BBGD: an online database for blueberry genomic data

Nadim W Alkharouf*1, Anik L Dhanaraj2, Dhananjay Naik2, Chris Overall3,4, Benjamin F Matthews3 and Lisa J Rowland2

Address: 1Department of computer and information sciences, Towson University, 7800 York Road, Towson, Maryland, 21252, USA, 2Fruit Laboratory, USDA/ARS, Henry A. Wallace Beltsville Agricultural Research Center, Bldg. 010A, BARC-West, 10300 Baltimore Ave., Beltsville, MD 20705, USA, 3Soybean Genomics and Improvement Laboratory, USDA/ARS, Henry A. Wallace Beltsville Agricultural Research Center, Bldg. 010A, BARC-West, 10300 Baltimore Ave., Beltsville, MD 20705, USA and 4George Mason University, School of Computational Sciences, Manassas, VA 20110, USA

Email: Nadim W Alkharouf* - nalkharouf@towson.edu; Anik L Dhanaraj - anik.luke.dhanaraj@monsanto.com; Dhananjay Naik - naikdhananjay@rediffmail.com; Chris Overall - coverall@gmu.edu; Benjamin F Matthews - matthewb@ba.ars.usda.gov; Lisa J Rowland - rowlandj@ba.ars.usda.gov

* Corresponding author

Abstract

Background: Blueberry is a member of the Ericaceae family, which also includes closely related cranberry and more distantly related rhododendron, azalea, and mountain laurel. Blueberry is a major berry crop in the United States, and one that has great nutritional and economical value. Extreme low temperatures, however, reduce crop yield and cause major losses to US farmers. A better understanding of the genes and biochemical pathways that are up- or down-regulated during cold acclimation is needed to produce blueberry cultivars with enhanced cold hardiness. To that end, the blueberry genomics database (BBDG) was developed. Along with the analysis tools and web-based query interfaces, the database serves both the broader Ericaceae research community and the blueberry research community specifically by making available ESTs and gene expression data in searchable formats and in elucidating the underlying mechanisms of cold acclimation and freeze tolerance in blueberry.

Description: BBDG is the world's first database for blueberry genomics. BBDG is both a sequence and gene expression database. It stores both EST and microarray data and allows scientists to correlate expression profiles with gene function. BBDG is a public online database. Presently, the main focus of the database is the identification of genes in blueberry that are significantly induced or suppressed after low temperature exposure.

Conclusion: By using the database, researchers have developed EST-based markers for mapping and have identified a number of “candidate” cold tolerance genes that are highly expressed in blueberry flower buds after exposure to low temperatures.

Background

Blueberry (Vaccinium corymbosum) is one of the major berry crops grown in the United States [1]. North America, in fact, is the world's leading blueberry producer, accounting for nearly 90% of world production at the present time. Total area devoted to growing commercial blueberries in North America is approximately 74,000 hectares. Blueberry is a high value crop, often times grown in acidic
and imperfectly drained soils that would otherwise be considered unfit for agricultural production [2]. Blueberry is also an important fruit crop because of its nutritional value. Of all fresh fruits and vegetables, blueberries are one of the richest sources of antioxidants [3]. Blueberry is a model organism for the heath family Ericaceae, which also includes the economically important, closely related cranberry as well as the economically important, more distantly related ornamentals, rhododendron, azalea, and mountain laurel. For all these related species, genomic studies, including EST generation and microarray analyses, are lacking or completely absent. Functional genomic studies on berry crops are lacking, especially studies dealing with the molecular impacts of low temperature on berry crop yield. Low temperature extremes reduce blueberry yields and impact the profitability and competitiveness of U.S. producers. Enhanced cold tolerance during the winter and early spring of elite varieties would be of great value to the blueberry industry. The Blueberry Genomics Database [4] is a public database that links blueberry expressed sequence tags (ESTs) with gene expression data and provides embedded analytical tools for data mining. BBGD was established in 2005 to serve as a sequence and microarray database for the blueberry community with its primary focus to store and analyze ESTs and microarray data generated from experiments aimed at studying cold acclimation and mid-winter hardiness of blueberry. The ultimate goal of these experiments is to apply the information to develop more cold hardy cultivars. The database allows for the correlation of expression levels and function by linking the EST data with the microarray results, since many of the ESTs were printed on the microarray slides. BBGD also serves as a means of novel gene discovery through EST analysis. Numerous applications to conduct statistical analysis on DNA microarray data have been integrated into BBGD for rapid data analysis. Analytical tools include t-tests to detect significantly induced/suppressed genes and online analytical processing (OLAP) to find correlations and relationships in data sets. These tools have been integrated into the database, thereby eliminating the need for third-party software.

