Review

ADVENT OF GENOMICS IN BLUEBERRY

Jose V. Die, Lisa J. Rowland

U.S. Department of Agriculture, Agricultural Research Service, Beltsville, MD

Abstract

Blueberry is a high value crop with recognized nutritional characteristics that has lead to an increase in consumer demand over the last several years. With its increasing agricultural and commercial importance, genetic and genomic tools have recently become available for use in characterizing its genetic diversity and in molecular breeding strategies. Here, we provide an overview of genomic research in blueberry, with a focus on EST/ transcriptome sequencing efforts. These resources are already providing novel insights into various biological processes from large-scale expression studies like microarrays, elucidation of phylogenetic relationships, and development of molecular markers and genetic linkage maps. Future blueberry breeding programs should benefit greatly from these new genomic tools.

Keywords

Bioinformatics; cDNA libraries; EST; NGS; transcriptome

1. Introduction

The United States is the world's largest producer of blueberries (*Vaccinium* spp.). In 2010, U.S. production reached 224,000 tons with a market value of $644 million. The production value increased three-fold during the past decade in the U.S. (USDA-NASS 2013). Worldwide production and demand has also dramatically increased with South America, Europe, and the Asian Pacific regions showing the greatest increases (Bañados 2008). Blueberry is a high value crop and is recognized as one of the most healthy and nutritious dietary sources among common fruits and vegetables (Hou 2003; Prior et al. 1998; Wang et al. 1996; Wu et al. 2004). Its many recognized nutritional characteristics are at least partly responsible for the steady increase in consumer demand over the last several years.

Due to its growing economic importance, there is an increase in efforts to better understand the biology of the blueberry crop, to address challenges in its production, and for crop improvement. Blueberry shows an incredible amount of genetic diversity that has yet to be efficiently characterized. Traditional breeding efforts have made undoubted progress in the short time since its domestication in the twentieth century (Mainland 2012). Efforts are currently focused on development of cultivars with broader soil adaptation and broader climatic adaptation in order to extend the growing areas, disease resistance, and high fruit quality. Molecular tools and high-throughput technologies can help to improve our understanding of the genetics underlying these traits. Molecular breeding could lead to rapid genetic improvement particularly when combining

64 certain traits for climatic adaptation with other important traits like fruit and nutritional quality.
65 Improvement via molecular breeding or marker-assisted selection is especially suitable for
66 blueberry because of its long generation times, high heterozygosity, inbreeding depression and
67 polyploidy, all of which tend to complicate genetic analyses, and can hamper traditional breeding
68 efforts. Major savings in time, labor, and land resources could be achieved if potentially low-
69 value genotypes could be eliminated at the seedling stage before field planting (Qu and Hancock
70 1997).
71
72 Large-scale sequencing projects have resulted in complete genomic sequences of plants ranging
73 from the 125 Mb *Arabidopsis thaliana (Arabidopsis Genome Initiative 2000)* to the 2.5 Gb *Zea*
74 *mays (Schnable et al. 2009)*. Although the blueberry whole-genome sequencing project is not yet
75 complete, the estimated blueberry genome is large (~1216 Mb; (Costich et al. 1993). This
76 estimate for blueberry is almost 10 times greater than that for *Arabidopsis* and 4 times greater
77 than that for the model legume *Medicago truncatula* (Young et al. 2011). The blueberry genome
78 is substantially larger (~3-6 times) than the genome sizes of other sequenced woody plants such
79 as grape or peach (Jaillon et al. 2007; Arús et al. 2012). *Vaccinium* spp. are of diploid, tetraploid
80 and hexaploid genome types. Thus, extensive genome rearrangements may have occurred during
81 their evolution. Recent findings suggest that the blueberry genome went through one or more
82 rounds of genome duplication during its evolution (Li et al. 2012). All these aspects highlight the
83 difficulty of genetic studies when working with complex genomes.
84
85 The achievement of *Arabidopsis* genome sequencing has ushered plant biology into the post-
86 genome era (*Arabidopsis* Genome Initiative 2000). The huge mass of genome data still being
87 generated is being analyzed in order to convert it into gene function data, adding value to the
88 nucleotide sequence collections. However, the main goal of this work is not only to understand
89 the biochemical and physiological functions of every gene and gene product but also determine
90 how they interact in an undoubtedly complex interplay and unravel the role they have in relevant
91 biological events. As the development of genetic, genomic and molecular tools for model
92 organisms have revealed fundamental understanding of basic biological processes, the challenge
93 has become transferring that information and technologies to important crop species that are less-
94 studied but are of notable agronomic importance. This is a new challenge that requires the

systematic application of global molecular approaches integrated through the increasing value of bioinformatics.

