http://www.pjbs.org

PJBS

ISSN 1028-8880

Pakistan Journal of Biological Sciences

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Pakistan Journal of Biological Sciences 10 (20): 3487-3496, 2007
ISSN 1028-8880

Biosystematics and Plant Proteomics: Role of Proteomics in Plant Phylogenetic Analysis

[1]Ch. Muhammad Ishtiaq, [1]Q. He, [1]J.P. Huang, [1]Yi Wang, [2]P.G. Xiao and [1]Yi Yu Cheng
[1]Department of Chinese Medicine Sciences and Engineering, College of Pharmaceutical Sciences, Institute of Bioinformatics, Zhejiang University, Hangzhou, 310028, People's Republic of China
[2]Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Beijing, 100094, People's Republic of China

Abstract: Since time immemorial, systematics has played significant role in every sphere of life. Biosystematics has evolved from folk taxonomy towards natural classification system and then culminated into homology based classification system. A good systematic approach is practical and predictive of phylogenetics of taxa incorporating different data. The morphological, chemical and molecular (genomics and proteomics) informations are used to explore the exact inter-relationship among the organisms. Proteomics is an essential and inevitable aspect in plant biology which can help in deciphering the functions of the genes that are or will be sequenced. Proteomics has proved to be a good tool in characterisation of individual lines and genetic distances among the genera, species, subspecies, verities and populations describing their phylogenetic inter-relationships. Two-dimensional electrophoresis (2-DE) is the major technique being applied for polypeptide characterization of each taxon for exploring phylogenetic or physiological relationships among organs, tissues or organisms. Moreover, proteomics can lead to unraveling the natural phenomena of plants development and their response to changing environment. These proteomic derived informations and their application in phylogenetic studies can be useful in agro-biotechnology development for better yield and safe use of food and medicines.

Key words: Plant phylogenetics, two-dimensional electrophoresis, plant proteomics, biosystematics, agro-biotechnology

INTRODUCTION

Systematics has played a vital and key role throughout the history of biology. The recognition of basic kinds of organisms, their properties and relationships in higher categories was earliest biological discipline. Advancements and developments in biology as a whole have interacted with systematics throughout its history. Some basic and earliest taxonomic treatments include (Mayer, 1982; Dupuis, 1984; Hull, 1988; Stevens, 1994).

In history of systematics, three revolutions have occurred (Khun, 1970). First and foremost classification called Folk Taxonomy (FT) was emerged from prehistory and organisms were classified according to their relationship with human affairs. The first Scientific Revolutionary Concept (SRC) was launched by the ancient Greeks which had shed profound effects on biological systematics and it was elegantly demonstrated by Hull (1988). In this approach taxa were defined by the possession of necessary and sufficient defining traits. This view was culminated to its apex in the work of Linnaeus, became untenable as the wealth of biological diversity due to the explorations of the 18th and 19th centuries.

In later years another approach based on Natural System (NS) was laid down (Stevens, 2000). In this approach taxa were recognized and delimited by overall resemblance in many characters although these characters were often selected for their importance in the biology of the group in question.

The third major evolution in the history of systematics was begun by Hennig (1965 and 1966) due to his careful examination of the idea of HOMOLOGY. It means that two similar organisms posses and share same characteristics. According to modern concept of homology, historical continuity of information (Van Valen, 1976; Roth, 1991), it was concluded that two organisms

Corresponding Author: Ch. Muhammad Ishtiaq and Yi Yu Cheng, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, 310028, People's Republic of China Tel: +86-571-87952509 Fax: +86-571-87951138

share a feature because of descent from a common ancestor that had that feature. Hennig stated that Homologous Similarities (HS) are useful for reconstructing the relative order of branching events in a system that is changing by descent with modification.

The traditional systematical approaches conducted from old times up to now can be categorized as:

- Pre-history---folk classifications.
- ancient greeks through linneaeus---essentialism.
- Natural system---overall resemblance; importance.
- Darwin---evolutionary language.
- Numerical phenetics---computer aided.
- Phylogenetic systematics (Cladistics)----------synapomorphies, monophyly.

Practicality is always a major goal of a systematics is simply to provide names that can be used easily and in a stable manner. The classification system should be informative and predictive. A good classification system comprises of some important aims and purposes such as:

- Practility------operationality, ease, stability.
- Information content-----optimal summarization of what is known about entities.
- Productivity------of unknown features of entities.
- Function in theories------capture entities acting in or resulting from natural processes.

A good classification should provide an optimal summarization of what should be predictive about the entities classified. Furthermore, a good classification should be predictive of unknown features of the entities. Finally, a good classification should have a role in process theory-it should relate directly to the natural processes acting on the entities being classified, capturing entities acting in, or resulting from, these natural processes. This last requirement is actually the most important and the other criteria flow from this one if it is successfully applied. In case of biology, the most powerful organizing process on which to base a predictive general-purpose classification is evolution via descent with modification.

The central role of phylogeny reconstruction in systematics, ecology and evolutionary biology has become widely realised in recent years (Donoghue, 1994; Harvey and Pagel, 1991; Martins, 1996). Explicit cladistic phylogenies now provide a critical basis for classification as well as for studies of speciation, biogeography, ecology and behaviour (among many other areas).

As discussed earlier, the fundamental advances leading to recent progress in systematics are in theory and method about how to use any data for phylogenetic reconstruction. During the 1980s, there was a great deal of philosophical and conceptual debate over the methodologies that should be applied to phylogenetic studies in general and how to best infer phylogenetic relationships from molecular data (Swofford *et al*., 1996) However, advancement that occurred in technology of molecular biology is the availability of new data at molecular level (genomics and proteomics). The incorporation of these new data has also helped to reinvigorate systematics and fueled the recent quantum leap forward in knowledge of phylogenetic relationships. In this article we will discuss only importance of molecular data (proteomics) in plant systematic studies.

