Genetics of proteome variation for QTL characterization: application to drought-stress responses in maize

Dominique de Vienne¹, Agnès Leonardi, Catherine Damerval and Michel Zivy

Station de Génétique Végétale, INRA/UPS/INA PG/CNRS URA 2154, Ferme du Moulon, 91190, Gif-sur-Yvette, France

Received 20 July 1998; Accepted 25 September 1998

Abstract

The proteome is emerging as an important concept of the post-genome era. Powerful nucleic acid approaches (EST, DNA chips, etc.) are still limited because DNA sequences and mRNA levels are not sufficient to predict the structure, function, amount, and activity of the proteins in the cell. The proteome can now be subjected to large-scale analysis, owing to spectacular progress in the techniques of identification of proteins excised from two-dimensional (2-D) gels. In addition, computer-based analysis of 2-D gels makes it possible to quantify the protein spot intensities, which are commonly genetically variable. The loci controlling these variations may be mapped on the genome (PQLs, Protein Quantity Loci). Beyond the interest for regulatory genetics and molecular biology, the PQL methodology can provide an additional tool for the difficult task of identifying QTLs (Quantitative Trait Loci), in the context of the candidate gene approach. The PQL methodology is presented with the example of the phosphoglycerate mutase variations in maize, and the candidate gene/protein approach is illustrated for traits responsive to drought stress.

Key words: 2-D PAGE, proteome, QTL, PQL, candidate gene, maize, drought stress.

Introduction

The concept of ‘proteome’ (for PROTEin complement expressed by a genOME, Wilkins et al., 1996) has emerged recently (see also Zuckermandl, 1986) as a consequence of questions raised in the context of genome and post-genome projects. Although the DNA/RNA methodologies have allowed inumerable and outstanding break-throughs in genetics and genomics, they have limitations, particularly in physiology, because the protein structure and function cannot be completely described from the gene sequence. After synthesis, the proteins usually undergo post-translational modifications, such as phosphorylation or removal of a signal peptide, which can alter their activity and/or location in the cell. In addition, due to large differences in mRNA and protein turnover, the mRNA may poorly reflect how much protein is present in the cell: a protein can be still abundant while the mRNA is no longer detectable because its synthesis has stopped. Finally, the molecular mechanisms involved in cell differentiation, development, adaptation to environment, etc., cannot be deduced only from nucleic acid sequences or quantity.

The proteome concept is tightly linked to two-dimensional polyacrylamide gel electrophoresis (2-D PAGE). For more than 20 years, this technique has been the most powerful to resolve complex mixtures of proteins (O’Farrell, 1975; Klose, 1975). Several hundreds of proteins can be revealed on single 2-D gels. Their possible post-translational modifications can be studied provided they modify their pi or Mr, and the spot intensities allow the protein amounts to be estimated by using dedicated softwares (see references in Appel et al., 1997). Thus, even though it does not give any direct access to protein activity, 2-D PAGE allows the detection of variations which may be physiologically relevant.

Compared to the original technique, various simplifications and improvements have increased the resolving power and reproducibility of 2-D PAGE: large size 2-D gels without stacking gels, apparatuses to run and stain simultaneously series of gels (Damerval et al., 1986; Granier and de Vienne, 1986), use of immobilized pH gradients in the 1st dimension (Righetti, 1990), etc. (reviewed in Klose and Kobalz, 1995; Humphery-Smith

¹ To whom the correspondence should be addressed. Fax: +33 1 69 33 23 40. E-mail: devienne@moulon.inra.fr

© Oxford University Press 1999
et al., 1997). New high throughput methods are now available for the routine identification of proteins excised from the gels: in addition to N-terminal and internal micro-sequencing, analysis of amino acid composition has been used (Touzet et al., 1996a; Maillet et al., 1996; Golaz et al., 1996), and above all mass spectrometry techniques are making great strides (Jungblut and Thiede, 1997; Shevchenko et al., 1996; Humphery-Smith et al., 1997). Large databases of proteins are already available for Escherichia coli, yeast and man (a list of web sites can be found at http://expasy.hcuge.ch/ch2d/2d-index.html).