One such important crop species is blueberry. Initial genomic efforts in blueberry focused on, the identification and study of a comprehensive collection of blueberry genes activated during cold acclimation. Expression profiling strategies comprised of high-throughput Expressed Sequence Tag (EST)-sequencing, microarray-based transcriptome profiling, and Next Generation-transcriptomes sequencing have now been conducted in order to study several aspects of blueberry biology reflecting the diversity of research interest. Various websites have been established to support such genomic experiments by sharing up-to-date bioinformatics and functional genomic technologies. The present review summarizes the current initiatives of genomic advancements in blueberry research. Although valuable information about changes in gene expression can be gained from different fields of research, we have focused on approaches utilizing analysis of the varying qualitative and quantitative changes in messenger RNAs. This review illustrates in a broad sense how recent biotechnological advances in plant genomics are generating advanced technologies and resources for blueberry which will lead to new strategies for crop improvement.

2. Blueberry cDNA libraries

The advent of high-throughput sequencing tools and bioinformatics allows a whole-genome analysis approach to the study of gene expression. The composition of mRNA populations in a given organ or tissue offers an overview of the transcribed genes and, thus, is an important tool in understanding the biochemical pathways involved in physiological responses. Generating ESTs is a valuable resource in order to study changes in mRNA populations or patterns of gene expression in response to a given environmental condition or during development. In this review, we will describe, among other things, cDNA and EST libraries that have been developed for blueberry. To avoid confusion, we will use the term ‘cDNA library’ to refer to the physical library, the collection of actual cDNA clones from a single transformation event, and the term

'EST collection' to refer to each group of cDNA sequences or ESTs that were generated during various research studies.

During the pre-omic era there was a total lack of detailed information on any cDNA libraries in blueberry. In fact it has only been a decade ago since the first blueberry EST sequences have become available. Although genomic research in blueberry, and in the *Ericaceae* family in general, is still quite new, significant progress has been made in the last few years (Rowland et al. 2012b). The initial identification of a collection of genes activated during cold acclimation in blueberry became possible due to several programs sponsored by the USDA/ARS. ESTs from these first cDNA libraries were released in 2003. The majority of EST collections were produced from 2003 to 2010, and most publicly available sequences during this timeframe were from cDNA libraries made from flower buds at different stages of cold acclimation. Interestingly the EST database has recently expanded to include sequences responsive to another abiotic stress (mineral soils) and sequences comprising new pathways (flavonoid metabolism) and from a wider tissue/organ set including roots and fruits at different stages of development. By March 2013, more than 22,400 blueberry ESTs have become available in the EST database of GenBank at NCBI, of which about 96.5% have been generated from standard libraries (Fig. 1). In terms of EST numbers, blueberry ranks at the 131st position among plants in general and first among plants in the *Ericaceae* family.

Fig. 1. Distribution of blueberry resources submitted to the dbEST division of Genbank. **A.** Number of EST collections and type of cDNA libraries from which they were generated. **B.** Number of total ESTs available over the past years. Retrieved March 1, 2013.

2.1 **Approach I:** Standard cDNA libraries and high-throughput EST-sequencing

In total, 13 EST collections from several standard cDNA libraries–have been generated and deposited in the EST database of GenBank (Table 1). Similar cloning procedures were followed for construction of the cDNA libraries: polyA$^+$-enriched RNA strategy was adopted in most

cases. Only one library (AL) was generated from total RNA. The unidirectional lambda cloning vector (Uni-ZAP^{TM}, Stratagene) and mass excision of aliquots of the libraries to convert phage clones into plasmid clones was the chosen method for most of the studies.

The selection of genotypes and types of organs studied has depended on the aim of the research. The first genomics-based research projects in blueberry were conducted to identify genes associated with cold acclimation. Thus, the first few thousand (about 2,600) ESTs were generated by USDA-ARS scientists from non-acclimated (NA) and cold acclimated (CA) flower bud libraries (Dhanaraj et al. 2007; Dhanaraj et al. 2004). Most were 5'-end ESTs although about 100 3'-end ESTs were also generated from the CA library. Here, the blueberry cultivar 'Bluecrop' was used because it is the industry standard and is fairly cold hardy. An aliquot of the NA library was also provided by Lisa Rowland (USDA-ARS, Beltsville, MD) to the Floral Genome Project. Researchers participating in this project generated another 1,758 5'-end ESTs from this library (Albert et al. 2005). Since then, another two libraries were generated with the aim of understanding the mechanisms of toxicity and Al^{3+} resistance in blueberry and woody perennials in general. Samples of blueberry roots (root apex to the elongation zone, ~3 cm) were collected at different timepoints over a 48h period after Al^{3+}-treatment. Two cDNA libraries were constructed, using the contrasting genotypes 'Brigitta' (Al^{3+}-resistant) and 'Bluegold' (Al^{3+}-sensitive) and a cDNA-amplified fragment length polymorphism (cDNA-AFLP) analysis was conducted to identify genes regulated by Al^{3+} (Inostroza-Blancheteau et al. 2011). Most recently, characterization of flavonoid biosynthesis during fruit development has been reported using the cv. 'Rubel'. This genotype is known to have superior antioxidant capacity and flavonoid content. Two cDNA libraries, one from pooled fruit mRNA from stage 5/6 (mid fruit development stage) and one from stage 7/8 (ripening stage), were constructed. From these, the largest number of ESTs from a single project so far in blueberry were generated-- >17,000 sequences distributed among 5 EST collections (Zifkin et al. 2012).

