ELSEVIER

Journal of Chromatography B, 782 (2002) 137–149

JOURNAL OF CHROMATOGRAPHY B

www.elsevier.com/locate/chromb

Review

Combining proteomic and genetic studies in plants

Hervé Thiellement[a,*], Michel Zivy[a], Christophe Plomion[b]

[a]*Unité Mixte de Génétique Végétale, INRA/CNRS, la Ferme du Moulon, F-91190 Gif-sur-Yvette, France*
[b]*INRA, Equipe de Génétique et Amélioration des Arbres Forestiers, F-33610 Cestas, France*

Abstract

Plant proteomics is still in its infancy, although numerous experiments have been undertaken since the end of the 1970s. In this review we focus on the interactions between proteomics and genetics. A given genome can express various proteomes according to differentiation, development, tissues, cells and subcellular compartments, and proteomes are modified in function of biotic and abiotic environment. These different proteomes and the way they respond to environment can be compared between genotypes, allowing the characterization of mutants or lines, the study of mutation pleiotropic effects, the genetic mapping of expressed genes. These comparisons also permit to hypothesize for ''candidate proteins'' that might be involved in the genetic variation of traits of economic or agronomic interest.

Keywords: Reviews; Proteomics; Genomics

Contents

*Corresponding author. Tel.: +33-1-6933-2334; fax: +33-1-6933-2340.
E-mail address: thiellement@moulon.inra.fr (H. Thiellement).

1570-0232/02/$ – see front matter
PII: S1570-0232(02)00553-6

1. Introduction

With the completion of the sequence of the first plant genome, of *Arabidopsis thaliana* in 2000 [1], plant biology has also turn the century by entering the so-called post genomic era and, as other life sciences, has developed new approaches. We will briefly describe and discuss in the following pages the developments in plant proteomics and their relationships to plant genetics. Several reviews have been published in the last few years that can be read as a useful complement to the present contribution [2–5]. As yet described in these reviews, the opportunity given to reveal several hundreds to few thousands of gene products on one single 2D gel—by the means of two-dimensional electrophoresis of denatured proteins—permitted to examine:

(1) variations in gene expression, according to the plant development and in response to various abiotic and biotic stresses or treatments, leading to the identification of regulated proteins;
(2) genetic variations: mutants lines have been characterized, genetic distances have been estimated, phylogenetic relationships have been established and factors controlling protein expression have been mapped.

The position of proteins on 2D gels depends on their primary sequence, and any mutation (amino acid substitution, insertion/deletion) that has an effect on the p*I* or on the mobility in SDS–PAGE will modify the position of the protein: they lead to position shift (PS) variants (Fig. 1). Mutations leading to the absence of the protein will produce presence/absence (P/A) variations (Fig. 1). The latter can also be due to PS: one of the two spots being masked by another spot or its p*I* being outside of the pH range of the isoelectrofocusing. In most cases, observed PS and P/A have been shown to be under monogenic control and indeed correspond to allelic variations (reviewed in Ref. [6]). Such markers are physiologically relevant in that they reveal loci whose transcripts are translated in the organ analysed. The other type of variations that can be observed in 2DE is the variation in amount of a same protein spot in different genotypes (genetically determined quantitative variations) or in different developmental stages or organs.

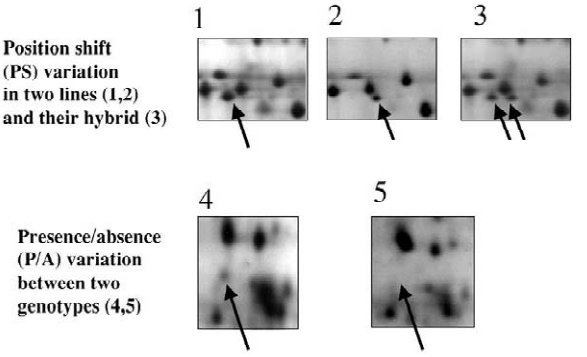

Fig. 1. The two qualitative spot variations detected by comparing proteomes from different genotypes.

Although the term ''proteome'' was introduced in a conference by Wilkins in 1994, to refer to the total protein complement of a genome, the roots of this modern concept date back to 1975 with high-resolution two-dimensional polyacrylamide gel electrophoresis. 2DE is still today the most resolutive technique for the analysis of complex protein mixtures, but it does not permit protein identification. The development of mass spectrometry techniques (peptide mass fingerprinting and peptide sequencing (reviewed in Ref. [7])), not only allowed to identify the proteins showing variations in function because of genetic variation or physiological changes, but also made it possible to undertake protein inventories in different plant structures (organs, tissues, cells, organelles, ribosomes). The topics discussed in this review, on genetically oriented plant proteomics, are summarized in Fig. 2.

2. Proteomes of a genome

2.1. Differentiation and development

The proteomes of the different organs of a plant are obviously different. They are often studied separately, e.g., in proteome databases [8], but comparisons between them are scarce, and most of them are actually related to the study of genetic variations.

Several studies have demonstrated that organ-spe-

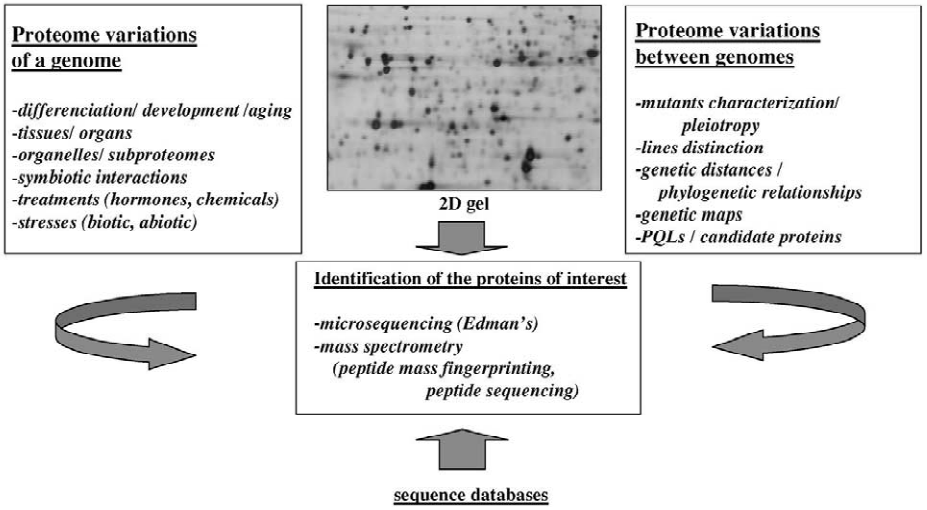

Fig. 2. Combining proteomic and genetic studies.

