Journal of Plant Biology, August 2006, 49(4) : 271-275

Taxonomic Discrimination of Cyanobacteria by Metabolic Fingerprinting Using Proton Nuclear Magnetic Resonance Spectra and Multivariate Statistical Analysis

Suk Weon Kim[1], Sung Hee Ban[1], Chi Yong Ahn[2], Hee Mock Oh[1, 2], Hoeil Chung[3], Soo Hwa Cho[3], Young Mok Park[5], and Jang Ryol Liu[4]*

[1]Biological Resources Center, [2]Environmental Biology Center,
[3]Department of Chemistry, Hanyang University, Seoul 133-791, Korea
[4]Plant Genomic Research Center, KRIBB, Daejeon 305-333, Korea
[5]Proteome Analysis Team, Korea Basic Science Institute, Daejeon 305-333, Korea

When whole-cell extracts are analyzed, proton nuclear magnetic resonance (^{1}H NMR) spectroscopy provides biochemical profiles that contain overlapping signals of the majority of the compounds. To determine whether cyanobacteria could be taxonomically discriminated on the basis of metabolic fingerprinting, we subjected whole-cell extracts of the cyanobacteria to ^{1}H NMR. The ^{1}H NMR spectra revealed a predominance of signals in the aliphatic region. Principal component analysis (PCA) of the data then enabled discrimination of the cyanobacteria. The hierarchical dendrogram, based on PCA of the aliphatic region data, showed that six cyanobacterial taxa were discriminated from two eukaryotic microalgal species, and that the six taxa could be subsequently divided into three groups. This agrees with the current taxonomy of cyanobacteria. Therefore, our overall results indicate that metabolic fingerprinting using ^{1}H NMR spectra and multivariate statistical analysis provide a simple, rapid method for the taxonomical discrimination of cyanobacteria.

Keywords: cyanobacteria, dendrogram, pattern recognition, principal component analysis, taxonomy

Cyanobacteria (blue-green algae) comprise a large and diverse group of organisms in the prokaryotic kingdom. They are widely distributed in freshwater, marine, and terrestrial habitats (Brock et al., 1994). As important ecosystem components, cyanobacteria constitute more than 30% of the total planetary biomass (Beardall et al., 2002), and are, therefore, a major source of food and energy.

Cyanobacterial identification and classification have relied upon morphological characters, chemical markers such as lipids, polyamines, and carotenoids, and macromolecular methods, such as protein electrophoresis patterns, phycobiliprotein patterns, immunological studies, and 16S rDNA sequence analysis (Wilmotte and Golubic, 1991; Henson et al., 2004; Song et al., 2004).

Since DNA sequencing has become readily available, the tendency has been to determine their classification based solely on 16S rDNA data, which is considered to reflect phylogenetic classification. However, such classification is now questionable because of recent studies demonstrating the widespread occurrence of horizontal gene transfer in bacterial genomes (Ragan, 2001). Therefore, any phylogenetic classification based on DNA sequence analysis should be confirmed by phenotypic studies that include comparisons of cell wall composition, fatty acid and protein profiling, and isoprenoid quinone.

Whole-cell extracts subjected to ^{1}H NMR (proton nuclear magnetic resonance) analysis provide an overview of hydrogen-containing compounds. Consequently, an NMR spectrum is more representative of the physiology of the cell (metabolite pools) than of the structure (comprising immobile components such as the cell wall). Therefore, NMR-based metabolic fingerprinting (pattern recognition of metabolic fingerprinting data using multivariate statistical analysis) has been used to discover new metabolic biomarkers from urine (Gavaghan et al., 2000) and to screen for drug toxicity (Robertson et al., 2000; Nicholson et al., 2002). Likewise, this ^{1}H NMR-based metabolic fingerprinting method has enabled researchers to discriminate various ecotypes of *Arabidopsis thaliana*, and has revealed the possible components that contribute most to the discrimination of those ecotypes (Ward et al., 2003).