Proteomics is the study of complete expressed proteins encoded by genome in any tissue (Patterson and Abersold, 2003). The term "*proteome*" PROTEin complement of the genOME was firstly coined by Marc Wilkins in 1994, a graduate student at Macquarie university, Australia (Wilkins *et al*., 1995).

The word Proteomics is chosen as analogue of genome and it depicts the total set of proteins expressed by a genome at particular time in any part/tissue of an organism under study. It can be said that proteomics is the systematic analysis of the proteins expressed by the genome.

In current years, proteomic is booming up very rapidly because of many reasons, as development in advancement in instrumentation, performance and data management. Due to the ongoing developments in bioinformatics and the wealth of informations generated by the different genome sequence programmes have evoked great expectations of the outcome of proteome research, in particular in relation to phylogenetic and systematic studies, developmental and growth patterns in different eco-climates and drug development. Proteomics is the endeavor to understand gene function and to characterize the molecular processes of the living cell through the large-scale study of proteins found in specific biological contexts. So, it might be expected that proteomics could lead to the discovery of novel markers of taxonomic importance, environmental-response related, diseases marker proteins and may prove novel targets for drug development.

The mapping of proteins encode by even a relatively small plant genome requires a technique with very high resolving power. This high throughput technique (2-DE) can identify the differentially expressed proteins which will ease the understanding of complex biological networks. Moreover, by proteomic approach comparison between different genera, species, subspecies, varieties and populations of different provenances can be drawn in

order to explore the phylogenetically and taxonomically important marker proteins. Similarly, proteomes of resistant and susceptible plant tissues can be compared to identify resistance-related proteins, or proteomes of red and white flowers can be compared to identify proteins involved in flower pigmentation. Moreover, this can lead to understanding of mechanism of how plants respond to the environmental stress by identifying proteins from normal and stress treated species. It will be leaping step for massive agronomic yield that is inevitable necessity of the hour due to rapid population explosion and changing environmental conditions.

In this study practical aspects and applications of proteomics in plant sciences, with particular emphasis on its use to explore the complex taxonomic and phylogenetic status of different taxa is discussed.

Proteomics and plant secondary metabolism: The proteomics can provide a promising approach for studying secondary metabolism in plants and plant cells. Many of these metabolites are used as fine chemicals such as drugs, anti-oxidants, flavours, fragrances, dyes and insecticides, but unfortunately the yields of such commercially important compounds are generally low and purification is costly. In order to increase the yield, several strategies may be followed, for example the production of plant cell suspension cultures and metabolic engineering. The best studied model system for the production of secondary metabolites is provided by *Catharanthus roseus* (Verpoorte *et al*., 1997). This medicinal plant is known for its accumulation of many alkaloids, among which are the effective anti-tumour drugs vinblastine and vincristine. Cell cultures of *C. roseus* have been widely used for studying the regulatory processes of alkaloid accumulation and it is quite feasible to culture *C. roseus* cells in commercial-scale bioreactors (Schiel and Berlin, 1987). However, alkaloid yields in suspension cell cultures are generally too low to allow commercialisation. One strategy to increase the alkaloid yields in cultured cells is metabolic engineering, e.g., by over expression of rate limiting enzymes (Verpoorte *et al*., 1999). Therefore the identification of proteins involved in the alkaloid biosynthesis is necessary. The proteomics approach has also been used to discover the biosynthetic pathway of isoflavonoid-derived phytoalexins synthesis in cell suspension cultures of *Phaseolus vulgaris* (Bell *et al*., 1986).

In maize Damerval and Guilloux (1998) studied the pleiotropic effects of the *Opaque*2 (O2) gene, which codes for a transcription factor. The comparison of 2-DE patterns of O2 and wild-type maize lines in several unrelated back grounds has permitted to identify specific targets of the O2 gene. Several enzymes belonging to various metabolic pathways were identified, confirming that O2 is a regulatory gene connecting different grain metabolism pathways.

Proteomics and plant-environment interaction mechanism: In plants, some research on biomarkers involved in tolerance to water deficit mechanism has been studied (Riccardi *et al*., 1998). Plants facing biotic and abiotic environmental stimuli react with a number of biochemical and physiological changes (Hirai and Saito, 2004).

The mechanism of plant-environment response involves the modification of their ability to detect the stimulus leading to the transduction of signal in the plant cell that triggers down the activation of appropriate genetic programmes (Knight and Knight, 2001). Protein modification (phosphorylation) is an important mechanism in response to environmental changes by the plant (Peck, 2003). And signal transduction involves the many networking of tissue or organ networking at many levels (Knight and Knight, 2001). Hence, proteomics is one of the tools that can help to understand the cross-talk between different signaling pathways.

Proteomics and agronomic studies: 2-DE approach has lead to investigate the mechanism of plant response to abiotic stresses i.e., high or low temperature, mainly because of possible applications to breeding programs of cultivated (agricultural) species. Generally, elevated temperatures induce synthesis of Heat Shock Proteins (HSPs). Their synthesis is correlated to the acquisition of thermal tolerance, i.e., the ability to withstand higher temperatures. Plants differ from other organisms in that they synthesize a great number of low-molecular mass HSPs (LMW-HSPs). About 48 HSPs in tomato cell cultures were identified and genetic variability of HSPs was explored out (Nover and Scharf, 1987). They declared that a high level of polymorphism was detected and more than one-third of the 35 detected HSPs were found qualitatively or quantitatively variable between three inbred lines, while only 13% of the other proteins were variable among the same genotypes. Later, it was found that thermal tolerance was not correlated with qualitative variation of HSPs but with the quantitative variation of two LMW-HSPs common to the seven tested genotypes (Madidi *et al*., 1993). 2-DE approach was used to investigate the induction of proteins by low temperature (cold induced proteins-CIPs) in rape seedlings