cific proteins are more variable between genotypes than organ-unspecific proteins, and that the level of genetic variability depends on the organ or tissue considered. In maritime pine, Bahrman and Petit [9] examined the genetic variation of proteins in three organs (needle, bud and pollen) of 18 unrelated trees. Of 902 polypeptides scored for their presence/absence variations, staining intensities and position shift variations (see Fig. 1), 27.2% were polymorphic. Among spots common to all three organs, only 18.1% were variable, while organ-specific polypeptides were characterized by a very high level of polymorphism (56% on average, ranging from 44.4% for spots specific to needles to 58.1% for pollen polypeptides and to 70.4% for those of the buds).

A high level of variability for organ-specific polypeptides was also found in maize. By comparing the proteins expressed in three organs (mesocotyl, second leaf's sheath and blade of 3-week-old seedlings) between two inbred lines, Leonardi et al. [10] found a genetic qualitative or quantitative variation for 3.6% of 357 proteins showing no variation between organs, while 22.6% were found variable among 629 proteins showing organ-specific variations. Later, this comparison was extended to two additional genotypes: 59% of the proteins showing a variation, mainly quantitative, between organs were genetically variables, to be compared to 18% of the stable proteins. The authors suggested [11] that the higher level of genetic variation of organ-specific proteins amounts might be related to a higher number of genes controlling their expression: the higher this number is, the higher the number of possible targets for mutations affecting protein amount. This hypothesis was supported by results in maritime pine [9] where position shift variants were twice as frequent and quantitative variants five times as frequent among organ-specific spots as among spots found in every analysed organ.

Reduced ''functional constraints'', as defined by Kimura [12], might also explain the increased level of allelic variability of organ-specific polypeptides that are expressed in a single cellular environment, compared to ''house keeping'' ubiquitous proteins, as suggested by Klose [13].

In other proteomic studies on development the objective is not to compare the differences between organs or to identify extensively all the proteins of a given organ, but to analyse the physiological events occurring during a precise phase of differentiation. A first example is given by Gallardo et al. [14], who studied the germination of seeds in *A. thaliana*. Changes in abundance were found for 39 proteins during germination sensu stricto and for 35 others during radicle protrusion. Variations of protein expression were also found during priming, i.e., pregermination followed by drying, a treatment that

allows faster germination. Eighty-four proteins were identified by MALDI-TOF. This analysis allowed the association of several proteins with different processes (e.g., imbibition of the seed, dehydration, mobilization of storage proteins, etc.). Some of these associations were already known, while others were noticed for the first time. A second example was given by Plomion et al. [15] for proteins involved in wood formation. They correlated the variation of a protein involved in ethylene (a gaseous phytohormone) biosynthesis with mechanical and biochemical wood related traits.

Although the objectives are different, these studies are similar to the study of proteome response to environment (see Section 3).

2.2. Subcellular compartments

Most of the available proteomes correspond to the abundant and soluble proteins, i.e., in no case are the cell proteins exhaustively separated and visualized. In particular, the nuclear and hydrophobic membrane-associated proteins are not extracted. In addition, the most basic and more generally, the low abundance proteins (e.g., transcription factors, protein kinases, etc.) are not seen on standard 2DE protein maps. The recovery of more proteins can be achieved in several ways among which the analysis of proteins after sub-fractionation according to cell type and subcellular compartments is of the foremost interest.

To be able to hypothesize the function of an unknown protein, one of the criteria to take into account is its localization inside the cell, i.e., to which compartment, to which organelle this protein belongs. Also, one of the limitation of proteomics being the low representation of the genome expression, with 2000–3000 gene products revealed in best 2D gels instead of the 10 000 to 15 000 genes expressed in the same tissue and same developmental stage, fractionation into the different subcellular compartments is becoming a necessity. A pioneer work by Granier et al. [16] has attributed to the proteins extracted from wheat etiolated seedlings their subcellular localization, using enriched fraction of mitochondria and chloroplasts.

One of the first organelle proteome projects in plants was the EU-founded collaborative program on plant plasma membrane. Santoni et al. [17] describe their objectives and their first accomplishments including the identification of proteins on their web site (http://sphinx.rug.ac.be:8080/ppmdb/index.html) with clickable spots on the protein map. Since it is difficult to reveal hydrophobic and basic proteins on 2D gels, significant work has been done to overcome these difficulties [18–21]. Using liquid-grown *Arabidopsis* calli, one partner of this project generated a database where the proteins are assigned to mitochondria, endoplasmic reticulum, Golgi/prevacuolar compartment or plasma membrane [22].

More recently, several papers described different plant ''subproteomes''. In the same issue of *Plant Physiology* of December 2001, two groups [23,24] undertook the description of the mitochondrial proteome of *Arabidopsis thaliana*, identifying 81 and 52 protein spots, respectively. In one case [23], mitochondrial proteins were sub-fractionated into soluble and membrane bound to better hypothesize their functional role.

The chloroplast, the organelle where the photosynthesis takes place, has also been the subject of different proteomic analysis and a short review on this specific organelle was published in 2000 [25]. A first description of the chloroplastic proteins was published in pea [26], but efforts are now directed towards the fully sequenced *Arabidopsis* and also towards the compartmentalization of the organelle, from the chloroplastic inner and outer envelope to the stroma and to the thylakoid membrane system [25]. A recently published study [27] described the protein composition of the lumenal compartment of the thylakoids, in *Arabidopsis* for easier identification, but the corresponding thylakoid lumen subproteome of spinach was compared also in the same study, both proteomes showing similarities in number and relative amounts of spots. A nice example of complete inventory is the study of ribosomal proteins of spinach chloroplasts, that allowed interesting comparisons with the structure of bacterial ribosomes [28,29]. New complexes can also be identified: Peltier et al. [30] identified the components of a protease complex in chloroplasts of *A. thaliana*. Interestingly, mass spectrometry data allowed them to improve the genome sequence annotation by correcting the intron and exon boundaries predicted from genomic sequences.

