14

DNA Microarray Applied to Data Mining of *Bradyrhizobium elkanii* Genome and Prospection of Active Genes

Jackson Marcondes and Eliana G. M. Lemos
Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista
Brazil

1. Introduction

One of the factors responsible for the expansion and competitiveness of soybean crop [*Glycine max* (L.) Merrill], *Fabaceae*, is its nitrogen fixation capacity through a symbiotic association with *Bradyrhizobium japonicum* (Jordan, 1982) and *Bradyrhizobium elkanii* (Kuykendall et al., 1992) bacteria. Because of this association, N fertilization is usually not required in soybean fields. The genome of *Bradyrhizobium japonicum* USDA 110 was completely sequenced (Kaneko et al., 2002) while the genome of *Bradyrhizobium elkanii* SEMIA 587 (Rumjanek et al., 1993) is currently being sequenced at our lab. Results obtained so far allowed the selection of clones from genomic DNA libraries for the development of DNA microarray containing 2654 *B. elkanii* genes.

The potential of high-throughput DNA microarray technology applied to study the transcriptional response of many organisms to genetic and environmental changes has been clearly demonstrated in the past few years (Dhamardi & Gonzalez, 2004). Recently, the design and use of a partial-genome microarray for transcriptome analysis of *B. japonicum* was reported. This study had focused in regulatory cascade that induces nitrogen fixation (*nif*) genes (Hauser et al, 2006).

During its life cycle, rhizobia face challenges that demand gene expression regulation. The expression of specific genes is followed by the development of plant-bacteria symbiosis. In the soil, the free-living bacteria are saprophytes competing with other bacteria for subsistence. When they eventually reach a plant's rhizosphere, although rich in nutrients, bacteria must evaluate the host compatibility and compete with other rhizobia and even individuals of the same species before entering roots and establish the symbiosis through nodulation and biological nitrogen fixation. At this stage, the environment is critically altered and bacteria must adapt to a new intracellular life stile in the host plant (Oke and Long, 1999; Loh and Stacey, 2003). These successive environment changes are reflected in metabolic alterations.

Rhizobia ecological and physiological studies are important for practical research applied to the commercial inoculants industry. The competition with native bacteria for nodulation is considered the most limiting factor in the use of inoculants in agriculture (López-Garcia et al., 2002). Besides the genetic differences, the physiological status of rhizobia in the inoculant seems to be different from rhizobia in the soil, in different populations. For

instance, in high titration *B. japonicum* cultures, a *quorum-sensing* factor inhibited the expression of *nod* genes through the induction of *nolA* (Loh et al., 2001).
Since high amounts of viable cells are required in commercial rhizobia inoculants, the physiological status of inoculants is near the stationary phase and they are usually obtained through rich cultures. Because of the importance of *B. elkanii* SEMIA 587 in the composition of commercial inoculants, we studied its metabolic behavior investigating gene expression profile through DNA microarrays in two culture conditions. The response to different nutritional conditions was observed at the lag, log and stationary phases in a complex Triptone-Yeast Medium (TY) and in a Rhizobium Defined Medium (RDM).

2. Experimental design viewing microarray data

Culture conditions and growth curves. *B. elkanii* 587 cultures were kept at 28°C with aeration through an orbital stirring regime at 160 rpm. The culture media TY (Beringer, 1974) and RDM (Vincent, 1970) were used in this study. Flasks containing 100 ml of culture medium were initially inoculated with 1.5 x 10^8 cells/ml for the bacterial cells harvesting at the lag, log, and stationary phases in both media. The representative times of each phase, lag, log or stationary were, respectively: T1 (24 h), T2 (48 h) and T5 (120 h) for cultures in TY medium; and T1 (24h), T5 (120h) and T8 (192 h) for cultures in RDM.
Isolation of total RNA and cDNA labeled synthesis. Cells were collected at previously determined culture ages for total RNA extractions. The whole volume of each culture was centrifuged at 4600 x g for 10 min at 4°C. After cell lysis, extractions were performed with trizol (Invitrogen) and chloroform following the manufacturer's recommendations. ***The*** cDNA direct labeling technique was used during reverse transcription using the fluorophores Cy3 and Cy5. cDNA samples obtained from the RNA isolated from bacteria cultured in RDM were labeled with Cy3, while those produced in TY medium were labeled with Cy5. For labeling, 15 μg of total RNA of each sample was added to 15 μg of randon hexamer primers $pd(N)_6$ (Amersham Bioscience); and 1 μl of the synthetic mRNAs controls spike Ref or Test (Amershan Biosciences) for Cy3- or Cy5-labeling, respectively. The RNA mixture was joined to 2 μl of dNTP mix (dATP, dGTP, dCTP, at 5 mM each; dTTP at 2 mM), 1 μl of Cy3-dUTP or Cy5-dUTP (Amersham Pharmacia Biotech) at 25 μM, and 200 U of SuperscriptII reverse transcriptase (Invitrogen). cDNA synthesis was performed at 37°C for 3 h in the dark. The sample containing labeled cDNA was then purified in a Microcon YM-100 column (Millipore).
Common Reference RNA (CRR) for the synthesis of fluorescently labeled cDNA. In addition to the previous protocol, a bulk containing 5 μg of RNA corresponding to each culture phase in RDM was prepared. This sample, totalizing 15 μg of RNA, was named CRR and converted into Cy3-flurescently labeled cDNA. This sample was used as reference against independent samples of the phases lag, log and stationary of the medium RDM, now labeled with Cy5 from 15 μg of each RNA template. The synthesis and purification of labeled cDNAs was performed as described above.
Construction of DNA microarrays and hybrizations. Briefly, *B. elkanii* 587 genomic DNA was isolated and randomly fragmented. Fragments between 0.6 and 2.0 Kb were recovered and cloned in pUC19/*Sma*I vector (Amershan) to produce a genomic library in *Escherichia coli* DH5α. The clones selected for microarray composition were amplified independently through PCR. For each clone, PCR reactions (20 μl) were performed with 25 ng of plasmidial

