Proteomic investigation of natural variation between *Arabidopsis* ecotypes

François Chevalier, Olivier Martin, Valérie Rofidal, Anne-Dominique Devauchelle, Samuel Barteau, Nicolas Sommerer and Michel Rossignol

Laboratoire de Protéomique, INRA, Montpellier, France

Two-dimensional (2-D) gel electrophoresis and peptide mass fingerprinting were used to investigate the natural variation in the proteome among 8 *Arabidopsis thaliana* ecotypes, of which 3 were previously shown to display atypical responses to environmental stress. Comparison of 2-D maps demonstrated that only one-quarter of spots was shared by all accessions. On the other hand, only 15% of the 25 majors spots accounting for half the total protein amount could be classified as major spots in all ecotypes. Identification of these major spots demonstrated large differences between the major functions detected. Accordingly, the proteomes appeared to reveal important variations in terms of function between ecotypes. Hierarchical clustering of proteomes according to either the amount of all anonymous spots, that of the 25 major spots or the functions of these major spots identified the same classes of ecotypes, and grouped the three atypical ecotypes. It is proposed that proteome comparison has the capacity to evidence differences in the physiological status of ecotypes. Results are discussed with respect to the possibility to infer such differences from limited comparisons of major proteins. It is concluded that classical proteomics could constitute a powerful tool to mine the biodiversity between ecotypes of a single plant species.

Keywords: *Arabidopsis* ecotypes / Hierarchical clustering / Peptide mass fingerprinting / Two-dimensional gel electrophoresis

Received	6/10/03
Revised	19/1/04
Accepted	21/1/04

1 Introduction

Proteomics is becoming an essential field to investigate plant growth and evolution. Very early, the exploration of proteome was successfully used to characterize the relationships between populations [1–4]. In recent years, such approaches allowed the establishment of distances for instance between different species of the Brassicaceae family [5], various wheat cultivars [6], or oriental and American ginseng [7]. On the other hand, to date, little attention was paid to the natural variation occurring among plant ecotypes by opposition to recent works in animal field [8, 9]. However, biodiversity constitutes a high potential resource for searching genes of interest, as it is well established for the model plant *Arabidopsis thaliana* [10–12]. On this model, large differences between ecotypes were observed for a variety of features such as, for instance, light and hormone sensitivity [13], seed size [14], the light-dependent hypocotyls growth [15], or the growth rate [16]. In a previous study on *Arabidopsis* root system architecture, we recently demonstrated that, whereas a majority of ecotypes responds to phosphate starvation by decreasing both growth of the primary root and initiation of lateral roots, other ecotypes use only one of these responses [17].

The present work was undertaken to explore the potential of proteomics to investigate natural variation within *Arabidopsis* ecotypes. For this purpose we selected a small number of ecotypes encompassing the four ones most commonly used in genetic or genomic studies and a variant from one of them as a control for close genetic distance (all of the five displaying the major response to phosphate starvation), as well as three ecotypes showing alternative responses to phosphate

Correspondence: Dr. François Chevalier, UR 1199 – Laboratoire de Protéomique, INRA, 2 place Viala, F-34060 Montpellier Cedex 1, France
E-mail: chevalie@ensam.inra.fr
Fax: +33-4-9961-3014

Abbreviations: MP, major protein; **TGS**, Tris-Glycine – SDS buffer

© 2004 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

starvation. Classical 2-DE and MALDI-TOF-MS were used to compare the distances between the proteomes of these ecotypes, with special emphasis to the proteins accounting for the majority of the protein amount in each ecotype.

2 Materials and methods

2.1 Plant material, culture conditions, and chemicals

Arabidopsis thaliana ecotypes Ws-1, Cvi-0, Col-0, Col-4, Ler-1, Be-0, and Ll-0 (Table 1) were progenies from NASC (http://nasc.nott.ac.uk). Plants were grown under hydroponic conditions [18] for 42 days before sample harvest. Urea, phosphoric acid, and acetic acid were from VWR (Fontenay-ss-Bois, France); CHAPS, Triton X-100, iodoacetamide, bromophenol blue and Comassie blue were from Sigma-Aldrich (St. Louis, MO, USA); glycerol, SDS, DTT, TGS, and Tris were from Euromedex (Mundolshein, France). IPG strips and buffer were from Amersham Biosciences (Buckinghamshire, UK), and acrylamide from Bio-Rad (Hercules, CA, USA).