(Meza-Basso *et al*., 1986). Guy and Haskell (1988) detected polypeptides associated with cold acclimation in spinach seedlings: the synthesis of these proteins was increased during the period of freezing tolerance acquisition and reduced during re-acclimation. Cabane *et al.* (1993) identified chilling-acclimation-related proteins in soybean, one of them being a heat shock protein (HSP70). In poplar two families of high-molecular mass polypeptide were shown to accumulate in response to chilling (Hausman *et al*., 2000). Quantitative changes for 26 proteins were detected in response to cold treatment of potato tubers (Berkel *et al*., 1994). Long-term chilling tolerance of tomato fruit acquired by heat shock treatment was shown to be correlated to the persistence of HSPs (Sabehat *et al*., 1996). Hence, discovery and understanding the role of HSPs and CIPs may be a fruitful step in improvement of agronomic and medicinal plant species.

Experimental steps in proteome analysis: For proteome analysis, the most common and basic approach applied is two-dimensional gel electrophoresis developed in 1975, (O'Farrell, 1975) coupled with MALDI-TOF and ESI-MS/MS and supported by online molecular data bases for identification and characterization of proteins. As 2-DE is a long-standing biochemical technique that is commonly being used to exploit plant proteomic research (Zolla *et al*., 2003; Ganeteg *et al*., 2004) and is especially applicable to the analysis of abundant changes in highly expressed cytosolic proteins (Ferro *et al*., 2000).

Use of 2-DE for phylogenetic studies: Two-dimensional electrophoresis analysis of proteins can be a good source of monogenic and codominant markers for population genetics analysis and variability studies and for genetic mapping (Kersten *et al*., 2002). The structure of genetic variability in natural populations has always been a subject of great interest to population geneticists, evolutionists and plant breeders. For such analysis type of Marker Identification (MI) is confronted with advantages and limitations, hence a variety of techniques needs to be available, among others is 2-DE. 2-DE allows revealing of far more markers compared to isozyme electrophoresis for which few assays are available and it was reported that the 2-DE revealed as many alleles per polymorphic loci as in isozyme studies (De Vienne *et al*., 1996).

For the last 20 years, experiments have been performed at various taxonomic levels to assess genetic differences using 2-DE protein patterns; some examples are detailed below. The cultivated wheats are allopolyploid species possessing either two (for hard wheat *Triticum turgidum* AABB) or three (for soft wheat *Triticum aestivum* AABBDD) genomes. Those genomes originate from a common ancestor that has diverged since and today there are numerous diploid or polyploid wild species with different so-called homoeologous genomes that still can be intercrossed, more or less easily. The number of bivalent chromosomes figures at meiosis led the wheat geneticists to hypothesize which of the present species may have given the different genomes of the cultivated wheats. The A genome originated from *Triticum urartu* and the D genome from *T. taushii* (Kerby and Kuspira, 1987). However, there was still a conflicting debate about which species of the *Sitopsis* section contributed the B genome and the cytoplasm of the cultivated wheats. To answer this question, 2-DE patterns of the different species of this section (*T. longissimum*, *T. sharonense*, *T. bicorne*, *T. searsii* and *T. speltoides*) were compared and compared with Chinese Spring (CS), a bread wheat cultivar. By analysing the number of spots found in common, similarity indices were computed between each pair of genotypes. From the resulting similarity matrix, dendrograms were drawn, reflecting the phylogenetic relationships in the *Sitopsis* section. Besides the perfect accordance between the 2-DE based dendrogram and the classical taxonomy, it was found that *T. speltoides* is the wild species, the most related to the B genome of cultivated wheats (Bahrman *et al*., 1988). This finding was confirmed by cytoplasmically encoded proteins: the large subunit of Rubisco and two forms of the b-ATPase (Bahrman *et al*., 1988) have the same allelic forms in CS and *T. speltoides*, while the other *Sitopsis* species show other alleles of these chloroplastic genes.

The study of phylogenetic relationships by comparing 2-DE patterns was extended to species of the *Triticum* genus possessing different genomes (Thiellement *et al*., 1989). In another experiment, proteome approach was used to understand the genetic affinity between the homologous genomes of Triticeae tribe, i.e., the A and D genomes of *Triticum*, the H genome of *Hordeum* (barley) and the S genome of *Secale* (rye). This technique well deciphered the controversial phylogeny of this tribe and dendrograms obtained still reflected well the known phylogenetic relationships (Zivy *et al*., 1995). In the experiment 1275 spots were scored and 198 spots were found to be common to all species. Different distance indices; Dj (Jaccard index), Dm (matching coefficient), Dnl (Neil and Li, 1979 index), Dm (corrected matching coefficient) were calculated and resulting matrices were used to construct phenograms according to UPGMA method. The results were supporting to the previous isozyme gene loci work of McIntyre on this tribe (McIntyre, 1988). Phenogram generated from one of algorithm (Dm) is presented in Fig. 1.

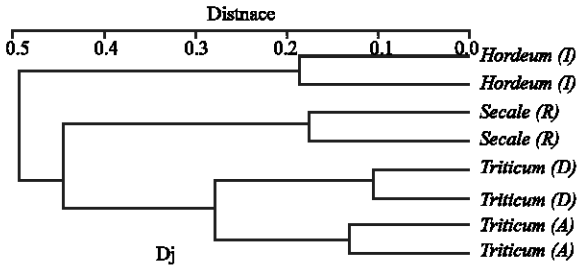

Fig. 1: Dendrogram computed from the matrix genetic distances between species of the Triticeae tribe, estimated from the numbers of common and noncommon spots

Jacobsen *et al.* (2001) used the 2-D gel electrophoresis for the classification of wheat varieties, he demonstrated this approach can differentiate the varieties efficiently, which otherwise cannot be isolated merely by MALDI-MS or ANN techniques.