DNA template, 200 μM dNTPs, 5 pmoles of the primers M13 forward and reverse (5'-CCCAGTCACGAGTTGTGTAAACG and 5'-AGCGGATAACAATTTCACAGG, respectively), $MgCl_2$ 1.5 mM, reaction buffer [1x] and 1.0 U of Taq DNA polymerase (Invitrogen). Amplification was performed with forty denaturing cycles at 96°C for 20 sec, annealing at 50°C for 30 sec and extension at 72°C for 4 min, followed by 5 min of final extension at 72°C. DNA samples were spotted in glassed slides CMT-GAPS2 (Corning) in duplicates using a GMS-417 Arrayer (Affymetrix). The resulting microarray contained 2,654 *B. elkanii* genes and 5 negative controls, human and plant genes. Hybridization was performed in the GeneTac Hybridization (Genetic Microsystems), to which microarray glass slides were attached. Each cDNA mix was distributed on the slide and hybridized at 42°C for 12 h. After the hybridization, slides were washed automatically and sequentially in 2x SSC/0.5% SDS, 0.5x SSC and 0,05x SSC, at 25°C. Each washing corresponded to a 15-min period, with 10 sec of flow and 20 sec of incubation, for 10 cycles. Slides were dried for 15 min. Three slides were used per experiment i.e., RNA extractions, labeled cDNA synthesis and hybridization were performed in triplicates in each experiment.

Microarray data validation through Quantitative Real-Time PCR. The genes *fixN* (symbiotic cytochrome oxidase) and *σA* (RNA polymerase primary sigma factor) were selected as target and endogenous control genes, respectively. Primers and probes were obtained through the system Assays-by-Design (Applied Biosystems). cDNAs were prepared as in the microarray experiments, except for the absence of fluorophores in the reaction. The resulting cDNAs were used in a 20 μL reaction in the presence of 1 μL of the assay (primers/probe) and 10 μL of TaqMan Universal PCR Master Mix [2x] (Applied Biosystems). The experiment was conducted in an ABI 7500 (Applied Biosystems), following the thermal cycling conditions automatically determined by the equipment. Data were analyzed by the program RQ Study (Applied Biosystems), through the algorithm $2^{-\Delta\Delta Ct}$, which calculates the fold change of gene expression of the target gene, normalized by the endogenous calibrator. The expression of the gene *fixN*, detected in *B. elkanii* during the log phase in RDM in DNA microarray experiments was confirmed by the analysis of relative quantification through Real-Time PCR. Gene expression values were coherent and very similar: 0.65 versus 0.67, for Real-Time PCR and DNA microarray, respectively.

3. Data analysis

Fluorescent signals were scanned by a GMS-418 Arrayer Scanner (Affymetrix). The location and identification of each gene in the array were defined in a text file created with the help of the program CloneTracker 2 (Biodiscovery). To adjust systematic differences in the relative intensity of each signal, quantified data were exported and transformed by the software GeneSight 5.5 (Biodiscovery). This normalization process was applied following the lowess (locally weighted linear regression) correction parameters, as a local normalization method (Quackenbush, 2002).

Three growth phases in two culture conditions were compared by the determination of the ratio (r) of median intensities of each ORF pair in the spots, in array triplicates. The ratios (Cy5/Cy3) were calculated so the Log r (base 2) of the absolute value of expression rates was positive for intensities that were higher in the medium TY, i.e., genes potentially expressed in a complex medium. On the other hand, values were negative for higher intensities in RDM, representing genes potentially expressed in a minimal medium. When

the CRR experiment data were processed, positive log r values indicated genes potentially expressed in each individual culture phase in RDM, while negative values designated the reference mixture.
Triplicate array data, which were independent in a same experiment, were processed by the statistical tool SAM (Significance Analysis of Microarrays) using Microsoft Excel. This analysis is based on a series of specific t-tests for each gene, adapted for the large-scale detection of differentially expressed genes (Tusher et al., 2001). Results were grouped functionally following the classification available on RhizoBase (www.kazusa.org.jp/ rhizobase). Furthermore, principal component analysis was performed as principal global expression patterns implemented in the Statistica data analysis software system v.7 (StatSoft Inc).

4. Global expression parameters and patterns.

The complex medium TY contained triptone and yeast extract as N, carbon and energy sources. The defined medium RDM contained glycerol as the only source of carbon and energy, and sodium glutamate as N source. Besides the metabolic differences, very distinct growth rates were observed in each medium (Fig. 1). In RDM, *B. elkanii* cells showed a growth rate nearly 4 times slower, with G = 68.2 h, in comparison to TY medium (G = 17.2 h).

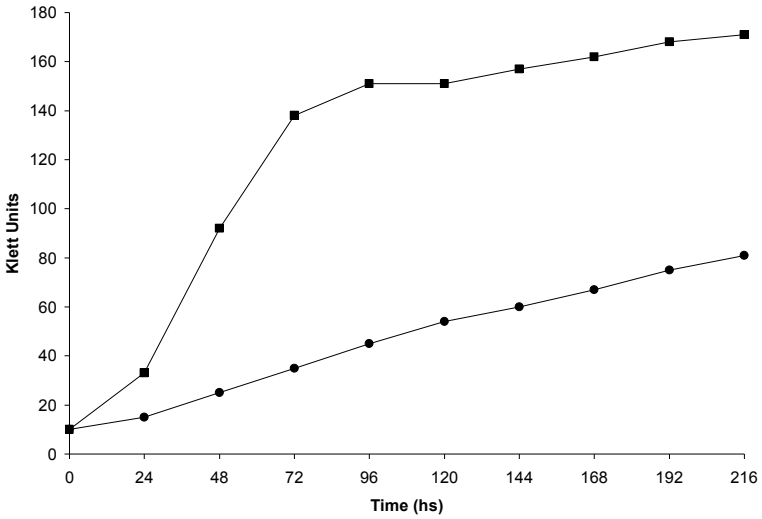

Fig. 1. Growth of *B. elkanii* SEMIA 587 on RDM medium (•) and TY medium (▪) for ten days. Cells were harvested for RNA extraction at 24, 120, and 192 hours for RDM, or 24, 48, and 120 hours for TY, meaning lag, log and stationary phases.

Based on a fold change defined as 1.5, a Log r ≥ 0.58 was considered as significantly higher indicator of gene expression in *B. elkanii* in our analyses (Fig. 2). The technical parameters of the SAM analysis shown in Table 1 demonstrate these data are statistically reliable. It can be observed that the highest number of significant genes was found in the log growth phase, exactly when cells present the highest metabolic activity, while no differentially expressed gene was found in RDM cultured cells at the stationary phase (Table 1). Because of the importance of this culture phase in obtaining cells for commercial inoculant production, this

result led us to analyze *B. elkanii* culture in RDM in detail, applying a mix of Reference RNA (CRR). Therefore, stationary phase data in RDM were obtained from independent analyses of this curve. Lag and log results were considered more interesting for further discussion.