Current phenotypic studies for bacterial classification, e.g., comparisons of cell wall composition, fatty acid and protein profiling, and isoprenoid quinone are time-consuming and tedious. The objective of our study was to determine whether metabolic fingerprinting of cyanobacteria is an available alternative to those conventional approaches. Here, we attempted to demonstrate that ^{1}H NMR-based metabolic fingerprinting of cyanobacterial whole-cell extracts could discriminate between cyanobacteria, thereby providing quantitative data for their classification. Using pattern recognition methods, Fourier transform infrared spectroscopy (FTIR) has previously been applied to distinguish closely related microbial strains (Helm et al., 1991; Naumann et al., 1991; Freeman et al., 1994; Goodacre et al., 1998), cyanobacteria (Kansiz et al., 1999), and flowering plants (Kim et al., 2004).

MATERIALS AND METHODS

Cyanobacteria

Six cyanobacteria (*Microcystis aeruginosa* UTEX2388, *M. aeruginosa* PCC7806, *Anabaena* sp. PCC7120, *Synechocystis* sp. PCC6803, *Synechoccus lividus* PCC6715, and *Oscillatoria planctonica*), plus two eukaryotic microalgae (*Navicula* sp. and *Scenedesmus* sp.), were used in this study (Table 1). Their cultures were not axenic, although the levels of contaminants were low (<1% by microscopic examination) and

*Corresponding author; fax +82-42-860-4608
e-mail jrliu@kribb.re.kr

Table 1. Eight cyanobacteria used in the experiment.

Number	Family	Taxa	Superkingdom
1	Chroococcaceae	*Microcystis aeruginosa* UTEX2388	Prokaryote
2	Chroococcaceae	*Microcystis aeruginosa* PCC7806	
3	Nostocaceae	*Anabaena* sp. PCC7120	
4	Chroococcaceae	*Synechocystis* sp. PCC6803	
5	Chroococcaceae	*Synechococcus lividus* PCC6715	
6	Oscillatoriaceae	*Oscillatoria planctonica*	
7	Naviculaceae	*Navicula* sp.	Eukaryote
8	Scenedesmaceae	*Scenedesmusp* sp.	

would not be expected to have a significant effect on the spectra. Each cyanobacterial strain was transferred to a liquid medium and cultured on a gyratory shaking incubator (150 rpm) under a cool-white fluorescent lamp (25°C, approximately 30 mol $m^{-2} s^{-1}$). At their exponential and stationary growth phases, the strains were collected by centrifugation (3000 rpm for 10 min), then rinsed twice with sterile distilled water before being frozen in liquid nitrogen container. Frozen cyanobacterial strains were freeze-dried and stored at -70°C until use.

^{1}H NMR Analysis

Freeze-dried cells (15 mg) were placed in sterile 2-mL microvial tubes equipped with screw-caps and neoprene o-rings. One mL of D_2O:CD_3OD (v:v, 80:20) containing 0.005% (w/v) TSP-d_4 (sodium salt of trimethylsilylpropionic acid) was added to each sample. The cells were homogenized six times using a mini bead beater (zirconia silica bead: 0.1 mm in diameter; 4600 rpm for 30 s). These cell mixtures were blended thoroughly and heated at 50°C in a water bath for 10 min. After cooling, the samples were centrifuged at 13,000 rpm for 5 min. The supernatant (700 μL) was transferred to a 5-mm NMR tubes. Each NMR sample was stored at 4°C until analysis. All NMR spectra were collected on a Varian UNIT 500 NMR spectrometer (VARIAN, USA) at 298°K; a pre-saturation pulse was used for water suppression. Each spectrum consisted of 60 scans of 32 k data points, and the ^{1}H NMR chemical shifts in the spectra were referenced to TSP-d_4 at 0.00 ppm. Whole-peak intensities in every 0.005 ppm of the ^{1}H NMR spectra (range of 0.5~12.5 ppm) were used as the variables, and ^{1}H NMR spectra were collected from three different runs.