3. Biotic and abiotic environment

3.1. Abiotic stresses

A great interest has been brought to plant response to abiotic stresses, mainly because of possible applications to breeding programs of cultivated species.

Elevated temperatures induce, in plants as in other organisms, the 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), e.g., Nover and Scharf [31] showed the induction of 48 HSPs in tomato cell cultures. As HSPs were (i) easy to identify as induced by HS treatment (no sequencing was necessary for identifying them), (ii) revealed on 2DE patterns, not only by radioactive labelling but also by silver staining, and (iii) a priori known to be correlated to thermal tolerance, genetic studies were early undertaken to investigate the correlation between HSP polymorphism and the genetic variation of tolerance to high temperatures. Zivy [32] investigated the genetic variability of HSPs in wheat. A high level of polymorphism was detected: 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 [33] 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.

Exposure to cold can induce freezing tolerance, but the situation was not the same as for high temperature and HSPs since no specific set of proteins was a priori known as induced by cold. Thus, 2DE was mainly used for the description of protein response to cold and for trying to identify induced polypeptides whose regulation could be correlated to cold acclimation: Meza-Basso et al. [34] showed the induction of proteins by low temperature in rape seedlings. Guy and Haskell [35] 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. [36] identified chilling-acclimation-related proteins in soybean, one of them being a heat shock protein (HSP70). In poplar [37], two families of high-molecular mass polypeptides were shown to accumulate in response to chilling treatment in cuttings as in in vitro raised shoots. Quantitative changes for 26 proteins were detected in response to cold treatment of potato tubers [38]. Long-term chilling tolerance of tomato fruit acquired by heat shock treatment was shown to be correlated to the persistence of HSPs [39]. A genetic approach was initiated by Danyluk et al. [40], who studied the response to cold treatment of three varieties of *Triticum aestivum*: one set of 18 proteins was transiently induced, and a second set of 53 induced proteins stayed at a high level of expression during the 4 weeks required to induce freezing tolerance. Thirty-four of the latter were expressed at a higher level in the most freezing tolerant line.

Water deficit induces numerous morphological, physiological and biochemical responses in plants: e.g., reduced growth of aerial parts, stomatal closure, leaf rolling, leaf senescence, osmotic adjustment. Many cellular functions are affected and different types of genes have been found to be induced by this stress (reviewed in Ref. [41]). Water deficit and high salinity have in common to decrease the osmotic potential, and some metabolic responses are common to both stresses. In several studies, induced proteins have been identified. Claes et al. [42] isolated a cDNA for a glycin-rich protein by using DNA probes synthesized according to microsequences of a polypeptide induced by salt stress in rice. Reviron et al. [43] identified a protease inhibitor induced by drought in rape leaves. Moons et al. [44–46] identified in rice a novel class of glycin-rich proteins, a series of group 3 LEA (late embryogenesis abundant) proteins, and a series of group 2 LEA proteins (also called dehydrins) by microsequencing and Western blotting. Among three rice varieties, the most tolerant to salinity synthesized higher amounts of these proteins. Rey et al. [47] showed the induction of a thioredoxin-like protein in potato chloroplasts by water stress, and Moons et al. [48] the induction of pyruvate orthophosphate dikinase in rice roots. Costa et al. [49] found 38 proteins induced by drought in needles of maritime pine. Among them, enzymes involved in the response to oxidative stress,

in heat shock response and in lignin biosynthesis were identified. Riccardi et al. [50] found 78 proteins affected by drought in the growing part of maize leaves, 50 of which being up-regulated. Microsequences were obtained for 19 of them, allowing the identification of proteins involved in several metabolic pathways in addition to hydrophilic proteins of different classes.

3.2. Plant–microbe interactions

In the last 2 years several groups have undertaken the proteomic approach of the symbiotic association between N-fixing bacteria and legumes, i.e., the root nodules. The symbiosis between soybean and *Rhizobium* was studied at the peribacteroid membrane interface, made up by the plant [51]. Of the 17 peribacteroid membrane proteins sequenced, only six are homologous to proteins of already known function. More recently, Saalbach et al. [52] examined pea root nodules and identified 46 proteins from the peribacteroid membrane and from the space between this membrane and the bacteroid one. Another symbiosis was examined by Natera et al. [53] between white clover *Melilotus alba* and the bacterium *Sinorhyzobium meliloti*. Differential expression of proteins between the nodules and the non-nodulated roots and between bacteroids and cultured bacteria led to the identification of a hundred of proteins among the few hundreds up or down regulated when are compared symbiotic and non-symbiotic metabolisms of the two partners.

Since most higher plants, including the nodule forming Fabaceae, are involved also in mycorrhizal symbioses, with fungi of the order Glomales, Bestel-Corre et al. [54] undertook a study to compare both symbiotic interactions, between *Medicago truncatula* and the arbuscular mycorrhizal fungus *Glomus mosseae*, and between *M. truncatula* and the bacterium *Sinorhyzobium meliloti*. No plant protein was found commonly induced by both symbionts, although such finding was expected. However, numerous proteins were identified, newly synthesized or up or down regulated, the identification of the plant proteins being greatly facilitated by the large EST (expressed sequence tag) library constructed in *Medicago truncatula*, the model plant for symbiotic interactions.

4. Polymorphism of genes encoding the proteins

As far as their products show an allelic variation, the genes encoding the revealed proteins can be mapped on the chromosomes, allowing expressed genes to be added to the genetic maps mainly established with non-coding DNA-markers. Qualitative variants such as P/A and PS (Fig. 1) have also been widely used to study relationships between genotypes, populations, species and genus.

4.1. Genetic relationships

The structure of genetic variability in natural populations has always been a subject of great interest to population geneticists, evolutionists and plant breeders. It is generally accepted that the choice of a molecular screening technology for analyzing the extent and distribution of genetic diversity in natural populations will depend on many factors (http://webdoc.gwdg.de/ebook/y/1999/whichmarker/). Because each type of marker presents advantages and limitations, a variety of techniques needs to be available, among others is 2DE. The strength of this technique is that protein loci sample the genome differently compared to most PCR-based techniques—2DE indeed reveals the genetic variability of the expressed genes only—and therefore provides a different level of information with respect to the diversity of questions being addressed. In addition, 2DE allows to reveal far more markers compared to isozyme electrophoresis for which few assays are available, and it was reported [6] that the 2DE revealed as many alleles per polymorphic loci as in isozyme studies.

