct a diploid and tetraploid map of blueberry had been initiated but lack of funds had left them insufficiently saturated for QTL mapping. One microarray experiment had been carried out in blueberry and had successfully identified many transcripts whose abundances increase with cold acclimation (Dhanaraj et al. 2007; Rowland et al. 2008).

The status of genomic tools for blueberry improvement is quite different now in 2012. Through our transcriptome sequencing project funded by SCRI, we have generated over 600,000 new ESTs from the highbush blueberry cultivar Bluecrop by next generation 454 sequencing. Sequences were generated from fruit at four stages of development, flower buds at four stages of cold acclimation, and leaves. The raw sequences have been deposited in the short read archive of GenBank (Accession #s SRX100856; SRX100859; SRX100861-SRX100867). The reads have been assembled into approximately 15,000 contigs and 124,000 singletons (Rowland et al. submitted). We have also developed a public website where the sequences and various assemblies, their annotations, and their frequencies in each of the libraries are available (http://bioinformatics.towson.edu/BBGD454). In addition to these new 454 sequences, ~17,000 new Sanger ESTs from fruit of the highbush cultivar Rubel have been deposited into the EST database of GenBank in the last few months (Zifkin et al. 2011), bringing the total number of sequences in the EST database to ~22,000. Furthermore, an effort to sequence the whole genome of blueberry is currently underway, being led by Allan Brown (of North Carolina State University and Plants for Human Health Institute, Kannapolis, NC). The diploid highbush selection being sequenced is, in fact, one of the parents of our diploid mapping population, called W85-20. Genetic linkage maps of our diploid and tetraploid populations have much better coverage than they did in 2008 and have now proven useful for mapping QTL for several important traits.

C. Recently Completed Research of Key Project Personnel/Preliminary Data

Since this is a renewal proposal, this is described in the following in depth Progress Report.

III. PROGRESS REPORT

The objectives of our original proposal "Generating Genomic Tools for Blueberry Improvement" were:
1) deeper sequencing of the highbush blueberry transcriptome,
2) developing molecular markers from the available expressed sequence tags (ESTs),
3) using the molecular markers in several well-defined mapping and genetic relationship studies including further saturating the diploid and tetraploid genetic linkage maps, mapping quantitative trait loci (QTL) for chilling requirement, cold tolerance, and fruit quality traits in highbush

blueberry, carrying out studies on the population genetic structure of wild lowbush blueberry, and studying the phylogeny of species within the *Cyanococcus* section, and
4) delivering research results and training in molecular breeding through our Outreach Program.

We have completed or nearly completed all of these objectives. Progress on each objective is described separately below.

A. Transcriptome Sequencing

Co-PIs and other Collaborators Involved on Project: Lisa J. Rowland, Nadim Alkharouf, James J. Polashock, Nahla V. Bassil, and Dorrie Main

We have generated over 600,000 new ESTs from the highbush blueberry cultivar Bluecrop by next generation 454 sequencing, which have been assembled overall into ~15,000 contigs and ~124,000 singletons (Rowland et al. submitted). The sequences were generated from nine different cDNA libraries—libraries were from blueberry fruit at four different stages of development (green, white, pink, and blue/ripe fruit), flower buds at four different stages of cold acclimation (0, 400, 800, and 1300 hours of low temperature exposure), and a combination of leaves and stems. The sequences were assembled in a variety of ways (each library was assembled separately, all fruit libraries were assembled together, all bud libraries were assembled together, and all libraries were assembled together for an overall assembly). Each assembly was annotated and functionally mapped to Gene Ontology (GO) terms. The raw sequences have been deposited in the short read archive of GenBank (Accession #s SRX100856; SRX100859; SRX100861-SRX100867). We have also developed a public website where the sequences and various assemblies, their annotations, and their frequencies in each of the libraries are available (http://bioinformatics.towson.edu/BBGD454).

The frequency of the most abundant sequences in each of the libraries has also been compared across all libraries to identify genes that are potentially differentially expressed during cold acclimation and fruit development. Real-time PCR was performed to confirm their differential expression patterns. Overall, 15 out of 18 (or 83%) of selected genes had differential expression patterns similar to what was predicted from their reads alone. The assembled sequences were also mined for SSRs and SSRs were tested on parents of our two blueberry mapping populations (described below under Marker Development section). This large collection of 454 ESTs has proven to be a valuable resource for marker development and identification of genes that are potentially differentially expressed. It should be useful in future studies for identifying genes that play important roles in flower bud development, cold acclimation, chilling unit accumulation, and fruit development in blueberry and related species.

B. Marker Development

Co-PIs and other Collaborators Involved on Project: Lisa J. Rowland, Nahla V. Bassil, Allan F. Brown, Emily Buck, Dorrie Main

SSR primer pairs were designed from four sources: Sanger ESTs from Lisa Rowland that were already in GenBank (447); the assembled EST sequences described above (7,705 VCB for

Vaccinium corymbosum 'Bluecrop'); genomic sequences from W85-20 obtained from Allan Brown (1,000 KAN for Kannapolis); and ESTs from three different genotypes ('Duke', 'Puru', and a breeding line) from Plant & Food Research obtained from Emily Buck (96). SSR primer pairs were designed with the SSR tool available at the Genome Database for the Rosaceae website (www.rosaceae.org) or using the Primer 3 website. Early in the project, EST-PCR primers were designed from Sanger ESTs already in GenBank. Many were designed from candidate cold tolerance genes (genes known to be differentially expressed during cold acclimation from microarray data (Dhanaraj et al. 2007)) and some from flowering-related genes. Later in the project, EST-PCR primers were designed from 454 ESTs from developing fruit. Primers were designed from the ends of the ESTs using the Primer 3 website as described previously (Rowland et al. 2003c).

Up to 364 of the 972 SSRs tested were polymorphic in parents of the tetraploid mapping population while 135 of 579 SSRs were polymorphic in parents of the diploid mapping population. Polymorphic SSRs were distributed out to the different labs doing the genotyping of the populations. SSR primers that gave monomorphic fragments, based on size, between the parents of the diploid mapping population were provided to Emily Buck (Plant & Food Research), who was able to convert 52 of these to HRM markers which utilized sequence variation. EST-PCR primers were tested on parents of the diploid mapping population and polymorphic markers added to the diploid map by the Rowland lab. The Rowland lab is currently testing the mapped EST-PCR markers on the tetraploid parents to determine if any can be mapped in that population as well.

For SNP development, short read sequences were generated for parents of the tetraploid mapping population, parents of the diploid mapping populations except for W85-20 (already sequenced by Allan Brown) and from another highbush blueberry cultivar, Emerald. Up to 36.7 Gb of sequence were generated with the Illumina GAIIx platform at Oregon State University's Center for Genome Research and Biocomputing. Current assembly of the reference blueberry genome, W85-20, was obtained from Allan Brown. A SOAP-based (Li et al., 2008) SNP detection pipeline was developed by Todd Mockler and Chris Sullivan by aligning the reads to the largest 28 scaffolds exceeding 20 Kb in length of the available version of the reference genome assembly.

C. Mapping and Genetic Relationship Studies

Co-PIs and other Collaborators Involved on Project: Lisa J. Rowland, Nahla V. Bassil, Frank A. Drummond, Chad Finn, James F. Hancock, James W. Olmstead, Emily Buck, Julie Graham, Susan McCallum, Ann Powell

Diploid mapping project. The interspecific diploid testcross population, developed by the Rowland lab, segregates for mid-winter cold tolerance and chilling requirement and has been described in detail elsewhere (Rowland et al. 1999). Briefly, a true diploid testcross population was generated by crossing a *V. darrowi* selection Fla4B (low chilling, cold sensitive, evergreen) x diploid *V. corymbosum* selection W85-20 (high chilling, cold hardy, deciduous) F1 hybrid to another diploid *V. corymbosum* W85-23. SSR, EST-PCR, and HRM markers were identified that were polymorphic between the original parents, Fla4B or W85-20, present in the F1, and absent

in the testcross parent W85-23, thus useful for developing a map of the F1 hybrid. These markers were expected to segregate 1:1 in the population. These markers were followed in the population, and JoinMap was used to construct the map. Presently, over 200 markers have been mapped to 15 linkage groups (see Appendix 1 for the actual map). These include a combination of old RAPD markers, and new EST-PCR, SSR, and HRM markers. Some of the mapped EST-PCR markers are from genes encoding such proteins as: peroxiredoxin, sucrose synthase, DNAJ, salt tolerance protein, galactinol synthase, reduced vernalization requirement, polyubiquitin, and a dehydrin. Before funding of this renewal would take place, we expect to have mapped over 300 markers.