Table 1. Origin of selected *Arabidopsis thaliana* ecotypes (http://nasc.nott.ac.uk)

Ecotype	NASC No.	Origin
Col-0	N1093	Columbia (USA)
Col-4	N933	Columbia (USA)
Be-0	N965	Bensheim (Germany)
Ll-0	N1339	Llagostera (Spain)
Rld-1	N 913	Netherlands (Koornneef group)
Cvi-0	N902	Cape Verdian Islands
Ws-1	N2223	Wassilewskija (Russia)
Ler-1	N1642	Landsberg (Poland)

2.2 Protein extraction

For each ecotype, one root sample corresponding to more than 150 root systems was grinded in liquid nitrogen, and the fine powder was mixed with 90% v/v acetone, 10% v/v TCA solution (100% w/v) and 0.07% v/v 2-mercaptoethanol. After incubation at −20°C for 30 min, insoluble material was pelleted at $42\,000 \times g$ with a TL 100 ultracentrifuge using a TLA 100.3 rotor (Beckman Coulter, Palo Alto, CA, USA). Pellets were washed three times with pure acetone containing 0.07% v/v 2-mercaptoethanol, air-dried, and solubilized in buffer containing 9 M urea, 4% w/v CHAPS, 0.05% v/v Triton X-100, and 65 mM DTT. Protein amount was estimated using the Bradford method. All manipulations were performed at 4°C.

2.3 Two-dimensional electrophoresis

For each ecotype sample, 2-DE was performed in triplicate from the same extract using 18 cm linear pH 4–7 IPG strips. 200 μg protein samples were supplemented with 0.5% v/v IPG buffer pH 4–7 and 0.002% w/v bromophenol blue. Strips were hydrated directly with protein solution. Isoelectric focusing was performed using an IPG-Phor device until 100 000 $kV \cdot h^{-1}$. Before the second dimension, the strips were reduced (50 mM Tris-HCl, pH 8.8, 6 M urea, 30% v/v glycerol, 2% w/v SDS, 130 mM DTT) and alkylated in the same buffer containing 130 mM iodoacetamide instead DTT for 15 min. Strips were then embedded using 0.6% w/v low-melt agarose in running buffer containing traces of bromophenol blue on the top of a 11% acrylamide gel. SDS-PAGE was carried out, using a 2-D electrophoresis DALT system, at 15 mA per gel overnight at 10°C. Gels were stained using colloidal Coomassie blue [19]. Images from stained gels were digitalized at 300 dpi with a GS 710 densitometer (Bio-Rad, Hercules, CA, USA) and analyzed using the Progenesis software (Nonlinear Dynamics, Newcastle upon Tyne, UK). Gel triplicates were matched to create an average gel with spots present at least on two of the three gels. Average gels corresponding to the different *Arabidopsis* ecotypes were compared and spots of interest were selected for subsequent protein identification by MALDI-TOF-MS analysis.

2.4 Protein identification by MALDI-TOF-MS

Spots were picked from preparative gels (500 μg proteins) using a spot picker robot (Perkin Elmer, Wellesley, MA, USA). Acrylamide pieces were collected in 96-well microplates with 50 μL acetic acid (1% v/v). Pieces of gel were then washed using a Multiprobe II robot (Perkin Elmer) in several steps with water, 25 mM ammonium carbonate, and acetonitrile. Proteins were digested with trypsin (12.5 μg/mL in 25 mM ammonium carbonate). Supernatants were mixed with equal volumes of matrix solution (α-cyano-4-hydroxycinnamic acid) and spotted onto targets. Peptide mass fingerprints were acquired using a Biflex III mass spectrometer (Bruker Daltonics, Bremen, Germany). Spectra were calibrated internally and annotated automatically. The MASCOT search engine software (Matrix Science, London, UK) was used to search NCBInr database. The following parameters were used for database search: mass tolerance of 100 ppm, a minimum of four matched peptides and one miscleavage allowed.

2.5 Data analysis

Protein amount in each spot was estimated from its volume after normalization with respect to the total volume of all spots detected in the gel. Euclidean or Manhattan

© 2004 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

distances between protein amounts among ecotypes were further used to compute a similarity matrix between ecotypes [20]. For hierarchical clustering, aggregation was made using the Ward criteria [20].