**Construction and content**

BBGD is a relational database built on SQLServer2000 and is housed at the Beltsville Agricultural Research Center in Beltsville, MD, USA. The database was implemented on a server running Windows 2000 Server and Internet Information Server (IIS 5.0). The web interface uses active server pages (ASP) and ASP.Net scripts, written in visual basic, to query the backend database. BBGD is divided into two separate but related entities, a sequence database and a microarray database. This allows for the correlation of gene function, deduced from the EST data, with expression levels, deduced from the microarray data. The bulk of the EST and microarray data held at the BBGD currently deals with identifying cold-responsive genes in blueberry flower buds.

**Utility and discussion**

The BBGD web site acts as a gateway to the microarray and EST sequencing projects that have been implemented to identify cold-responsive genes in blueberry (fig. 1). In addition, it provides a wealth of general information on blueberry for the public.

**Microarray experiments**

DNA microarrays allow for the measurement of mRNA expression levels for thousands of genes at a time. The microarray portion of the database stores microarray experiments that measure gene expression changes in blueberry flower buds across a number of time points after low temperature exposure in the field and the cold room environment [5]. A list of slides printed during one of the experiments is illustrated in table 1. A list of genes that were printed on the slides are available as a supplement [see Additional file 1]. A number of web based applications have been implemented that allow users to query the microarray experiments from anywhere using a web browser and an internet connection. Users can choose to query a specific time point (fig. 2), across all or selected time points, or query by gene name, ID or GenBank accession number. Users can also conduct advanced queries to find genes that have similar expression in different experiments and/or biological samples. Results from cluster analysis and online analytical processing (OLAP) [6,7] are also displayed.

**Sequence database**

The sequence database provides access to EST sequences stored at BBGD. With the development of high-throughput DNA sequencing technologies, EST analysis has become a rapid and relatively inexpensive way to identify genes, proteins, and metabolic pathways through homology with other sequence data repositories such as GenBank. Analysis of ESTs can provide an overall picture of transcripts involved in organ or tissue development. In BBGD, all relevant information about every EST is stored; including the cloning vector and bacterial host strain, insert size, dbEST ID, GenBank accession number, and Blast results, which include E-value, score and identity percentage. Perl scripts were written to extract this information from the Blast results and are available through the authors. The database also contains results of EST analysis [8] and contig assembly for the libraries that were sequenced, along with graphical representations and charts. Table 2 depicts the libraries that were locally constructed; contig assembly and analysis results are provided on the web site. SeqMan from DNAStar Inc (Madison, WI) was used for contig assembly and clustering. Like the
microarray portion, the sequence database allows users to query by gene name, ID or GenBank accession number. Users can also browse a specific library in a table format (fig. 3). The database was extremely useful in the identification and characterization of transcripts that are highly expressed during cold acclimation [8,9] and in the development of EST-based markers for mapping cold tolerance in blueberry [10]. A Blast application (fig. 4) was developed that provides users with a means to Blast a sequence of interest against the ESTs stored in BBGD and/or a collection of sequence data comprising EST and genomic sequences from all plant species (kingdom Viridiplantae). Since the blueberry genome has not been sequenced, EST sequences such as the ones stored in BBGD will play an important role in gene identification and discovery, as they have in other organisms [11-14].

Conclusion
As a result of BBGD and the associated analysis tools, genes potentially involved in cold acclimation in blueberry were identified [5,8]. From the microarray experiments a number of genes were found to be up-regulated across all measured time points. To name a few, among them were stress/defense related genes dehydrins and GRPF1, cell structure genes, and auxin-mediated signaling pathway genes such as protein kinase PINOID [5]. From the EST analysis alone, monooxygenase, dehydrins, beta amylase, galactinol synthase, and heat shock proteins,
Table 1: List of slides printed during one of the microarray experiments stored in BBGD. Probe combination refers to the RNA samples that were used in hybridization, and combination order refers to the dye that was used to label each sample.