For the last 20 years, experiments have been performed at various taxonomic levels to assess genetic differences using 2DE 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* (reviewed in Ref. [55]). 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, 2DE 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 2DE 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 [56]. This finding was confirmed by cytoplasmically encoded proteins: the large subunit of Rubisco and two forms of the β-ATPase [57] 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 2DE patterns was extended to species of the *Triticum* genus possessing different genomes [58]. In another experiment between different genus of the Triticeae tribe, i.e., the A and D genomes of *Triticum*, the H genome of *Hordeum* (barley) and the S genome of *Secale* (rye), the dendrograms obtained still reflected well the known phylogenetic relationships [59], although the number of spots found in common dropped down to 30% between *Triticum* and *Hordeum*, indicating that the limits of the method may have been reached.

In European oaks the relationships between the two closely related species *Quercus petraea* and *Quercus robur* were examined by Barreneche et al. [60]. They used 2DE 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.

In the Brassicaceae family of plants, to which *Arabidopsis* belongs, are encountered several species of agronomic interest such as rape (*Brassica napus*), cabbages (*B. oleracea*), mustards (*B. juncea* and *B. nigra*) and radishes (*Raphanus* sp.). Distance indices calculated by counting common and distinct spots among the representatives of these species led, as in Triticeae, to dendrograms reflecting well the genetic relationships between them [61].

In maritime pine (*Pinus pinaster*), Bahrman et al. [62] used 2DE to study the relationships between seven provenances of the natural range, and to evaluate the genetic variability existing within and between geographical origins. Taking advantage of the possibility to distinguish between allelic forms of protein loci in the megagametophyte (a nutritive haploid tissue surrounding the embryo of conifer seeds), a total of 968 spots were scored, from which 84% were variable. Based on this information, three main groups (namely Atlantic, Mediterranean and North African) could be distinguished, a genetic structure that was in agreement with terpene data. In another study, Petit et al. [63] showed that proteins revealed by 2DE displayed a similar level of genetic differentiation among populations than terpenes and isozymes, indicating the absence (or similar level) of selection acting on the three types of loci.

David et al. [64] studied the genetic differentiation of 11 wheat populations originating from a single one. All 11 evolved independently during 8 years in different locations. Thirty-nine out of 162 polypeptides taken into account in this analysis proved to be polymorphic between the populations. 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.

4.2. Characterization of genotypes and mutant lines

Back in the 1980s, Zivy et al. [65,66] distinguished three different wheat varieties by qualitative and quantitative differences of proteins revealed by 2DE. The use of alloplasmic lines (the same nuclear genome introduced by repeated backcrosses in alien cytoplasms) permitted to reveal genetic differences for cytoplasmically encoded proteins (e.g., the large subunit of Rubisco). Improvements of the 2DE technique at the level of extraction procedures [67,68] and larger gels has increased the number of genetic variations detected [69].

Although only a fraction of the mutations in the coding sequence are detectable on 2DE gels, the great number of gene products allowed many authors to characterize and distinguish unambiguously genotypes, even when belonging to related populations. This was published in wheat [70–72], barley [73–75], sugarcane [76] or in pepper [77,78] for instance.

Looking at the 2DE pattern of a mutant compared to the 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*, 2DE patterns of a series of mutants affected in the first steps of development were examined by Santoni et al. [79]. The amount of one protein, characterized as an isoform of actin, was found correlated to the length of the hypocotile. In tomato [80], the comparison of 2DE 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. In experiments with moss, Kasten et al. [81] 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 uncharacterised mutant was likely a cytokinin overproducer [82]. This hypothesis was later confirmed by cytokinin dosage [83]. The comparison of several late-flowering mutants of *Arabidopsis* revealed protein spots that appeared or disappeared compared to the wild type [84]. However, in the F2 offsprings of the crosses between mutants and wild type, none of the variable protein spots co-segregated with the flowering phenotype. In addition to the necessity of genetic confirmations, these experiments demonstrated that 2DE analysis can reveal the genetic heterogeneity of the ecotypes or lines used in mutagenesis experiments. Also it must be noticed that such proteome analysis examined only a fraction of the proteins.

The power of proteome analysis in mutant characterization is even more evidenced by studying pleiotropic mutations. Gottlieb and de Vienne [85] compared two near-isogenic lines of pea differing by the *r* gene that determines round (*RR*) or wrinkled (*rr*) seeds. In the proteomes of the mature seeds, nearly 10% of the spots differed in amounts, confirming the numerous known physiological differences between the two types of seeds. In maize Damerval and de Vienne [86] studied the pleiotropic effects of the *Opaque2* (*O2*) gene, which codes for a transcription factor. The comparison of 2DE patterns of *O2* and wild-type maize lines in several unrelated backgrounds 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 [87].

It is obvious today that proteomics is indeed useful and powerful for distinguishing genotypes, even in closely related backgrounds. Because of the level where this analysis takes place, i.e., relatively far from the DNA sequence, the differences observed may be numerous when the genetic differences are few. However, this allows one to decipher the multiple effects of a single mutation. In addition, only the protein level is pertinent for looking at posttranslational modifications that are also the subject of genetic variations.

4.3. Genetic maps

The last 10 years have seen a dramatic increase of molecular methods for mapping quantitative trait loci (QTLs) (e.g., Ref. [88]). While DNA-based tech-

niques are considered as the technique of choice for quickly saturating a genome, 2DE not only provide useful molecular markers for mapping the expressed genome, but also may provide ''candidate proteins'' to understand the biological function of QTLs (see Section 5.2). Despite this advantage, genetic mapping of protein markers has been reported in very few plant species (reviewed in Ref. [6]). Genetic localization of protein markers has mainly been reported for wheat, maize and maritime pine, although few PS loci were also mapped in other crops including barley [89] and pea [6]. In wheat, Colas des Francs and Thiellement [90] reported chromosomal localization of 35 proteins comparing euploid and ditelosomic lines. In maritime pine, Bahrman and Damerval [91] and Gerber et al. [92] reported linkage analysis for 119 and 65 loci, respectively. Plomion et al. [93,94] and Costa et al. [95] used a three-generation inbred pedigree of this species to map 68 proteins from haploid and diploid tissues. In maize, a composite linkage map showing the distribution of 65 PS loci was presented [6]. In pine and maize, protein loci were found on each chromosome, interspersed with other markers (RFLPs in maize, RAPDs and AFLPs in pine).