Based on sequence homology with available databases, a large percentage of these ESTs has been assigned into functional categories. However, an important percentage of them have not shown any significant homology with known proteins. In such cases gene function will need to be inferred on the basis of further experiments such as gene expression analysis over time course or developmental stages or transformation or mapping experiments.

The above-mentioned projects have been complemented by efforts by the New Zealand Institute for Plant & Food Research (PFR) that has also generated four cDNA libraries encompassing over 9,000 unigenes from floral and fruit tissues. Three different genotypes (Duke, Puru and E118B4-41) have been used to investigate genes in the anthocyanin and related pathways. However, they are not publicly available and must be accessed using the in-house PFR BioView platform (Buck et al. 2012).

2.2 **Approach II:** Subtracted cDNA libraries

Although EST collections are a valid and reliable source of gene expression data, they also have limitations. Random sampling of cDNA clones gives preferential access to more highly abundant transcripts because these clones will be present in the libraries at a higher frequency than those representing less abundant transcripts, whereas weakly expressed genes, important regulatory transcription factors, or genes differentially expressed in response to a given condition can be more difficult to identify. Thus, other strategies are necessary in order to identify rarer classes of transcripts that may play a key role in the processes under study.

The suppression subtractive hybridization (SSH) approach was developed in order to both normalize (equalization between higher and lower transcript frequency) and enrich libraries in sequences differentially represented between two samples (Diatchenko et al. 1999). There is a large number of examples demonstrating that this approach is a good strategy to identify differentially expressed genes and this method has been successfully used in woody perennials (Şahin-Çevik 2013; Gulyani and Khurana 2011; Legay et al. 2011; Leida et al. 2010).

The SSH approach has been chosen by two blueberry research groups to generate cDNA libraries aimed at identifying potential regulatory genes during abiotic stress or novel markers of pathogenic interactions (Table 2). With the goal to supplement their previous work on cold acclimation, (Naik et al. 2007) constructed the first blueberry subtracted libraries. Complementary DNAs were prepared from RNAs isolated from floral buds of field-grown plants

of the highbush blueberry cultivar ‘Bluecrop’ at 0 h and 400 h of cold acclimation, and forward (SL) and reverse (RL) SSH libraries were developed. About 565 ESTs from the SL library (where 400 h cDNA was used as tester and 0 h cDNA was used as driver) and 170 ESTs from the RL library (0 h cDNA used as tester, 400 h cDNA used as driver) were generated and deposited in GenBank. By comparing the genes from the two libraries in combination with real-time PCR experiments, several genes up- or down-regulated at 400 h were identified. While lower gene expression in 400 h flower buds may simply reflect the fact that plants are not actively growing at this time point, some transcription factors from the reverse library could correspond to negative regulators of the cold acclimation pathway that need to be turned down to allow fine cold hardiness control in woody plants. This approach has led to the identification of many proteins related to signal transduction and transcription factors that are now available for further analyses and are useful candidates to explore such hypotheses.

The SSH approach has also been used for the study of anthracnose fruit rot resistance. This disease is caused by the fungal pathogen *Colletotrichum acutatum* (Verma et al. 2006; Polashock et al. 2005) and is a major constraint to blueberry production due to its severe economic impact (Milholland 1995). Five time points of ripe fruit following inoculation with *C. acutatum* (0, 24, 48, 96, 144 h) were combined in the preparation of a forward library (where the resistant cultivar ‘Elliott’ served as the tester and the susceptible cultivar ‘Jersey’ as the driver) and a reverse library (where ‘Jersey’ served as the tester and ‘Elliott’ as the driver). By screening clones from the forward library against the reverse library, 34 sequences were shown to be differentially expressed in the resistant genotype. These sequences are available in the GenBank database. It is important to note that a subset of the ESTs generated from libraries representing plant tissues infected by a eukaryotic pathogen may not correspond to host plant genes. However, the authors verified that no PCR products were observed in reactions with fungal genomic DNA and primers designed from these ESTs indicating that all the 34 ESTs deposited in GenBank originated from blueberry. The establishment of a *C. acutatum*–blueberry pathosystem is of interest at the molecular level because results have provided evidence for an active resistance response in ripe fruit of blueberry (Miles et al. 2011).

One advantage of SSH libraries is that the approach facilitates gene discovery. However, due to restriction enzyme hydrolysis in the cloning procedure, cDNAs fragments are generally shorter than those in standard libraries. Consequently, protein prediction and sequence annotation can be difficult in some cases. However, even after bioinformatic analysis, high quality sequences of considerable length can be identified that do not have significant homology to any other sequences in the public databases. Important percentages of sequences without significant homology to any known gene have been reported in the SSH blueberry libraries (Table 2). Although the percentage of novel genes discovered from these experiments should definitely drop as blueberry whole genome sequencing projects continue, the fact that no putative homologs have yet been identified in other plant genomes indicates that these genes could be extremely interesting for further functional studies and could reveal novel activated woody perennial pathways.