All the individuals of the mapping population have been evaluated for cold hardiness over two years and for chilling requirement for one year according to our previously described protocols (Arora et al. 2000; Rowland et al. 1998). Chilling requirement is being determined again this year. A correlation test was run to compare the two years' cold hardiness data. The Pearson correlation coefficient was 0.663, indicating a good correlation between the two years' data. Composite Interval Mapping (Hackett et al. 2001; Bradshaw et al. 2004) of the cold hardiness trait (using both years' data separately and averaged) was conducted using MapQTL. Two significant QTL for cold hardiness were identified that were consistent between the two years. They were both located on linkage group 6 (Appendix 1). One QTL, with a maximum LOD of 4.73 explaining 29.9% of the phenotypic variance, was located between the EST-PCR marker Berc 230 (2.7 cM away) and the RAPD marker OPU1c. The other QTL, with a maximum LOD of 3.36 explaining 23.9% of the phenotypic variance, was located between the RAPD marker OPP12b (7cM away) and the EST-PCR marker SL256.

Tetraploid mapping project. The tetraploid mapping population is an F1 population that was generated from a cross between the northern highbush cv Draper and the southern highbush cv Jewel. The population is segregating for many traits including cold hardiness, chilling requirement, fruit stem scar size, fruit color, yield, soluble solids, total titratable acids, etc. SSR markers were first screened on parents and a subset of progeny to identify markers that were present in one or both parents and segregating 1:1 or 3:1, respectively. The whole population was then genotyped with these 'mappable' markers. Since highbush blueberry is an autotetraploid (Krebs and Hancock 1989), the software TetraploidMap (Hackett et al. 2007) was used to develop the maps of the parent plants. Appendix 2 shows the current map of the first parent 'Draper', comprised of 287 markers on 15 linkage groups ranging from 30 to 191 cM.

The mapping population was established at five field locations: Manor, Georgia; Interlachen, Florida; Grand Junction, Michigan; Corvallis, Oregon; and Invergowrie, Scotland. Locations were selected for their varying average annual chill accumulation and winter temperatures. The population was phenotyped for 19 traits (relating to four main categories: general plant characteristics, flower and fruit development, yield, and fruit quality) in 2011 in Georgia, Florida, Oregon, and Michigan. Phenotyping will be carried out again at these four locations and for the first time in Scotland this year. The population was also evaluated for chilling requirement in a greenhouse study at Michigan and an evaluation of cold tolerance is underway this year, too, in Michigan. So far, correlation tests have been performed to compare phenotypic data from two of the locations, Georgia and Florida. Initial QTL analyses have been carried out as well using Genstat on the Georgia and Florida data and the chilling requirement data from the

Michigan greenhouse study. QTL for all the traits have been identified in at least one of the locations, and analyses are still ongoing. The locations of some of the QTL (fruit stem scar size, soluble solids, total titratable acids, yield, fruit weight, and proportion vegetative bud break) are shown on the map (Appendix 2).

Lowbush genetic relationship studies. Lowbush blueberry production in the U.S. is from managed wild fields in Maine. Individual plants of wild lowbush blueberry are quite variable in terms of yield and we have been investigating possible causes of these yield differences. Lowbush is predominantly outcrossing (Aalders and Hall 1961; Hall et al. 1979) and self-fertility is poor due to early-acting inbreeding depression (Hokanson and Hancock 2000). Lowbush blueberry is pollinated by rented honey bees, which tend to fly short distances, thus it is thought that plants would most likely be pollinated by themselves or by near neighbors. It has been hypothesized that, if the distribution of individuals in fields is such that closely related individuals tend to cluster together in patches, then crosses between these plants could result in low yields due to inbreeding (Myra et al. 2004). In this project, we have tested this hypothesis in several ways. Initially, we investigated the spatial genetic structure (or patterns of genetic relationship) of fields of lowbush blueberry by sampling seven plants along a linear transect per field in four fields in Maine and genotyping them using EST-PCR markers (Bell et al. 2009). In this first study, we did not find evidence of spatial genetic structure within fields (a correlation between genetic distance and physical distance), although plants within a field were more related to each other than to plants in other fields further away.

Subsequently, we investigated spatial genetic structure on a finer local scale with more sampling intensity (100 closely spaced individual plants within a single field). Again we found that, in general, most of the individuals were not highly related to their near neighbors, but we did find evidence of a few patches of highly related individuals (Bell et al. submitted). In another study, we investigated whether genetic relationship has an effect on yield. We used EST-PCR markers to examine genetic relatedness of individuals within two wild blueberry fields in Maine, in combination with genetic crosses, to determine if relationship of parents affects yield. We found no evidence of yield being affected by genetic relationship except in the case of self-crosses, which generally resulted in lower yields. Notably, we found that the better selfers also generally yielded more in outcrosses (Bell et al. 2010). Most recently, we have undertaken a study on gene flow in lowbush blueberry. Here we collected open pollinated fruit from two lowbush clones, extracted the seed, and germinated the seedlings. The seedlings were genotyped, as well as the mother plants and the surrounding 5-6 nearest neighbors. Initial results from a paternity analysis of the seedlings suggest that none of the seedlings resulted from self-crosses (probably due to lack of germination of seed from self crosses). In addition, only about 15% of the seedlings appeared to have been sired by one of the nearest neighbors, or, stated another way, about 85% of the seedlings appeared to have been sired by a plant outside of the near neighbor group. Thus, it appears that honeybees move pollen further than one would expect from observations on foraging habits alone. Therefore, the few patches of genetically similar individuals we found in the local scale study probably cannot explain the yield differences seen in lowbush blueberry because 1) bees are moving pollen further than expected and 2) genetic relationship does not appear to have a dramatic effect on yield anyway except in the case of self-crosses.

Phylogenetic study of species within the *Cyanococcus* section. EST-PCR markers were also used to examine evolutionary relationships among the different blueberry species within the section *Cyanococcus* of the *Vaccinium* genus (Rowland et al. in press). At least three wild representatives of each of the blueberry species at each ploidy level were included in the study for a total of more than 50 genotypes. Species included: the diploids *V. boreale*, *V. corymbosum*, *V. darrowii*, *V. elliottii*, *V. myrtilloides*, *V. pallidum*, *V. tenellum*; the tetraploids *V. angustifolium*, *V. corymbosum*, *V. hirsutum*, *V. myrsinites*, *V. pallidum*, *V. simulatum*; and the hexaploids *V. constablaei* and *V. virgatum*. UPGMA (unweighted pair group method analysis) and NJ (neighbor joining) trees were constructed and were nearly identical based on 249 polymorphic EST-PCR markers. Interestingly, almost all the *V. elliottii* representatives grouped together and separate from the *V. caesariense* (or diploid *V. corymbosum*) and *V. atrococcum* (also known as *V. fuscatum* Ait.) representatives, providing support for considering *V. elliottii* a separate species from *V. corymbosum*. Sequences from the nuclear granule bound starch synthase (*waxy*) gene are also being used to examine evolutionary relationships among the species. This work is being done in collaboration with Ann Powell at the University of Evansville in Indiana and should be completed this year.

D. Outreach Program

Our main objectives of our Outreach program were to provide various forms of training in molecular breeding (including mapping, QTL identification, marker-assisted selection, and bioinformatics) to breeders and industry representatives, high school interns, and undergraduate and graduate students through participation in our project. We also proposed to develop a website to make our new transcriptome sequencing results publicly available. To date, we have provided training to 11 high school interns, who participated directly in the mapping projects in the Rowland lab. They learned to extract DNA from the mapping plants, design primers from ESTs, set up PCRs, and run gels to identify markers for mapping. They then followed segregation of the markers in the diploid mapping population, entered the data in an EXCEL spreadsheet, and participated in running the mapping programs to add their markers to the current map. One of the students also participated in analyzing the transcriptome sequencing data, identifying potentially differentially expressed genes from the number of reads of the most highly abundant sequences. Students' research results were presented at their high school science fairs and written up in thesis style. Students also presented their research at the annual Beltsville Agricultural Research Center (BARC) Poster Day along with other new scientists and postdocs at BARC. In addition, twelve undergraduate students, seven graduate students, and two postdocs received training while contributing to this project in various PIs' laboratories. They participated in a variety of areas including phenotyping and genotyping the mapping populations, population genetic studies on lowbush blueberry, phylogenetic study of *Cyanococcus* species, and bioinformatic analyses. One of the high school interns and one college student traveled to New Zealand and worked on the project there as well, where they were trained in DNA extractions, primer design, and the HRM technique and analysis. In addition, a video was made for lowbush blueberry grower education to present the concepts of outcrossing, self-compatibility, and clonal habit (Rose and Drummond 2010).