3 Results

3.1 Comparison of 2-D protein maps from *Arabidopsis* ecotypes

The first goal of this work was to evaluate the capacity of classical proteomics as a molecular tool to characterize the natural biodiversity in *Arabidopsis thaliana*. Therefore, eight different ecotypes (Table 1), originating from contrasted habitats and previously shown to display different developmental strategies [17] were selected, and root soluble protein extracts were separated by 2-DE. Based on preliminary work, 18 cm IPG (pH 4–7) and a second dimension covering the 15–150 kDa range were used to get large insights into the proteomes of these ecotypes. In order to obtain an information as quantitative as possible, gels were stained with colloidal Coomassie blue. Triplicate gels were first matched to create an average gel containing those spots observed at least two times among the three gels; thereafter, average gels from the different ecotypes were matched using the ecotype Col-0 as reference (Fig. 1). Over an average of *ca.* 250 spots

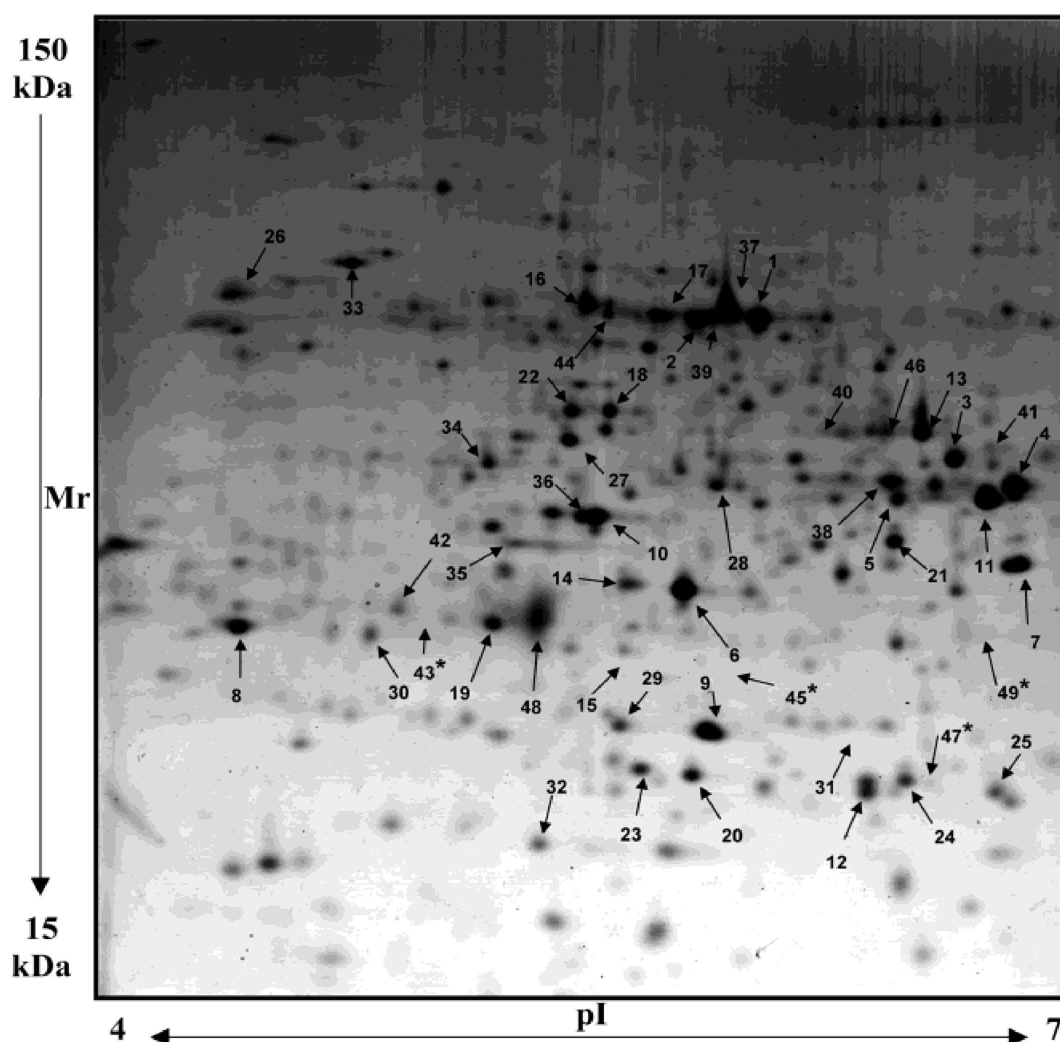

Figure 1. 2-D PAGE reference map of *Arabidopsis thaliana* soluble root proteins from ecotype Col-0. 2-D PAGE conditions: pH 4–7 IPG (first dimension) and 11% SDS-PAGE (second dimension). Proteins spots are visualized by colloidal Coomassie blue staining. Arrows indicate major spots analyzed by MALDI-TOF-MS. Spot numbers with star superscripts refer to spots from other ecotypes and not detected in ecotype Col-0.