3. Genome-wide sequencing: Next Generation Sequencing technologies

Next generation sequencing (NGS) platforms produce vastly more data than was ever possible with EST sequencing capillary technology. It is a versatile technology that is being applied in a variety of ways and under continuing evolution. These instruments are revolutionizing genomics and genome science and the combination of significantly lower cost and increased speed of sequencing has resulted in an explosive growth of available data. In 2009, the International Nucleotide Sequence Database Collaboration (INSDC) started the international public archival resource 'Sequence Read Archive (SRA)' for next-generation sequencing data (Kodama et al. 2012). Undoubtedly, NGS use has been adopted by the plant biology research community and is now widely used to characterize plant genomes and transcriptomes. It has been stated that the unprecedented level of sensitivity and high-throughput nature will make NGS technologies the method of choice for gene expression analysis in plant genomics (Jain 2012).

Recently, the generation of over 600,000 blueberry NGS transcriptome sequences using 454-GS FLX Titanium technology (454 Life Sciences, Roche Diagnostic) has been reported (Rowland et al. 2012a). This represents the first transcriptome of the species. NGS sequences were obtained from 9 cDNA libraries (leaves, flower buds at different stages of cold acclimation, and fruit at

different stages of development) prepared using the industry standard highbush cultivar 'Bluecrop' (Table 3). Approximately 15,000 contigs and 124,000 singletons have been annotated and functionally mapped to Gene Ontology terms. The assembled sequences have also been mined for SSRs. The raw sequences are publicly available in the SRA database of NCBI. The developmental series experiments are a valuable resource, providing insights into cellular processes and transcription regulation in relation to flower bud and fruit development (Rowland et al. 2012a). The same strategy has also been used to study molecular mechanisms involved in the biosynthesis of anthocyanins by another research group (Li et al. 2012). The cv. 'Northland' was chosen because of its superior antioxidant capacity and anthocyanin content. Two libraries were prepared from the skin and pulp of blue fruits collected 50 days after full bloom. *De novo* assembly generated a collection of 34,464 unigenes using Illumina RNA-Seq technology (Illumina Genome Analyzer IIx, Illumina). Through comparative transcript profiling, over 90 differentially expressed genes have been identified regulating the fruit metabolism and anthocyanin content during ripening (Li et al. 2012).

These new collections of sequences represent a substantial improvement to the limited genomic resources in blueberry that have existed so far. They are undoubtedly valuable resources for the scientific community and will serve as a platform to accelerate the knowledge gained in the past few years on flower bud development, cold acclimation, chilling unit accumulation/ vernalization, flowering, fruit development, and nutritional quality traits. Moreover, they are important tools for development of molecular markers and genetic linkage maps in blueberry and closely related crops, and future blueberry breeding programs should benefit from these genomic resources.

4. Development of bioinformatic tools

One important challenge in using high-throughput sequencing technologies is the downstream computational analysis and interpretation of such large data sets. Development of bioinformatic tools is highly important for managing data and generating transcription profiles. Web-accessible databases for storage, evaluation and mining of expression profiles offer data integration by linkages between physiological conditions and expression patterns, suggesting gene function. In

addition, the integration of decentralized data across the internet can facilitate collaboration between geographically isolated research groups by providing access to data of common interest.

Several websites are now available that include access to blueberry sequences. One is the Plant Genome Network website (PGN: http://www.pgn.cornell.edu/) hosted by the Floral Genome Project (FGP). A main objective of this consortium is to uncover patterns of conservation and divergence of the floral transcriptome among angiosperms. In addition to a general-purpose EST analysis pipeline and web-based database, PGN was designed to provide an EST processing and annotation service for smaller EST projects that may not have the informatic resources to generate a public database. The PGN provides public access to EST library statistics, unigene build details, EST chromatograms, and permits FGP taxon-specific BLAST searches. The database holds 1,549 unigenes from 1,758 ESTs from blueberry flower buds (Albert et al. 2005).

The Blueberry Genomics Database (BBGD: http://bioinformatics.towson.edu/BBGD/) has also been established and currently houses EST and microarray data. The primary focus is to store and analyze EST and microarray data for the identification of genes associated with cold acclimation and freeze tolerance in blueberry. The database provides embedded analytical tools for data mining with numerous applications to conduct statistical analysis and is hosted by the Bioinformatics server at Towson University in Maryland (Alkharouf et al. 2007).

More recently, the Genome Database for *Vaccinium* has been created (http://www.vaccinium.org/). The objective is to house and integrate genomic, genetic and breeding data for blueberry, cranberry and other *Vaccinium* spp. This database will include the ongoing genome sequencing project that is being lead by Dr. A. Brown at North Carolina State University. The draft genomic sequence of a diploid V. corymbosum blueberry selection 'W8520' has been recently generated and is currently being assembled (A. Brown, personal communication). The database aims to integrate the annotation of transcripts, traits, maps, and markers being generated by various *Vaccinium* researchers.