In European oaks, the relationships between the two closely related species *Quercus petraea* and *Quercus robur* were examined by Barreneche *et al.* (1996). They used 2-DE for studying the genetic differentiation between the two species by comparing 23 oaks from six European countries covering partly the natural geographic range of white oaks in Europe. Total proteins from seedlings were analyzed and 530 polypeptide spots scored, among which 101 were polymorphic. The dissimilarity between the two species was 0.36, whereas the within species dissimilarities were 0.35 for *Q. petraea* and 0.33 for *Q. robur*. Such very close interspecific and intraspecific distances confirmed the low level of genetic differentiation between both species, already reported with isozymes, RAPDs (random amplified polymorphic DNAs) and chloroplastic DNA approaches.

In other research, Brassicaceae taxa were classified and their phylogenetic distances were determined by use of proteomic approach. Their ploidy level was also derived from their genetic distances. It is described that 2D-PAGE is an efficient tool to deduce genetic relationships in taxa of dicot Brassicaceae. Moreover, the proximity distances between *Arabidopsis* and the species of interest of same family i.e., Brassicaceae are also measured by proteomic technique (Marques *et al*., 2001). In this experiment, they have used only those 750 spots which showed high optic density. The obtained three gels of each species were matched by Melanie 3.0 software package (Appel *et al*., 1991) and generated a reference gel image for each accession. For more validity and reliability, they performed coelectrophoresis by mixing equal amounts of sample from each accession and napusD (internal reference marker-IRM) and obtained gels were used to elucidate the position of each spot by comparison with IRM. Then mixed gels were matched and compared with each other and a global gel with 2273 reproducible spots was created. They spots (present/absent) were formulated in a matrix form, which was used to construct a phenetic tree. The criteria of spots comparison was based on counting the number of spots commonly ($n_{xy)}$ or absent (n_{oo}) and specifically present in one (n_{xo}) or the other (n_{oy}) of the two considered accessions. They conducted two types of comparisons, first one on the basis of all the spots, the second one on the basis of the 676 napus D (IRM) spots.

Four distance indices were calculated:

- Jaccard index $Dj = 1-n_{xy}/n_{xy}+n_{xo}+n_{oy})$;
- Nei and Li (1979) index $Dnl = 1-2n_{zy}/ (N_x+N_y)$, N_x and N_y are the number of spots counted in accession x and accession y.
- Matching coefficient $Dm = 1-(n_{xy+} n_{oo})/N_t$, where N_t is the total number of spots taken into account in comparison (2273) for the minimum global gel);
- Corrected matching coefficient (Zivy *et al*., 1995) $Dm = 2DmN_{max}/(N_x+N_y)$.

where, N_{max} is the highest number of spots found in one of the nine reference gels. The four calculated indices predicted correlation with each other except Dm (with little correlation).

The phenetic trees were constructed by using three algorithms: Fitch-Margoliash (FM) based on maximum of likelihood (Fitch and Margoliash, 1967), Neighbor-Joining (NJ) and unweighted pair group method algorithm (UPGMA) (Sneath and Sokal, 1973) to depict the interrelationships among the varies taxa. Moreover, they also conducted Factorial Correspondence Analysis (FCA) to visualize the genetic proximities between the genotypes and in particular among the tetraploids. In the phenetic tree it was observed that two lines of *Arabidopsis* species were first merged and were far from other analysed taxa. This feature was used as test case for the methodology used in this phylogenetic studies. Another reliability test was performed to support the experimental findings, in which they used only 676 spots of reference synthetic map napusD. In limited data of napusD the two *Arabidopsis* species were merged first than *Raphanus* and three *Brassica* taxa and obtained tree conserved the main features (Fig. 2 and 3).

The analysis methodology used in this experiment showed that NJ and FM algorithms produced more or less similar phenetic trees while UPGMA generated more different tree from the first two. It was proposed that UPGMA method is less useful for the proteome

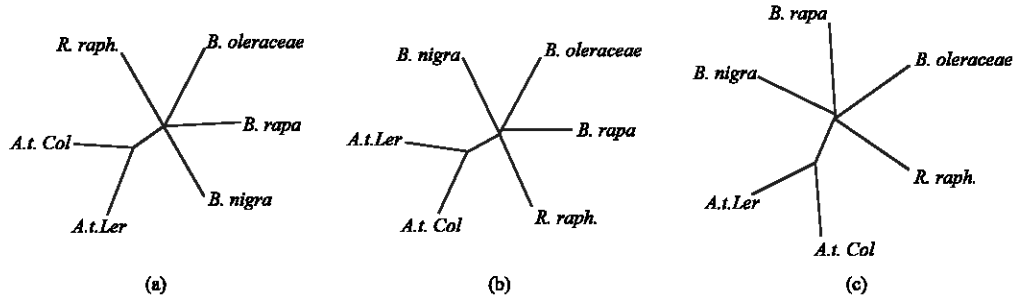

Fig. 2: Phenetic trees constructed from distance matrix calculated according to the Jaccard index on all the spots of six diploids. (a) With the FM algorithm; (b) With the NJ algorithm and (c) With the UPGMA algorithm