We also held two major meetings during our project, one in Kalamazoo, MI (in conjunction with the 11th North American Blueberry Research and Extension Workers' Conference) and one

stand-alone two-day meeting (A Workshop on Blueberry Genomics and Breeding) in East Lansing, MI. At these meetings, we shared our results, discussed our expectations, and provided training in marker-assisted breeding to blueberry breeders, industry representatives, advisory board members, and molecular geneticists. Our renewal proposal, in large part, is based on these interactions. Finally, as part of this project, we developed a public website where the transcriptome sequences and various assemblies, their annotations, and their frequencies in each of the libraries are available (http://bioinformatics.towson.edu/BBGD454).

IV. RATIONALE AND SIGNIFICANCE

The fastest growing fruit crop in the world is blueberry, with a production increase of nearly 50% over the last five years. Blueberry is a high value small fruit crop that thrives on acidic, imperfectly drained sandy soils, which might otherwise be considered worthless for agricultural crop production. With the discovery in recent years of its many health-promotive properties, demand for blueberries by the consumer is at an all-time high. The predominant cultivated blueberry species are *V. corymbosum* L. (highbush blueberry), *V. virgatum* Ait. (rabbiteye blueberry), and native stands of *V. angustifolium* Ait. (lowbush blueberry). Highbush cultivars are further separated into northern and southern types depending on their chilling requirements and winter hardiness.

Highbush and rabbiteye blueberries are now a major international crop, with about 40,500 hectares planted in North America, 16,200 in South America, 7,300 in Europe, 2,000 in China and Japan, and 1,200 in Australia and New Zealand (Brazelton 2009). Overall world production of highbush and rabbiteye now exceeds 300 metric tons. Of that total, about 75% is northern highbush, 10% is southern highbush, and 15% is rabbiteye. Lowbush blueberries are also harvested from over 60,700 hectares of natural stands in eastern North America, with annual production around 200 metric tons.

Much of the increased consumption of blueberries is thought to be due to greater awareness of their health benefits. Blueberries have been shown to be one of the richest sources of anthocyanins and antioxidants of all fresh fruits and vegetables (Prior et al. 1998). In recent studies, anthocyanins have been linked to many health benefits such as reducing eyestrain, improving night vision, helping to prevent macular degeneration, and exhibiting anti-cancer activity (Cho et al. 2004; Kalt et al. 2007). Antioxidants in general have been linked to fighting aging, cancer, and heart disease.

Using traditional breeding methods, great strides have been made in improving highbush blueberries since their domestication in the early twentieth century. Breeding efforts have focused on the development of cultivars with broader climatic adaptation, season extension, disease and pest resistance, mechanical handling tolerance, and high fruit quality (Galletta and Ballington 1996; Hancock and Retamales 2011). With regard to broader climatic adaptation, many breeding programs in the U.S. have worked on development of low-chilling highbush hybrid cultivars (predominantly *V. corymbosum* but with the very low-chilling diploid species *V. darrowii* in their parentage) suitable for growing in the southern U.S., because they are earlier ripening than the rabbiteye cultivars grown there (Hancock and Draper 1989; Galletta and Ballington 1996). Always a consideration in breeding cultivars for the northern and midwestern

U.S. is that the cultivars have an adequate level of cold tolerance. Indeed, several years ago a survey was performed of all the blueberry research and extension scientists in the U.S., and lack of mid-winter cold hardiness and susceptibility to spring frosts were identified as two of the most important genetic limitations of existing cultivars (Moore 1993). This is probably even more of a concern today with use of more diverse germplasm in breeding programs, especially if southern-adapted germplasm comprises a significant part of the genetic background, as it does for some of the newer releases (Ehlenfeldt et al. 2006).

High fruit quality has remained a high priority, with the fruit characteristics most sought after in blueberries being flavor, large size, light blue color, a small scar (because this is an entry site for pathogen infection), easy fruit detachment for hand or machine harvest, firmness, and a long storage life. A major goal in northern highbush breeding programs has been to combine late ripening with high fruit quality, while in southern highbush populations the combination of early ripening with high fruit quality has been paramount. In this effort to extend the highbush blueberry fruiting season, however, a major obstacle that has been faced by both the northern and southern highbush breeders has been generating very early and late harvested genotypes that have a pleasant sugar to acid ratio. Some early and late harvested genotypes do emerge with sweet flavors, but a high percentage of them are very tart. An understanding of the linkage relationships between harvest date and fruit quality QTL could aid significantly in developing breeding strategies and improving efficiency through marker assisted breeding. Specific breeding for improved phytonutrient content has not yet been undertaken, but rapid improvement could probably be made in this area since much variability exists in these traits among available germplasm (Ehlenfeldt and Prior 2001; Connor et al. 2002a and b; Taruscio et al. 2004; Mladin et al. 2006).

While blueberry breeders have been highly successful using traditional approaches, the incorporation of marker–assisted techniques could result in more cost-effective cultivar development strategies, principally by allowing breeders to better select parent plants for crosses and eliminate offspring with undesirable traits before they are field planted for long-term evaluation. The chilling requirement and cold tolerance of progeny genotypes could also be more accurately predicted, allowing breeders to evaluate them in appropriate environments. Marker-assisted breeding could significantly aid in genetic improvement particularly when combining certain traits like climatic adaptation and season extension with other important traits like fruit quality. Woody perennials, like blueberry, are especially suitable for improvement via marker-assisted breeding because of their long generation times, high heterozygosity, self-incompatibility, inbreeding depression, and polyploidy of commercial types.

Unlike with highbush and rabbiteye blueberry, there has been little traditional breeding of lowbush blueberry. Traditional breeding methods were undertaken in the 1970's in Nova Scotia and some lowbush cultivars were released. However, given that this rhizomatous species has proven very difficult to propagate and that it is very slow growing, with the time to fruiting via current propagation practices at ~ 6 years, breeding efforts have been essentially abandoned. More recent work by Jamieson (2008) to develop hybrid seed families that produce high yielding plants has provided an alternative to traditional cultivar development. Lowbush plants grown from seeds produce rhizomes more freely than plants propagated from cuttings. In addition, seed propagated families have the advantages of lower cost and the ability to store and germinate seed

as needed, in addition to more rapid plant establishment. Therefore, we believe that identification of elite genotypes with high self-fertility and productivity to generate hybrid seed families could be utilized to increase production. Seedlings from such seed families could be used to replace poor performing clones and fill in bare areas of existing production fields, as well as establish new industries where lowbush are not traditionally grown such as Europe.

With the information we have obtained in the first phase of this SCRI-funded project, we are clearly at the cusp of incorporating marker-assisted breeding strategies in active highbush breeding programs with a real, demonstrable impact on the blueberry industry. Molecular markers for candidate genes and QTL are now available that show high promise in the breeding of highbush blueberries, and several platforms now exist for the discovery of many more useful genes and markers. We also have a strategy, based upon high yielding seed families, worth testing for feasibility of increasing productivity of lowbush blueberry in existing fields and establishing new fields. A close-knit team of blueberry molecular geneticists and breeders has been assembled to move blueberry improvement to the next generation.

When we began our USDA-SCRI funded project three years ago, only a few thousand ESTs were publically available for work on blueberry, focused primarily on cold acclimation research (Dhanaraj et al. 2004, 2007). Another ~16,000 ESTs had been developed from blueberry fruit by Plant & Food Research (formerly HortResearch) in New Zealand, but they were not publically available. Only a few hundred robust molecular markers like EST-SSRs (Boches et al. 2005) and EST-PCRs (Rowland et al. 2003c) had been generated for mapping. The first mapping populations and genetic linkage maps had been developed but were not sufficiently saturated for QTL mapping (Rowland et al. 1999, 2003b; Brevis et al. 2007).

Today, through our USDA-SCRI support, we have generated 600,000 ESTs from different organs of blueberry: fruit at different stages of development (green, white, pink, and blue stage), flower buds at different stages of cold acclimation (0', 400', 800', and 1300' of chilling), and leaves and stems, using 454 sequencing. These sequences have been assembled into about 15,000 contigs and 124,000 singletons (Rowland et al. submitted), and are publicly available in the short read archive of GenBank and on our website (http://bioinformatics.towson.edu/BBGD454). If desired these assembled sequences can be used by scientists now and later (after assembly with other newly available ESTs) to develop their own custom microarrays for gene expression studies. From the 454 sequences, we identified 15,886 new blueberry EST-SSR loci. Diploid and tetraploid blueberry maps have been generated, each with over 200 SSR and EST-PCR markers, and QTL have been identified for a wide array of traits including chilling requirement, cold tolerance, rate of development and fruit quality (for a detailed description, see the Progress Report section). Studies have been carried out to investigate the spatial genetic structure (patterns of genetic relationship) of lowbush blueberry on both a local scale (within a field) and on a wide scale (across several cold hardiness zones in Maine and extending beyond Maine), and a gene flow study on lowbush blueberry is near completion. A diploid blueberry genome sequence is emerging now, and is available to our team of scientists for gene discovery, marker development, and integration with our genetic linkage maps and identification of candidate genes near QTL of interest. The identification of SNP markers is also well on its way enabling the development of the first SNP chip for blueberry.