In plants, moderate databases exist for maize (Touzet et al., 1997). Large databases of proteins are already available for Arabidopsis thaliana (Tsutiga et al., 1994, 1996; Kamo et al., 1995), and the development of a database of plasma membrane proteins of Arabidopsis thaliana is in progress (http://sphinc.rug.ac.be:8080/).

Among various applications (see Damerval et al., 1988, for a review in plants), the proteome analysis has proved to be a powerful tool in yeast for analysing the effects of transcription factors, by comparing wild-type and mutant strains for changes in expression of known-function proteins (Boy-Marcotte et al., 1998). In the same line, Damerval and Le Guilloux (1998) discovered novel targets of the transcription factor Opaque-2 by comparing maize kernel proteomes of near-isogenic lines for the encoding gene. The comparison of 2-D patterns of different mutants of Arabidopsis thaliana allowed Santoni et al. (1997) to suggest that a developmental mutant was over-accumulating cytokinins, an hypothesis which was later confirmed.

This paper will focus on the genetic variations of the maize proteome, and their applications for addressing physiological questions, in particular in the context of QTL (Quantitative Trait Loci) characterization.

Quantitative genetics of the maize proteome

Polygenic variations of individual proteins

Visual comparison of 2-D gels in various species has long revealed that protein spots displayed reproducible intensity differences between genotypes (Klose and Feller, 1981; Zivy et al., 1983; Bahrmann et al., 1985). With the use of computer packages allowing spot intensities to be compared between gels, estimates of the extent of this variation have become possible. For example, in maize, 190 coleoptile silver-stained protein spots were quantified in a set of 21 lines with three replicates for each line. Analyses of variance revealed that as many as 124 spots (65.3%) displayed significant differences between lines ($P < 0.05$). The ratios of the maximum to minimum spot intensities in the 21 lines was higher than three for 40% of these proteins (Fig. 1) (data from Burstin et al., 1994). Such an observation is common in Higginbotham et al., 1991; Jorgensen and Nguyen, 1995), and similar results were obtained in wheat (Zivy et al., 1984), sugarcane (Ramagopal, 1990), maritime pine (Bahrmann and Petit, 1995), mouse (Klose, 1982), and man (Goldman et al., 1987).

Damerval et al. (1994) showed that the variable proteins could be considered as any macroscopic quantitative trait, and that the marker-based methods for mapping the loci controlling their variation (QTL, Quantitative Trait Loci) could apply (Soller et al., 1976). For every individual of an F$_2$ progeny derived from a cross between distant maize lines, 72 coleoptile polypeptides were automatically quantified. An RFLP genetic linkage map allowed QTLs for 42 polypeptides to be mapped (PQLs, for Protein Quantity Loci). One to five independent PQLs were found to affect single polypeptides, 27% of which resulted in more than a doubling of the polypeptide spot intensity. Dominance was observed for half of the PQL, with a high amount dominant over a low amount in most cases, which is consistent with the observation that the relative abundances of proteins in the F$_1$ hybrids, on average, deviate towards the high parental values (Leonardi et al., 1988; de Vienne et al., 1988).

The PQL approach applied to the phosphoglycerate mutase variations

The example of the phosphoglycerate mutase (PGAM, EC 5.4.2.1) may illustrate the power of the PQL approach. This monomeric enzyme catalyses the interconversion of 3-phosphoglycerate and 2-phosphoglycerate, and in plants is independent from the cofactor 2,3-bisphosphoglycerate (Carreras et al., 1982). RFLP mapping using a cDNA probe of this enzyme (gift of Puigdomenech, Barcelona, Spain), allowed us to integrate 4 loci in the maize genetic map previously published.

