

# Online Analytical Processing (OLAP): A Fast and Effective Data Mining Tool for Gene Expression Databases

Nadim W. Alkharouf,<sup>1,2</sup> D. Curtis Jamison,<sup>2</sup> and Benjamin F. Matthews<sup>1</sup>

<sup>1</sup>*Soybean Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD 20705, USA*

<sup>2</sup>*School of Computational Sciences, George Mason University, Fairfax, VA 22030, USA*

Received 27 July 2004; revised 26 November 2004; accepted 7 December 2004

Gene expression databases contain a wealth of information, but current data mining tools are limited in their speed and effectiveness in extracting meaningful biological knowledge from them. Online analytical processing (OLAP) can be used as a supplement to cluster analysis for fast and effective data mining of gene expression databases. We used Analysis Services 2000, a product that ships with SQLServer2000, to construct an OLAP cube that was used to mine a time series experiment designed to identify genes associated with resistance of soybean to the soybean cyst nematode, a devastating pest of soybean. The data for these experiments is stored in the soybean genomics and microarray database (SGMD). A number of candidate resistance genes and pathways were found. Compared to traditional cluster analysis of gene expression data, OLAP was more effective and faster in finding biologically meaningful information. OLAP is available from a number of vendors and can work with any relational database management system through OLE DB.

## INTRODUCTION

Until recently, data mining required expensive and cumbersome data mining software or a database expert who could accurately translate a request for information into a functional, preferably efficient, query. Database warehouses and online analytical processing (OLAP) offer an attractive and readily available alternative.

As compared to a database, a data warehouse has faster retrieval time, internally consistent data, and a construction that allows users to slice and dice (ie, extract a single item (slice) and compare items in a cross-tabulated table (dice)). The primary difference between a data warehouse and a traditional transaction database lies in the volatility of the data. The information in a transaction database is constantly changing, whereas data in a data warehouse is stable; its information is updated at standard intervals (monthly or weekly). A perfect data warehouse would be

updated to add values for the new time period only, without changing values previously stored in the warehouse. Thus, microarray databases can be data warehouses, because the data in them is consistent and stable. Gene expression values in any given experiment remain the same and usually only new data from new experiments is added. Data warehousing software is incorporated in most of the major relational database management systems such as SQLServer2000 and Oracle 9i.

OLAP represents a class of software that enables decision support and reporting based upon a data warehouse [1]. A schematic view of how OLAP software interacts with the data warehouse is shown in Figure 1. OLAP allows for the fast analysis of shared multidimensional information. It is fast because most system responses to users are delivered within 5 seconds, with the simplest analysis taking no more than 1 second and very few taking more than 20 seconds. However, speeds vary by OLAP vendor and system hardware. The key feature of OLAP is that it provides a multidimensional, conceptual view of the data, including full support for hierarchies and multiple hierarchies.

OLAP's underlying structure is the cube [2]. A cube is defined by any number of data dimensions; it is not limited to three; and sometimes an OLAP cube may have fewer than three dimensions. The data dimensions describe an OLAP cube just as width, height, and depth

---

Correspondence and reprint requests to Benjamin F. Matthews, Soybean Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD 20705, USA; E-mail: matthewb@ba.ars.usda.gov

This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

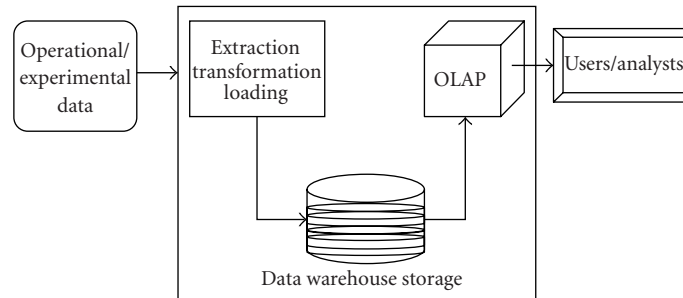


FIGURE 1. OLAP, cubes and where they fit in a data warehousing solution. OLAP provides efficient and easy-to-use reporting tools and graphical interface, to enable users to mine a data warehouse for hidden information.

describe a geometrical cube. Where it is appropriate, dimensions can be organized into any number of levels (hierarchies).