In the wake of the release of the first blueberry 454 transcriptome (Rowland et al. 2012a), another database was created, which is an extension of the original BBGD website, and is also

hosted on the Bioinformatics server at Towson University (http://bioinformatics.towson.edu/BBGD454/). The database houses the 454 sequences, their assemblies and annotations, as well as their frequencies in each of the libraries. The web-based interface was developed to allow researchers to search or browse the data and aid in its analysis and interpretation (Rowland et al. 2012a). Eventually the goal is to integrate the original BBGD website and this new site so both will be accessible through one URL.

5. Exploitation of functional resources

With databases such as the above coming on board, and advances in computational molecular biology, it is possible to now mine and analyze large EST datasets efficiently and exhaustively. Our laboratory has been working toward increasing our understanding of the genetic control of cold hardiness in blueberry to ultimately use this information to develop more cold hardy cultivars for the industry. Using such a strategy, computational analysis revealed that ~2.4% randomly picked clones from the reverse SSH library described previously (Table 2; (Naik et al. 2007) had significant homology to members of the *CBF* gene family, transcription factors quickly induced in response to cold and drought stress (Wisnieswski et al. 2013). Cloning followed by overexpression of the *V. corymbosum CBF* in *Arabidopsis* resulted in induction of *COR* (cold-regulated) gene expression and constitutive freezing tolerance in transgenic plants. This indicates that the cold acclimation pathway of blueberry retains and utilizes functional components of the *CBF* system (Polashock et al. 2010). A further and relevant step in this research was the overexpression of the *CBF* gene under the control of the *CaMV* 35S promoter in transgenic blueberry lines. Southern highbush blueberry cv. 'Legacy' plants overexpressing the CBF gene (isolated from the northern highbush cv. 'Bluecrop') showed an increase in freezing tolerance, suggesting the potential manipulation of the CBF system for improvement of freezing tolerance in woody fruit crops (Walworth et al. 2012). Most recently, the *CBF* gene from *V. myrtillus* was overexpressed in transgenic *Arabidopsis* lines, also resulting in constitutive freezing tolerance, further supporting the importance of the *CBF* transcription factors during cold acclimation of multiple species within the genus *Vaccinium* (Oakenfull et al. 2013).

Another obvious application of EST sequencing is the production of microarrays. About 2,500 clones from the CA and NA flower bud libraries have been used to construct the first cDNA microarray of blueberry (Dhanaraj et al. 2007). Transcript profiling analysis was conducted at multiple times during cold acclimation under field and cold room conditions. This combination of EST sequencing and microarray construction has proven to be a powerful tool for successfully identifying genes involved not just in cold stress but in cold acclimation. Moreover, interesting differences in expression between cold room *vs.* field conditions and between cold tolerant and cold sensitive genotypes have been reported (Dhanaraj et al. 2007; Rowland et al. 2008). With the generation of NGS data, the development of a new larger blueberry microarray would allow the study of expression of thousands of genes covering a wider range of tissues/organs and biotic and abiotic stresses (Rowland et al. 2012b).

Another use of EST databases is in the development of molecular markers. ESTs from the standard CA, NA, and subtracted flower bud libraries, as well as the recent 454 sequences of blueberry have been used in the development of EST-polymerase chain reaction (EST-PCR) markers (Rowland et al. 2003) and SSRs (Boches et al. 2005; Rowland et al. 2012a). These markers have already been utilized in several DNA fingerprinting/genetic relationship studies such as studies on the population genetics of wild lowbush blueberry (Bell et al. 2009; Bell et al. 2012) and genetic relationship studies on highbush (Rowland et al. 2003; Boches et al. 2006) and rabbiteye blueberry (Rowland et al. 2010). They are also currently being used in several large mapping efforts on diploid and tetraploid blueberry (Rowland et al. 2012b).

With the rapid expansion of available transcriptome sequences, opportunities for digital analysis of gene expression will continue to expand. By combining electronic expression computation and experimental analysis we are currently identifying candidate genes for several horticulturally significant traits. Through an approach based on digital transcript profiles at early times during cold acclimation under field conditions coupled with experimental transcriptomic analysis, the list of cold-responsive genes can be narrowed down to what might be the key players in the cold acclimation pathway. Similarly, data mining approaches may allow genes controlling molecular aspects of dormancy in blueberry to be identified. Some of these candidate genes could

401 eventually be mapped as EST-PCR markers in our mapping populations to determine if they map
402 to the same regions as QTL for these traits.