Data Normalization and Statistical Analysis

PCA and hierarchical analysis of these NMR data were conducted using Matlab (vers. 6.5). The NMR spectra were automatically converted to ASCII files via ADVASP Lite (short version of ADVASP package; Umatek, USA). Spectral intensities were scaled to TSP-d_4 and integrated regions of equal width corresponding to the region of 12.0 to 0 ppm. The residual proton signal corresponding to TSP-d_4 (δ 0.0) was removed at this stage. The generated ASCII file was imported into Matlab (vers. 6.5) for PCA analysis according to NIPALS algorithm (Wold, 1966). ^{1}H NMR spectra for the total metabolite (0 to 12 ppm), carbohydrate (3 to 6 ppm), aliphatic (1 to 3 ppm), and aromatic (6 to 9 ppm) regions from each spectrum were divided, and the resulting reduced data set was then normalized to unit variance. Following this processing, PCA and hierarchical clustering analysis were conducted separately for those four regions. A hierarchical dendrogram was generated to display the relationship among cultures from PCA of the ^{1}H NMR spectral data using the unweighted-pair group method and arithmetic mean (UPGMA) analysis with a multivariate statistical package (MVSP 3.13; Kovach Computing Services, UK).

RESULTS AND DISCUSSION

Representative ^{1}H NMR spectra of the total regions from six cyanobacterial cultures and two eukaryotic microalgal cultures are presented in Figure 1. The spectra obtained from the cyanobacteria showed a preponderance of signals comprising several sharp peaks in the aliphatic region. To investigate the spectral region representing the most relevant taxonomic relationship among cyanobacteria, our NMR spectral data were analyzed in four different ways: total (0 to 12 ppm), carbohydrate (3 to 6 ppm), aliphatic (1 to 3 ppm), and aromatic regions (6 to 9 ppm). Our displays included a PCA plot of spectral data for the aliphatic region from the exponential growth phase (Fig. 2A) and the stationary phase (Fig. 2C). The PCA plot of the aliphatic region from the former phase (Fig. 2A) revealed that the prokaryotic taxa were clustered in a group whereas the two eukaryotic taxa were located at the opposite extreme. However, a PCA plot of the aliphatic region from the stationary phase (Fig. 2C)

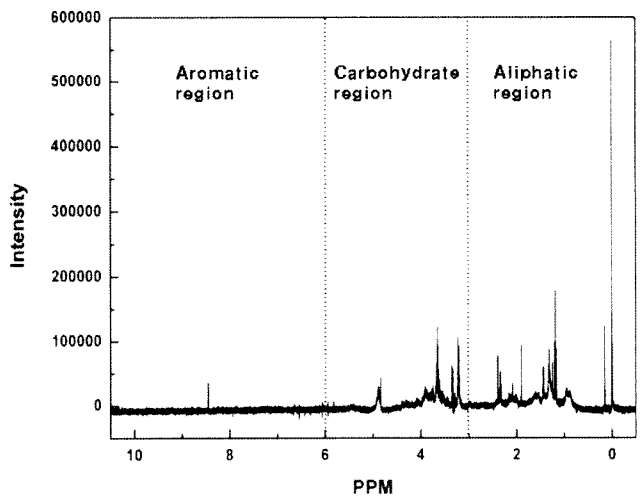

Figure 1. Representative ^{1}H NMR spectra of *Scenedesmus* sp. at exponential growth phase.