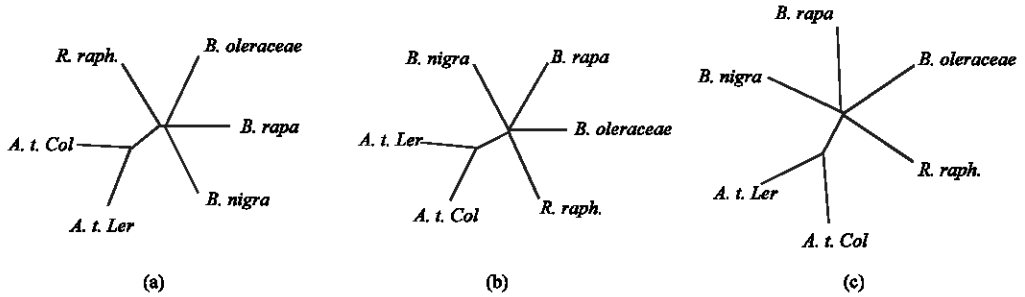

Fig. 3: Phenetic tree constructed from the distance matrix calculated according to the Jaccard index on the restricted set of the 676 Drakkar spots. (a) With the FM algorithm; (b) With the NJ algorithm and (c) With the UPGMA algorithm

comparisons in this experiment. These proteomic based taxonomic studies of *Brassicaceae* were claimed to be in accordance with previous established taxonomy. In proteomic studies like in the *Triticeae* it was emphasized that more spot number used for constructing phenetic trees was better than tree constructed by using less spots because more spots can provide more informations and precision accordingly (Thiellement *et al*., 1989). From this study on *Brassicaceae* it can be predicted that proteomics approach can prove to be a good taxonomic technique to investigate inter-relationships among the different taxa of other plants.

In maritime pine (*Pinus pinaster*) Bahrman *et al*. (1994) used 2-DE method to study the relationships between seven provenances of the natural range and to evaluate the genetic variability existing within and between geographical origins. From the megagametophyte tissue, a total of 968 spots were scored which can differentiate between main groups; Atlantic, Mediterranean and North African and this genetic structure that was in agreement with terpene data taxonomy.

David *et al*. (1997) studied the genetic differentiation of 11 wheat populations originating from a single one. Multivariate analysis showed that all populations differentially evolved from the original one and that natural selection rather than random drift was responsible for these differentiations.

Wheat varieties are well discriminated on basis of 2-PAGE than with MALDI-TOF-MS because former can differentiate the varieties possessing similar protein with same molecular weight yet well isolated due to isoelectric points (Jacobsen *et al*., 2001). Zivy *et al*., (1983 and 1984) demonstrated that three wheat varieties can be distinguished by qualitative and quantitative differences of proteins revealed by 2-DE.

Proteomics for differentiation of wild and mutant types: The 2-DE patterns of mutant compared to wild type may permit either to evaluate the effects of a mutation, for instance examine the pleiotropy of an already described mutant, or to look for the protein(s) encoded or influenced by the mutated gene. In the model plant *Arabidopsis thaliana*, 2-DE patterns of a series of mutants affected in the first steps of development were examined by Santoni *et al*. (1997).

In tomato the comparison of 2-DE patterns between the wild-type and a Fe-deficient mutant led to the identification of several enzymes whose amounts were different and that were involved in anaerobic metabolism and stress defense (Herbik *et al*., 1996).

In an experiments on moss Kasten *et al.* (1997) compared the proteins of a chloroplastic mutant with the wild type whether supplemented with cytokinin or not. It was concluded that the hormone affects both nuclear and cytoplasmically encoded proteins.

The analysis of the protein patterns of a series of *Arabidopsis* mutants affected in early development and of hormone-treated wild types led to a biochemical classification that was consistent enough to predict that an uncharacterized mutant was likely a cytokinin overproducer (Santoni *et al.*, 1997). This hypothesis was later confirmed by cytokinin dosage (Faure *et al.*, 1998).

The comparison of several late-flowering mutants of *Arabidopsis* revealed protein spots that appeared or disappeared compared to the wild type (Tacchini *et al.*, 1995). However, in the F2 offsprings of the crosses between mutants and wild types, none of the variable protein spots co-segregated with the flowering phenotype. In addition to the necessity of genetic confirmations these experiments demonstrated that 2-DE analysis can reveal the genetic heterogeneity of the ecotypes or lines used in mutagenesis experiments.

The power of proteome analysis in mutant characterization is even more evidenced by studying pleiotropic mutations. Gottlieb and de Vienne (1988) compared two near-isogenic lines of pea differing by the r gene that determines round (RR) or wrinkled (rr) seeds. In proteomes of mature seeds nearly 10% of spots differed in amounts, confirming numerous known physiological differences between two types of seeds.

Significance of proteomics vs DNA based study: The possibility to study the genetic variation (in pedigrees and in natural populations) at the protein level may in this respect be extremely useful. Proteins act directly on biochemical processes and thus must be closer to the build up of the phenotype, compared to DNA-based markers. Therefore, 2-DE appears as a very interesting technique to understand the variability in trait expression. In this context, proteins certainly constitute more informative markers compared to DNA markers (Gerber *et al.*, 1997). In maritime pine authors had demonstrated the rationale of this approach. For several traits i.e., seed weight and growth related traits it was detected that there was a significant protein-trait association among the 84 protein loci genotyped on 18 unrelated trees, suggesting some of these proteins are responsible for the trait variation itself.

Candidate proteins for gene confirmation: The identification of the genes responsible for genetic variation in agronomically important traits has mainly been investigated using linkage mapping with anonymous markers in segregating progeny where linkage disequilibrium can be maximized (Tanksley, 1993). Validation of a candidate gene needs further confirmation before implementation of such information in breeding programs. The study of the genetic variation of proteins can provide such a validation tool which had been exemplified in maize.

Moreover, intensity of protein spot predicts the quantity of spots expressed by the specific genome set can be revealed on 2-DE gels. Proteomic approach can unravel the differences among common proteins at quantity levels which may depict the genome differences among the analyzed taxa.