| Slide number | Biological sample | Probe Combination       | Combination Order | Time point (hours) |
|--------------|-------------------|-------------------------|-------------------|-------------------|
| 24           | 1                 | CAColdRm-/CAColdRm-     | 647/555           | 0                 |
| 25           | 1                 | CAColdRm+/CAColdRm-     | 647/555           | 500               |
| 26           | 1                 | CAColdRm+/CAColdRm-     | 647/555           | 1000              |
| 27           | 1                 | CAColdRm+/CAColdRm-     | 555/647           | 0                 |
| 28           | 1                 | CAColdRm+/CAColdRm-     | 555/647           | 500               |
| 29           | 1                 | CAColdRm+/CAColdRm-     | 555/647           | 1000              |
| 36           | 1                 | CAField+/CAField-       | 647/555           | 0                 |
| 37           | 1                 | CAField+/CAField-       | 647/555           | 67                |
| 38           | 1                 | CAField+/CAField-       | 647/555           | 399               |
| 39           | 1                 | CAField+/CAField-       | 647/555           | 779               |
| 40           | 1                 | CAField+/CAField-       | 555/647           | 1234              |
| 41           | 1                 | CAField+/CAField-       | 555/647           | 0                 |
| 42           | 1                 | CAField+/CAField-       | 555/647           | 67                |
| 43           | 1                 | CAField+/CAField-       | 555/647           | 399               |
| 44           | 1                 | CAField+/CAField-       | 555/647           | 779               |
| 45           | 1                 | CAField+/CAField-       | 555/647           | 1234              |
| 46           | 1                 | CAField+/CAField-       | 647/555           | 0                 |
| 47           | 1                 | CAField+/CAField-       | 647/555           | 67                |
| 48           | 1                 | CAField+/CAField-       | 647/555           | 399               |
| 49           | 1                 | CAField+/CAField-       | 647/555           | 779               |
| 50           | 1                 | CAField+/CAField-       | 647/555           | 1234              |
| 51           | 1                 | CAField+/CAField-       | 555/647           | 0                 |
| 52           | 1                 | CAField+/CAField-       | 555/647           | 67                |
| 53           | 1                 | CAField+/CAField-       | 555/647           | 399               |
| 54           | 1                 | CAField+/CAField-       | 555/647           | 779               |
| 55           | 1                 | CAField+/CAField-       | 555/647           | 1234              |
| 56           | 1                 | CAField+/CAField-       | 647/555           | 67                |
| 57           | 1                 | CAField+/CAField-       | 647/555           | 0                 |
| 58           | 1                 | CAField+/CAField-       | 647/555           | 67                |
| 59           | 1                 | CAField+/CAField-       | 647/555           | 779               |
| 60           | 1                 | CAField+/CAField-       | 555/647           | 1234              |
| 61           | 1                 | CAField+/CAField-       | 555/647           | 0                 |
| 62           | 1                 | CAField+/CAField-       | 555/647           | 67                |
| 63           | 1                 | CAField+/CAField-       | 555/647           | 399               |
| 64           | 1                 | CAField+/CAField-       | 555/647           | 779               |
| 65           | 1                 | CAField+/CAField-       | 555/647           | 1234              |
| 66           | 1                 | CAField+/CAField-       | 647/555           | 0                 |
| 67           | 1                 | CAField+/CAField-       | 647/555           | 67                |
| 68           | 1                 | CAField+/CAField-       | 647/555           | 399               |
| 69           | 1                 | CAField+/CAField-       | 647/555           | 779               |
| 70           | 2                 | CAField+/CAField-       | 555/647           | 0                 |
| 71           | 2                 | CAField+/CAField-       | 647/555           | 67                |
| 72           | 2                 | CAField+/CAField-       | 647/555           | 399               |
| 73           | 2                 | CAField+/CAField-       | 647/555           | 779               |
| 74           | 2                 | CAField+/CAField-       | 647/555           | 1234              |
| 75           | 2                 | CAField+/CAField-       | 555/647           | 0                 |
| 76           | 2                 | CAField+/CAField-       | 555/647           | 67                |
| 77           | 2                 | CAField+/CAField-       | 555/647           | 399               |
| 78           | 2                 | CAField+/CAField-       | 555/647           | 779               |
| 79           | 2                 | CAField+/CAField-       | 555/647           | 1234              |
among others, were identified as being highly expressed during cold acclimation, demonstrating how analysis of ESTs was an effective strategy to identify candidate cold acclimation-responsive transcripts in blueberry [8]. Blueberry is an important small fruit crop, and these types of studies on cold acclimation in flower buds will go a long way toward achieving our ultimate goal of producing more cold hardy cultivars.

Future perspectives
Work is underway to add data on the current status of the blueberry genetic linkage maps and EST-PCR markers being used for mapping.

Availability and requirements
BBGD is accessible from http://psi081.ba.ars.usda.gov/BBGD/ Sequence and microarray data can be downloaded and the manager of the database can be contacted by email at nalkharouf@towson.edu.

Table 2: List of the locally constructed libraries stored in BBGD and the number of ESTs in each.

| Library                    | Number of ESTs |
|----------------------------|----------------|
| Cold acclimated            | 1312           |
| Non-acclimated             | 1241           |
| Forward subtractive library| 586            |
| Reverse subtractive library| 287            |

Figure 2
A snap shot showing the results obtained when querying BBGD for genes that had an expression ratio of >= 2 fold. Among the fields that are returned are results from standard deviation calculations (STDEV) and T-test calculations summary (significantly induced/suppressed or no change). Users can get the sequence by clicking on the clone ID and/or can query PubMed for relevant articles by clicking on the gene name.