© 2004 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

for each ecotype, one-quarter was found in all the ecotypes (Fig. 2), and half of them was shared by at least 75% of ecotypes. On the other hand, about 10% of all spots appeared to be specific for one ecotype.

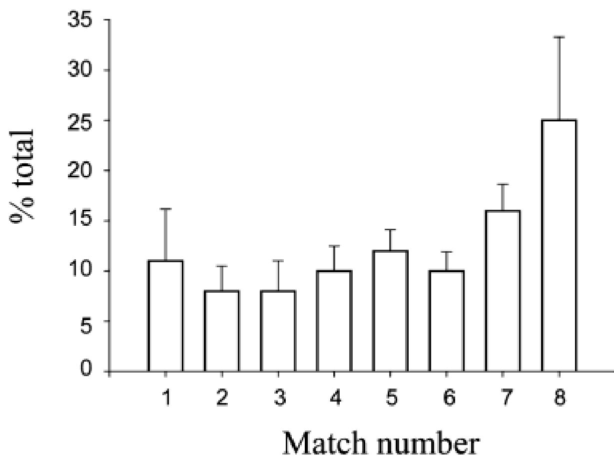

Figure 2. Distribution of spots according to the number of ecotypes where they were detected.

As simple gel comparison seemed to reveal specific features between the proteomes of *Arabidopsis* ecotypes, a classification was made by hierarchical clustering (Fig. 3a). For this purpose, the amount of each spot was estimated by its normalized volume as obtained by image analysis. Euclidian distances were then computed for all spots to build the similarity matrix for ecotypes, and clustering was performed using the Ward's method to link the variables. A first cluster corresponding to the two closest ecotypes was observed for Col-0 and Col-4. Another cluster grouped the ecotypes Be-0 and Ll-0, as well as the ecotype Rld-1, whereas the three last ecotypes defined a third cluster.

3.2 Comparison of *Arabidopsis* ecotypes from their major protein patterns

Above comparison of total 2-D maps indicated both the occurrence of contrasted proteomes between ecotypes and proximity between some of them. Additional analysis showed that the 25 most abundant spots from each *Arabidopsis* ecotype cumulated half the total protein amount detected on the p*I* and MW range investigated (50.2% ± 6.0% of the total signal computed by image analysis). In a first attempt to get more insights into the molecular basis of this structure, focus was given on these major proteins (MPs). Over the eight ecotypes, the 25 MPs from each defined a total subset of 49 differently matching spots accounting individually for between 0.7% and 6.7% of the total protein amount in their respective 2-D map.

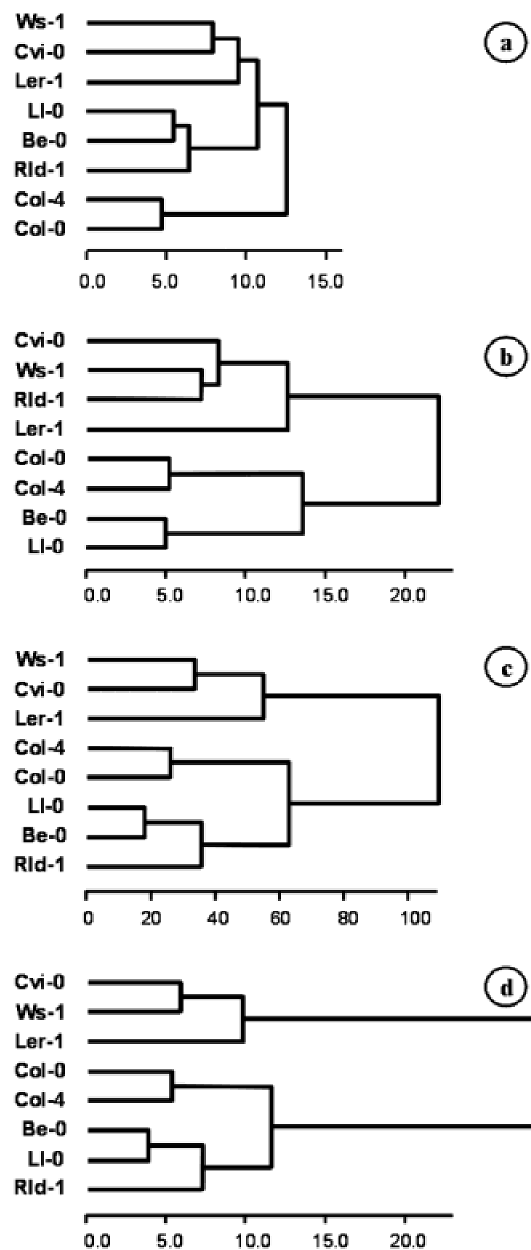

Figure 3. Classification of the proteomes from the eight ecotypes. Hierarchical clustering was performed using (a), (b), (d) the Euclidian distance or (c) the Manhattan distance as metric and the Ward criterion for linkage. (a) Spot abundance (all spots); (b), (c) abundance of MP spots; (d) function of MP spots.