The experiments outlined in this proposal will further pave the way to the molecular breeding of highbush blueberries. With two additional years of data collection and validation, we feel our breeder cooperators will be able to use marker-assisted selection to produce early and late ripening blueberry varieties with exceptional fruit quality and appropriate climatic adaptations. We will also make selections of high yielding, self-compatible lowbush clones with excellent fruit quality, make crosses among them, and begin evaluating seedlings in the field and greenhouse. Many of our stakeholders are breeders that are integrally involved in the project as co-PDs, and our new research plan involves working directly with them to identify markers linked to their traits of interest in their own breeding populations. We have already had two major meetings where a group of about 20 blueberry breeders, industry representatives, and molecular geneticists sat down together to share their research results and expectations. This proposal, in large part, comes from these interactions.

A critical aspect of the Specialty Crop Research Initiative is the expected delivery of research results through technology transfer, outreach, teaching, and other forms of training. In the first phase of this project, we provided training to 11 high school interns, who participated directly in the mapping projects. Twelve undergraduate students, seven graduate students, and two postdocs received training while contributing to this project in the PIs' laboratories also. One of the high school interns and one college student traveled to New Zealand and worked on the project there as well. Early in our project we also formed an advisory panel. At two major meetings held during our initial project period, one in Kalamazoo, MI and one in East Lansing, MI, we shared our results, received input, and provided training in marker-assisted breeding to breeders, industry representatives, advisory board members, and students.

Besides working directly with breeders and their breeding populations, our new research plan includes a strong social science component. Its major goals include helping molecular geneticists better target their science and outreach to address breeders' needs and concerns, clarifying areas of confusion, and identifying obstacles limiting adoption of molecular-assisted breeding. We will also hold another major meeting to share our results and get input again from breeders, industry groups, advisory board members, etc., to assess how our project is going and how we can improve. We will also provide a demonstration of our lowbush work to lowbush growers in Maine. High school interns and college students will continue to be involved in carrying out the research. A new collaboration will ensue which involves working with undergrads at Davidson College in North Carolina, who are studying genomics and bioinformatics. They will utilize our websites, our whole genome sequence, transcriptome sequences, and genetic linkage maps as training tools in their classroom. The scientists on our project will provide the students with a list of potential projects to work on, like searching the literature for candidate genes for the various traits we are most interested in, finding these genes in the whole genome and transcriptome sequences, and, once found, designing primers to the genes or SSRs near the genes to determine if they map to any of the QTL for that particular trait. In this way, they will make a major contribution to our project while gaining experience working on a real-life problem. We feel at the conclusion of the activities outlined in this proposal that, in addition to actually implementing marker-assisted selection in blueberry breeding programs, molecular geneticists will have a clear vision of what the breeders need for future work, and a new generation of students will have experience in the application of genomic information to real breeding problems.

V. APPROACH

A. Highbush Blueberry

Objective 1. Complete the assembly and annotation of all available transcriptome sequences of blueberry.

Overview

We have already assembled and annotated our 454 blueberry transcriptome sequences from the previous project. However, there are now new publicly available ESTs from blueberry, and we would like to assemble our existing 454 blueberry sequences with all the other blueberry transcript sequences present in GenBank, including the newly deposited 17,134 ESTs from two fruit libraries (Zifkin et al. 2011).

Methods

We will assemble all the publicly available blueberry transcriptome sequences using the de-novo assembly tools, Cap3 (Huang and Madan 1999) and SeqMan Pro (www.DNAStar.com) to cluster the ESTs into unigene sets. Assembly programs do vary in the way they assemble reads, so using more than one as verification is always advantageous. The unigene sequences will then be annotated using BLASTX and domain finding tools. The sequences will also be assigned to the three major gene ontology (GO) categories (biological process, cellular component and molecular function). Annotation and GO assignment will be done using the Blast2Go software package (Conesa and Götz 2008).

A database, built on MySQL, will be designed to hold and analyze the data. We will also develop an online user interface to allow scientist to view and query the data from the web. The user interface and search capabilities will be built using ASP.NET and visual studio (Microsoft, Redmond, OR). The latest blueberry mapping data will be added to the database as well. Dr. Alkharouf has extensive experience in DNA sequence analysis and building online genomic databases. He designed and built the Soybean Genomic and Microarray Database (Alkharouf et al. 2004) and the Blueberry Genomic Database (Alkharouf et al. 2007).

Deliverables

1. All the blueberry transcriptome sequences assembled into one annotated unigene set. This can be used by scientists in construction of custom microarrays. This can also be used to map the transcript sequences to the genome sequence.
2. An expanded publicly available website where scientists can view and query the transcript data as well as the mapping data.

Objectives 2 and 3. Add simple sequence repeat (SSR) markers, developed from the scaffolds of the whole genome sequence assembly, and other markers (developed from candidate genes of traits of interest) to the current diploid and tetraploid genetic linkage maps of blueberry. Anchor the maps to the whole genome sequence and align the two maps to each other.

Overview

Genomic sequencing of the diploid blueberry selection W85-20 (*V. corymbosum*) was initiated with NC State funding in 2009. This sequencing effort represents a de novo hybrid approach that includes both large Roche insert libraries of varying sizes (3, 8, and 20Kb) and short Illumina reads. To date, the Roche 454 assembly (assembled with Newbler version 2.6) has resulted in 13,757 scaffolds that consist of 393 Mbp in length (N50 is 145Kbp). The largest 500 scaffolds represent approximately 1/3 of the blueberry genome (153 Mbp) with an average scaffold size of 306,341 bp. The Illumina assembly utilizing multiple assemblers (Myra, Soap deNovo and others) has resulted in an assembly of over 100,000 contigs spanning a total length of over 300 Mbp. Efforts are underway now to integrate these two assemblies, and the expectation is the draft sequence will be made publically available in 2012. The Roche-454 assembly is available through the Genomic Database for Vaccinium (Vaccinium.org) but is currently password protected.

Over 43,000 SSRs (di-, tri-, and tetra-nucleotide repeats) have been identified from the 454 assembly and 1000 primer pairs representing SSRs on the 500 largest scaffolds (2 SSR primer pairs per scaffold) were synthesized in 2011. Initial screening of 100 of these primer pairs (SSRs designated as KAN) on parent plants of our diploid and tetraploid mapping populations has demonstrated the usefulness of the markers in further saturating these genetic linkage maps.

The interspecific diploid mapping population segregates for mid-winter cold tolerance and chilling requirement, and has been described in detail elsewhere (Rowland et al. 1999). Briefly, a true diploid testcross population was generated by crossing a *V. darrowi* selection Fla4B (low chilling, cold sensitive, evergreen) x diploid *V. corymbosum* selection W85-20 (high chilling, cold hardy, deciduous) F1 hybrid to another diploid *V. corymbosum* W85-23. The current map (Appendix 1) was constructed using JoinMap and is currently comprised of over 200 markers on 15 linkage groups, including RAPD, SSR, EST-PCR, and HRM markers. We expect to have over 300 markers mapped before funding of this renewal would begin.

The tetraploid mapping population is an F1 population that was generated from a cross between the northern highbush cv. Draper and the southern highbush cv. Jewel. The population is segregating for many traits including cold hardiness, chilling requirement, fruit stem scar size, fruit color, yield, soluble solids, total titratable acids, etc. Since highbush blueberry is an autotetraploid, the maps of the parent plants are being constructed using the software program TetraploidMap. The current map of the first parent 'Draper' is comprised of 287 markers on 15 linkage groups (Appendix 2).

Methods

We plan to continue screening the other 900 SSR primer pairs and genotype the polymorphic SSRs in the diploid and tetraploid mapping populations to allow us to align the 500 largest 454 scaffolds in our current assembly to the genetic linkage maps. With our current genomic assembly, this will represent 1/3 of the genome but, as the assembly improves, we expect the total genome size represented by these markers will considerably improve as well. Whenever possible, 2 SSRs/scaffold will be utilized to identify both the location and orientation of scaffolds. The SNP markers described below (Objective 4) will also be used to align a significant portion of the blueberry genome to the genetic linkage maps.

In addition, we plan to continue adding EST-PCR and HRM markers to the diploid map for certain genes, such as candidate genes of our traits of interest. Attempts will also be made to map EST-PCR markers in the tetraploid population. Markers (SSRs, EST-PCRs, SNPs) that can be mapped in both the diploid and tetraploid populations will be used to align the two maps to each other and determine the degree of synteny and colinearity.