Abbreviations
EST – Expressed Sequence Tag; BBGD – Blueberry Genomics Database.

Authors’ contributions
NWK designed and developed the database, he also developed most of the web interface. CO developed the blast application, while ALD, DN, BFM and LJR contributed to the generation of the EST and microarray data. All authors have read and approved the final manuscript.
The Blast application allows users to cut and paste a sequence or to select a file containing a FASTA sequence and Blast the in-house sequences stored in BBGD or a collection of EST and genomic sequences from all plant species (kingdom Viridiplantae).

**Figure 4**
The Blast application allows users to cut and paste a sequence or to select a file containing a FASTA sequence and Blast the in-house sequences stored in BBGD or a collection of EST and genomic sequences from all plant species (kingdom Viridiplantae).
Acknowledgements

We thank Dr. Mark Tucker at the Soybean Genomics and Improvement Laboratory, USDA-ARS, and Dr. Johar Ali at the Genomics Sequencing Center in Vancouver, Canada, for their thorough review of the manuscript. NWK, ALD and DN were supported by the Beltsville agricultural research center (BARC) of the United States department of agriculture (USDA). Publication costs were covered through an agreement between the university of Maryland system and BioMed Central publishers.

References

1. Moore JN: The blueberry industry of North America. Acta Hort 1993, 346:15-26.
2. Galletta GJ, Ballington JR: Blueberries, cranberries, and lingonberries. In Fruit Breeding: Vine and Small Fruits Crops Volume II. Edited by: Janick J, Moore JN. Wiley, New York, USA; 1996:1-107.
3. Prior RL, Cao G, Martin A, Sofic E, McEwen J, O’Brien C, Lischner N, Ehlenfeldt M, Kalt W, Kremer G, Mainland CM: Antioxidant capacity as influenced by total phenolic and anthocyanin content, maturity, and variety of Vaccinium species. Journal of Agricultural and Food Chemistry 1998, 46:2686-2693.
4. Blue Berry Genomics Database (BBGD) [http://psi081.ba.ars.usda.gov/BBGD/]
5. Dhanaraj AL, Alkharouf NW, Beard HS, Chouikha IB, Matthews BF, Wei H, Arora R, Rowland LJ: Major differences observed intrancript profiles of blueberry during cold acclimation under field and cold room conditions. Planta. DOI: 10.1007/s00425-006-0382-1
6. Codd EF, Codd SB, Salley CT: Providing OLAP (on-line analytical processing) to user-analysis: An IT mandate. Technical Report, EF Codd & Associates 1993.
7. Alkharouf N, Jamison C, Matthews BF: Online Analytical Processing (OLAP): A fast and effective data mining tool for gene expression databases. Journal of Biomedicine and Biotechnology 2005, 2:181-188.
8. Dhanaraj AL, Slavin JP, Rowland LJ: Analysis of gene expression associated with cold acclimation in blueberry floral buds using expressed sequence tags. Plant Science 2004, 166:863-872.
9. Dhanaraj AL, Slavin JP, Rowland LJ: Isolation of a cDNA clone and characterization of expression of the highly abundant, cold acclimation-associated 14 kDa dehydrin of blueberry. Plant Science 2005, 168:949-957.
10. Rowland LJ, Mehra S, Dhanaraj AL, Ogden EL, Slavin JP, Ehlenfeldt MK: Development of EST-PCR markers for DNA fingerprinting and genetic relationship studies in blueberry (Vaccinium, section Cyanococcus). Journal of the American Society for Horticultral Science 2003, 128:682-690.
11. Alkharouf N, Khan R, Matthews BF: Analysis of expressed sequence tags from roots of resistant soybean infected by the soybean cyst nematode. Genome 2004, 47:380-388.
12. Kim S, Ahn KP, Lee YH: Analysis of genes expressed during rice Magnaporthe grisea interactions. Mol Plant-Microbe Interact 2001, 14:1340-1346.
13. Kruger WM, Pritsch C, Shao S, Muehlbauer G: Functional and comparative bioinformatics analysis of expressed genes from wheat spikes infected with Fusarium graminearum. Mol Plant-Microbe Interact 2002, 15:445-455.

Additional File 1

List of genes printed on microarray slides. The supplemental table lists all the genes that were printed on the microarray slides that were involved in the blueberry cold hardiness experiments [5].
Click here for file [http://www.biomedcentral.com/content/supplementary/1471-2229-7-5-S1.xls]