Fig. 1. Distribution of the ratios of maximum to minimum spot intensities for 124 coleoptile silver-stained proteins significantly variable in 21 maize lines. The spots were quantified using the KEPLER software package (Large Scale Biology Corp., Rockville, Maryland, USA).
(Causse et al., 1996), one on chromosome 2 near the marker umc14, two on chromosome 3 (Fig. 2), and one on chromosome 6 near bnl8.23. On the other hand, the protein was identified on 2-D gels from coleoptiles of etiolated seedlings by micro-sequencing (Touzet et al., 1996), and by amino acid composition on 2-D gels from blades of the 7th leaf and from endosperms 35 d after pollination (unpublished data). A position shift of the protein was observed between the parental lines ‘Io’ and ‘F-2’ in the three organs, which segregated 1:2:1 in the F2 progeny. As discussed by de Vienne et al. (1996), such positional variations are very generally due to polymorphism of structural genes of proteins. Interestingly, the position shift of this protein was cosegregating with one of the RFLP revealed on chromosome 3 using the cDNA probe, indicating that, in the three organs analysed, this gene is expressed (Fig. 2). The other genes can be pseudogenes, genes coding other forms of PGAM not yet detected on the 2-D gels, or genes expressed in other organs or under other environmental conditions. The amount of PGAM was quantified in the individuals of the progeny (in the heterozygotes the volumes of the two spots were added). Only one major PQL was detected, the most likely position of which corresponded to the expressed locus on chromosome 3 (Fig. 2) (full paper in preparation). Thus, combining proteome analysis and PQL mapping allowed the identification of the genomic region wherein the molecular basis of the positional and quantitative variations lies. This approach represents a new tool in molecular genetics, since the continuous variation can now be subjected to genetic analysis.

PQLs as a tool for characterizing QTLs—the example of responses of maize to water deprivation

Functional and positional candidate genes

In plant literature, QTLs have been mapped for a large variety of traits, including developmental and architectural traits, yield components, stress responses, disease and pest resistances ( Tanksley, 1993; Stuber, 1995; McCouch and Doerge, 1995, for reviews), and even biochemical and molecular traits (see above, and Prioul et al., 1997). However, very few have been isolated and characterized at the molecular level. It should be noted that ‘characterizing QTL’ does not necessarily mean ‘cloning novel genes’. Of course, large-effect QTLs for which there is no convincing candidate gene could be isolated by transposon tagging or by positional cloning in small genome species; to our knowledge, there is so far no publication on QTL cloning with such strategies, although works are in progress in some laboratories. On the other hand, as a consequence of the genome projects, a growing number of known-function genes are stored in databases and physically and/or genetically mapped, which may provide the geneticist with putative candidate genes for the traits under study. Actually in the literature the concept of candidate genes corresponds to two non-exclusive situations (Table 1): (i) when the physiological bases of a trait are known, it is possible to make assumptions about the known-function genes whose polymorphism may affect the trait’s variation. For example, in maize, many of the regulatory and enzyme genes of the flavonoid pathway have been cloned, and represent ‘functional’ candidate genes for the concentration of maysin in the silks, a C-glycosyl flavone synthesized along a branch of the pathway. A genetic analysis suggested that one of them, pl, encoding a transcription activator, is a large-effect QTL for this trait (Byrne et al., 1996); (ii) in most