Deliverables/Use of research results
1. Further saturated maps of the diploid and tetraploid populations
2. Alignment of the maps to each other and to the whole genome sequence

By aligning the genomic sequence to the genetic linkage maps, breeders will be in a position to identify additional markers in regions where QTL are identified. By saturating these regions with markers, the location and effect of individual QTL can be better estimated. In addition, the tighter linkage of markers to QTL will greatly reduce the amount of linkage drag (the transfer of undesirable, yet linked DNA) that occurs through marker-assisted selection. A further benefit of aligning the genomic sequence to the genetic linkage maps is that scientists will be able to identify candidate genes in regions of QTL. This allows for the possibility of designing gene specific markers based on candidate genes as well as increasing our understanding of the biological processes that control the various desirable traits.

Objective 4. Develop a single nucleotide polymorphism (SNP) chip and add SNPs to the genetic linkage maps.

Overview
A SNP chip allows for typing of thousands of SNPs simultaneously. Therefore, use of a SNP chip for genotyping the individuals of a mapping population can result in a densely saturated map in a very short period of time. To develop a SNP chip for genotyping our blueberry mapping populations, short read sequences from parents of the tetraploid and diploid mapping populations and one additional highbush blueberry cultivar Emerald were generated. A pipeline was developed for SNP detection using SOAP and SoapSNP. The pipeline uses SOAP to align the data against the current assembled genome.

Methods
Next the pipeline will be used to sort the alignment data first by scaffold and then by position. Once the data are sorted, we will be able to process it through SoapSNP and format the output by cultivar. We expect the assembly of the W85-20 genome in the next 6 months to generate a good reference for which to align the short read sequences. The latest version of the assembly that contains 13,000 scaffolds will be tested. The pipeline will allow us to quickly identify SNPs that are heterozygous in one parent and homozygous in the other and therefore mappable in each of the populations. We will target these SNPs that are also polymorphic in ‘Emerald’ to assess these SNPs in the blueberry core collection housed at the USDA-ARS, NCGR in Corvallis, Oregon. Presently, we plan to construct and use an Illumina GoldenGate chip to genotype 1152 SNPs in 480 samples. Samples include the tetraploid population, the diploid population, and accessions from the core collection and other important blueberry cultivars.

Deliverables
1. SNP chip
2. Further saturated maps of the diploid and tetraploid populations using SNPs
3. Enhanced ability to align the maps to each other and to the whole genome sequence
5. Further genotyped blueberry core collection at Corvallis, Oregon using SNPs

Objective 5. Increase marker coverage in the vicinity of QTL (identified in the previous project) associated with chilling requirement, harvest date, and fruit quality traits (soluble solids, pH, titratable acidity, and picking scar), and validate and/or identify new QTL linked to these traits in additional actual breeding populations.

Overview
One of the major obstacles that has been faced by northern and southern highbush blueberry breeders has been generating very early and late ripening genotypes that have a pleasant sugar to acid ratio and small picking scar. A small scar is essential for the fruit to maintain high quality throughout the wholesale fresh market shipping chain. Some early and late harvested genotypes do emerge with sweet flavors and small scars, but a high percentage of them are very tart and/or have a large scar. An understanding of the linkage relationships between harvest date and fruit quality QTL could aid significantly in developing breeding strategies and improve efficiency through marker assisted breeding (MAB). Bottom line – useful parents could be more efficiently identified using the molecular markers and progeny that do not harbor the gene(s) for the desired traits could be selected out in the greenhouse before field planting.

In the first phase of this research, a segregating F1 population of 'Draper' x 'Jewel' was planted at five locations (Gainesville FL, Waycross GA, Invergowrie, Scotland, Corvallis, OR and Benton Harbor, MI), and was evaluated in 2011 for a wide array of developmental and fruit quality characteristics. This population will be evaluated for the same characteristics at all locations in 2012. A linkage map was developed of this population using SSR markers and then marker linkages with these horticultural traits were identified in a QTL analysis. Using the phenotypic data collected in 2011, QTL for most of these traits were identified at multiple sites but their utility needs to be further verified in additional families. More tightly linked, flanking markers also need to be added to the linkage groups to limit recombination between the SSR/trait combinations in breeding populations.

Specific goals and methods
1) Increase marker coverage in the vicinity of the QTL associated with chilling requirement, harvest date, and fruit quality traits (soluble solids, pH, titratable acidity, and picking scar).

The emerging diploid blueberry genome sequence assembly will be searched for the mapped markers (SSR primer sequences) nearest the QTL associated with chilling requirement, harvest date, soluble solids, pH, titratable acidity, and scar. Scaffolds will be identified that contain these sequences, and additional markers will be developed from other SSRs and gene sequences present on these scaffolds and mapped in the Draper x Jewel population to increase marker coverage in the vicinity of these QTL. As more and more scaffolds from the genome sequence assembly are located on the map (Objectives 2 and 3), likewise more SSRs and gene sequences near the QTL can be identified and added to the genetic linkage map. Furthermore, with the

addition of SNPs to the maps (Objective 4), many SNPs near the QTL of interest will be identified. Using the genome and transcriptome sequence assemblies, potential candidate genes for the QTL regions will also be identified and mapped (Objective 11—work with Davidson College).

2) Verify in additional families the QTL for chilling requirement, cold hardiness, harvest date, and fruit quality traits (soluble solids, pH, titratable acidity, and picking scar) found in the analysis of the Draper x Jewel population. Also, search for new QTL for chilling requirement, cold hardiness, harvest date, and fruit quality traits (soluble solids, pH, titratable acidity, and picking scar) in these additional families.

Each cooperating breeding program will select two to four families of 100 individuals that are already in the field and appear to be segregating for these traits.

--Low chill, southern highbush populations with the emphasis on finding QTL for early ripening genotypes with high SS/TA ratios and small scar.

--High chill, northern highbush populations with the emphasis on finding QTL for late ripening genotypes with high SS/TA ratios and small scar.

We expect to select about eight families of southern and northern highbush, for a total of 16 populations. Flowering and harvest dates will be evaluated throughout the growing season in 2013 and 2014 for each individual, by estimating the percentage of open flowers and fully blue fruit on a 3-4 day schedule. When each individual's fruit is 30% blue, fruit weight, soluble solids, pH, and titratable acidity will be evaluated on a sample of 25 fruit. The fruit scar will be characterized objectively (diameter) and subjectively (rating of any tearing or "wetness" of scar). For validation of the Draper x Jewel QTL in other populations, 24 individuals [2 parents plus 22 progeny (11 of each extreme phenotype)] from the 16 populations will be genotyped first using SSR markers that were found to be located near the QTL of interest in the Draper x Jewel population. Thus, it will be determined in each of the new populations if there is an association between the presence of a particular SSR allele and the trait in question, as there was in the Draper x Jewel population. To search for new QTL for these traits in the breeding populations, the same 24 individual plants from each population will be further genotyped using a group of about 150 SSR markers previously mapped in the Draper x Jewel population and located at about 10 cM intervals across all the linkage groups. If there appears to be an association between the trait(s) in question and regions of the map not identified in the Draper x Jewel population, then a larger group of individuals from that population will be genotyped for more markers near the region in question to more precisely locate the potential QTL.

Deliverables:
1. Identification and verification of QTL associated with chilling requirement, harvest date, fruit quality traits (soluble solids, pH, titratable acidity, and picking scar) in each cooperators breeding population

2. Identification of markers close enough to the traits of interest to use in marker-assisted breeding

Objective 6. Identify and test candidate genes near QTL for controlling traits of interest.

Overview
Scientists involved on the project and students at Davidson College (Objective 11) will identify candidate genes for QTL of traits of interest by searching the literature for work done in other plants and by identifying genes near QTL from the whole genome assembly. To test whether these candidate genes actually control the traits in question, we will examine expression of the genes under appropriate conditions. For example, expression of candidate genes for chilling requirement would be examined during chill unit accumulation of genotypes with low and high chilling requirements. Expression of candidate genes for ripening time would be examined during fruit ripening of genotypes with early and late ripening times. A few selected genes may be cloned and used to transform Arabidopsis or tomato to test gene function. There is not time during this grant period to test gene function by transformation of blueberry.

Methods
Expression of selected candidate genes will be examined under appropriate conditions as described above by real-time PCR as we have done previously with the CBF gene, a transcription factor that turns on a family of cold-responsive genes during cold acclimation (Polashock et al. 2010). In this case, we compared CBF transcript levels in flower buds of cold sensitive and cold hardy blueberry varieties during cold acclimation and found CBF levels to be higher in the more cold hardy variety. If we believe transformation of Arabidopsis or tomato would help to establish the function of a particular gene, we will clone it and overexpress it in Arabidopsis or tomato as we have done previously with the blueberry CBF gene in Arabidopsis (Polashock et al. 2010).