In relational database systems, OLAP cubes are constructed from a fact table and one or more dimension tables. A fact table is the relational table in the warehouse that stores the detailed values for measures (the thing you are measuring). For example, this could be the values for the relative change in gene expression. The dimension tables however are more abstract, containing only one row for each leaf (lower) member in the fact table. They are used to create summaries and aggregates of the data in the fact table. Ad hoc calculations and statistical analysis can also be achieved, but are vendor specific. Analysis Services 2000 (used here) is capable of such ad hoc calculations on complex data.

The relationship between two dimensions can be modeled using a grid as shown in Table 1. Dimensions are the labels along the axes of the grid and each of the cells is a fact. Facts correspond to the cross product of each dimension of the cube. The data in the cell is a measure, a numerical value. A cube is designed to aggregate, analyze, and find trends in the measures. For example, if the cube describes relative gene induction, the measure is the average relative expression level of a gene under experimental conditions compared to control conditions, and the cube is used to compute this average for the dimensions chosen. In other words, the measure is the number that you would find in the grid cell.

Dimensions are organized into smaller units by using levels where necessary. Levels may also contain other levels, depending on how they are configured in the cube. For example, in Table 1 which represents a two-dimensional cube from our data warehouse designed to identify soybean cyst nematode (SCN) resistance-associated genes in soybean cultivars Peking (P) and Kent (K), the biosamples are considered one level under  $K + /K-$  (Kent infected with SCN versus uninfected), which in turn is another level (along with  $P + /P-$ ; Peking infected with SCN versus uninfected) under the dimension probe combination. A fact describes the combination of the various dimensions, for example, probe combination =  $P + /P-$ ,

TABLE 1. The organization of a cube with two dimensions. In this example, probe combination and genes are dimensions;  $P + /P-$ ,  $K + /K-$ , biosample 1, biosample 2, A01A10, SSH1B07, D09H12, and B03C02 are levels of the respective dimension. The cells containing various figures are facts. Individual data in the fact cells are the values of the measures. In this example, there are two measures used in the cube, one is the fold induction, the second is the result of the  $t$  test (1 significantly induced, -1 significantly suppressed, 0 unchanged).

Fold induction/ $t$ test	Probe combination			
	$P + /P-$	$K + /K-$		
		Biosample 1	Biosample 2	
Genes	A01A10	1.2	1.5	0.76
		1	-1	1
	SSH1B07	0.34	2.3	-0.98
		0	1	-1
	D09H12	-1.6	1.4	0.03
		-1	1	0
	B03C02	2	1.8	-2.1
		1	1	-1

gene = A01A10, time = 6 hours yields a specific fact about the induction of gene A01A10 in  $P + /P-$  6 hours after SCN infection (assuming we added a third dimension of time). This representation is just like the  $(x, y, z)$  coordinate system in mathematics. Depending on the way the cube is being used, the fact may show a measure of the induction of a gene at a specific biosample or the result of the  $t$  test or some other differential gene expression test.

The meaning of the measure depends on how the cube is defined. The value represents an aggregation for the defined grouping. The measures inside the cube are always numeric. The mathematical operations of count and sum are the primary reason why data warehouses are useful. Calculated measures, such as average, can be calculated from those two basic measures. These are called aggregations. Once dimensions are organized and a cube is being processed, the aggregations are calculated. Generally,