Table 1. QTL characterization: cloning versus candidate genes

<table>
<thead>
<tr>
<th>Cloning</th>
<th>Candidate gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insertional mutagenesis or positional cloning (small genome species)</td>
<td>Gene QTL? Gene?</td>
</tr>
<tr>
<td>Functional gene</td>
<td>positional QTL</td>
</tr>
<tr>
<td>Candidate gene</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2. Genetics of the phosphoglycerate mutase (PGAM) variations in a maize F2 progeny. Two of the four RFLP loci mapped using a cDNA probe for this enzyme are on chromosome 3, one near umc10 and the other 16 cM apart. The position shift of the PGAM observed on 2-D gels from coleoptile, leaf and endosperm cosegregates with the second cDNA locus. The dotted line indicates the threshold of the LOD score value computed with the software MAPMAKER QTL.
cases there is no sufficient physiological information, and candidate genes may be proposed \textit{a posteriori}, due to the observation of an apparent coincidence between a QTL and a mutation/gene previously known on the genetic or physical map of the species. Such ‘positional’ candidate genes are commonly mentioned in the QTL literature, but usually the question as to whether the association is due to fortuitous genetic linkage or to an actual physiological relationship is not examined further. To our knowledge, the best achieved results in that direction have been published by Doebley’s group. In progeny derived from maize × teosinte crosses, Doebley and Stec (1993) found that QTL(s) for architecture differences between maize and teosinte mapped in a region of chromosome 1 known to contain \textit{Tb1}, a gene controlling the branched aspect of maize. Further genetic (Doebley et al., 1995) and molecular (Doebley et al., 1997) experiments confirmed the very likely role of \textit{Tb1} as one of the QTLs for maize domestication.

\textbf{Candidate genes and proteins}

Whether the candidate gene is functional or positional, there is usually no direct or simple strategy to demonstrate that its polymorphism actually explains a part of the variation of the trait. A complementation test such as the one used by Doebley et al. (1995) is very powerful, but unfortunately will remain restricted to QTLs with clear-cut effects. It is worth noting that transformation, when possible, will pose particular problems: the goal is not substituting a deficient by an active allele, or inactivate a transcript by an antisense strategy, but to compare the effects of two (or more) active alleles (the polymorphism at QTLs is not generally expected to be of all-or-null type, a view consistent with the first examples of identified QTLs). Other non-exclusive strategies include searching for the effects of the candidate gene in different genetic backgrounds, comparing allele sequences, and detailed functional analysis of the candidate gene.

In this connection, measurement of the variation of quantity or activity of the product of a positional candidate gene may represent an essential step for the validation. If the enzyme, regulatory factor or other product displays no genetic variability of quantity or activity, the candidate can hardly be retained. In other words a necessary (but not sufficient) condition for retaining a positional candidate is that a QTL for the quantity (PQL) and/or activity of its product is detected in the chromosomal region exhibiting the apparent co-location of the structural gene and the trait’s QTL (Fig. 3a).

This approach can also result in identifying ‘candidate proteins’, that is to say proteins whose genetic factors controlling quantity/activity appeared colocated with trait’s QTL(s), while the structural gene does not (Fig. 3b). Given the large confidence interval of PQL/QTL positions, co-locations in a single region could just be due to fortuitous linkage. But the same associations observed on two (or more) different chromosomes would be more likely to reveal possible physiological link between the protein and the trait. That hypothesis could be tested by studying the consequence of over- or under-expression of the protein in transformed plants. Such proteins would provide an access to the cloning of QTLs through the search for the molecular bases of their variations of quantity and/or activity (for example, by molecular analyses of the regulatory regions).

\textbf{PQLs of proteins responsive to mild water stress in maize}

The candidate gene/protein approach was applied to agronomic and physiological traits responsive to mild water deprivation in maize. The 100 individuals of a recombinant inbred line (RIL) population derived from a cross between a sensitive line (‘Io’, an American dent line from the Iodent group) and a tolerant line (‘F-2’, a flint line from the Institut National de la Recherche Agronomique, France) were characterized in the field for agronomic traits (yield, anthesis-silking interval [ASI], leaf senescence), in a greenhouse for physiological and morphological traits (photosynthesis, water status, carbon metabolism, abscisic acid [ABA] content, survival, growth during application of the stress) and for the amounts of individual proteins affected by water-stress as revealed on 2-D gels of 6th leaf proteins. Microsequences of internal fragments of 19 of these putative candidate proteins allowed the identification, or tentative identification, of 16 of them (Riccardi et al., 1998), among which...
Fig. 4. Apparent colocation of PQLs of proteins affected by drought stress and QTLs of traits responsive to mild water stress in maize. (a) The structural gene of ASR1 maps in a region of chromosome 10 containing both a locus controlling ASR1 amount and QTLs for leaf senescence, ABA content and ASI. (b) The three PQLs of protein P71 are associated with QTLs of growth on chromosomes 1, 4 and 8. (+) and (−) refer to high and low alleles from the parental line ‘Io’.