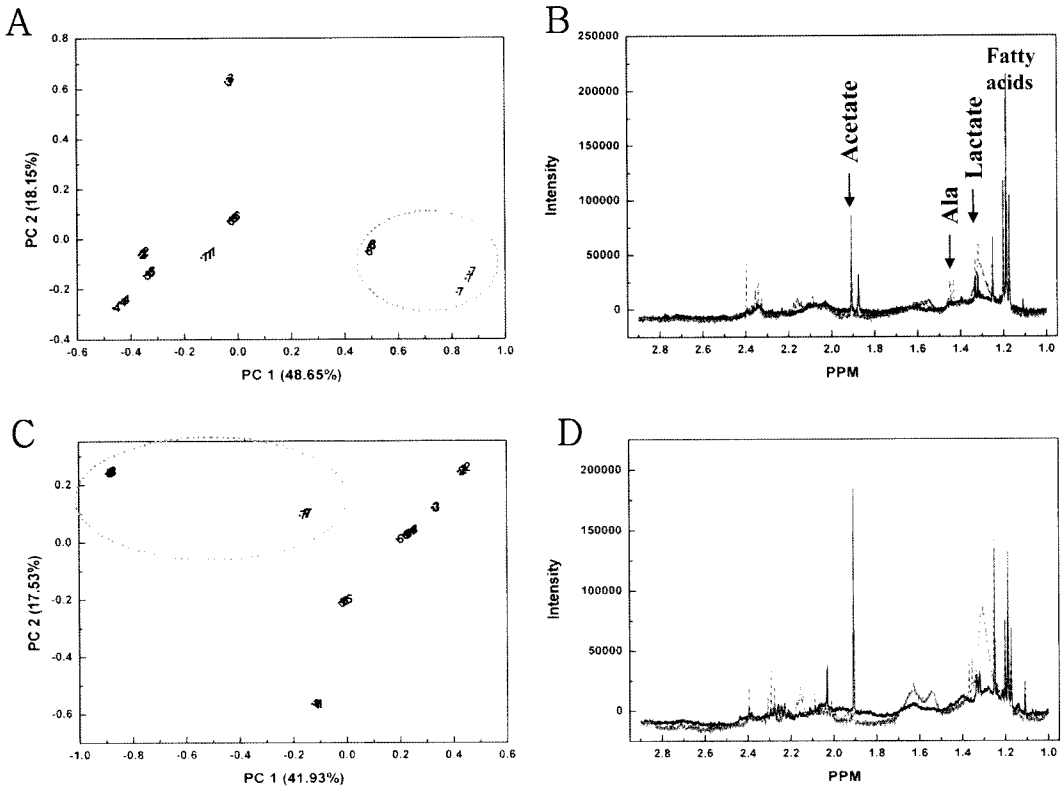

Figure 2. Two-dimensional PCA plot and enlarged view of NMR aliphatic region spectrum from eight cyanobacteria. **A**, PCA plot for exponential growth phase; **B**, Enlarged view of NMR aliphatic region spectrum at exponential growth phase; **C**, PCA plot for stationary phase; **D**, Enlarged view of NMR aliphatic region spectrum at stationary phase. The first two principal components in each plot are displayed. Percentages in parentheses account for partition of each principal component of total variation. Averaged spectra of prokaryotic taxa (solid line) and eukaryotic taxa (dashed line) are presented. Each number in PCA plot indicates individual cyanobacterium: 1, *M. aeruginosa* UTEX2388; 2, *M. aeruginosa* PCC7806; 3, *Anabaena* sp. PCC7120; 4, *Synechocystis* sp. PCC6803; 5, *S. lividus* PCC6715; 6, *O. planctonica*; 7, *Navicula* sp; and 8, *Scenedesmus* sp. Duplicated numbers indicate three replicates for each sample. Arrows show peaks assigned to compounds from NMR spectrum database.

did not show a discrete discrimination pattern. Replicated samples of each cyanobacterial culture were also grouped into discrete clusters, indicating that the PCA was able to discern the cyanobacteria of one taxon from another.

Figure 2B presents an enlarged view of the aliphatic region from our ^{1}H NMR spectra. These data revealed peaks as the major contributor to the discrimination of cyanobacteria between prokaryotic and eukaryotic cultures, with two major peaks being found at 2.396 ppm and 1.3 ppm. Relatively clear peaks in the aliphatic region could be assigned to particular amino acids (e.g., acetate at 1.907 ppm; alanine doublet at 1.47 ppm; lactate at 1.317 ppm). In addition, the ^{1}H NMR aliphatic-region data gave evidence of metabolic changes during the exponential growth phase against the stationary phase (Fig. 2D). Because of the limit of the deconvolution of ^{1}H NMR spectra by reference, we were unable to fully assign these peaks to specific compounds. However, this region contained many peaks attributable to amino acids, thereby providing us with possible clues as to the variance of free amino acids between prokaryotic and eukaryotic cyanobacteria.