The use of marker-assisted selection (MAS) in breeding programs relies on the presence of linkage disequilibrium between marker loci and quantitative trait loci (QTLs). Because linkage disequilibrium decreases at each generation due to recombination, the efficiency of MAS will quickly decline unless markers are found that are physically linked to the QTLs, or in the extreme case, being the QTLs themselves [96]. 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, 2DE 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. In maritime pine, Gerber et al. [97] demonstrated the rationale of this approach. For several traits among which seed weight and growth related traits, they detected significant ''protein–trait'' associations among the 84 protein loci genotyped on 18 unrelated trees, suggesting some of these proteins to be responsible for the trait variation itself.

5. Polymorphism of genes controlling protein expression

5.1. Protein quantity loci

Damerval et al. [98] were the first in 1994 who investigate the genetic determinism of quantitative variation of proteins separated by 2DE using a QTL (quantitative trait loci) detection strategy. They used a linkage map constructed with RFLPs and PS loci segregating in a F2 progeny of maize to locate by interval mapping [99], the ''PQLs'' (protein quantity loci), that explain part of the spot intensity variation. For the 72 proteins analyzed, 70 PQLs were detected for 42 proteins, 20 of them having more than one PQL. PQLs were found to be distributed all over the genome. PQLs controlling the accumulation of needle proteins were also detected in a F2 progeny of maritime pine [100] and the same conclusions were drawn.

The question arises as to whether or not the variability of genetic expression and its consequences in terms of protein quantity variation may also play a role in the phenotypic variability.

5.2. Candidate proteins

During the last 10 years, 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 [101]. While successful, it is important to point out the limitation of such a QTL analysis. QTLs can only be localized approximately, i.e., the confidence interval of a QTL is large on the genetic map [102]. This means that QTL mapping does not allow identifying the underlying genes. Thus, trait dissection analysis, although already useful in marker-assisted breeding, can be viewed as a first step towards the identification of the genes controlling part of the quantitative trait variation. In addition, based on the total map length of the genome and on DNA content, the average physical

equivalent of 1 cM distance varies in huge proportions, e.g., 230, 460, 2500, 3500 and 12 000 kb in *Arabidopsis*, *Eucalyptus*, maize, wheat and pine, respectively. This means that the identification of the gene corresponding to the QTL by positional cloning [103] would be very difficult in most crop plants and forest trees.

The accessibility of the biological meaning of a QTL can proceed from the candidate gene (reviewed in Ref. [104]) and the candidate protein (reviewed in Ref. [105]) approaches. Co-localizations between QTLs and known-function genes have been already reported [104]. The main limitation of this approach is that a co-localization between a candidate gene and a QTL can be purely fortuitous.

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 was exemplified in maize. Using a population of recombinant lines, de Vienne et al. [105] found a co-localization between a candidate gene for drought-stress response (the *Asr1* gene, an ABA/water stress/ripening related protein) and QTLs for leaf senescence and anther-silking interval (a symptomatic trait of drought effect), as well as a major PQL controlling the expression of the ASR1 protein under drought condition. The PQL strategy was also applied in maritime pine for the glutamine synthetase (GS), an enzyme involved in nitrogen assimilation. This candidate gene for biomass production co-localized with a QTL for early growth and a PQL controlling the amount of GS [106].

The PQL strategy can also result in identifying proteins whose genetic factors controlling protein quantity and/or activity co-localized with agronomic trait's QTLs, while the structural gene does not. In this connection, de Vienne et al. showed [105] that three PQLs controlling the quantity of a single leaf protein and three QTLs of height growth in maize were co-localized.

5. Conclusion

Even if proteomic studies in plants have been undertaken for 20 years, plant proteomics is still in its infancy. It is only during the last few years that the application of mass spectrometry, together with the availability of one fully sequenced plant genome and, in many agronomic species, of thousands of sequenced cDNAs (ESTs), have permitted to go further than using the 2DE as a source of genetic markers.

Every experiment published in plant proteomics today is accompanied by a list of the identified proteins, and biochemical and biological hypothesis and models can readily be tested. As in medicine and in microbiology, the mass of data is exponentially growing and bioinformatic tools are urgently needed to cope with such a challenge.

Unfortunately only a very few web sites are, at the beginning of 2002, freely available to academic scientists, with organized 2DE databases, showing references protein maps with clickable spots linked to protein database like Swissprot (http://sphinx.rug.ac.be:8080/ppmdb/index.html, http://www.expasy.ch/cgi-bin/map2/def?ARABIDOPSIS (see Fig. 3), http://www.gartenbau.uni-hannover.de/genetik/AMPP).

The transcriptomic tools, such as high-density cDNA filters, cDNA microarrays or DNA chips [107–109], are usefully complemented by proteomics, since the amounts of a protein and of its mRNA are not well correlated [110–112], and since proteins turn over and posttranslational modifications cannot be studied at the DNA or RNA level.

Moreover, new approaches in proteomics are being developed such as: (i) isotope-coded affinity tags [113], (ii) two-dimensional liquid chromatography [114], (iii) protein chips for studying protein–protein interactions or protein interactions with other molecules (e.g., Ref. [115]) and (iv) recovery of multisubunit complexes using non-denaturing electrophoresis [116], that continued development of proteomic technology and will lead to the description of biologically complex network of protein regulation.