403

404

405 ## 6. Concluding remarks

406 During the past decade, considerable effort has been made to generate genomic tools to better
407 understand the biology of blueberry to address challenges in its adaptation to new environments
408 and for crop improvement. The advent of high-throughput sequencing tools and bioinformatics
409 allows a whole-genome analysis approach to gene expression. Currently, about 22,400 blueberry
410 ESTs are available through the dbEST division of Genbank. One microarray has been
411 constructed, two NGS transcriptomes have been recently reported (SRA at the NCBI), and whole
412 genome sequencing and assembly is underway. Throughout recent years it has become clear that
413 each method has inherent limitations and none of them alone suffices to unequivocally assign a
414 function to a gene of interest and understand significant interactions between them. Full benefit
415 of available information and functional characterization of blueberry genes will be obtained
416 provided that complementary integrated approaches are utilized. Proteomics and metabolomics
417 will have a strong relevance here, especially in the field of fruit nutritional quality. Some years
418 ago there was a bias toward sequences derived from flower buds; however, the current list of
419 organs, tissues and conditions from which the sequences have been obtained, reflects the
420 continual broadening of the biological questions addressed by the blueberry research community.
421 Interesting genes including putative novel genes and genes of current unknown function, as well
422 as various new potential regulators, may be assigned to specific processes. Detailed
423 characterization will help us to discover the fine networks underlying biochemical pathways of
424 interest. Research groups are actively using information generated over the last few years to
425 develop molecular markers for studies of genetic diversity, spatial genetic structure, and gene
426 flow in blueberry, as well as to identify QTL associated with cold hardiness, chilling
427 requirement, and fruit quality traits for marker-assisted breeding. Breeding programs will likely
428 benefit from the recent genomics advances in order to develop new high quality cultivars with
429 elevated anthocyanin contents and high antioxidant capacity combined with early or late ripening
430 and appropriate climatic adaptation.

Internet Resources

Database of Expressed Sequenced Tags (NCBI) http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html

Sequence Read Archive database (NCBI) www.ncbi.nlm.nih.gov/sra

Blueberry Genomics Database http://bioinformatics.towson.edu/BBGD/

The Plant Newtwork Website http://www.pgn.cornell.edu/

Genome Database for *Vaccinium* *http://www.vaccinium.org/*

Transcriptome Database for Blueberry http://bioinformatics.towson.edu/BBGD454

References

Albert V, Soltis D, Carlson J, Farmerie W, Wall PK, et al. (2005) Floral gene resources from basal angiosperms for comparative genomics research. BMC Plant Biology 5 (1):5

Alkharouf N, Dhanaraj A, Naik D, Overall C, Matthews B, Rowland L (2007) BBGD: an online database for blueberry genomic data. BMC Plant Biology 7 (1):5

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408 (6814):796-815. doi:http://www.nature.com/nature/journal/v408/n6814/suppinfo/408796a0_S1.html

Arús P, Verde I, Sosinski B, Zhebentyayeva T, Abbott A (2012) The peach genome. Tree Genetics & Genomes 8 (3):531-547. doi:10.1007/s11295-012-0493-8

Bañados M Expanding blueberry production into non-traditional production areas : northern Chile and Argentina, Mexico and Spain. In: IX International VAccinium Symposium, 2008. pp 439-445

Bell DJ, Drummond FA, Rowland LJ (2012) Fine-scale spatial genetic structure associated with *Vaccinium angustifolium* Aiton (Ericaceae). International Journal of Modern Botany 2 (4):72-82. doi:10.5923/j.ijmb.20120204.02

Bell DJ, Rowland LJ, Zhang D, Drummond FA (2009) Spatial genetic structure of lowbush blueberry, Vaccinium angustifolium, in four fields in Maine. Botany 87 (10):932-946. doi:10.1139/b09-058

Boches PS, Bassil NV, Rowland LJ (2005) Microsatellite markers for Vaccinium from EST and genomic libraries. Molecular Ecology Notes 5 (3):657-660. doi:10.1111/j.1471-8286.2005.01025.x

Boches PS, Bassil NV, Rowland LJ (2006) Genetic diversity in the highbush blueberry evaluated with microsatellite markers. Journal of the American Society for Horticultural Science 131 (5):674-686

466 Buck EJ, Scalzo J, Wiedow C, Hurst R, Allan AC, McGhie TK, Bassil NV, Rowland LJ (2012)
467 Progress in Blueberry Research in New Zealand. International Journal of Fruit Science 12
468 (1-3):304-315. doi:10.1080/15538362.2011.619444

469 Costich DE, Ortiz R, Meagher TR, Bruederle LP, Vorsa N (1993) Determination of ploidy level
470 and nuclear DNA content in blueberry by flow cytometry. Theoret Appl Genetics 86
471 (8):1001-1006

472 Dhanaraj A, Alkharouf N, Beard H, Chouikha I, Matthews B, Wei H, Arora R, Rowland L
473 (2007) Major differences observed in transcript profiles of blueberry during cold
474 acclimation under field and cold room conditions. Planta 225 (3):735-751.
475 doi:10.1007/s00425-006-0382-1

476 Dhanaraj AL, Slovin JP, Rowland LJ (2004) Analysis of gene expression associated with cold
477 acclimation in blueberry floral buds using expressed sequence tags. Plant Science 166
478 (4):863-872. doi:http://dx.doi.org/10.1016/j.plantsci.2003.11.013

479 Diatchenko L, Lukyanov S, Lau Y-FC, Siebert PD (1999) Suppression subtractive hybridization:
480 A versatile method for identifying differentially expressed genes. In: Sherman MW (ed)
481 Methods in Enzymology, vol Volume 303. Academic Press, pp 349-380.
482 doi:http://dx.doi.org/10.1016/S0076-6879(99)03022-0