Cyanobacteria at the stationary phase also were subjected to NMR, using metabolic fingerprinting. However, dendrograms based on PCA scores at that phase did not agree with our results for the exponential growth phase (data not shown). A similar approach involving FTIR has been taken with bacterial cells collected from the exponential to the stationary phase (Helm et al., 1991; Naumann et al., 1991; Curk et al., 1994; Goodacre et al., 1996).

Dendrograms were constructed based on PCA scores of the NMR total (Fig. 3A), aliphatic (Fig. 3B), carbohydrate (Fig. 3C), and aromatic regions (Fig. 3D). There, most of the cultures could be divided into two major groups -- cyanobacterial and eukaryotic cultures. However, PCA scores for the aromatic region failed to discriminate between these two types, perhaps because of the low level of signals in that region.

Cyanobacteria are currently divided into five subsections, based on their morphological and developmental features (Rippka and Herdman, 1992). In general, molecular phylogenies support the monophyly of subsections II, IV and V, whereas subsections I and III are not considered monophyletic (Turner, 1997). Our dendrogram based on total regions was able to group *Oscillatoria*, belonging to subsection III, with *Synechococcus*, *Synechocystis*, and *Microcystis*, which belong to subsection I. In contrast, dendrograms based on the aliphatic and carbohydrate regions could be divided into three distinct groups, a conclusion that agrees with the current taxonomy of cyanobacteria. However, the dendrogram based on the carbohydrate region failed to group two strains of *Microcystis* at the highest degree of similarity. The dendrogram based on the aliphatic region showed that the cyanobacterial cultures were further divided into *Anabaena* and the other five cultures. Furthermore, *O. planctonica* was divided into a cluster separate from the other four. Among those, *Synechococcus* and *Synechocystis* could be clustered

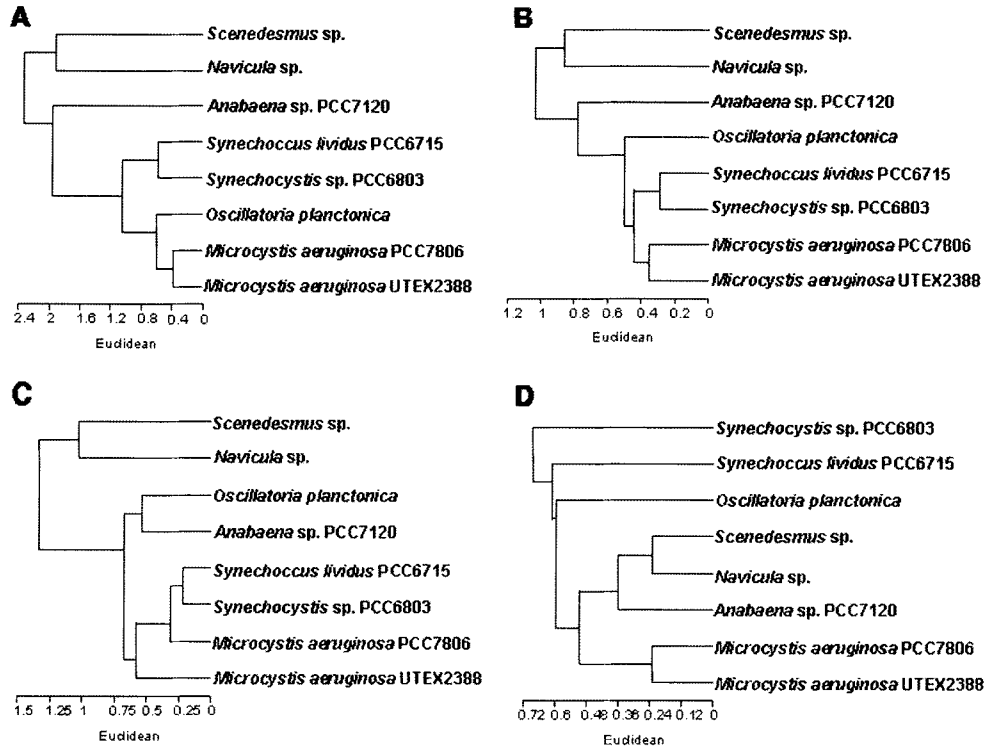

Figure 3. Hierarchical dendrograms based on PCA of NMR data from six cyanobacterial and two eukaryotic microalgal cultures. **A**, Dendrogram based on PCA of NMR total region; **B**, Aliphatic region; **C**, Carbohydrate region; **D**, Aromatic region.