Condition	Input parameters		Output parametes					
	Fold Change	ΔValue	FSN	FDR	q-value (%)	Data confiability (%)	# Significant Genes	# Differentially expressed genes
Lag phase	1.5	0.13	0.87	1.06	0.2879	99.71	61	24 Cy5-TY
								37 Cy3-RDM
Log phase	1.5	0.13	0.78	0.68	0.3588	99.64	91	23 Cy5-TY
								68 Cy3-RDM
Stationary phase	1.5	0.4	0.65	0.97	0.1284	99.87	60	60 Cy5-TY
								No genes

FSN: False Significant Number.
FDR: False Discovery Rate.

Table 1. Applied parameters by SAM tool for data analysis and results.

Global gene expression patterns based on five parameters, two media and three phases, and functional categorization were analyzed (Fig. 3). The two main axes of principal component analysis accounted for 78% and 9,7% of the total variability. RDM cultures showed a higher number of differentially expressed genes in lag and log phases. This is possibly due to a more active cellular metabolism in these adaptation conditions, synthesis of enzymes and exponential growth, which requires additional gene expression for the exploration of the nutrient sources available in this medium. These results were matched with a higher amount of gene groups observed in cells grown in TY at the stationary phase, a situation in which cells were subjected to a higher stress caused by faster growth limiting available resources. The predominance of linkage protein transporters was observed in both media, while in the functional group related to transcription no differential gene expression was detected in any of the conditions. Additional data information and large tables of genes can be viewed in the supplementary material.

In general, specific genes that is going to be discussed below are categorized according to functional groups for each medium and in the different phases, respectively in the following tables (Tables 2, 3, and 4).

5. Biosynthesis of amino acids, translation apparatus and regulatory roles

Differently from cells that showed a more dynamic growth rate in TY medium, cells in RDM probably contained less ribosomes and synthesized proteins more slowly. In addition, cells grown in defined medium are more deprived of amino acids than cells grown in rich media (Tao et al., 1999). Considering the lag and log phases, these differences reflected in a higher number of differentially expressed genes and increased the expression levels related to amino acid biosynthesis and translation in RDM. A higher occurrence of genes related to the biosynthesis of amino acids of the aspartate and branched-chain families. In RDM, duringthe log phase, a *glnD* gene that encodes an uridylyltransferase was detected. This

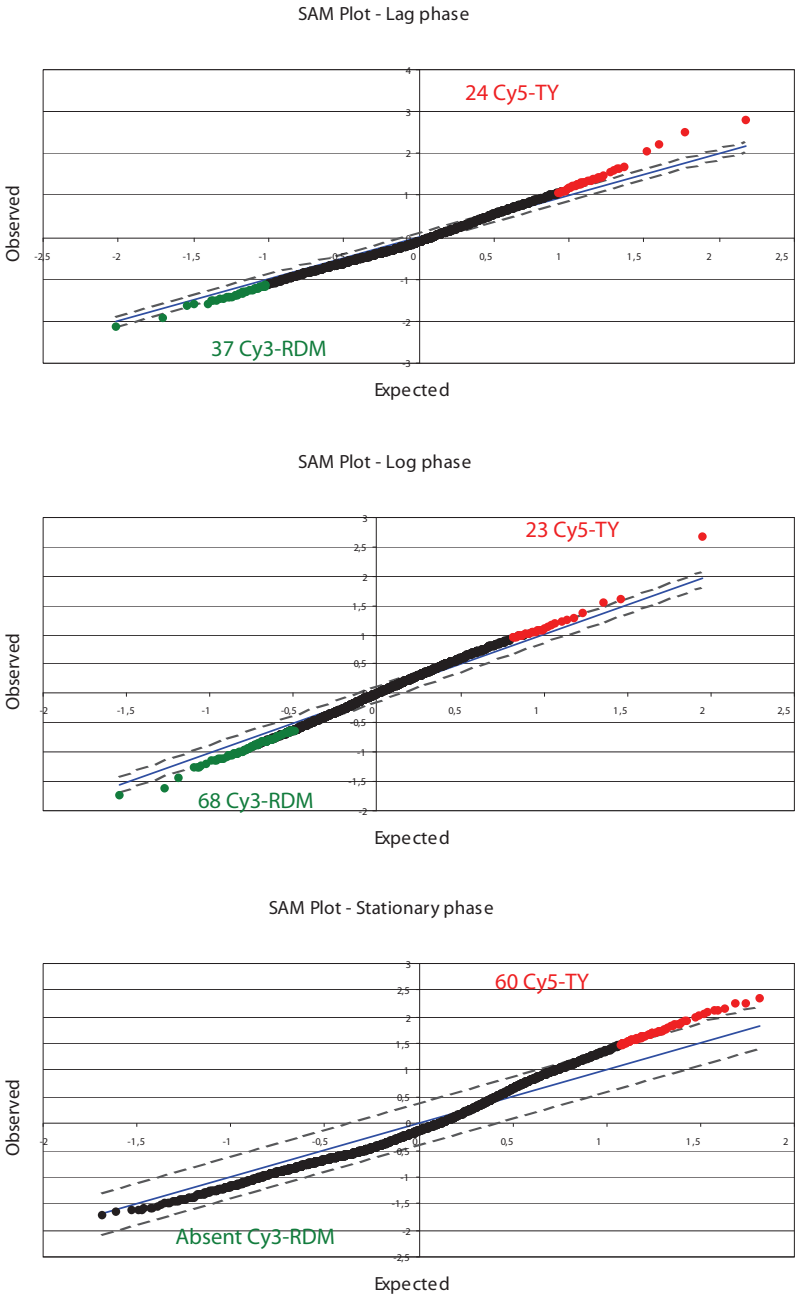

Fig. 2. Gene distribution by SAM tool. SAM plots represent the observed d(i) versus expected dE(i) relative difference of differentially expressed genes. The continuous blue line indicates when d(i) = dE(i). The dotted lines indicate the Δ distance from continuous line.

enzyme has a key role in the adenylation/deadenylation of glutamine synthetase, via uridylation/deuridylation of an adenyltransferase. Glutamine synthetase, on its turn, is involved with the production of biologically active nitrogenated compounds of and it can be considered the key enzyme in the control of N metabolism (Shatters et al., 1989).