Recently, Gil-Agusti *et al.* (2005) had used 2-DE as classification tool for green coffee species. They obtained fingerprints of the expressed genomes in form of polypeptides, which were used as differentiatal markers for coffee species. They found that proteomic approach is suitable to characterize each genome via differential expression of protein profiles (a total of 30 proteins were reported to be differentially expressed). This number of differential proteins is limited as they used CBB staining, which is less sensitive; this number may be enhanced by application of classical silver staining procedure. All in all, it is evident that proteomic approach can be a valuable taxonomic tool for plants (Fig. 4).

In brief the 2-D electrophoresis patterns are good indicators of genetical or physiological changes (Damerval *et al.*, 1987) which presents its importance as tool for phylogenetic studies.

Concluding perspectives: It is worth pointing out that proteomics is a burgeoning field in plant biology and agricultural biotechnology (Thiellement *et al.*, 1999). Proteomic in conjunction with genetics, physiology and molecular biology has become an essential tool in understanding the mechanism of plant growth and development. These breakthroughs will enhance the improvement process of crops and crops management by furnishing new tools to plant breeders. In case of medicine and/or main stream biology proteome approaches are of real importance in plant sciences.

In the near future, proteomics will allow us to confront many of the problems related to phylogeny and its methodology. Protein expression patterns from any one of the development stages of pluricellular species provide a complete figure of the total proteome needed to create a phylogenetic hypothesis. With more complete sequence information on these taxa more consistent phylogenies will be constructed that will result in a much more realistic view of the tree of life.

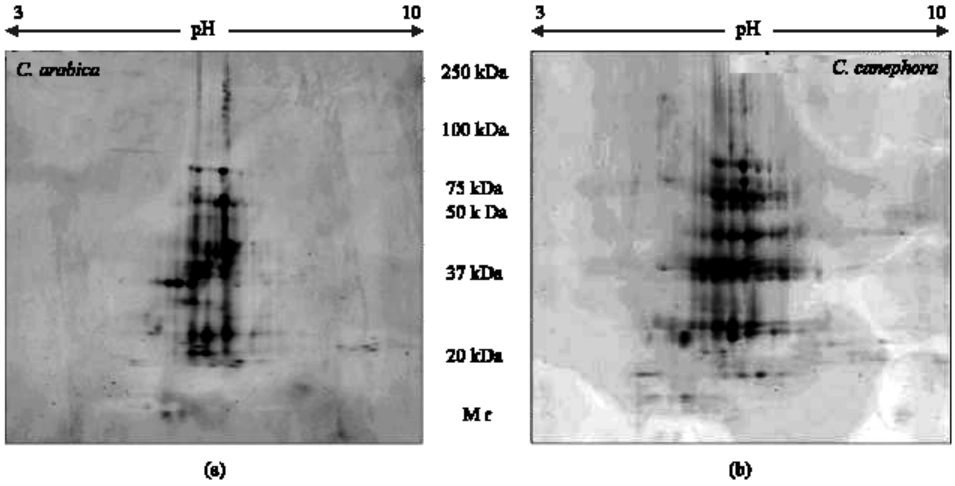

Fig. 4: Representative maps of (a) *Coffea arabica* (var. Colombia) and (b) *C. canephora* (var.) Indiano Robusta). First dimension: IPG pH 3-10 linear gradient. Second dimension: SDS-PAGE in a 7% T to 20%T porosity gradient

Proteomics can prove to be a useful separation system for resolving some current problems of systematics, phylogenetics, within-species variations and eco-physiology. This technique well satisfies the requirements of the taxonomists and naturalists. The electrophoretic analysis of polypeptides on PAGE or 2-D can still play important role in botany and can be applied profitably in taxonomic and phylogenetic studies on diverse groups of Angiosperms and other taxa. Further, improvements in 2-DE technologies and proteomic methodologies are still press need for better understanding of the structure, function of proteins and the mechanism of protein-protein interaction in changing global environmental scenario.

REFERENCES

Appel, R.D., D.F. Hochstrasser, M. Funk and J.R. Vargas *et al*., 1991. The MELANIE project: From a biopsy to automatic protein map interpretation by computer. Electrophoresis, 12: 722-735.

Bahrman, N., M. Zivy and H. Thiellement, 1988. Genetic polymorphism in maritime pine (*Pinus pinaster* Ait.) assessed by two-dimensional gel electrophoresis of needle, bud and pollen proteins. Heredity, 61: 473-480.

Bahrman, N., M. Zivy, C. Damerval and P. Baradat, 1994. Organisation of the variability of abundant proteins in seven geographical origins of maritime pine (*Pinus pinaster* Ait.), Theor. Applied Genet., 88: 407-411.

Barreneche, T., N. Bahrman and A. Kremer, 1996. Segregation of AFLP markers in *Betula pendula* (Roth). For. Genet., 3: 89-92.

Bell, J.N., T.B. Ryder, V.P. Wingate, J.A. Bailey and C.A. Lamb, 1986. Differential accumulation of plant defence gene transcripts in a compatible and an incompatible plant pathogen interaction. Mol. Cell Biol., 6: 1615-1623.

Berkel, J. van, F. Salamini and C. Gebhardt, 1994. Transcripts accumulating during cold storage of potato (*Solanum tuberosum* L.) Tubers are sequence related to stress-responsive genes. Plant Physiol., 104: 445-452.

Cabane, M., P. Calvet, P. Vincens and A.M. Boudet, 1993. Characterization of chilling-acclimation-related proteins in soybean and identification of one as a member of the heat shock protein (HSP 70) family. Planta, 190: 346-353.

Damerval, C., Y. Hebert and D. De Vienne, 1987. Is the polymorphism of protein amounts related to phenotypic variability? A comparison of two-dimensional electrophoresis data with morphological traits in maize. Theor. Applied Genet., 74: 194-202.