together while the two strains of *M. aeruginosa* were clustered individually. In that dendrogram, *Synechococcus* and *Synechocystis* exhibited the highest degree of similarity, followed by the two strains of *M. aeruginosa*.

Kansiz et al. (1999) have used metabolic fingerprinting of FTIR spectroscopy to discriminate five taxa of cyanobacteria. However, their dendrogram does not support the current taxonomy, i.e., three strains that belong to subsection I are clustered into two different groups with strains belonging to subsections III and IV. However, in our study, the dendrogram of six taxa were based on metabolic fingerprinting using ^{1}H NMR data in the aliphatic region spectra. Our results were in agreement with the current taxonomy, i.e., those six taxa could be divided into three groups, thereby indicating that metabolic fingerprinting with ^{1}H NMR spectra and multivariate statistical analysis provides a simple, rapid method for the taxonomical discrimination of cyanobacteria. Therefore, this method can be used as an alternative to time-consuming, tedious phenotypic studies of cyanobacteria, such as comparisons of cell wall composition, fatty acid and protein profiling, and isoprenoid quinone.

Recent studies have demonstrated that horizontal gene transfer occurs between unrelated bacterial taxa via recombination or transduction events. Most strikingly, Sneath (1993) has suggested that parts of the 16S rRNA sequences are transferred between genomic species in *Aeromonas*. Furthermore, the highly conserved state of 16S rRNA makes it a prime candidate as a vector for gene transfer (Doolittle, 1999; Ragan, 2001). Because evidence for its recombination (Wang et al., 1997; Yap et al., 1999) is especially compelling, one must be cautious in making monophyletic interpretations, especially those based on the analysis of single sequences. Other multiple-character tests should support any DNA sequence-based phylogenetic classification. Therefore, we conclude that our metabolic fingerprinting method, using ^{1}H NMR, can be used to justify the phylogenetic classification of cyanobacteria, especially when based solely on 16S rDNA sequence analysis.

ACKNOWLEDGEMENTS

This work was supported by a grant to YMP from the Korea Research Council of Fundamental Science and Technology; and by grants to JRL from the Marine Bio 21 Program, the Korea Science and Engineering Foundation through the Plant Metabolism Research Center of the Kyung Hee University, and the Crop Functional Genomics Center of the 21st Century Frontier Research Program, funded by the Korean Ministry of Science and Technology.

Received April 4, 2006; accepted May 25, 2006.

LITERATURE CITED

Beardall J, Heraud P, Roberts S, Shelley K, Stojkovic S (2002) Effects of UV-B radiation on inorganic carbon acquisition by the marine microalga *Dunaliella tertiolecta* (Chlorophyceae). Phycologia 41: 268-272

Brock TD, Madigan MT, Martinko JM, Parker J (1994) Biology of Microorganisms. Ed 7, Prentice Hall, Englewood Cliffs

Curk MC, Peladan F, Hubert JC (1994) Fourier transform infrared (FTIR) spectroscopy for identifying *Lactobacillus* species. FEMS Microbiol Lett 123: 241-248

Doolittle WF (1999) Phylogenetic classification and the universal tree. Science 284: 2124-2129

Freeman R, Goodacre R, Sisson PR, Magee JG, Ward AC, Lightfoot NF (1994) Rapid identification of species within the *Mycobacterium tuberculosis* complex by artificial neural network analysis of NMR data. J Med Microbiol 40: 170-173