In relation to the genes directly associated to translation roles, there was a predominance of aminoacyl-tRNA synthetase-encoding genes, e.g. *cysS*, *leuS* and *lysS*. Although no differential gene expression related to transcription was observed, several transcription regulators were observed among genes with regulatory roles. Members of the transcriptional regulation families TetR, AraC and LysR were predominant in RDM during lag and log phases, while the family Crp only predominated in TY medium. In the log phase, member of the families MarR, AsnC and AraC were found in TY.

Bacteria developed mechanisms to neutralize and expel toxic compounds from cells. Stressful conditions and nutrient source changes result in the activation and repression of genes through direct interaction of regulatory proteins with effector molecules or DNA transcriptional elements. The regulation families MerR and MarR comprise transcriptional activators and repressors, respectively. In *E. coli* these regulators control multidrug resistance pumps (Brooun et al., 1999). In rhizobia, MerR and LysR regulators also act in the repression of nodulation genes (Loh & Stacey, 2003). The regulatory protein AraC, in *E. coli* plays both positive and negative control of the arabinose operon.

Regulatory systems of two components were also detected in both culture media and in RDM only during the log phase. This system comprises the main mechanisms through

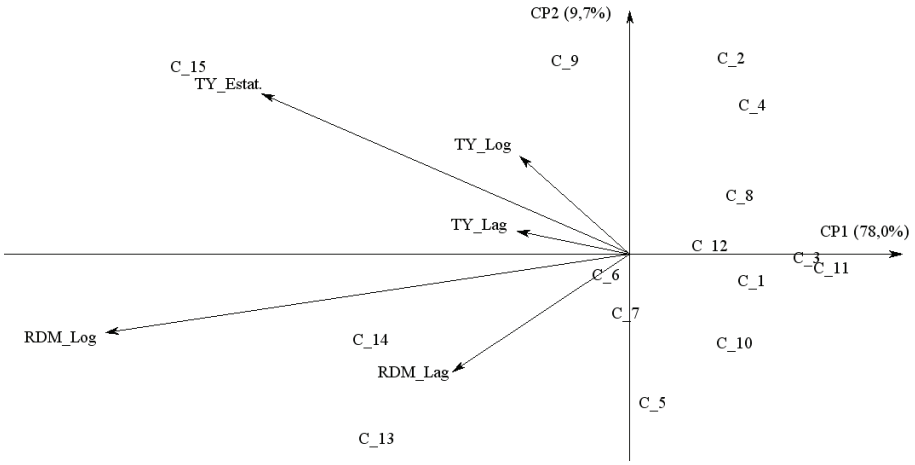

Fig. 3. Global analysis of gene expression profiles by principal component analysis showed in a bi-plot graph. It represents the projection on the two main axes of principal gene analysis based on liquid media type, growth phases, and functional groups. Vector size: number of genes. C_1 to C_15: functional categories, respectively: Amino acid biosynthesis; Biosynthesis of cofactors, prosthetic groups, and carriers; Cell envelope; Cellular processes; Central intermediary metabolism; Energy metabolism; Fatty acid, phospholipid and sterol metabolism; Purines, pyrimidines, nucleosides, and nucleotides; Regulatory functions; DNA replication, recombination, and repair; Transcription; Translation; Transport and binding proteins; Other categories; Hypothetical.

which bacteria detect environment signals and control general cellular processes; therefore it is important for environment adaptation. Typically, they consist of two individual proteins, a sensory histidine kinase and a response regulator (Pao & Saier, 1995; Stock et al., 2000).

6. Nucleotide biosynthesis, DNA replication and cell division

While genes encoding enzymes related to purine nucleotide biosynthesis were predominant in cells grown in defined medium, as *garS, KsgA* and blr7256 (adenylate cyclase), only one gene, *pyrG*, related to pyrimidines, was detected in TY medium. Ribonucleotides are later converted into deoxyribonucleotides through reduction processes. The biosynthesis of pyrimidine ribonucleotides is simpler than the biosynthesis of purine ribonucleotides, especially in the rich medium TY, when the nucleotide biosynthesis through the saving pathway is active (Tao et al., 1999). This pathway involves the use of purine and pyrimidine compounds pre-produced by biodegradation during cell nutrition. Hence, the presence of high quality nutrients in the culture medium results in high concentrations of nucleoside triphosphates (Gaal et al., 1997; Keener & Nomura, 1996).
Differentially expressed genes related to replication, recombination and DNA repair were predominant in cells grown in defined medium, especially during the log phase. The genes *topA* and *gyrA*, which encode topoisomerase I and II, respectively, were detected. DNA topoisomerases avoid DNA supercoiling during replication acting on the rotational tension of the molecule and offering a coiling support for the relaxation of the DNA helix (Zechriedrich et al., 2000). In rich medium, still during the lag phase, the cell division-related gene *ftsK* was detected. This gene seems to have a direct role in the later septation process in *E. coli* (Yu et al., 1998).

7. Biosynthesis of cofactors, prosthetic groups and carriers

Differently from the tendency observed for amino acid and nucleotides biosynthesis genes in RDM, genes associated with the biosynthesis of cofactors, prosthetic groups and carriers were predominant in the rich medium TY. This can be explained by the fact that cells grown in defined medium received most of the vitamins available in RDM, as nicotinic acid, biotin and thiamin. On the other hand, the synthesis of these vitamins must be activated in rich media. In TY, the most expressed gene in the log phase was *thiE*, which is involved in thiamin biosynthesis, followed by the genes *hemN*, and *nadD*, which are related to the biosynthesis of porphyrin and nicotinamide, respectively. The gene *pcaB*, which was also expressed in rich medium, is involved in the production of succinate and acetyl-CoA. Differentially expressed genes involved in the biosynthesis of cobalamin, *cobD*, and porphyrin, *hemE* were found in cells grown in RDM in lag and log phases, respectively.
In both media, transport and binding proteins of the family ABC (ATP Binding Cassette) were predominant. These transporters are part of a special class of membrane proteins characterized by ATP binding and the presence of large multiple transmembrane domains. Several members of this family are active transporters that act in the modulation of cellular events such as uptake, metabolism, cellular effectivity and toxicity (Glavinas et al., 2004). In RDM, other symporters and antiporters responsible for the flow of ions that regulate intracellular pH and metallic ions that work as enzyme cofactors were detected. In RDM, $FeSO_4$ was present as iron source, while TY probably contained a complex mixture of iron sources.