**ASR1**, an ABA/water stress/ripening induced protein, increased by a factor 2.1 in ‘Io’, but was never detected in ‘F-2’. ASR1 was initially described in tomato (Iusem *et al.*, 1993), and is also induced by water stress in *Solanum chacoense* (Silhavy *et al.*, 1995) and *Pinus taeda* (Chang *et al.*, 1995). As Iusem *et al.* (1993) found the protein in the nucleus, it was hypothesized that ASR1 could be involved in the protection of DNA structure or in gene regulation during water loss by changing DNA topology (Silhavy *et al.*, 1995).

A genetic map of more than 200 RFLP markers (Causse *et al.*, 1996, and unpublished data) was used for mapping PQLs and QTLs for agronomic or physiological traits. One to 5 PQLs were found for 47 induced or repressed proteins, some of them displaying apparent coincidences with QTLs which could be physiologically relevant. The RFLP observed with the cDNA of ASR1 (the maize cDNA was isolated by M Hoefer *et al.*, unpublished results) co-segregated with the presence or absence of the protein. This locus mapped on chromosome 10, in a region exhibiting a QTL for xylem sap ABA content, a QTL for leaf senescence, and a QTL for ASI (Fig. 4a). The latter traits were highly correlated in the progeny \((r = 0.58, P < 0.0001)\). The most parsimonious interpretation is to consider that the polymorphism of the structural gene of ASR1 would be responsible for the presence or absence variation of the protein, which in turn would affect pleiotropically the other responsive traits.
Experiments are in progress to verify this hypothesis. Another drought-induced protein (P71, pl 5.5 and M, 39 kDa), the sequence of which was not determined, had three PQLs with apparent co-location with QTLs of growth under stress on chromosome 1, 4 and 8 (Fig. 4b). The high-alleles of PQLs are associated with the low-alleles for growth, i.e. the more abundant the protein, the more the growth is reduced during the stress. Further experiments are needed to confirm the role of this protein in the trait variation, and that of other induced proteins. In this connection, it is worth noting that the quantity of a protein is only one component of the protein efficiency. QTLs for enzyme activity, which integrates both quantity and specific activity, can also be used as tools for candidate gene/protein validation (Causse et al., 1995; Prioul et al., 1997; Prioul et al., 1999).

Conclusion

In spite of the lack of sensitivity for the detection of low-abundance proteins, 2-D PAGE remains the most powerful technique to reveal the proteome, i.e. to give an image of the genome translated in any given developmental or environmental context. Owing to genome projects and physiological studies, a function can be assigned to an increasing number of proteins. The marker-based PQL methodology makes it possible to map the loci involved in the numerous individual protein variations and, further, to analyse the pleiotropic effects and epistatic interactions operating in the cell. This ‘molecular quantitative genetics’ may help in analysing metabolic and regulatory bases of the variation of macroscopic traits.

References


Humphery-Smith I, Cordwell SJ, Blackstock WP. 1997. Proteome research: complementarity and limitations with respect to the RNA and DNA worlds. Electrophoresis 18, 1217–1242.


Klose J. 1982. Genetic variability of soluble proteins studied by two-dimensional electrophoresis on different inbred mouse strains and on different mouse organs. Journal of Molecular Evolution 18, 315–328.