483 Gulyani V, Khurana P (2011) Identification and expression profiling of drought-regulated genes
484 in mulberry (Morus sp.) by suppression subtractive hybridization of susceptible and
485 tolerant cultivars. Tree Genetics & Genomes 7 (4):725-738. doi:10.1007/s11295-011-
486 0369-3

487 Hou D (2003) Potential mechanisms of cancer chemoprevention by anthocyanins. Current
488 Molecular Medicine 3 (3):149-159

489 Inostroza-Blancheteau C, Aquea F, Reyes-Díaz M, Alberdi M, Arce-Johnson P (2011)
490 Identification of Aluminum-Regulated Genes by cDNA-AFLP Analysis of Roots in Two
491 Contrasting Genotypes of Highbush Blueberry (Vaccinium corymbosum L.). Mol
492 Biotechnol 49 (1):32-41. doi:10.1007/s12033-010-9373-3

493 Jaillon O, Aury J-M, Noel B, et al. (2007) The grapevine genome sequence suggests ancestral
494 hexaploidization in major angiosperm phyla. Nature 449 (7161):463-467.
495 doi:http://www.nature.com/nature/journal/v449/n7161/suppinfo/nature06148_S1.html

496 Jain M (2012) Next-generation sequencing technologies for gene expression profiling in plants.
497 Briefings in Functional Genomics 11 (1):63-70. doi:10.1093/bfgp/elr038

498 Kodama Y, Shumway M, Leinonen R (2012) The sequence read archive: explosive growth of
499 sequencing data. Nucleic Acids Research 40 (D1):D54-D56. doi:10.1093/nar/gkr854

500 Legay G, Marouf E, Berger D, Neuhaus J-M, Mauch-Mani B, Slaughter A (2011) Identification
501 of genes expressed during the compatible interaction of grapevine with Plasmopara

502 viticola through suppression subtractive hybridization (SSH). Eur J Plant Pathol 129
503 (2):281-301. doi:10.1007/s10658-010-9676-z

504 Leida C, Terol J, Martí G, Agustí M, Llácer G, Badenes ML, Ríos G (2010) Identification of
505 genes associated with bud dormancy release in Prunus persica by suppression subtractive
506 hybridization. Tree Physiology 30 (5):655-666. doi:10.1093/treephys/tpq008

507 Li X, Sun H, Pei J, Dong Y, Wang F, Chen H, Sun Y, Wang N, Li H, Li Y (2012) De novo
508 sequencing and comparative analysis of the blueberry transcriptome to discover putative
509 genes related to antioxidants. Gene 511 (1):54-61.
510 doi:http://dx.doi.org/10.1016/j.gene.2012.09.021

511 Mainland CMM (2012) Frederick V. Coville and the History of North American Highbush
512 Blueberry Culture. International Journal of Fruit Science 12 (1-3):4-13.
513 doi:10.1080/15538362.2011.619117

514 Miles TD, Day B, Schilder AC (2011) Identification of differentially expressed genes in a
515 resistant versus a susceptible blueberry cultivar after infection by Colletotrichum
516 acutatum. Molecular Plant Pathology 12 (5):463-477. doi:10.1111/j.1364-
517 3703.2010.00687.x

518 Milholland RD Anthracnose fruit rot (ripe rot). In: Caruso FL, D.C. R (eds) Compendium of
519 Blueberry and Cranberry diseases, St Paul, MN, 1995. APS Press, p 17

520 Naik D, Dhanaraj AL, Arora R, Rowland LJ (2007) Identification of genes associated with cold
521 acclimation in blueberry (Vaccinium corymbosum L.) using a subtractive hybridization
522 approach. Plant Science 173 (2):213-222.
523 doi:http://dx.doi.org/10.1016/j.plantsci.2007.05.003

524 Oakenfull RJ, Baxter R, Knight MR (2013) A C-Repeat Binding Factor Transcriptional
525 Activator (CBF/DREB1) from European Bilberry (*Vaccinium myrtillus*) Induces Freezing
526 Tolerance When Expressed in *Arabidopsis thaliana*. PLoS ONE 8 (1):e54119.
527 doi:10.1371/journal.pone.0054119

528 Polashock JJ, Arora R, Peng Y, Naik D, Rowland LJ (2010) Functional Identification of a C-
529 repeat Binding Factor Transcriptional Activator from Blueberry Associated with Cold
530 Acclimation and Freezing Tolerance. Journal of the American Society for Horticultural
531 Science 135 (1):40-48

532 Polashock JJ, Ehlenfeldt MK, Stretch AW, Kramer M (2005) Anthracnose Fruit Rot Resistance
533 in Blueberry Cultivars. Plant Disease 89 (1):33-38. doi:10.1094/pd-89-0033

534 Prior RL, Cao G, Martin A, Sofic E, McEwen J, O'Brien C, Lischner N, Ehlenfeldt M, Kalt W,
535 Krewer G, Mainland CM (1998) Antioxidant Capacity As Influenced by Total Phenolic
536 and Anthocyanin Content, Maturity, and Variety of Vaccinium Species. Journal of
537 Agricultural and Food Chemistry 46 (7):2686-2693. doi:10.1021/jf980145d