The future of biology will be the development of knowledge and data at every step of the connected worlds between the gene sequence and the living being. In addition to genomics, transcriptomics and proteomics, new levels of regulation are now amenable to analysis: RNomics [117], where the role of transcribed but not translated RNAs are studied, and metabolomics [118] where the products of metabolism are directly measured, one level of regulation

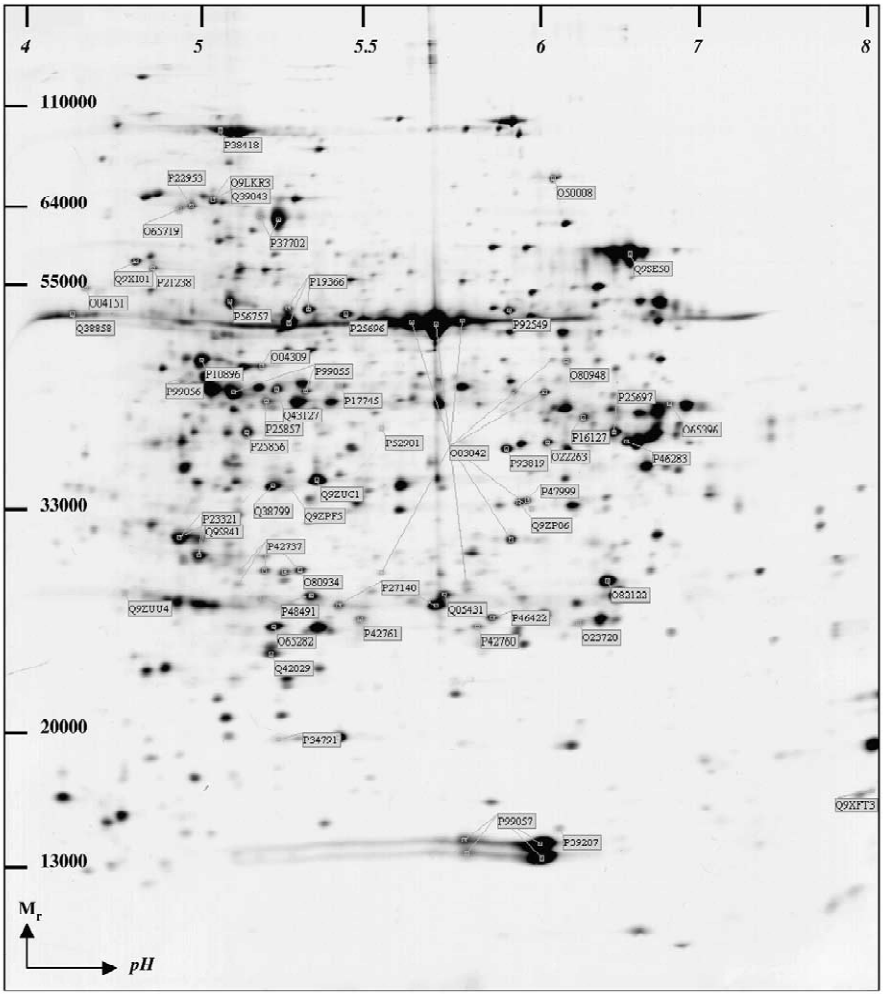

Fig. 3. The 2DE map of *Arabidopsis thaliana* leaf proteins, that can be found at http://www.expasy.ch/cgi-bin/map2/def?ARABIDOPSIS. As it happens for tissues with very abundant proteins, the large subunit of Rubisco (O03042) is found in many spots. Besides three main spots representing the native protein, several smaller polypeptides can be found on the gel resulting from in vivo or in vitro proteolysis.

ahead and one beyond than the one studied in proteomics.

References

[1] Arabidopsis Genome Initiative, Nature 408 (2000) 796.
[2] H. Thiellement, N. Bahrman, C. Damerval, C. Plomion, M. Rossignol, V. Santoni, D. de Vienne, M. Zivy, Electrophoresis 20 (1999) 2013.
[3] M. Zivy, D. de Vienne, Plant Mol. Biol. 44 (2000) 575.
[4] K.J. van Wijk, Plant Physiol. 126 (2001) 501.
[5] M. Rossignol, Curr. Opin. Biotechnol. 12 (2001) 131.
[6] D. de Vienne, J. Burstin, S. Gerber, A. Leonardi, M. Le Guilloux, A. Murigneux, M. Beckert, N. Bahrman, C. Damerval, M. Zivy, Heredity 76 (1996) 166.
[7] J. Li, S.M. Assmann, Plant Physiol. 123 (2000) 807.
[8] A. Tsugita, M. Kamo, T. Kawakami, Y. Ohki, Electrophoresis 17 (1996) 855.
[9] N. Bahrman, R. Petit, J. Mol. Evol. 41 (1995) 231.
[10] A. Leonardi, C. Damerval, D. de Vienne, Genet. Res. Cambridge 52 (1988) 97.
[11] D. de Vienne, A. Leonardi, C. Damerval, Electrophoresis 9 (1988) 742.