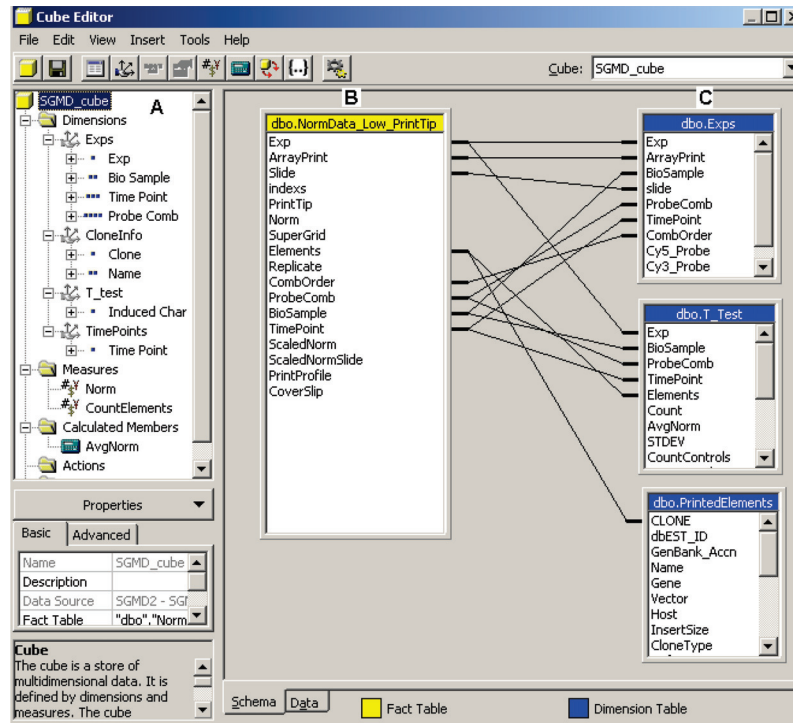


FIGURE 2. A snapshot of a multidimensional cube of gene expression data constructed in Microsoft's Analysis Services 2000 (shipped with SQLServer2000). (A) shows the dimensions of the cube and their associated levels, (B) is the fact table, and (C) shows the dimension tables.

aggregations are calculated immediately after the cube is initially populated or when there is a change in the content of the cube.

OLAP has been used to make some important discoveries in the biomedical field. For instance, Dzeroski et al [3] used OLAP on a database of patients with Y chromosome deletions and found correlations between deletion patterns and patient populations, as well as clinical phenotype severity. OLAP has also been used in the health management field. For example, Silver et al [4] used OLAP to make business decisions that improved operational efficiency of hospitals while maintaining high levels of patient care. Hristovski et al [5] found OLAP to be a suitable data mining tool for public health. However, to the best of our knowledge, OLAP has not been applied to gene expression databases.

We applied OLAP technology to our microarray warehouse, the soybean genomics and microarray database (SGMD) [6], to mine a time-course experiment aiming at discovering genes expressed in soybean roots upon infection by the SCN. SCN is the major pest of soybean and causes an estimated loss of 1\$ billion in the United States per year. The discovery of genes expressed under these conditions will provide scientists with information and tools to develop soybean cultivars that are resistant to SCN. Using OLAP we identified numerous candidate genes and associated pathways in a susceptible soybean

cultivar (Kent) after infection with SCN [7, 8]. In comparison to traditional gene expression data mining methods, such as *k*-means and self-organizing maps (SOM) clustering, OLAP performed significantly better at finding candidate genes for further study.

## METHODS

### Cube construction

We used Analysis Services 2000 (Microsoft, Redmond, Wash), a product that comes with SQLServer2000, to build a multidimensional cube of gene expression experiments conducted over time (Figure 2). Our fact table contained rows of data describing clones and their fold induction at each time point for each biosample and probe combination ( $P + /P-, K + /K-$ ). The measures from this fact table were the normalized log ratio from Lowess print-tip normalization [9], called norm, and the count of unique clones printed (called CountElements). A calculated measure, named AvgNorm, was created to represent the average normalized log ratio from the two measures mentioned above. Four dimensions were created. The first was experiments (exps), which had four levels, exp, biosample, time point, and probe combination. A second dimension, called CloneInfo, had two levels, the clones ID's and their names. The third (*t* test) and fourth (TimePoints) dimensions had one level each,

induced char (which refers to the results of the *t* test) and time point, respectively (Figure 2). Cubes are very flexible, new dimensions and measures can be added and removed to customize the data analysis process, that is, the cube can be configured to answer the scientific question at hand.