Deliverables
1. Determination of function of candidate genes.

B. Lowbush Blueberry

Objective 7. Identify and propagate high yielding, good selfing clones (genotypes) with good fruit quality.

Overview
Lowbush is predominantly outcrossing (Aalders and Hall 1961; Hall et al. 1979) and self-fertility is poor due to early-acting inbreeding depression (Hokanson and Hancock 2000). Our previous research has indicated, however, that a fair number of individual genotypes (clones) are self-fertile (Bell et al. 2010). In addition, we found a strong positive correlation between yield potential and the degree of self-fertility observed in clones (Bell et al. 2010). This is particularly important for improving yields in existing lowbush blueberry fields and for initiation of new fields by seed. Therefore, a logical strategy for selecting lowbush blueberry genotypes to be the foundation of seed families (Jamieson 2008) that can be used for planting is to use self-fertile high yielding clones with high fruit quality. However, we do not know if F1 generations from crosses of self-fertile clones will be characterized by viable seed, self-fertile plants, and good fruit quality and quantity.

Methods

Presently, we do not have any means of identifying clones with high-levels of self-fertility (> 75% fruit-set) with genetic markers. In the first year of this study we plan on identifying 15-20 self-fertile clones at the University of Maine Blueberry Hill Research Farm in Jonesboro, Maine. This will be performed by selfing 80-100 geo-referenced clones via field hand pollinations as described by Bell et al. (2010), including clones previously found to be good selfers. The level of selfing will be determined by assessing the proportion fruit-set obtained in each of the self-crosses. Leaf tissue will also be collected from an equal number of good selfers and poor selfers (15-20 of each). DNA will be extracted from each of the leaf samples and the clones will be genotyped using approximately 50-100 single locus SSR markers (as determined from our highbush mapping studies). The SSRs will be used to determine each clone's level of heterozygosity (defined as the average number of SSR alleles per locus) as described in Brevis et al. (2008). We have unpublished data that suggests that good selfers may have higher levels of heterozygosity than poor selfers. Thus, we will determine if there is a positive correlation between selfing ability and heterozygosity.

At harvest, fruit quality characteristics will be measured on open- and self-pollinated fruit of each self-fertile clone such as: proportion of mature seeds / fruit, fruit weight, color, flavor, soluble solids, pH, and titratable acidity, as well as yield (from 1 m^2 rakes). Fruit (both open-pollinated and self-pollinated) will also be harvested from each of these clones for extraction of seed. Seed will be extracted and germinated as we have done previously (for the earlier lowbush gene flow study) and proportion germination will be recorded for each clone. Seedlings will be grown in the greenhouse under optimal conditions to assess growth rate and survival from self-crosses as compared to open pollinations.

Deliverables

1. Identification of 15-20 self-fertile lowbush blueberry clones.
2. Evaluation of relationship between heterozygosity and self-fertility using SSRs.
3. Identification of high yielding, self-fertile clones with good fruit quality.
4. Assessment of seed viability and seedling growth rate and survival from self-crosses as compared to open pollinations.

Objective 8. Carry out a diallel crossing design of clones to generate seed families for field testing.

Overview

Attempts were made to develop lowbush blueberry cultivars through traditional breeding in the 1970's in Nova Scotia and some lowbush cultivars were released. However, the cultivars, propagated through cuttings, tended to produce few rhizomes and spread very slowly. Therefore, these varieties were not widely planted. Because lowbush plants grown from seeds produce rhizomes more freely, Jamieson (2008) developed several hybrid seed families of lowbush blueberry that consistently produce high yielding plants as an alternative to planting cultivars. In addition to more rapid plant establishment, seed propagated families also have the advantages of lower cost and the ability to store and germinate seed as needed. Seedlings from such seed families could be used successfully to replace poor performing clones and fill in bare areas of existing production fields.

Seed families of high yielding clones might be best derived by crossing high yielding, good selfing clones. We have previously found that self-fertility of clones is a good predictor of outcross yields (Bell et al. 2010). Whether or not seedlings from crosses of high yielding, good selfing clones would tend to be high yielding, good selfers too, is not known.

Methods.
In year 1, we will obtain seeds from two seed families described by Jamieson (2008), KBF-10, reported to be a potentially good family for filling in bare areas of fields, and KBF-7, reported to be a good family for fresh market fruit production. We will germinate the seeds and grow the seedlings in the greenhouse in both Maine and Scotland. The Scottish Crop Research Institute is interested in testing the feasibility of establishing a lowbush blueberry industry in Scotland while we are interested in testing the seedlings for filling in bare areas of established fields in Maine. In year 2, we will plant the seedlings for field-testing in replicated trials at the Blueberry Hill Research Farm in Maine and in Scotland. In year 2, we will also select 5-10 self-fertile clones with high yield and good fruit characteristics from the Blueberry Hill Farm (Objective 7, above) and carry out a complete diallel crossing design (including selfs). Seeds will be extracted from fruit and germinated for proportion germination assessment and future performance trials of F1 seedlings.

Deliverables
This work will allow the future evaluation of F1 seedlings from various seed families. It will allow the identification of the best genotypes for development of seed families for future propagation.

Objective 9. Assess tetraploid genetic linkage maps (developed for highbush blueberry) for use with lowbush blueberry.

Overview
Genetic linkage maps have now been developed from a tetraploid highbush blueberry population (Draper x Jewel) and an interspecific diploid population [(*V. darrowi* x *V. corymbosum*) x *V. corymbosum*]. Lowbush blueberry has not been the focus of genetic mapping largely due to the lack of breeding effort in this crop species. However, we have shown that many of the EST-PCR and EST-SSR markers derived from highbush blueberry work well for characterizing lowbush blueberry (Bell et al. 2008). Use could possibly be made of the highbush genetic linkage map and QTL for cold hardiness, high fruit quality, etc. in lowbush blueberry if there is conservation of the markers and good synteny between the genomes.

Methods
We propose following a subset of the mapped highbush SSR markers in one of the lowbush F1 families, KBF-7 or KBF-10, from Jamieson (2008). Leaf tissue will be collected from ~100 seedlings of the F1 population. DNA will be extracted, and a subset of ~90 mapped SSR markers that are polymorphic in the lowbush population (about 6 per linkage group x the current 15 linkage groups) will be followed in the population. Assuming *V. angustifolium* is an autotetraploid like *V. corymbosum*, TetraploidMap will be used to construct a map of the lowbush population. The map will be compared to the tetraploid highbush map. Linkages of markers, marker positions, and orders will be compared to determine the degree of synteny

between the two maps. In a separate study, a group of lowbush blueberry clones from different cold hardiness zones are currently being evaluated for cold hardiness and chilling requirement to learn more about the variation of these traits in wild populations in Maine. Because this data will be available to us, markers found to be linked to cold hardiness and chilling requirement in the diploid and tetraploid highbush mapping populations will be followed in this collection of lowbush clones. Thus, we will determine if the markers identified in the highbush population are associated with these traits in this lowbush collection. This work will enable us to evaluate the utility of the highbush map for use with lowbush blueberry and learn more about the evolution of the blueberry genome.

Deliverables
1. Determination of the usefulness of a tetraploid genetic linkage map derived from *V. corymbosum* for use in understanding plant response and adaptation in lowbush blueberry.

C. Social Science/Outreach Component

Objective 10. Carry out survey of breeders, growers, and molecular geneticists.

Overview
Participants at breeder/scientist meetings have identified a major area of concern in taking marker-assisted breeding (MAB) technology to its users: information needs. More specifically, molecular geneticists need several types of information from and about breeders in order to make choices about the science so that it is tailored to its end uses and to effectively craft outreach. Social science research to fill this critical gap will answer the following questions:

- How well do breeders understand the science (e.g., how it works, what it can do)?
- How and when will breeders want to apply it (e.g., what traits, what stage)?
- What do breeders see as the practical, logistical barriers to adoption, and why do they see these as such?
- How well do molecular geneticists understand breeders' needs?

Empirical knowledge of these should produce two outcomes: (1) better decisions about how to develop the science (e.g., which type of markers to use, how dense to make the map) and (2) more targeted outreach. The social science component of this project will focus on assessing barriers to adoption by studying not only breeders' perceptions of the questions above, but also molecular geneticists' perception of these same issues. A comparison will provide more objective information on perceptual differences between the two groups that will limit adoption of MAB if not addressed. Since there are no breeders of lowbush blueberries, the main information need for this crop is growers' perceived barriers to adoption of new seed families for future propagation.