Access	Gene (Genic product)	Expression	Medium
	Amino acid biosynthesis		
blr6298	probable asparagine synthetase [glutamine-hydrolyzing] (*asnB*)	0.65	***TY***
blr6299	probable asparagine synthetase [glutamine-hydrolyzing] (*asnB*)	-0.79	***RDM***
	Regulatory functions		
blr4281	transcriptional regulatory	1.14	
bll5328	two-component hybrid sensor and regulator	0.96	***TY***
blr5832	transcriptional regulatory protein MarR family	0.67	
blr7984	transcriptional regulatory protein TetR family	-0.62	***RDM***
	Translation		
blr3790	cysteinyl-tRNA synthetase (*cysS*)	-0.87	***RDM***
	Purines, pyrimidines, nucleosides, and nucleotides		
blr4125	5'-phosphoribosyl-5-aminoimidazole synthetase (*garS*)	-0.86	***RDM***
bll4102	rRNA-adenine N6,N6-dimethyltransferase (*ksgA*)	-0.96	
	DNA replication, recombination, and repair		
bll0572	putative DNA topoisomerase I	0.67	***TY***
bll0875	recombinase	-0.97	***RDM***
	Cellular processes		
blr0616	cell division protein (*ftsK*)	0.89	***TY***
	Biosynthesis of cofactors, prosthetic groups, and carriers		
blr3257	cobalamin biosynthesis protein(*cobD*)	-0.87	***RDM***
	Central intermediary metabolism		
blr6768	glycogen branching enzyme (*glgB*)	-0.99	
bll2065	carbonic anhydrase (*icfA*)	-0.85	***RDM***
bll7271	carbon monoxide dehydrogenase large chain	-1.11	
	Energy metabolism		
blr3728	cytochrome D ubiquinol oxidase subunit	1.17	***TY***
blr0150	cytochrome O ubiquinol oxidase subunit I (*cyoB*)	1.15	
bll3782	cytochrome C oxidase (*coxP*)	-0.85	***RDM***
	Fatty acid, phospholipid and sterol metabolism		
blr2947	probable acyl-CoA dehydrogenase	1.19	***TY***
bll6043	enoyl-CoA hydratase/isomerase family protein	0.91	
bll3855	putative acyl-CoA dehydrogenase	-0.67	***RDM***

Table 2. Genes showing significant values for lag phases

8. Central intermediary and energy metabolism

Cells grown in rich medium did not present marked expression of genes involved in carbon, nitrogen and energy metabolism. On the other hand, following the expression tendency of genes involved in the biosynthesis of units for macromolecule production, cells grown in defined medium showed a more pronounced expression of genes related to central intermediary and energy metabolism. Glycerol was the only carbon source used for cells

Access	Gene (Genic product)	Expression	Medium
	Amino acid biosynthesis		
bll5902	threonine dehydratase	1.26	***TY***
bll7862	putative homoserine O-acetyltransferase	-0.64	***RDM***
bll6037	probable SgaA serine-glyoxylate aminotransferase	-0.61	
	Regulatory functions		
bll3466	transcriptional regulatory protein Crp family	0.70	***TY***
blr3467	two-component response regulator	-0.65	
blr7678	transcriptional regulatory protein AraC family	-0.69	***RDM***
blr5548	transcriptional regulatory protein LysR family	-0.81	
	Translation		
blr0627	leucyl-tRNA synthetase (*leuS*)	-0.68	
blr1133	lysyl-tRNA synthetase (*lysS*)	-0.69	***RDM***
blr6589	aminopeptidase P	-0.70	
	Others		
bll0916	bifunctional uridylyltransferase/uridylyl-removing enzyme (*glnD*)	-0.62	***RDM***
	Purines, pyrimidines, nucleosides, and nucleotides		
bll4805	CTP synthase (*pyrG*)	0.66	***TY***
blr7256	putative adenylate cyclase	-0.93	***RDM***
	DNA replication, recombination, and repair		
blr5111	DNA topoisomerase I (*topA*)	-0.63	
bll4696	DNA gyrase subunit A (*gyrA*)	-0.67	***RDM***
bll5057	topoisomerase II (*gyrA*)	-1.14	
	Biosynthesis of cofactors, prosthetic groups, and carriers		
bll7942	probable 3-carboxy-cis,cis-muconate cycloisomerase (*pcaB*)	0.82	
bll7086	anaerobic coproporphyrinogen III oxidase (*hemN*)	0.79	***TY***
blr0430	nicotinate-nucleotide adenylyltransferase (*nadD*)	0.60	
blr6658	thiamine-phosphate pyrophosphorylase (*thiE*)	0.90	
bll2399	uroporphyrinogen decarboxylase (*hemE*)	-0.65	***RDM***
	Transport and binding proteins		
blr0971	ABC transporter ATP-binding protein	0.60	***TY***
bll0731	glycerol-3-phosphate ABC transporter membrane spanning protein	-0.63	
blr4039	ABC transporter substrate-binding protein	-0.64	***RDM***
blr4115	putative symporter	-0.67	
blr3904	probable iron transport protein	-0.69	
bll3739	probable Na+/H+ antiporter	-0.78	
	Central intermediary metabolism		
bll0199	probable nitrile hydratase regulator	-0.62	
blr2806	nitrite extrusion protein	-0.64	***RDM***
blr0725	nitrogen regulatory IIA protein (*ptsN*)	-0.76	
bsr1750	ferredoxin (*fer3*)	-0.85	

Access	Gene (Genic product)	Expression	Medium
	Energy metabolism		
bll8141	phosphoenolpyruvate carboxykinase (*pckA)*	-0.65	***RDM***
blr4655	Phosphoenolpyruvate synthase (*ppsA*)	-0.74	
blr3958	putative acetyl-coenzyme A synthetase	-1.00	
bll4782	pyruvate dehydrogenase beta subunit (*pdhB*)	-0.64	
	Fatty acid, phospholipid and sterol metabolism		
bll6363	putative acyl-CoA dehydrogenase	0.81	***TY***
bll4711	fatty acid CoA ligase	0.81	
blr1046	long-chain-fatty-acid-CoA ligase	-0.59	***RDM***

Table 3. Genes showing significant values for log phases.

grown in defined medium, and it is used in a very efficient way. Metabolism occurs via glycerol kinase and glycerol phosphate dehydrogenase, producing glyceraldehyde-3-phosphate, which is then converted into pyruvate (Arias & Martinez, 1976). During the log phase in RDM, specifically, a gene bll0731 related to an ABC transporter for glycerol-3-phosphate was detected with higher expression level. Genes involved in pyruvate and acetyl-CoA metabolism were also differentially expressed in RDM.