### Microarray data

Gene expression levels of approximately 6000 soybean genes were measured at seven time points after SCN infection [7]. The standard reference design was used for these microarray experiments. The reference (control) sample was RNA extracted from soybean (cultivar Kent) roots which is SCN susceptible, not infected with SCN, and our treatment samples were RNA extracted from Kent cultivar 6 hours, 12 hours, 24 hours, 2 days, 4 days, 6 days, and 8 days after infection with SCN. Reverse labeling of probes was conducted because the two dyes (Cy3 and Cy5) may not have the same labeling efficiencies and do not have exactly the same correspondence between mRNA concentration and fluorescent intensities. Each gene was printed in triplicate on glass slides. Two replicated slides (one of which is the dye swap) were used for each time point. Two biological samples were also used to account for biological variation and inherent variation in the extraction of mRNA, generating a total of  $7 \times 2 \times 2 = 28$  slides and 12 data points for each gene. Self-self hybridized slides were generated for *t* test analysis. *t* tests were used to determine differentially expressed genes at each of the time points [9]. Details on slide printing, hybridization, and scanning protocols are described in Alkharouf et al [7]. OLAP was used to produce lists of common significantly induced/suppressed genes at the early (6, 12, and 24 hours), mid (2 and 4 days), and late (6 and 8 days) time points. We used results of the *t* test to determine significance ( $P \leq .05$ ) and chose a cutoff value of 1.5 fold for extra stringency. In addition, *k*-means and 2D SOM clustering were applied on the time series data set. *k*-means was done using *J*-Express version 2.0 (MolMine; <http://www.molmine.com>) setting  $K = 20$ , initialization method to Forgy, and distance metric to Euclidean. SOM was done using the 2D SOM algorithm from GeneSight version 3.5.2 (BioDiscovery; <http://www.biodiscovery.com>), setting the number of horizontal clusters to 5, the number of vertical clusters to 5, distance metric to Euclidean and clustering by genes only.

## RESULTS

OLAP was used to drill down, slice, and dice the time series data and find lists of genes induced and suppressed in each of the specified time intervals (Table 2). OLAP was used to find commonly induced or suppressed genes at two or more time points and in one or more biosamples. OLAP was very quick and efficient in providing those reports. On average OLAP only took 2 to 5 seconds to return

a result of a query after the cube was constructed (running on a 1.8 GHz Pentium 4 workstation with 1 GB RAM). This is a fraction of the time needed to produce similar reports from complex SQL queries and multiple-table joins. For instance selecting statistically induced genes common to the 6-, 12-, and 24-hour time points, which requires 3-table joins, took almost 25 seconds to achieve, whereas the same report took only 1 second with OLAP running on the same system.

A common technique for viewing multidimensional output is to view the output as a two-dimensional "slice" of a cube. This is the way the Microsoft SQLServer2000 analysis services display output. This is a simple and informative technique to view the reports in a spreadsheet-like manner. Multidimensional extensions (MDX) can also be used to query cubes instead of using the user interface mentioned above. MDX is a syntax designed for querying multidimensional objects and data and is more flexible than the user interface. It was used to query the cube and obtain the results shown in Table 2. MDX has a similar syntax to SQL, but is designed to work with multidimensional cubes instead of relational tables. The SQLServer2000 analysis services manager has an interface that accepts MDX queries.

The OLAP reports highlighted a number of genes and defense pathways that were triggered in soybean in response to SCN infection (Table 2). These are discussed in detail in [7]. The key findings in the study were that the nematodes elicit the activation of a transcription factor (WRKY) that shuts down a defense pathway known as the salicylic acid inducible pathway, thereby rendering the plants more susceptible to nematode infection.

OLAP found a number of candidate resistance genes that *k*-means and SOM did not (Table 2), whereas cluster analysis did not reveal any new information that OLAP did not identify by MDX queries. For instance OLAP found trehalose-6-phosphate synthase (TPS) induced at the mid time points, whereas cluster analysis did not. TPS is a key enzyme of sugar metabolism and its induction at the mid time points, where the nematode has formed the syncytium (feeding site), may be an indicator of the parasite's success in utilizing the plants metabolite synthesis apparatus for its own sustenance. Metabolic profiling experiments conducted in collaboration with the Noble foundation also show increased levels of trehalose in Kent 48 hours after infection with SCN (unpublished data). OLAP also found jasmonic acid (JA) inducible genes, such as pathogenesis-related protein PR-6 and chalcone synthase, induced at the early and mid time points whereas cluster analysis did not. The JA signaling pathway is known to be induced in plants after wound damage or parasitic infection [18].

Generally, we found OLAP a lot more powerful for determining genes induced at specific time intervals but not at other time points. This was hard to do using cluster analysis, because the algorithms are designed to group genes with similar profiles, not necessarily to identify

TABLE 2. Genes found to be induced at different time intervals using OLAP, *k*-means, and SOM clustering. Many of the key candidate resistance genes were identified by OLAP and not cluster analysis, in particular those genes induced at specific time intervals and not others. Cluster analysis did not reveal any other genes that OLAP did not.