[12] M. Kimura, in: The Neutral Theory of Molecular Evolution, Cambridge University Press, Cambridge, 1983.
[13] J. Klose, J. Mol. Evol. 18 (1982) 315.
[14] K. Gallardo, C. Job, S.P. Groot, M. Puype, H. Demol, J. Vandekerckhove, D. Job, Plant Physiol. 126 (2001) 835.
[15] C. Plomion, C. Pionneau, J. Brach, P. Costa, H. Bailleres, Plant Physiol. 123 (2000) 959.
[16] F. Granier, H. Thiellement, F. Ambard-Bretteville, R. Rémy, Electrophoresis 7 (1986) 476.
[17] V. Santoni, D. Rouquie, P. Doumas, M. Mansion, M. Boutry, H. Degand, P. Dupree, L. Packman, J. Sherrier, T. Prime, G. Bauw, E. Posada, P. Rouze, P. Dehais, I. Sahnoun, I. Barlier, M. Rossignol, Plant J. 16 (1998) 633.
[18] M. Chevallet, V. Santoni, A. Poinas, D. Rouquie, A. Fuchs, S. Kieffer, M. Rossignol, J. Lunardi, J. Garin, T. Rabilloud, Electrophoresis 19 (1998) 1901.
[19] V. Santoni, P. Doumas, D. Rouquie, M. Mansion, T. Rabilloud, M. Rossignol, Biochimie 81 (1999) 655.
[20] V. Santoni, T. Rabilloud, P. Doumas, D. Rouquie, M. Mansion, S. Kieffer, J. Garin, M. Rossignol, Electrophoresis 20 (1999) 705.
[21] V. Santoni, S. Kieffer, D. Desclaux, F. Masson, T. Rabilloud, Electrophoresis 21 (2000) 3329.
[22] T.A. Prime, D.J. Sherrier, P. Mahon, L.C. Packman, P. Dupree, Electrophoresis 21 (2000) 3488.
[23] A.H. Millar, L.J. Sweetlove, P. Giege, C.J. Leaver, Plant Physiol. 127 (2001) 1711.
[24] V. Kruft, H. Eubel, L. Jansch, W. Werhahn, H.P. Braun, Plant Physiol. 127 (2001) 1694.
[25] K. van Wijk, Trends Plant Sci. 5 (2000) 420–425.
[26] J.B. Peltier, G. Friso, D.E. Kalume, P. Roepstorff, F. Nilsson, I. Adamska, K.J. van Wijk, Plant Cell 12 (2000) 319.
[27] M. Schubert, U.A. Petersson, B.J. Haas, C. Funk, W.P. Schroder, T. Kieselbach, J. Biol. Chem. 277 (2002) 8354.
[28] K. Yamaguchi, A.R. Subramanian, J. Biol. Chem. 275 (2000) 28466.
[29] K. Yamaguchi, K. von Knoblauch, A.R. Subramanian, J. Biol. Chem. 275 (2000) 28455.
[30] J.B. Peltier, J. Ytterberg, D.A. Liberles, P. Roepstorff, K.J. van Wijk, J. Biol. Chem. 276 (2001) 16318.
[31] L. Nover, K.D. Scharf, Eur. J. Biochem. 139 (1984) 303.
[32] M. Zivy, Theor. Appl. Genet. 74 (1987) 209.
[33] S. El Madidi, M. Zivy, in: H. Chlyah, Y. Demarly (Eds.), Le progrés génétique passe-t-il par le repérage et l'inventaire des génes? London:AUPELF-UREF, 1993, p. 173.
[34] L. Meza-Basso, M. Alberdi, M. Raynal, M.L. Ferrero-Cadinanos, M. Delseny, Plant Physiol. 82 (1986) 733.
[35] C.L. Guy, D. Haskell, Electrophoresis 9 (1988) 787.
[36] M. Cabane, P. Calvet, P. Vincens, A.M. Boudet, Planta 190 (1993) 346.
[37] J.F. Hausman, D. Evers, H. Thiellement, L. Jouve, Plant Cell Rep. 19 (2000) 954.
[38] J. van Berkel, F. Salamini, C. Gebhardt, Plant Physiol. 104 (1994) 445.
[39] A. Sabehat, D. Weiss, S. Lurie, Plant Physiol. 110 (1996) 531.
[40] J. Danyluk, E. Rassart, F. Sarhan, Biochem. Cell Biol. 69 (1991) 383.
[41] J. Ingram, D. Bartels, Annu. Rev. Plant Physiol. Plant Mol. Biol. 47 (1996) 377.
[42] B. Claes, R. Dekeyser, R. Villaroel, M. van den Bulke, G. Bauw, M. van Montagu, A. Caplan, Plant Cell 2 (1990) 19.
[43] M.P. Reviron, N. Vartanian, M. Sallantin, J.C. Huet, J.C. Pernollet, D. de Vienne, Plant Physiol. 100 (1992) 1486.
[44] A. Moons, G. Bauw, E. Prinsen, M. Van Montagu, D. Van der Straeten, Plant Physiol. 107 (1995) 177.
[45] A. Moons, A. De Keyser, M. Van Montagu, Gene 191 (1997) 197.
[46] A. Moons, J. Gielen, J. Vandekerckhove, D. Van der Straeten, G. Gheysen, M. Van Montagu, Planta 202 (1997) 443.
[47] P. Rey, G. Pruvot, N. Becuwe, F. Eymery, D. Rumeau, G. Peltier, Plant J. 13 (1998) 97.
[48] A. Moons, R. Valcke, M. Van Montagu, Plant J. 15 (1998) 89.
[49] P. Costa, N. Bahrman, J.M. Frigerio, A. Kremer, C. Plomion, Plant Mol. Biol. 38 (1998) 587.
[50] F. Riccardi, P. Gazeau, D. de Vienne, M. Zivy, Plant Physiol. 117 (1998) 1253.
[51] S. Panter, R. Thomson, G. de Bruxelles, D. Laver, B. Trevaskis, M. Udvardi, Mol. Plant Microbe Interact. 13 (2000) 325.
[52] G. Saalbach, P. Erik, S. Wienkoop, Proteomics 2 (2002) 325.
[53] S.H. Natera, N. Guerreiro, M.A. Djordjevic, Mol. Plant Microbe Interact. 13 (2000) 995.
[54] G. Bestel-Corre, E. Dumas-Gaudot, V. Poinsot, M. Dieu, J.F. Dierick, T.D. van, J. Remacle, V. Gianinazzi-Pearson, S. Gianinazzi, Electrophoresis 23 (2002) 122.
[55] K. Kerby, J. Kuspira, Genome 29 (1987) 722.
[56] N. Bahrman, M. Zivy, H. Thiellement, Heredity 61 (1988) 473.
[57] N. Bahrman, M.-L. Cardin, M. Seguin, M. Zivy, H. Thiellement, Heredity 60 (1988) 87.
[58] H. Thiellement, M. Seguin, N. Bahrman, M. Zivy, J. Mol. Evol. 29 (1989) 89.
[59] M. Zivy, S. el Madidi, H. Thiellement, Electrophoresis 16 (1995) 1295.
[60] T. Barreneche, N. Bahrman, A. Kremer, For. Genet. 3 (1996) 89.
[61] K. Marques, B. Sarazin, L. Chane-Favre, M. Zivy, H. Thiellement, Proteomics 1 (2001) 1457.
[62] N. Bahrman, M. Zivy, C. Damerval, P. Baradat, Theor. Appl. Genet. 88 (1994) 407.
[63] R.J. Petit, N. Bahrman, P. Baradat, Heredity 75 (1995) 382.
[64] J.L. David, M. Zivy, M.-L. Cardin, P. Brabant, Theor. Appl. Genet. 95 (1997) 932.
[65] M. Zivy, H. Thiellement, D. de Vienne, J.-P. Hofmann, Theor. Appl. Genet. 66 (1983) 1.
[66] M. Zivy, H. Thiellement, D. de Vienne, J.-P. Hofmann, Theor. Appl. Genet. 68 (1984) 335.
[67] C. Colas des Francs, D. de Vienne, H. Thiellement, Plant Physiol. 782 (1985) 178.
[68] N. Bahrman, D. de Vienne, H. Thiellement, J.P. Hofmann, Biochem. Genet. 23 (1985) 247.
[69] C. Damerval, D. de Vienne, M. Zivy, H. Thiellement, Electrophoresis 7 (1986) 52.
[70] N.G. Anderson, S.L. Tollaksen, F.H. Pascoe, L. Anderson, Crop Sci. 25 (1985) 667.

