

Database

Open Access

BBGD: an online database for blueberry genomic data

Nadim W Alkharouf*¹, Anik L Dhanaraj², Dhananjay Naik², Chris Overall^{3,4}, 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

Published: 30 January 2007

Received: 18 September 2006

BMC Plant Biology 2007, **7**:5 doi:10.1186/1471-2229-7-5

Accepted: 30 January 2007

This article is available from: <http://www.biomedcentral.com/1471-2229/7/5>

© 2007 Alkharouf et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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 (BBGD) 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: BBGD is the world's first database for blueberry genomics. BBGD 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. BBGD 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, account-

ing 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



Figure 1
A snap shot of BBGD's main web page that shows the wealth of information that is available through the web site and the general capabilities of the database.

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-	647/555	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-_Tifblue	647/555	0
47	1	CAField+/CAField-_Tifblue	647/555	67
48	1	CAField+/CAField-_Tifblue	647/555	399
49	1	CAField+/CAField-_Tifblue	647/555	779
50	1	CAField+/CAField-_Tifblue	647/555	1234
51	1	CAField-/CAField-_Tifblue	555/647	0
52	1	CAField+/CAField-_Tifblue	555/647	67
53	1	CAField+/CAField-_Tifblue	555/647	399
54	1	CAField+/CAField-_Tifblue	555/647	779
55	1	CAField+/CAField-_Tifblue	555/647	1234
66	1	CAField+/CAField-_Tifblue	647/555	67
67	1	CAColdRm-/CAColdRm-_Tifblue	647/555	0
68	1	CAColdRm+/CAColdRm-_Tifblue	647/555	500
69	1	CAColdRm+/CAColdRm-_Tifblue	555/647	500
70	2	CAColdRm-/CAColdRm-_Tifblue	647/555	0
71	2	CAColdRm+/CAColdRm-_Tifblue	647/555	500
72	2	CAColdRm-/CAColdRm-_Tifblue	555/647	0
73	2	CAColdRm+/CAColdRm-_Tifblue	555/647	500
56	2	CAField-/CAField-_Tifblue	647/555	0
57	2	CAField+/CAField-_Tifblue	647/555	67
58	2	CAField+/CAField-_Tifblue	647/555	399
59	2	CAField+/CAField-_Tifblue	647/555	779
60	2	CAField+/CAField-_Tifblue	647/555	1234
61	2	CAField-/CAField-_Tifblue	555/647	0
62	2	CAField+/CAField-_Tifblue	555/647	67
63	2	CAField+/CAField-_Tifblue	555/647	399
64	2	CAField+/CAField-_Tifblue	555/647	779
65	2	CAField+/CAField-_Tifblue	555/647	1234
30	2	CAColdRm-/CAColdRm-	647/555	0
31	2	CAColdRm+/CAColdRm-	647/555	500
32	2	CAColdRm+/CAColdRm-	647/555	1000
33	2	CAColdRm-/CAColdRm-	555/647	0
34	2	CAColdRm+/CAColdRm-	555/647	500
35	2	CAColdRm+/CAColdRm-	555/647	1000
14	2	CAField-/CAField-	647/555	0
15	2	CAField+/CAField-	647/555	67
16	2	CAField+/CAField-	647/555	399
17	2	CAField+/CAField-	647/555	779
18	2	CAField+/CAField-	647/555	1234
19	2	CAField-/CAField-	555/647	0
20	2	CAField+/CAField-	555/647	67
21	2	CAField+/CAField-	555/647	399
22	2	CAField+/CAField-	555/647	779
23	2	CAField+/CAField-	555/647	1234

BioSample: 12
 Probe combination: CAField+/CAField-
 Time: 399 hrs.

ElementID	GenBank_Accn	Name	Count	FoldChange	STDEV	T-Test Result
CA36	CF810660	dehydrin 1	18	4.0185	0.7596582	S. Induced
CA63	CF810877	early light-induced protein	12	4.5264	0.3225069	S. Induced
CA76	CF810917	Early light inducible protein	18	6.5314	0.9537108	S. Induced
CA111	CF810442	LEA 1P	18	2.9204	0.4879327	S. Induced
CA156	CF810476	Putative beta amylase	18	4.2098	0.702749	S. Induced
CA278	CF810581	Galactinol synthase	9	2.7522	0.5145313	S. Induced

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.

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.

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.

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

Cold Acclimated library clones:

CloneID	dbESTID	GenBank_Accn	CloneName (Best Blast hit)
CA1	20189951	CF810506	No high scores
CA1.1			Unknown
CA10	20189886	CF810441	phospholipid-hydroperoxide glutathione peroxidase
CA101	20189864	CF810419	Expressed Protein
CA102	20189867	CF810422	hypothetical protein
CA1028	20189865	CF810420	NA
CA1029	20189866	CF810421	Lipid transferase protein precursor
CA1030	20189868	CF810423	Putative protein(also has hi scores for Frigida protein)
CA1031	20189869	CF810424	hypothetical protein
CA1032	20189870	CF810425	Dehydrin 1
CA1034	20189871	CF810426	hypothetical protein
CA1035	20189872	CF810427	Heat shock protein

Figure 3
 The sequence database allows users to browse a specific library in table format. This snap shot represents part of the table that was returned by the database. 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.

Sequence

Upload a file:

OR

Paste a sequence (FASTA):

Options

BLAST Type: Database: Output Type:

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*).

Additional material

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 blue berry cold hardiness experiments [5].

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2229-7-5-S1.xls>]

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

- Moore JN: **The blueberry industry of North America.** *Acta Hort* 1993, **346**:15-26.
- 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.
- Prior RL, Cao G, Martin A, Sofic E, McEwen J, O'Brien C, Lischner N, Ehlenfeldt M, Kalt W, Krewer 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.
- Blue Berry Genomics Database (BBGD)** [<http://psi081.ba.ars.usda.gov/BBGD/>]
- Dhanaraj AL, Alkharouf NW, Beard HS, Chouikha IB, Matthews BF, Wei H, Arora R, Rowland LJ: **Major differences observed in transcript profiles of blueberry during cold acclimation under field and cold room conditions.** *Planta*. DOI: 10.1007/s00425-006-0382-1
- Codd EF, Codd SB, Salley CT: **Providing OLAP (on-line analytical processing) to user-analysis: An IT mandate.** *Technical Report, EF Codd & Associates* 1993.
- 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.
- Dhanaraj AL, Slovin 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.
- Dhanaraj AL, Slovin 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.
- Rowland LJ, Mehra S, Dhanaraj AL, Ogden EL, Slovin 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 Horticultural Science* 2003, **128**:682-690.
- 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.
- Kim S, Ahn KP, Lee YH: **Analysis of genes expressed during rice Magnaporthe grisea interactions.** *Mol Plant-Microbe Interact* 2001, **14**:1340-1346.
- 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.
- Ewing RM, Kahla AB, Poirrot O, Lopez F, Audic S, Claverie JM: **Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression.** *Genome Res* 1999, **9**:950-959.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