Time	GeneID	GeneName	OLAP	<i>k</i> -means	SOM	Comments
Induced at all time points	BM139889	Proline-rich glycoproteins	✓	—	—	Cell wall proteins that are found activated during pathogen attack [10] to reinforce the cell wall
	BM107775	Peroxidase	✓	✓	✓	Involved in detoxification and is activated during the hypersensitive response in plants against pathogen attack [11]
	BM139591	Cytochrome P450 monooxygenase	✓	—	—	Photosynthesis-related gene
	BM107779	Photosystem II core proteins	✓	✓	✓	Involved in plant photosynthesis and energy production
	BM107798	4-coumarate-CoA ligase	✓	✓	✓	Involved in phenylpropanoid metabolism and the synthesis of secondary metabolites that are known to be involved in plant defense [12]
	BM108156	Transcription factor WRKY6	✓	✓	✓	Believed to suppress PR-1 genes, thereby inferring susceptibility to pathogen attack in plant species [13]
Induced at the early time points only	CA850582	Trypsin inhibitor proteins	✓	—	—	Proteinase inhibitors
	BM107847	Germin-like protein	✓	—	—	Known to have antimicrobial activity, activated in plants during pathogen infection [14]
	CA851099	Pathogenesis-related protein PR-6	✓	—	—	Proteinase inhibitors known to be induced by jasmonic acid [15]
Induced at the mid time points only	DUP21F10	Trehalose-6-phosphate synthase (TPS)	✓	—	—	Synthesizes trehalose, is thought to be an important regulator of sugar metabolism [16]
	BM108164	Pyrophosphatase	✓	—	—	Metabolism-related gene
	BM108095	Sali3-2 protein	✓	—	—	Induced by aluminum in soybean roots [17]
	BM107806	Chalcone synthase	✓	—	—	Induced by the jasmonic acid signaling pathway [18]
Induced at the late time points only	BM108193	Glutamate dehydrogenase	✓	—	—	Metabolism-related gene
	CA853854	Geranylgeranyl hydrogenase	✓	—	—	Metabolism-related gene
	BM107804	Tyrosine-phosphatase	✓	—	—	Metabolism-related gene
Commonly induced at the early and mid time points	CA850882	Stress-induced gene SAM-22	✓	—	—	A stress-induced PR-10 protein, which is a ribonuclease protein found activated in plants after viral infection [15]
	BM107930	Heat shock protein 70	✓	—	—	Helps new or distorted proteins fold into shape, found induced in a number of plant species after pathogen infection [19]
	BM107821	Lectin-chitin	✓	✓	✓	Cell wall protein

TABLE 2. Continued

Time	GeneID	GeneName	OLAP	<i>k</i> -means	SOM	Comments
Commonly induced at the early and late time points	BM107803	Beta-glucosidase	✓	—	—	Metabolism-related gene
	CA852009	Fructose-biphosphate aldolase	✓	—	—	Metabolism-related gene
	BM107809	Sucrose synthase	✓	✓	✓	Metabolism-related gene
	BM108104	ATP-synthase	✓			Metabolism-related gene
Commonly induced at the mid and late time points	BM108223	Lipoxygenase	✓	✓	✓	Involved in jasmonic acid synthesis and is implicated in plant responses against pathogens [18]
	BM108233	Ubiquitin	✓	—	—	Plays an important role in marking proteins for proteolytic degradation, one of the key events in the systematic defense mechanism of a plant against pathogen invasion [20]
	CA853086	Metallothionein	✓	✓	✓	A member of the aquaporin (AQP) water channel family, induced in rice upon infection with <i>Magnaporthe grisea</i> [21]

genes induced uniquely at one time point, but not at others. This explains why none of the genes found uniquely induced at the early, mid, or late time points were identified by cluster analysis (Table 2). Finding these genes is important for the dissection of the metabolic effects of the nematode invasion across time.