In a rich medium, carbohydrate metabolism through glycolysis happens through the Entner-Doudoroff pathway with a simultaneous operation of the Embden-Meyerhof-Parnas pathway, although fructose-1,6-biphosphato aldolase is present in reduced levels and the 6-phosphogluconate dehydrogenase ($NADP^+$) is absent. However, a 6-phosphogluconate dehyidrogenase (NAD) is present in *Bradyrhizobium*, suggesting the operation of a new pathway (Kuykendall, 2005). So far, none of these enzymes was found in *B. elkanii* genome.

Still in RDM, the carbohydrate limitation may have induced the expression of genes involved in the production of energetic metabolites through lipid metabolism. In this class, some genes related to the biosynthesis of coenzyme A present in the category of fatty acid, phospholipid and sterol metabolism were detected in defined medium. On the other hand, in TY, the gene *zwf* (glucose-6-phosphate 1-dehydrogenase), which is involved in glycolysis, was detected during the stationary phase. This gene may probably have been induced in this phase for the use of poly-β- hydroxybutyrate (PHB). PHB is a carbon-storage polymer, and is production begins with the condensation of two acetyl-CoA molecules, synthesis of acetoacetyl-Coa and the reduction to produce D-β-hydroxybutyril-Coa, which is finally incorporated to a PHB molecule (Tombolini et al., 1995).

During the lag phase, the presence of the cytochromes oxidase *coxP*, detected in RDM, and *cyoB* and bll3728 in TY medium suggest that aerobic respiration increased in these culture conditions. However, since the only source of aeration of the liquid culture of both media was orbital agitation, it is likely that with time and the population density increase, a microaerobic environment was created. This oxygen limitation may have simulated a symbiotic condition inducing the expression of some symbiotic and N fixation-related genes (Table 5) at that point and upcoming phases. A symbiotic cytochrome-c oxidase, *fixN*, was detected in RDM during the log phase (Table 5). *fixNOQP* genes encode a cytochrome-c oxidase specifically required for bacteroid respiration (Fischer, 1994). This idea corroborates with the detection of genes encoding superoxide dismutase and catalase in TY medium (Table 5) in lag and stationary phases, respectively. However, these cellular detoxification genes may just be acting in the process of tolerance in the adaptation and stress period in each phase, lag and stationary.

Access	Gene (Genic product)	Expression	Medium
	Amino acid biosynthesis		
blr0650	imidazoleglycerol-phosphate dehydratase(*hisB*)	0.94	*TY*
	Regulatory functions		
bll3466	transcriptional regulatory protein Crp family	1.41	*TY*
blr1163	transcriptional regulatory protein MarR family	1.37	
blr1096	phosphate regulon, two-component response regulator (*phoB*)	1.00	
bll0904	two-component response regulator (*regR*)	0.96	
bll2814	transcriptional regulatory protein AsnC family	0.89	
blr8099	GTP-binding protein	0.78	
	Translation		
bll5087	glutamyl-tRNA (amidotransferase)	0.95	*TY*
blr1133	lysyl-tRNA synthetase (*lysS*)	0.89	
blr3130	serine protease DO-like precursor	0.88	
bll5368	serine protease	-0.69	*RDM*
	Purines, pyrimidines, nucleosides, and nucleotides		
blr7737	putative adenylate cyclase	0.92	*TY*
blr7371	carbamoylphosphate synthase small chain (*carA)*	1.07	
bll4805	CTP synthase(*pyrG*)	1.06	
	DNA replication, recombination, and repair		
bll0572	putative DNA topoisomerase I	0.93	*TY*
	Biosynthesis of cofactors, prosthetic groups, and carriers		
bll7086	anaerbic coproporphyrinogen III oxidase (*hemN*)	1.68	*TY*
blr0430	nicotinate-nucleotide adenylyltransferase (*nadD*)	1.21	
blr2404	probable gamma-glutamyltranspeptidase precursor (*ggt*)	0.77	
	Transport and binding proteins		
blr3568	ABC transporter permease protein	1.24	*TY*
bll6293	HlyB/MsbA family ABC transporter	1.08	
bll1193	integral inner membrane metabolite transport protein (*mtbA*)	0.85	
bll7341	RhtB family transporter (*rhtB*)	-0.56	*RDM*
	Energy metabolism		
blr6760	glucose-6-phosphate 1-dehydrogenase (*zwf*)	1.10	*TY*
bll0322	probable trehalose-6-phosphate synthase (*otsA*)	1.05	
bll0323	probable trehalose-phosphatase (*otsB*)	0.95	
bll5920	UDP-glucuronic acid epimerase (*lspL*)	0.93	
bll2210	multicopper oxidase (*copA*)	1.13	
	Fatty acid, phospholipid and sterol metabolism		
blr3955	acyl-CoA dehydrogenase	1.26	*TY*
blr0981	acyl-CoA dehydrogenase	1.23	
bll2948	putative long-chain-fatty-acid--CoA ligase	0.79	
blr0139	putative acyl-CoA dehydrogenase	-0.54	*RDM*

Table 4. Genes showing significant values for stationary phases.

Access	Phase	Gene (Genic product)	Expression	Medium
		Nitrogen fixation		
blr7496	***lag***	nitrogen fixation protein	0.89	**TY**
blr1770	***log***	Molybdenum processing protein (*nifQ*)	0.74	
blr2763		cytochrome-c oxidase (*fixN*)	-0.67	**RDM**
blr1756	***estat***	nitrogenase metalloclusters biosynthesis protein (*nifS*)	0.91	**TY**
blr1756		nitrogenase metalloclusters biosynthesis protein(*nifS*)	-0.63	**RDM**
		Hydrogenase		
bll6933	***estat***	HupK protein (*hupK*)	0.88	**TY**
		Detoxification		
bll7774	***lag***	superoxide dismutase	1.13	**TY**
blr0778	***estat***	catalase	1.14	
		Symbiosis		
blr2025	***lag***	acyl transferase (*nodA*)	1.00	**TY**
bll1631	***log***	GDP-mannose 4,6-dehydratase (*noeL)*	-0.68	**RDM**
blr3959		nodulation protein N (*nodN)*	-1.00	

Table 5. Genes related to symbiosis and nitrogen fixation showing significant expression ratios

An unexpected result was the expression of a *hupK* gene (Table 5) in TY medium during the stationary phase, because *B. elkanii* is classified as Hup^-, differently from *B. japonicum*, which is Hup^+ (Minamisawa, 1989). This gene is part of an operon that encodes the hydrogenase system. This hydrogen uptake system oxidizes the hydrogen produced by nitrogenase and this H_2 recycling reduces energy losses (Evans et al., 1988).