Damerval, C. and M.L. Guilloux, 1998. Characterization of novel proteins affected by the O_2 mutation and expressed during maize endosperm development. Mol. Gen. Genet., 257: 354-361.

David, J.L., M. Zivy, M.L. Cardin and P. Brabant, 1997. Protein evolution in dynamically managed populations of wheat: Adaptive responses to macro-environmental conditions. Theor. Applied Genet., 95: 932-941.

De Vienne, D., J. Burstin, S. Gerber, A. Leonardi, M. Le Guilloux, A. Murigneux, M. Beckert, N. Bahrman, C. Damerval and M. Zivy, 1996. Two-dimensional electrophoresis as a source of monogenic and codominant markers for population genetics and mapping the expressed genome. Heredity, 76: 166-177.

Donoghue, M.J., 1994. Progress and prospects in reconstructing plant phylogeny. Ann. Missouri Bot. Gard., 81: 405-418.

Dupius, C., 1984. Willi Hennings impact on taxonomic thought. Ann. Rev. Ecol. Syst., 15: 1-24.

Faure, J.D., P. Vittorioso, V. Santoni, V. Fraisier, E. Prinsen, I. Barlier, H. Van Onckelen, M. Caboche and C. Bellini, 1998. The PASTICCINO mutants of *Arabidopsis thaliana* are involved in the control of cell division and differentiation. Development, 125: 909-918.

Ferro, M., D. Seigneurin-Berny, N. Rolland, A. Chapel, D. Salvi and J. Garin *et al*., 2000. Organic solvent extraction as a versatile procedure to identify hydrophobic chloroplast membrane proteins. Electrophoresis, 21: 3517-3526.

Fitch, W.M. and E. Margoliash, 1967. Construction of phylogenetic trees. Science, 155: 279-284.

Ganeteg, U., C. Kulheim, J. Andersson and S. Jansson, 2004. Is each light harvesting complex protein important for plant fitness? Plant Physiol., 134: 502-509.

Gerber, S., M. Lascoux and A. Kremer, 1997. Relation between protein markers and quantitative traits in maritime pine (*Pinus pinaster* AIT.). Silvae Genet., 46: 5, 286.

Gil-Agusti, M.T., N. Campostrini, L. Zolla, C. Ciambella, C. Invernizzi and P.G. Righetti, 2005. Two-dimensional mapping as a tool for classification of green coffee bean species. Proteomics, 5: 710-718.

Gottlieb, L.D. and D. de Vienne, 1988. Assessment of pleiotropic effects of a gene substitution in pea by two-dimensional polyacrylamide gel electrophoresis. Genetics, 119: 705-710.

Guy, C.L. and D. Haskell, 1988. Detection of polypeptides associated with the cold acclimation process in spinach. Electrophoresis, 9: 787-796.

Harvey, P.H. and M.D. Pagel, 1991. The Comparative Method in Evolutionary Biology. Oxford.

Hausman, J.F., D. Evers, H. Thiellement and L. Jouve, 2000. Compared responses of poplar cuttings and *in vitro* raised shoots to short-term chilling treatments. Plant Cell Rep., 19: 954-960.

Hennig, W., 1965. Phylogenetic systematics. Ann. Rev. Entomol., 10: 97-116.

Hennig, W., 1966. Phylogenetic Systematics. Urbana, University of Illinois Press.

Herbik, A., A. Giritch, C. Horstmann, R. Becker, H.J. Balzer, H. Baumlein and U.W. Stephan, 1996. Iron and copper nutrition-dependent changes in protein expression in a tomato wild type and the nicotianamine-free mutant chloronerva. Plant Physiol., 111: 533-540.

Hirai, M.Y. and K. Saito, 2004. Post-genomics approaches for the elucidation of plant adaptative mechanism to sulphur deficiency. J. Exp. Bot., 55: 1871-1879.

Hull, D.L., 1988. Science as a Practice. An Evolutionary Account of the Social and Conceptual Development of Science. Chicago.

Jacobsen, S., L. Nesic, M. Petersen and I. Sondergaard, 2001. Classification of wheat varieties: Use of two-dimensional gel electrophoresis for varieties that cannot be classified by matrix assisted laser desorption/ionization-time of flight-mass spectrometry and an artificial neural network. Electrophoresis, 22: 1242-1245.

Kasten, B., F. Buck, J. Nuske and R. Reski, 1997. Cytokinin affects nuclear-and plastome-encoded energy-converting plastid enzymes. Planta, 201: 261-272.

Kerby, K. and J. Kuspira, 1987. The phylogeny of the polyploid wheats *Triticum aestivum* (bread wheat) and Triticum turgidum (macaroni wheat). Genome, 29: 722-737.

Kersten, B., L. Burkle, E.J. Kuhn, P. Giavalisco, Z. Konthur, A. Lueking, G. Walter, H. Eickhoff and U. Schneider, 2002. Large-scale plant proteomics. Plant Mol. Biol., 48: 133-141.

Khun, T.S., 1970. The Structure of Scientific Revolutions. Cambridge, Massachusetts.

Knight, H. and M.R. Knight, 2001. Abiotic stress signalling pathways: Specificity and cross-talk. Trends Plant Sci., 6: 262-267.

Madidi, S.El., M. Zivy, H. Chlyah and Y. Demarly, 1993. Le progre´s ge´ne´tique passe-t-il par le repe´rage et l'inventaire des ge´nes? London:AUPELF-UREF, pp: 173.

Marques, B., S.L. Chane-Favre, M. Zivy and H. Thiellement, 2001. Comparative proteomics to establish genetic relationships in the Brassicaceae family. Proteomics, 1: 1457-1462.