Methods
While there is a large literature on adoption and diffusion in agriculture, as yet there has been little work on plant breeders' adoption of scientific advances. To better understand obstacles to breeder adoption of MAB, this study will use Perceived Attributes Theory, which is part of the larger Diffusion of Innovations Theory (Rogers 2003). Perceived Attributes Theory emphasizes the importance of five characteristics of a new technology that affect adoption: trialability,

observability, relative advantage, complexity, and compatibility. These five characteristics will guide construction of survey and interview questions. To gather information on breeder and molecular geneticist perceptions, this study will use a mixed methods approach (Creswell 2008). The researcher will send an email survey to all blueberry breeders and genomic scientists. There are approximately 15 breeders and 5 genomic scientists. Survey questions will be based on observations from a November 2011 breeder/molecular geneticist meeting (East Lansing, MI) and phone interviews with key informant breeders and molecular geneticists. For lowbush blueberry, we will conduct a survey at grower meetings in Maine. Survey data will be analyzed statistically using techniques such as chi-square and regression analysis. The researcher will supplement the quantitative survey with in-depth, semi-structured phone interviews of all breeders and key informant lowbush growers. These will be recorded (with permission), transcribed, and entered in a QSR NVivo database for inductive coding. The interviews will ask questions on what breeders see as the main uses of the technology and obstacles to adoption, as well as questions about sources of confusion regarding the technology. The interviews will assist in documenting additional perceptions not covered by survey statements.

Deliverables/Use of results

Results will ultimately be used to better target molecular genetic research to meet the needs of blueberry breeders and to aid in implementation of marker-assisted breeding to blueberry breeding programs. By providing empirical information on differences between groups, molecular geneticists/genomicists will be better able to target their science and outreach to address breeder needs and concerns and to clarify areas of confusion. It will also provide more objective information on breeders' constraints that limit adoption, thereby enabling genomic scientists to better tailor their science and outreach and to ensure more widespread adoption. Finally, information on potential obstacles to adoption of lowbush seed families will help scientists and extension specialists craft more effective outreach.

Objective 11. Deliver research results and training to our blueberry stakeholders and provide training in mapping and bioinformatics to high school and college students.

Overview

To deliver research results and training to our highbush, rabbiteye, and lowbush blueberry stakeholders, we will hold two separate meetings, one focused more on marker-assisted breeding for development of highbush and rabbiteye cultivars and one focused on the development of lowbush hybrid seed families. We will also provide training to high school interns in mapping and to undergraduate college students in bioinformatics by participation in our project.

Methods

Within the first few months of funding, we will contact the members of our advisory board to determine if they want to continue on as part of our advisory panel or be replaced. If there are members that want to be replaced, we will decide on a list of candidates and contact them to fill the vacancies. We will continue to include breeders, molecular geneticists, and industry representatives on our board. The main purpose of the board will be to evaluate and make recommendations on our research results and results from our Outreach Program.

For the delivery of research results and training to our highbush and rabbiteye blueberry stakeholders, we will hold another major meeting during the second year of the renewal, similar to our previous meeting (A Workshop on Blueberry Genomics and Breeding) in East Lansing, MI. Like before, we will invite the major blueberry breeders in the U.S., as well as industry representatives, advisory board members, and everyone involved on the renewal to attend and participate. Again, Dr. Jim Hancock at Michigan State University will take the lead on organizing the meeting. At the meeting, we will share our results, including the results from the survey described in Objective 10 above, discuss our expectations, encourage attendees to give their input, and provide more training in marker-assisted breeding. Because the emphasis of this meeting is on marker-assisted breeding, it is geared more toward our highbush and rabbiteye blueberry stakeholders.

For the lowbush blueberry growers, we will organize a separate workshop where results are presented on the development and planting of lowbush hybrid seed families. Most lowbush blueberry growers do not consider yield enhancement by genotype introduction or planting fields *de novo*. This is because almost all of the lowbush blueberry growers in North America produce crops on naturally colonized fields of wild genotypes. However, there are fields in Maine that have been planted with selected genotypes using plantlets grown in tissue culture (Drummond, personal obs.). There are also many growers that have filled in bare spots in their fields with sods of wild plants, although currently the plants are selected without regard to yield potential and/or fruit quality. Therefore, if high yielding seed families are produced, grower adoption of planting seedlings from these families to enhance their field's yield is a possibility. This production practice, however, would have to be introduced, explained, and demonstrated to growers.

The timeline of this proposed project is not long enough to finish the development of commercially available lowbush seed families for planting. However, we believe that it is important to initiate grower education of this potential future technology at an early phase of its development. Dr. Frank Drummond will organize a field workshop in mid-July in 2014 at the annual Lowbush Blueberry Summer Field Day. This field day is held at the University of Maine Blueberry Hill Farm in Jonesboro, Maine. Growers from Maine, the Canadian Maritimes, Quebec, Massachusetts, New Hampshire, and Michigan attend this meeting. The workshop will involve a presentation explaining pollination of lowbush blueberry, outcrossing, self-infertility and self-fertility, and the yield advantage of clones that self. A field demonstration will show the yields of self-fertile clones compared to obligate outcrossing self-infertile clones that are self-pollinated by hand. Seedlings from the seed families of Jamieson (2008) will also be available for the growers to examine. In the fall of 2014, a University of Maine Cooperative Extension wild blueberry fact sheet will be written by Dr. Drummond and published on the Maine wild blueberry web site (http://extension.umaine.edu/blueberries/). This fact sheet will be used to describe the uses and advantages of planting self-fertile clones.

In addition to these two meetings, we will continue efforts to try to interest a new generation of students in agricultural research. High school students will be trained in marker-assisted breeding through research internships by involvement in mapping studies and evaluation of mapping populations for horticulturally significant traits. Lisa Rowland at the USDA/ARS, Beltsville, MD, has been mentoring students from area schools like Eleanor Roosevelt and Flowers High Schools for over 17 years in the use of molecular markers in agricultural research. She will

continue this work with students, involving at least two high school students per year in the diploid mapping project. Generally, students will learn to extract DNA from the mapping plants, design primers from ESTs, set up PCRs, and run gels to identify markers for mapping. They will then follow the segregation of markers in the diploid mapping population, enter the data in an EXCEL spreadsheet, and participate in running the mapping programs to add their markers to the current genetic linkage map. They will also help with phenotyping the mapping population. Students' research results will be presented at local science fairs and written up in thesis style. Students will also present their research at the annual BARC Poster Day.

We will also work with undergraduate students at Davidson College in North Carolina and Towson University in Maryland, who are studying genomics and bioinformatics. Through a bioinformatics class designed by Dr. Malcolm Campbell, the students at Davidson College will utilize our websites, our whole genome sequence, transcriptome sequences, and genetic linkage maps as training tools. The scientists on our project will provide the students with a list of potential projects to work on, such as searching the literature for candidate genes for the various traits we have been following in our mapping populations (chilling requirement, cold hardiness, fruit scar size, soluble solids, titratable acidity, ripening time, flowering time, etc.), finding these genes in the whole genome and transcriptome sequences, designing primers to the candidate genes or SSRs near the genes so the scientists can map the candidate genes or scaffolds that contain the candidate genes, and then determining if any of the candidate genes map to any of the QTL for that particular trait. Other possible projects, which will depend on adequate assembly of the whole genome sequence, are the identification and annotation of genes near the QTL that have already been identified, in an attempt to identify the best candidate genes for the various QTL. Undergraduate students studying bioinformatics at Towson University will help Dr. Nadim Alkharouf in the assembly, annotation and GO assignments of the transcriptome sequences. Dr. Alkharouf will take three of his best undergraduate students in the MB3-Bioinformatics track at Towson University and work with them on this project. There will also be the opportunity for some students to travel to Plant & Food Research to participate in marker development, screening and mapping, as well as bioinformatic analysis. In these ways, undergrad students will make a major contribution to our project while gaining experience working on real-world problems.

Deliverables

1. Workshop on Blueberry Genomics and Breeding.
2. Lowbush blueberry grower workshop at the 2014 annual Lowbush Blueberry Summer Field Day.
3. Extension Fact Sheet on self-fertility and choice of self-fertile clones for planting into existing or new fields.
4. Internships for high school students to participate in mapping project.
5. Training for bioinformatics students at Davidson College and Towson University using our research project.

D. Possible Pitfalls/Limitations

Identification of candidate genes near QTL of interest depends on adequate assembly of the whole genome sequence and on good congruence between the diploid and tetraploid highbush

genomes. Since tetraploid highbush is an autotetraploid thought to have evolved from the doubling of the diploid highbush genome, and interspecific crosses between all the blueberry species of the same ploidy level within the section *Cyanococcus* result in fertile hybrids, we believe there will be good synteny between the genomes of all the species in the *Cyanococcus* section. Efforts to assemble the diploid highbush genome are continuing and the assembly is expected to be published this year before funding of the renewal would take place. Regardless, the assembled sequence will be available to scientists and students working on this project through our co-PD, Dr. Allan Brown.