9. Conclusion

The roles of genes analyzed in this work were determined through their homology with proteins which sequences are stored in data banks. A number of genes was classified in the 'hypothetical' category, because no homology was found with known genes or roles. Data reflect the expression rates of the relative transcript levels of individual genes, with no indication of regulation mechanisms.

Cells grown in a richer medium (TY), with higher availability of carbon and energy sources, grew faster and expressed a group of genes involved in cellular processes, regulation roles and energetic metabolism-related pathways. Cells grown in defined medium (RDM) face the need of synthesizing all their building blocks (monomers) from a single carbon and energy source. This process reflected not only in the activation of biosynthesis pathways, but also in the high expression of cellular process regulators, central intermediary and energetic metabolism and the metabolism of fatty acids.

Among the genes with significant rates, *fixN* was selected for Real-Time PCR validation. The relative expression result validated DNA microarray results, demonstrating the analyses and differential expression rates are trustworthy for genes expressed in the media RDM and TY. The apparent alteration of an aerobic respiratory metabolism to anaerobic seems to be one of the most peculiar results. The respiratory chain in *B. japonicum* has several branches,

synthesizing a number of terminal oxidases appropriate for different environment conditions.

The induction of genes involved in symbiosis an N fixation during *Bradyrhizobium elkanii* culture must be considered to evaluate commercial inoculants preparations. The simulation of an environment with low oxygen tension may be taking place in commercial packages during the storage of the liquid product. Thus, the same process inducing these genes observed in our cultures might happen in liquid inoculants formulas. Two hypotheses are suggested for these conditions, and both must be experimentally analyzed and studied. First, cells in the inoculants contain a group of genes readily activated for symbiosis and N fixation at the moment of inoculation, which is favorable for an interaction with plants. Second, when bacterial cells are released from the package to be inoculated in plants, they undergo a new series of gene expression induction, retarding their interaction with plants and therefore the symbiosis process itself. Although the first hypothesis is very attractive, the second seems to be more adequate.

Finally, facing the new technologies and platforms for DNA sequencing that allows the analysis of gene expression rates, DNA microarrays still seem an attractive and important option once the same array has the advantage of being applied to different biological situations not always covered in sequencing experiments.

Probably, the DNA microarray technique will remain a viable alternative for many years primarily to search for genes related to environmental and metagenomics analysis.

10. Acknowledgements

The authors are thankful to 'Fundação de Amparo à Pesquisa do Estado of São Paulo' (FAPESP) for the financial support and the fellowship granted to J. M and also to Prof. Dr. Antônio S. Ferraudo for help with statistical analysis.

11. References

Arias, A.; Martinez-de-Drets, G. (1976). Glycerol metabolism in *Rhizobium*. *Canadian Journal of Microbiology*, v. 66, pp. 150-153, ISSN: 1480-3275.

Beringer, J. E. (1974) R factor transfer in *Rhizobium leguminosarum*. *Journal of General Microbiology*, v. 84, pp. 188-198, ISSN: 0022-1287.

Brooun, A.; Tomashek, J.J.; Lewis, K. (1999). Purification and ligand binding of EmrR, a Regulator of multidrug transporter. *The Journal of Bacteriology*, v. 181, pp. 5131-5133, ISSN: 0021-9193.

Dharmadi, Y.; Gonzalez, R. (2004). DNA microarrays: Experimental issues, data analysis, and apllication to bacterial systems. *Biotechnology Progress,* v. 20, pp. 1309-1324, ISSN: 8756-7938.

Evans, H.J.; Russell, S.A.; Hanus, F.J.; Ruiz-Argéso, T. (1988). The importance of hydrogen recycling in nitrogen fixation by legumes. *In: World crops: cool season food legumes,* Summerfield, J.R. (Ed), pp. 777-791, Kluwer Academic Publishers, ISBN: 90-247-3641-2:501–511, Boston, MA.

Fischer, H.M. (1994). Genetic regulation of nitrogen fixation in rhizobia. *Microbiological Reviews*, v. 58, pp. 352-386, ISSN:0146-0749.

Gaal, T.; Bartlett, M.S.; Ross, W.; Turnbough Jr, C.L.; Gourse, R.L. (1997). Transcription regulation by initiating NTP concentration: rRNA syntesis in bacteria. *Science*, v. 278, pp. 2092-2097, ISSN: 0036-8075.

Glavinas, H.; Krajcsi, P.; Cserepes, J.; Sarkadi, B. (2004). The role of ABC transporters in drug resistance, metabolism and toxicity. *Current Drug Delivery*, v. 1, pp. 27-42, ISSN: 1567-2018.

Hauser, F.; Lindemann, A.; Vuilleumier, S.; Patrignani, A.; Schlapbach, R.; Fischer, H.M.; Hennecke, H. (2006). Design and validation of a partial-genome microarray for transcriptional profiling of the *Bradyrhizobium japonicum* symbiotic gene region. *Molecular Genetics and Genomics*, v. 275, pp. 55-67, ISSN: 1617-4615.

Jordan, D. (1982). Transfer of *Rhizobium japonicum* Buchanan 1980 to *Bradyrhizobium* gen. nov., a genus of slow growing root-nodule bacteria from leguminous plants. *International Journal of Systematic Bacteriology*, v. 32, pp. 136-139, ISSN: 0020-7713.

Kaneko, T.; Nakamura, Y.; Sato, S.; Minamisawa, K.; Uchiumi, T.; Sasamoto, S.; Watanabe, A.; Idesawa, K.; Iriguchi, M.; Kawashima, K.; Kohara, M.; Matsumoto, M.; Shimpo, S.; Tsuruoka, H.; Wada, T.; Yamada, M.; Tabata, S. (2002). Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA research*, v. 9, pp. 189-197, ISSN, 1340-2838.