[71] B.D. Dunbar, D.S. Bundman, B.S. Dunbar, Electrophoresis 6 (1985) 39.
[72] P. Picard, M. Bourgoin-Greneche, M. Zivy, Electrophoresis 18 (1997) 174.
[73] A. Görg, W. Postel, M. Baumer, W. Weiss, Electrophoresis 13 (1992) 192.
[74] H. Saruyama, N. Shinbashi, Theor. Appl. Genet. 84 (1992) 947.
[75] T. Abe, R.S. Gusti, M. Ono, T. Sasahara, Genes Genet. Syst. 71 (1996) 63.
[76] S. Ramagopal, Theor. Appl. Genet. 79 (1990) 297.
[77] A. Posch, B.M. van den Berg, W. Postel, A. Gorg, Electrophoresis 13 (1992) 774.
[78] A. Posch, B.M. van den Berg, C. Duranton, A. Gorg, Electrophoresis 15 (1994) 297.
[79] V. Santoni, C. Bellini, M. Caboche, Planta 192 (1994) 557.
[80] A. Herbik, A. Giritch, C. Horstmann, R. Becker, H.J. Balzer, H. Baumlein, U.W. Stephan, Plant Physiol. 111 (1996) 533.
[81] B. Kasten, F. Buck, J. Nuske, R. Reski, Planta 201 (1997) 261.
[82] V. Santoni, M. Delarue, M. Caboche, C. Bellini, Planta 202 (1997) 62.
[83] J.-D. Faure, P. Vittorioso, V. Santoni, V. Fraisier, E. Prinsen, I. Barlier, H. Van Onckelen, M. Caboche, C. Bellini, Development 125 (1998) 909.
[84] P. Tacchini, A. Fink, H. Thiellement, H. Greppin, Physiol. Plant 93 (1995) 312.
[85] L.D. Gottlieb, D. de Vienne, Genetics 119 (1988) 705.
[86] C. Damerval, D. de Vienne, Heredity 70 (1993) 38.
[87] C. Damerval, M. Le Guilloux, Mol. Gen. Genet. 257 (1998) 354.
[88] M.J. Kearsey, A.G.L. Farquhar, Heredity 80 (1998) 137.
[89] M. Zivy, P. Devaux, J. Blaisonneau, R. Jean, H. Thiellement, Theor. Appl. Genet. 83 (1992) 919.
[90] C. Colas des Francs, H. Thiellement, Theor. Appl. Genet. 71 (1985) 31.
[91] N. Bahrman, C. Damerval, Heredity 63 (1989) 267.
[92] S. Gerber, F. Rodolphe, N. Bahrman, P. Baradat, Theor. Appl. Genet. 85 (1993) 521.
[93] C. Plomion, P. Costa, N. Bahrman, Silvae Genet. 46 (1997) 161.
[94] C. Plomion, N. Bahrman, C.E. Durel, D.M. O'Malley, Heredity 74 (1995) 661.
[95] P. Costa, D. Pot, C. Dubos, J.M. Frigerio, C. Pionneau, C. Bodenes, E. Bertocchi, M.T. Cervera, D.L. Remington, C. Plomion, Theor. Appl. Genet. 100 (2000) 39.
[96] W. Zhang, C. Smith, Theor. Appl. Genet. 83 (1992) 813.
[97] S. Gerber, M. Lascoux, A. Kremer, Silvae Genet. 46 (1997) 286.
[98] C. Damerval, A. Maurice, J.M. Josse, D. de Vienne, Genetics 137 (1994) 289.
[99] E.S. Lander, D. Botstein, Genetics 121 (1989) 185.
[100] P. Costa, C. Plomion, Silvae Genet. 48 (1999) 146.
[101] S.D. Tanksley, Annu. Rev. Genet. 27 (1993) 205.
[102] B. Mangin, B. Goffinet, A. Rebaï, Genetics 138 (1994) 1301.
[103] J.M. Rommens, M.C. Iannuzzi, B.S. Kerem, M.L. Drumm, G. Melmer, M. Dean, R. Rozmahel, J.L. Cole, D. Kennedy, N. Hidaka, M. Zsiga, M. Buchwald, J.R. Riordan, L.C. Tsui, F.S. Collins, Science 245 (1989) 1059.
[104] S. Pflieger, V. Lefebvre, M. Causse, Mol. Breed. 7 (2001) 275.
[105] D. de Vienne, A. Leonardi, C. Damerval, M. Zivy, J. Exp. Bot. 50 (1999) 303.
[106] C. Plomion, N. Bahrman, P. Costa, C. Dubos, J.-M. Frigério, S. Gerber, J.-M. Gion, C. Lalanne, D. Madur, in: S. Kumar, M. Fladung (Eds.), Molecular Genetics and Breeding of Forest Trees, The Haworth Press, Binghamton, New York, USA, 2002, in press.
[107] A. Castellino, Genome Res. 7 (1997) 943.
[108] B. Lemieux, A. Aharoni, M. Schena, Mol. Breed. 4 (1998) 277.
[109] A. Marshall, J. Hodgson, Nat. Biotechnol. 16 (1998) 27.
[110] L. Anderson, J. Seilhamer, Electrophoresis 18 (1997) 533.
[111] P.A. Haynes, S.P. Gygi, D. Figeys, R. Aebersold, Electrophoresis 19 (1998) 1862.
[112] N.L. Anderson, N.G. Anderson, Electrophoresis 19 (1998) 1853.
[113] S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, Nat. Biotechnol. 17 (1999) 994.
[114] A.J. Link, J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik, J.R. Yates 3rd, Nat. Biotechnol. 17 (1999) 676.
[115] H. Zhu, M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R.A. Dean, M. Gerstein, M. Snyder, Science 293 (2001) 2101.
[116] H. Schagger, W.A. Cramer, G. von Jagow, Anal. Biochem. 217 (1994) 220.
[117] J.S. Mattick, EMBO Rep. 2 (2001) 986.
[118] O. Fiehn, Plant Mol. Biol. 48 (2002) 155.