Gavaghan CL, Holmes E, Lenz E, Wilson ID, Nicholson JK (2000) An NMR-based metabonomic approach to investigate the biochemical consequences of genetic strain differences: Application to the C57BL10J and Alpk:ApfCD mouse. FEBS Lett 484: 169-174

Goodacre R, Timmins EM, Burton R, Kaderbhai N, Woodward AM, Kell DB, Rooney PJ (1998) Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks. Microbiology 144: 1157-1170

Goodacre R, Timmins EM, Rooney PJ, Rowland JJ, Kell DB (1996) Rapid identification of *Streptococcus* and *Enterococcus* species using diffuse reflectance absorbance Fourier transform infrared spectroscopy and neural networks. FEMS Microbiol Lett 140: 233-239

Helm D, Labischinski H, Schallen G, Naumann D (1991) Classification and identification of bacteria by Fourier transform infrared spectroscopy. J Gen Microbiol 137: 69-79

Henson BJ, Hesselbrock SM, Watson LE, Barnum SR (2004) Molecular phylogeny of the heterocystous cyanobacteria (subsections IV and V) based on *nifD*. Intl J Syst Evol Microbiol 54: 493-497

Kansiz M, Heraud P, Wood B, Burden F, Beardall J, McNaughton D (1999) Fourier transform infrared microspectroscopy and chemometrics as a tool for the discrimination of cyanobacteria. Phytochem 52: 407-417

Kim SW, Ban SH, Chung H, Cho SH, Chung HJ, Choi PS, Yoo OJ, Liu JR (2004) Taxonomic discrimination of flowering plants by multivariate analysis of Fourier transform infrared spectroscopy data. Plant Cell Rep 23: 246-250

Naumann D, Helm D, Labischinski H (1991) Microbiological characterizations by FT-IR spectroscopy. Nature 351: 81-82

Nicholson JK, Connelly J, Lindon JC, Holmes E (2002) Metabonomics: A generic platform for the study of drug toxicity and gene function. Nat Rev Drug Discov 1: 153-161

Ragan M (2001) Detection of lateral gene transfer among microbial genomes. Curr Opin Genet Dev 11: 620-626

Rippka R, Herdman H (1992) Pasteur Culture Collection of Cyanobacteria: Catalogue and Taxonomic Handbook, Catalogue of Strains. Vol 1, Institut Pasteur, Paris

Robertson DG, Reily MD, Sigler RE, Wells DF, Paterson DA, Braden TK (2000) Metabonomics: Evaluation of nuclear magnetic resonance (NMR) and pattern recognition technology for rapid *in vivo* screening of liver and kidney toxicants. Toxicol Sci 57: 326-337

Sneath PHA (1993) Evidence from *Aeromonas* for genetic crossing-over in ribosomal sequences. Intl J Syst Bacteriol 43: 626-629

Song E-K., Zulfugarov IS, Kim J-H, Kim E-H, Lee W-S, Lee C-H (2004) Selection and Characterization of transposon tagging mutants of *Synechocystis* sp. PCC 6803 sensitive to high-light and oxidative stress. J Plant Biol 47: 289-299

Turner S (1997) Molecular systematics of oxygenic photosynthetic bacteria. Plant Syst Evol 11: 13-52

Wang Y, Zhang Z, Ramanan N (1997) The actinomycete *Thermobispora bispora* contains two distinct types of transcriptionally active 16S rRNA genes. J Bacteriol 179: 3270-3276

Ward JL, Harris CH, Lewis J, Beale MH (2003) Assessment of ^{1}H NMR spectroscopy and multivariate analysis as a technique for metabolite fingerprinting of *Arabidopsis thaliana*. Phytochem 62: 949-957

Wilmotte A, Golubic S (1991) Morphological and genetic criteria in the taxonomy of Cyanophyta/Cyanobacteria. Archiv Hydrobiol (Suppl 92) Algol Stud 64: 1-24

Wold H (1966) Estimation of principal components and related models by iterative least squares, *In* KR Krishnaiah, ed, Multivariate Analysis. Academic Press, New York, pp 391- 420

Yap WH, Zhang Z, Wang Y (1999) Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. J Bacteriol 181: 5201-5209