Keener, J.; Nomura, M. (1996). Regulation of ribosome synthesis. In: *Escherichia coli and Salmonella: cellular and molecular biology*, Neidbardt, F.C.; Curtiss III, R.; Ingraham, J.L.; Lin, E.C.C.; Low, K.B.; Magasanik, B.; Reznikoff, W.S.; Riley, M.; Schacchter, M.; Umbarger, H.E. (Eds), pp. 1417-1431, ASM Press, ISBN: 1555811647, Washington, DC.

Kuykendal, L.D. (2005). Genus I. *Bradyrhizobium* Jordan 1982, 137^{VP}. In: *Bergey's Manual of Systematic Bacteriology, The Proteobacteria, part C (The Alpha-, Beta-, Delta-, and Epsilonproteobacteria),* Brenner, D.J., Krieg, N.R.; Staley, J.T.; Garrity, G.M. (Eds), pp. 438-442, Springer, ISBN: 978-0-387-24145-6, New York, NY.

Kuykendall, L.D.; Saxena, B.; Devine, T.E.; Udell, S.E. (1992). Genetic diversity in *Bradyrhizobium japonicum* Jordan 1982 and a proposal for *Bradyrhizobium elkanii* sp. *nov*. *Canadian Journal of Microbiology,* v. 38, pp. 501-505, ISSN: 1480-3275.

Loh, J.; Stacey, G. (2003). Nodulation gene regulation in *Bradyrhizobium japonicum*: a unique integration of global regulatory circuits. *Applied and Environmental Microbiology*, v. 69, pp. 10-17, ISSN: 0099-2240.

Loh, J.T.; Yuen-Tsai, J.P.Y.; Stacey, M.G.; Lohar, D.; Welborn, A.; Stacey, G. (2001). Population density-dependent regulation of the *Bradyrhizobium japonicum* nodulation genes. *Molecular Microbiology*, v. 42, pp. 37-46, ISSN: 0950-382X.

López-Garcia, S.L.; Vázquez, T.E.E.; Favelukes, G.; Lodeiro, A.R. (2002). Rhizobial position as a main determinant in the problem of competition for nodulation in soybean. *Environmental Microbiology*, v. 4, pp. 216-224, ISSN: 14622912.

Minamisawa, K. (1989). Comparison of extracellular polysaccharide composition, rhizobitoxine production and hydrogenase phenotype among various strains of *Bradyrhizobium japonicum*. *Plant and Cell Physiology,* v. 30, pp. 877-884, ISSN: 0032-0781.

Oke, V. &Long, S.R. (1999). Bacteroid formation in the *Rhizobium*-legume symbiosis. *Current Opinion in Microbiology*, v. 2, pp. 641-646, ISSN: 1369-5274.

Pao, G.M.; Saier Jr, M.H. (1995). Response regulators of bacterial signal transduction systems: selective domain shuffling during evolution. *Journal of Molecular Evolution,* v. 40, pp. 136-154, ISSN: 0022-2844.

Quackenbush, J. (2002). Microarray data normalization and transformation. *Nature Genetics*, v. 32, pp. 496-501, ISSN: 1061-4036.

Rumjanek, N.G.; Dobert, R.C.; Berkum, P.; Triplett, E.W. (1993). Common soybean inoculant strains in Brazil are members of *Brayrhizobium elkanii*. *Applied and Environmental Microbiology,* v. 59, pp. 4371-4373, ISSN: 0099-2240.

Shatters, R.G.; Somerville, J.E.; Kahn, M.L. (1989). Regulation of glutamine synthetase II activity in 104A14. *The Journal of Bacteriology*, v. 171, pp. 5087-5094, ISSN: 0021-9193.

Stock, A.M.; Robinson, V.L.; Goudreau, P.N. (2000). Two-component signal transduction. *Annual Review of Biochemistry*, *v.* 69, pp. 183-215, ISSN: 0066-4154.

Tao, H.; Bausch, C.; Richmond, C.; Blattner, F.R.; Conway, T. (1999). Functional genomics: Analysis of *Escherichia coli* growing o minimal and rich media. *The Journal of Bacteriology*, v. 181, pp. 6425-6440, ISSN: 0021-9193.

Tombolini, R.; Buson, S.; Squartini, A.; Nuti, M.P. (1995). Poly-beta-hydroxybutyrate (PHB) biosynthetic genes in *Rhizobium meliloti*-41. *Microbiology*, v. 141, pp. 2553-2559, ISSN: 1350-0872.

Tusher, V.; Tibshirani, R.; Chu, G. (2001). Significance analysis of microarrays applied to transcriptional responses to ionizing radiation. *Proceedings of the National Academy of Sciences, USA*, v. 98, pp. 5116-5121, ISSN: 0027-8424.

Vincent, J.M. (1970). *A manual for the practical study of root-nodule bacteria,* pp. 164, Blackwell Scientific, ISBN: 0632064102, Oxford, OXF.

Yu, X.; Tran, A.H.; Sun, Q.; Margolin, W. (1998). Localization of cell division protein FtsK to the *Escherichia coli* septum and identification of a potential N-terminal targetint domain. *The Journal of Bacteriology*, v. 180, pp. 1296-1304, ISSN: 0021-9193.

Zechiedrich, E.L.; Khodursky, A.B.; Bachellier, S.; Schneider, R.; Chen, D.; Lilley, D.M.J.; Cozzarelli, N.R. (2000). Roles of topoisomerases in maintaining steady-state DNA supercoiling in *Escherichia coli*. *The Journal of Biological Chemistry*, v. 275, pp. 8103-8113, ISSN 1083-351X.

Knowledge-Oriented Applications in Data Mining
Edited by Prof. Kimito Funatsu

ISBN 978-953-307-154-1
Hard cover, 442 pages
Publisher InTech
Published online 21, January, 2011
Published in print edition January, 2011

The progress of data mining technology and large public popularity establish a need for a comprehensive text on the subject. The series of books entitled by 'Data Mining' address the need by presenting in-depth description of novel mining algorithms and many useful applications. In addition to understanding each section deeply, the two books present useful hints and strategies to solving problems in the following chapters. The contributing authors have highlighted many future research directions that will foster multi-disciplinary collaborations and hence will lead to significant development in the field of data mining.

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821