In terms of speed, OLAP took approximately 1.2 minutes to generate all the reports summarized in Table 2 and are shown in their entirety on <http://psi081.ba.ars.usda.gov/SGMD/Publications/OLAP/>. In contrast, it took 5 times longer (approximately 6.5 minutes) to do one of the cluster analysis methods (including the time it takes to export the data from the database to the respective clustering software in the required format). If one were to also measure the time it takes to interpret the OLAP reports versus the clustering results, OLAP would be even at a more advantage point, because it makes the results easier to interpret. Results of the cluster analysis can also be accessed from the web site mentioned above.

## DISCUSSION

Gene expression data is valuable for the understanding of gene regulation and biological networks. A main goal of gene expression data analysis is to determine what genes are expressed as a result of a certain cellular state, that is, what genes are expressed in diseased cells that are not expressed in healthy cells. Microarray experiments

profile hundreds to thousands of genes at a time generating large data sets that are only getting bigger as more advances in genomics and microtechnologies are made. As these data sets become larger, however, the need for fast and effective database mining tools becomes more obvious and necessary. Data warehouses and OLAP provide tools to construct, populate, view, and access microarray data in an efficient and fast manner. The fundamental unit of OLAP software is the cube, which is a repository of integrated information from the existing data sources.

In our cube design the data sources were the relational tables in SGMD, a gene expression database [6]. Microarray databases are in fact data warehouses because of their consistent and stable data, and little if any modifications to the database model need to be made to use OLAP. OLAP proved to be more efficient than standard relational database queries that rely on time-consuming multitable joins. Although the results obtained from OLAP and these standard SQL queries are the same, the time it takes to execute an OLAP query was found to be 25 times greater than standard SQL queries.

OLAP provides a different view of the data compared to cluster analysis and provides additional insights into the data as shown in Table 2. OLAP identified a number of candidate resistance genes that cluster analysis did not. One reason is the large number of genes of an unknown function that makes such cluster analysis difficult to interpret. OLAP avoids this issue because it allows for the

categorization of genes into categories of known and unknown functions, thereby reducing the complexity of the problem by allowing investigators to analyze genes with a known function first. Another benefit of OLAP is that the values of the clustered elements do not all have to be the same unit, as they are in cluster analysis. This is useful when searching for trends across a heterogeneous data set. In OLAP, you can set any type or number of dimensions to drill your data with, thereby identifying trends that cannot be identified using cluster analysis.

OLAP's main advantage is that it is flexible and can be customized to answer the scientific question at hand if some prior knowledge is known about the data sets, whereas cluster analysis is mostly used as an initial data mining tool with no prior knowledge and is used mainly for grouping genes based on similar expression profiles. The genes that are clustered together however can vary considerably because of the different similarity metrics that are used. Another issue with clustering is that a gene can be characterized in more than one way, while it can belong to only one cluster. OLAP allows scientists, especially those not trained in the computational sciences, to mine their data sets to not only group genes based on their expression profiles but to also ask specific scientific questions such as "give me the genes induced at a certain time point, that is, not induced at all other time points, or the genes induced at time point A that are also induced at time points B, and C," for instance. The answers to these questions can provide valuable insights into the relationships between genes and pathways that cluster analysis cannot answer.

In the case of our data set, for instance, seeking resistance genes induced at specific time points yielded a number of candidate resistance genes and gave us insights into the metabolic changes in soybean when infected with SCN. Thus OLAP is an automation of the manual analysis that most biologists would always perform rather than relying on visually appealing but scientifically uninformative cluster analysis. We are not suggesting that OLAP is better than cluster analysis, but only that the two methods are useful and quite different. We are suggesting however that OLAP can be considered as a supplement or even an alternative to cluster analysis when clustering methods are not suitable to analyze a data set, such as small time-course data sets as ours.

The implementation of OLAP technology to gene expression analysis is not difficult given the right tools. OLAP can be applied to any gene expression database built on any of the major relational database management systems (Oracle, Sybase, MySQL, or even Access), through the use of OLE DB (an industry standard technology for database connectivity). OLAP reports can also be obtained using Excel's (Microsoft, Redmond, Wash) pivot tables, a feature that allows one to cross-tabulate columns in Excel. This might work well for small data sets. OLAP's ability to drill through the data and find common/unique genes given different criteria, along with its

flexibility, make it an important data mining tool in gene expression analysis, one that holds great promise in our view.

This study also demonstrates that databases and database applications may not be used solely for the storage and retrieval of expression data but that they can act as tools for doing exploratory data analysis as well. In fact databases can eliminate the need for third-party software, because most of the analysis, even time series analysis, and can be done within the database itself.