538 Qu L, Hancock JF (1997) Randomly Amplified Polymorphic DNA- (RAPD-) based Genetic
539 Linkage Map of Blueberry Derived from an Interspecific Cross between Diploid
540 Vaccinium darrowi and Tetraploid V. corymbosum. Journal of the American Society for
541 Horticultural Science 122 (1):69-73

542 Rowland L, Alkharouf N, Darwish O, Ogden E, Polashock J, Bassil N, Main D (2012a)
543 Generation and analysis of blueberry transcriptome sequences from leaves, developing
544 fruit, and flower buds from cold acclimation through deacclimation. BMC Plant Biology
545 12 (1):46

546 Rowland LJ, Bell DJ, Alkharouf N, Bassil NV, Drummond FA, Beers L, Buck EJ, Finn CE,
547 Graham J, McCallum S, Hancock JF, Polashock JJ, Olmstead JW, Main D (2012b)
548 Generating Genomic Tools for Blueberry Improvement. International Journal of Fruit
549 Science 12 (1-3):276-287. doi:10.1080/15538362.2011.619452

550 Rowland LJ, Dhanaraj AL, Naik D, Alkharouf N, Matthews B, Arora R (2008) Study of Cold
551 Tolerance in Blueberry Using EST Libraries, cDNA Microarrays, and Subtractive
552 Hybridization. HortScience 43 (7):1975-1981

553 Rowland LJ, Mehra S, Dhanaraj AL, Ogden EL, Slovin JP, Ehlenfeldt MK (2003) Development
554 of EST-PCR Markers for DNA Fingerprinting and Genetic Relationship Studies in
555 Blueberry (Vaccinium, section Cyanococcus). Journal of the American Society for
556 Horticultural Science 128 (5):682-690

557 Rowland LJ, Ogden EL, Ehlenfeldt MK (2010) EST-PCR markers developed for highbush
558 blueberry are also useful for genetic fingerprinting and relationship studies in rabbiteye
559 blueberry. Scientia Horticulturae 125 (4):779-784.
560 doi:http://dx.doi.org/10.1016/j.scienta.2010.05.008

561 Şahin-Çevik M (2013) Identification and expression analysis of early cold-induced genes from
562 cold-hardy Citrus relative Poncirus trifoliata (L.) Raf. Gene 512 (2):536-545.
563 doi:http://dx.doi.org/10.1016/j.gene.2012.09.084

564 Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, et al. (2009) The B73 Maize Genome:
565 Complexity, Diversity, and Dynamics. Science 326 (5956):1112-1115.
566 doi:10.1126/science.1178534

567 USDA-NASS (2013) Noncitrus Fruits and Nuts 2012 Preliminary Summary.
568 http://usda01.library.cornell.edu/usda/current/NoncFruiNu/NonFruiNu-01-25-2013.pdf.
569 Accessed 01/03/2013

570 Verma N, MacDonald L, Punja ZK (2006) Inoculum prevalence, host infection and biological
571 control of Colletotrichum acutatum: causal agent of blueberry anthracnose in British
572 Columbia. Plant Pathology 55 (3):442-450. doi:10.1111/j.1365-3059.2006.01401.x

573 Walworth A, Rowland L, Polashock J, Hancock J, Song G-q (2012) Overexpression of a
574 blueberry-derived CBF gene enhances cold tolerance in a southern highbush blueberry
575 cultivar. Molecular Breeding 30 (3):1313-1323. doi:10.1007/s11032-012-9718-7

576 Wang H, Cao G, Prior RL (1996) Total Antioxidant Capacity of Fruits. Journal of Agricultural
577 and Food Chemistry 44 (3):701-705. doi:10.1021/jf950579y

578 Wisnieswski M, Nassuth A, Teulières C, Marque C, Rowland LJ, Cao PB, Brown A (2013)
579 Genomics of cold hardiness in woody plants. Critical Reviews in Plant Sciences:in press

580 Wu X, Beecher GR, Holden JM, Haytowitz DB, Gebhardt SE, Prior RL (2004) Lipophilic and
581 Hydrophilic Antioxidant Capacities of Common Foods in the United States. Journal of
582 Agricultural and Food Chemistry 52 (12):4026-4037. doi:10.1021/jf049696w

583 Young ND, Debelle F, Oldroyd GED, Geurts R, Cannon SB, et al. (2011) The Medicago genome
584 provides insight into the evolution of rhizobial symbioses. Nature 480 (7378):520-524.

585 Zifkin M, Jin A, Ozga JA, Zaharia LI, Schernthaner JP, Gesell A, Abrams SR, Kennedy JA,
586 Constabel CP (2012) Gene Expression and Metabolite Profiling of Developing Highbush
587 Blueberry Fruit Indicates Transcriptional Regulation of Flavonoid Metabolism and
588 Activation of Abscisic Acid Metabolism. Plant Physiology 158 (1):200-224.
589 doi:10.1104/pp.111.180950

590

591

592