Martins, E.P., 1996. Phylogenies, spatial autoregression and the comparative method: A computer simulation test. Evolution, 50: 1750-1765.

Mayer, E., 1982. The Growth of Biological Thought. Cambridge, Massachusetts.

Meza-Basso, L., M. Alberdi, M. Raynal, M.L. Ferrero Cadinanos and M. Delseny, 1986. Changes in protein synthesis in rapeseed (*Brassica napus*) seedlings during a low temperature treatment. Plant Physiol., 82: 733-738.

McIntyre, C.L., 1988. Variation at isozyme loci in Triticeae. Plant Syst. Evol., 160: 123-142.

Nei, L. and W.H. Li, 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Nat. Acad. Sci. USA., 76: 5269-5273.

Nover, L. and K.D. Scharf, 1984. Synthesis, modification and structural binding of heat-shock proteins in tomato cell cultures. Eur. J. Biochem., 139: 303-313.

O'Farrell, P.H., 1975. High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem., 250: 4007-4021.

Patterson, S.D. and R.H. Abersold, 2003. Proteomics: The first decade and beyond. Nat. Genet. Suppl., 33: 311-332.

Peck, S.C., 2003. Early phosphorylation events in biotic stress. Curr. Opin. Plant Biol., 6: 334-338.

Riccardi, F.P., D. Gazeau, de Vienne and M. Zivy, 1998. Protein changes in response to progressive water deficit in maize. Quantitative variation and polypeptide identification. Plant Physiol., 117: 1253-1263.

Roth, V.L., 1991. Homology and hierarchies: Problems solved and unresolved. J. Evol. Biol., 4: 167-194.

Sabehat, A., D. Weiss and S. Lurie, 1996. The correlation between heat-shock protein accumulation and persistence and chilling tolerance in tomato fruit. Plant Physiol., 110: 531-537.

Santoni, V., M. Delarue, M. Caboche and C. Bellini, 1997. A comparison of two-dimensional electrophoresis data with phenotypical traits in *Arabidopsis* leads to the identification of a mutant (cril) that accumulates cytokinins. Planta, 202: 62-69.

Schiel, O. and J. Berlin, 1987. Large scale fermentation and alkaloid production of cell suspension cultures of *Catharanthus roseus*. Plant Cell Tiss. Org. Cult., 8: 153-162.

Sneath, P.H. and R.R. Sokal, 1973. Testing Statistical Significance of Geographic Variation Patterns, Numerical Taxonomy. W. H. Freeman, San Francisco, pp: 573.

Stevens, P.E., 1994. The Development of Biological Systematics: Antoine-Laurrent de Jussieu, Nature and the Natural System. New York.

Stevens, P.F., 2000. Botanical systematics, 1950-2000: Change or progress, or both? Taxon, 49: 635-659.

Swofford, D.L., G.J. Olsen, P.J. Wadedell and D.M. Hillis, 1996. In: Phylogenetic Inference. Hillis, D.M., C. Moritz and K.B. Mable (Eds.), Molecular Systematics. Sinauer Associates, Sunderland, MA, USA., pp: 407-514.

Tacchini, P., A. Fink, H. Thiellement and H. Greppin, 1995. Analysis of the leaf proteins in late flowering mutants of *Arabidopsis thaliana*. Physiol. Plant, 93: 312-316.

Tanksley, S.D., 1993. Mapping phylogeneis. Annu. Rev. Genet., 27: 205-233.

Thiellement, H., M. Seguin, N. Bahrman and M. Zivy, 1989. Homoeology and phylogeny of the A, S and D genomes of the triticinae. J. Mol. Evol., 28: 89-94.

Thiellement, H., N. Bahrman and C. Damerval, 1999. Proteomics for genetic and physiological studies in plants. Electrophoresis, 20: 2013-2026.

Van Valen, L., 1976. Ecological species, multi-species and oaks. Taxon, 25: 233-239.

Verpoorte, R., R. van der Heijden and P.R.H. Moreno, 1997. Biosynthesis of Terpenoid Indole Alkaloids in *Catharanthus roseus* Cells. In: The Alkaloids. Cordell, G.A. (Ed.), Vol. 49. Academic Press: San Diego, CA, pp: 221-299.

Verpoorte, R., R. van der Heijden, H.J.G. Ten Hoopen and J. Memelink, 1999. Metabolic engineering of plant secondary metabolite pathways for the production of new chemicals. Biotechnol. Lett., 21: 467-479.

Wilkins, M.R., J.C. Sanchez, A.A. Gooley, R.D. Appel, S.I. Humphrey and K.L. Williams, 1995. Progress with Proteome Projects: Why all proteins expressed by a genome should be identified and how to do it. Biotechnol. Genet. Eng. Rev., 13: 19-50.

Zivy, M., H. Thiellement, D. de Vienne and J.P. Hofmann, 1983. Study on nuclear and cytoplasmic genome expression in wheat by two-dimensional gel electrophoresis. Theor. Applied Genet., 66: 1-7.

Zivy, M., H. Thiellement, D. de Vienne and J.P. Hofmann, 1984. Study on nuclear and cytoplasmic genome expression in wheat by two-dimensional gel electrophoresis. Theor. Applied Genet., 68: 335-345.

Zivy, M., S.E.L., Madidi and H. Thiellement, 1995. Distance indices in a comparison between the A, D, I and R genomes of the Triticeae tribe. Electrophoresis, 16: 1295-1300.

Zolla, L., A.M. Timperio, W. Walcher and C.G. Huber, 2003. Proteomics of light-harvesting proteins in different plant species. Analysis and comparison by liquid chromatography-electrospray ionization mass spectrometry, Photosystem II. Plant Physiol., 131: 198-214.