E. Tentative Schedule

Year 1

Objective 1: All available blueberry transcriptome sequences will be assembled into contigs, and annotation of the sequences will begin.
Objectives 2 and 3: We will begin adding SSRs, developed from the scaffolds of the whole genome sequence assembly, and other EST-PCR and HRM markers (developed from candidate genes of traits of interest) to the diploid and tetraploid genetic linkage maps of blueberry.
Objective 4: The SNP chip will be developed and used to genotype the diploid and tetraploid mapping populations and the accessions from the core collection.
Objective 5: The diploid blueberry genome sequence will be searched for the mapped markers (SSR primer sequences) nearest the QTL associated with chilling requirement, harvest date, soluble solids, pH, titratable acidity, and scar identified in the Draper x Jewel population. Scaffolds will be identified that contain these sequences, and additional markers will be developed from other SSRs and gene sequences present on these scaffolds and mapped in the Draper x Jewel population to increase marker coverage in the vicinity of these QTL. The additional breeding populations will be phenotyped for the first year.
Objective 6: Real-time PCR will be performed to examine expression of some of the candidate genes.
Objective 7: Lowbush clones will be evaluated for self-fertility. Self-fertile clones will be evaluated for yield and fruit quality. High yielding, self-fertile clones with good fruit quality will be identified. DNA will be extracted from good selfers and poor selfers and genotyping for heterozygosity study will begin.
Objective 8: Seed families from Jamieson will be germinated.
Objective 10: Survey of breeders, growers, and molecular geneticists will be performed.
Objective 11: High school students will participate in research internships. Undergrad students will participate in bioinformatics class at Davidson College to identify candidate genes for QTL and in assembly of transcriptome sequences at Towson University. Scientists will attempt to map candidate genes or SSRs near candidate genes.

Year 2

Objective 1: Annotation of transcriptome sequences will be completed. All assemblies, BLASTX results, and annotation tables will be made publicly available through the Blueberry Genomic Database website. Mapping results will also be made available through the website.
Objectives 2, 3, and 4: We will continue adding SSRs, developed from the scaffolds of the whole genome sequence assembly, and other EST-PCR and HRM markers to the diploid and

tetraploid highbush genetic linkage maps. Using these SSRs and SNPs, we will anchor the sequence to the genetic linkage map and align the two maps.
Objective 5: Phenotyping of the breeding populations will be completed. Individuals with extreme phenotypes will be identified and genotyped using SSRs near QTL of interest in the Draper x Jewel population. For populations where QTL cannot be validated, we will search for new QTL using a subset of mapped markers.
Objective 6: We will continue real-time PCR to examine expression of more of the candidate genes. Selected genes will be introduced into Arabidopsis or tomato to test gene function.
Objective 7: Correlation of heterozygosity with self-fertility will be determined in lowbush clones.
Objective 8: Complete diallel will be performed using self-fertile clones. Seeds will be extracted and germinated. Seedlings from Jamieson seed families will be planted in the field.
Objective 9: Individuals from one of the Jamieson seed families will be genotyped with a subset of markers mapped in the tetraploid highbush population. Clones evaluated for cold hardiness and chilling requirement will be genotyped for markers linked to these traits in the highbush populations.
Objective 10: Survey results will be analyzed and presented at meetings.
Objective 11: High school students will participate in research internships. Undergrad students will participate in bioinformatics class at Davidson College to identify candidate genes for QTL and in annotation of unigenes at Towson University. Scientists will attempt to map candidate genes or SSRs near candidate genes. Meetings will be planned and held.

F. Coordination of Work

Lisa Rowland. Dr. Rowland (USDA/ARS, Beltsville, MD) is the Project Director, and, as such, she will hold regular conference calls and make sure the work is coordinated between the various labs. The Rowland lab will participate in adding SSRs (from genomic scaffolds) to the diploid mapping population and EST-PCR markers to the diploid and tetraploid mapping populations. Her lab will receive marker data from the various labs for the diploid population and update the map. Her lab will also participate in genotyping 2-3 of the northern highbush breeding populations for validation of QTL and identification of new QTL. Her lab will help with collecting tissue and RNA extractions needed for expression studies of candidate genes. Her lab will participate in genotyping lowbush clones for heterozygosity study. She will work with high school interns on the mapping project and with undergrad students at Davidson College on the identification of candidate genes for QTL.

Nahla Bassil. The Bassil lab (USDA/ARS, Corvallis, OR) will participate in adding SSRs (from genomic scaffolds) to the tetraploid mapping population. Her lab will also participate in development and genotyping of the mapping populations and core collection with the SNP markers. Her lab will also genotype 2 -3 of the northern highbush breeding populations for validation of QTL and identification of new QTL.

Allan Brown. The Brown lab (NCSU, Kannapolis, NC) will participate in screening SSRs (from genomic scaffolds) against the parents of the diploid and tetraploid highbush mapping populations and adding SSRs to the diploid mapping population. His lab will also participate in development of the SNP chip. Dr. Brown will lead the effort to anchor the genome sequence to

the maps. Dr. Brown will also work with undergrad students at Davidson College on identification of candidate genes for QTL.
Frank Drummond. The Drummond lab (University of Maine, Orono, ME) will evaluate lowbush clones for self-fertility and evaluate self-fertile clones for yield, fruit quality, and germination rate of seed. His lab will collect leaf tissue from good selfers and poor selfers and provide tissue to Dr. Rowland's lab for genotyping to determine level of heterozygosity. His lab will carry out a diallel of good selfers with high yields and good fruit quality. His lab will also germinate seed from the Jamieson seed families and plant seedlings for field trials. His lab will extract DNA from about 100 seedlings of one of the seed families and from clones already evaluated for cold hardiness and chilling requirement and provide DNA to the Scottish Crop Research lab. Dr. Drummond will also arrange the Field Day meeting with lowbush growers.

James Hancock. The Hancock lab (MSU, East Lansing, MI) will assist in phenotyping the breeding populations and in genotyping 2-3 of the northern highbush breeding populations for validation of QTL and identification of new QTL. Dr. Hancock will also arrange the meeting on marker-assisted breeding.

James Olmstead. The Olmstead lab (University of Florida, Gainesville, FL) will assist in phenotyping the breeding populations and in genotyping the southern highbush breeding populations for validation of QTL and identification of new QTL. The Olmstead lab will also participate in adding SSRs (from genomic scaffolds) to the tetraploid mapping population.

James Polashock. The Polashock lab (USDA/ARS, Chatsworth, NJ) will carry out real-time PCR analyses on selected candidate genes for QTL of interest. His lab will also transform Arabidopsis and/or tomato with selected genes to further test their function.

Nadim Alkharouf. The Alkharouf lab (Towson University, Towson, MD) will assemble all available blueberry transcriptome sequences (Sanger ESTs and next generation short reads) and annotate the sequences. They will upload the data to the Blueberry Genomics Database website and also expand the website to include the mapping data. Dr. Alkharouf will oversee undergrad students' involvement in the project.

Chad Finn. The Finn lab (USDA/ARS, Corvallis, OR) will assist in phenotyping the breeding populations.

Samuel Hanes. Dr. Hanes (University of Maine, Orono, ME) will carry out the survey of blueberry breeders, growers, and molecular geneticists and analyze the results.

Todd Mockler. Dr. Mockler will carry out the bioinformatic analyses essential for SNP detection.

Collaborators and/or contractors

Malcolm Campbell. Dr. Campbell will develop and teach the bioinformatics course at Davidson College. He will oversee the undergrad students' involvement in the blueberry project.

Emily Buck and Roy Storey. The New Zealand labs (Institute for Plant & Food Research Ltd) will participate in adding SSRs (from genomic scaffolds) and HRM markers to the diploid mapping population, assist in the diploid map construction, and develop and screen candidate gene-based markers. They will also assist in various bioinformatic analyses such as transcriptome assembly, SNP detection, and anchoring the genome sequence to the maps.

Julie Graham. The Graham lab (Scottish Crop Research Institute) will add SSRs (from genomic scaffolds) to the diploid and tetraploid mapping populations and will oversee construction of the tetraploid map. They will also germinate lowbush seed families and plant seedlings for field-testing. They will also assist with genotyping some of the lowbush blueberry populations.

G. Hazards

We have standard operating procedures and material safety data sheets in place for work with hazardous substances such as ethidium bromide, chloroform, phenol/chloroform, etc., as is mandated by federal and state governments. All transformation work with Arabidopsis or tomato will be conducted in approved lab and greenhouse facilities.