## REFERENCES

- [1] Codd F, Codd SB, Salley CT. *Providing OLAP (Online Analytical Processing) to User-Analysts: An IT Mandate*. San Jose, Calif: Codd EF & Associates;1993. Technical Report.
- [2] Gray J, Bosworth A, Layman A, Pirahesh H. *Data cube: A Relational Aggregation Operator Generalizing Group-by, Cross-Tabs and Sub-totals*. Washington, DC: Microsoft Corporation; 1995. 95-22. MSR Technical Report.
- [3] Dzeroski S, Hristovski D, Peterlin B. Using data mining and OLAP to discover patterns in a database of patients with Y chromosome deletions. *Proc AMIA Symp.* 2000;215–219.
- [4] Silver M, Sakata T, Su HC, Herman C, Dolins SB, O'Shea MJ. Case study: how to apply data mining techniques in a healthcare data warehouse. *Healthc Inf Manag.* 2001;15:155–164.
- [5] Hristovski D, Rogac M, Markota M. Using data warehousing and OLAP in public health care. *Proc AMIA Symp.* 2000;369–373.
- [6] Alkharouf NW, Matthews BF. The soybean genomics and microarray database. *Nucleic Acids Research.* 2004;32:398–400.
- [7] Alkharouf N, Chouikha I, Beard H, et al. Expression of soybean genes during invasion of susceptible roots by the soybean cyst nematode. *Mol Plant Microbe Interact.* In press.
- [8] Khan R, Alkharouf N, Beard H, et al. Resistance mechanisms in soybean: gene expression profile at an early stage of soybean cyst nematode invasion. *Nematology.* 2004;36(3):241–248.
- [9] Yang YH, Dudoit S, Luu P, et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research.* 2002;30(4):e15.
- [10] Esquerre-Tugaye M, Campargue C, Mazau D. The response of plant cell wall hydroxyproline-rich glycoproteins to microbial pathogens and their elicitors. In: Datta SK, Muthukrishnan S, eds. *Pathogenesis-Related Proteins in Plants*. Boca Raton, Fla: CRC Press; 1999:157–170.
- [11] Low PS, Merida JR. The oxidative burst in plant defense: function and signal transduction. *Physiol Plant.* 1996;96:533–542.

- [12] Ryan CA, Jagendorf A. Self defense by plants. *Natl Acad Sci.* 1995;92(10):4075.
- [13] Maleck K, Levine A, Eulgem T, et al. The transcriptome of *Arabidopsis thaliana* during systemic acquired resistance. *Nature Genetics.* 2000;26(4):403–410.
- [14] Schenk PM, Kazan K, Wilson I, et al. Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. *Natl Acad Sci.* 2000;97(21):11655–11660.
- [15] Van Loon LC. Occurrence and properties of plant pathogenesis-related proteins. In: Datta SK, Muthukrishnan S, eds. *Pathogenesis-Related Proteins in Plants.* Boca Raton, Fla:CRC Press;1999:1–19.
- [16] Eastmond PJ, Li Y, Graham IA. Is trehalose-6-phosphate a regulator of sugar metabolism in plants? *Exp Bot.* 2003;54(582):533–537.
- [17] Ragland M, Soliman KM. Sali5-4a and sali3-2: two genes induced by aluminum in soybean roots. *Plant Physiology.* 1997;114(3):555–560.
- [18] Creelman RA, Mullet JE. Jasmonic acid distribution and action in plants: regulation during development and response to biotic and abiotic stress. *Proc Natl Acad Sci.* 1995;92(10):4114–4119.
- [19] Puthoff DP, Nettleton D, Rodermeil SR, Baum TJ. *Arabidopsis* gene expression changes during cyst nematode parasitism revealed by statistical analyses of microarray expression profiles. *Plant.* 2003;33(5):911–921.
- [20] Kepinski S, Leyser O. Ubiquitination and auxin signaling: a degrading story. *Plant Cell.* 2002;14:81–95.
- [21] Kim S, Ahn IP, Lee YH. Analysis of genes expressed during rice-*Magnaporthe grisea* interactions. *Mol Plant Microbe Interact.* 2001;14(11):1340–1346.