Only 15% of these spots was ranked as MPs in all ecotypes and one-half contributed to the MPs in half or more the ecotypes, indicating the occurrence of large differences in expression level throughout the ecotypes (Fig. 4). Simultaneously, one-quarter of the MPs was not detected in at least another ecotype, whereas the remaining was observed at lower abundance and not classified as MP.

© 2004 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

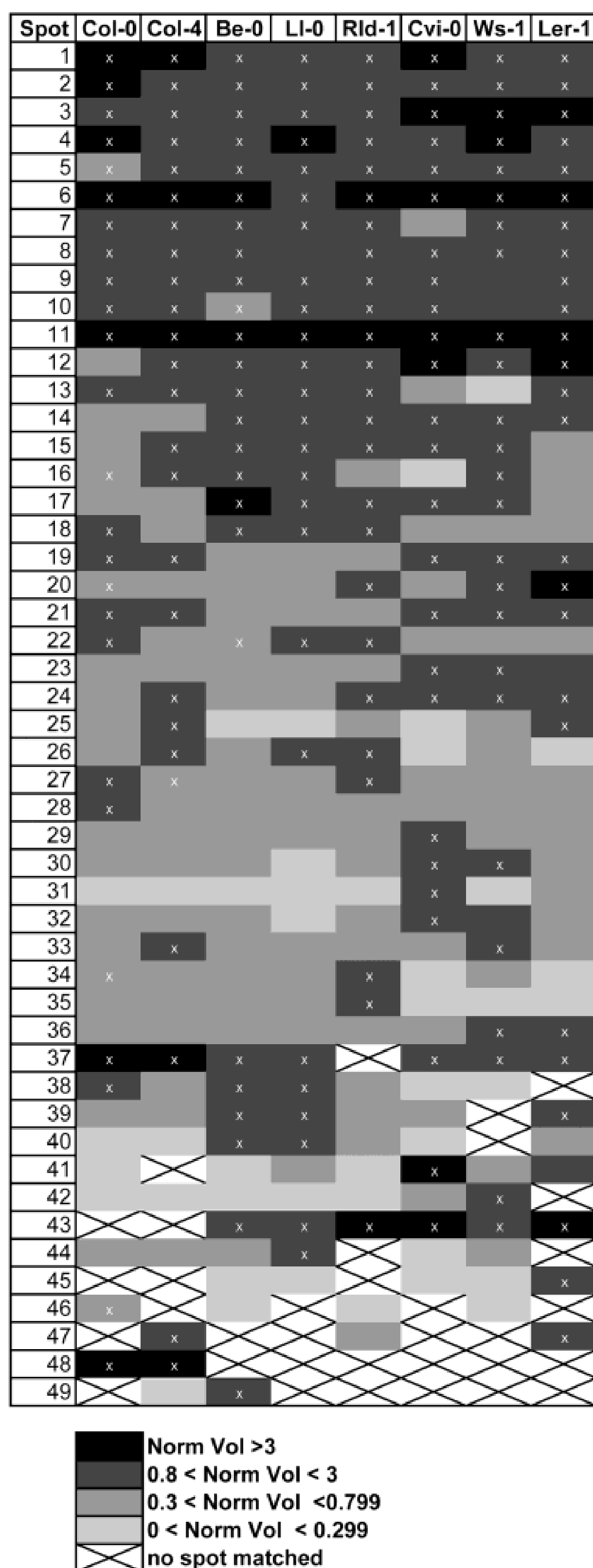

Figure 4. Expression of the 49 spots ranked as MPs in the eight ecotypes. Cross within cells indicate that the spot was ranked as MP in the corresponding ecotype.

As large differences in proteome features appeared to be detectable soon at the MPs' level, the different ecotypes were also classified by taking into account only the MPs. Figure 3b shows that, when using Euclidian distances as metric and the Ward's criterion for clustering, both the previous Col-0/Col-4 cluster and Be-0/Ll-0 cluster were still obtained. By opposition to the classification over the whole proteomes, however, the Rld-1 ecotype was no longer clustered with Be-0/Ll-0, but with the three remaining ecotypes. Nevertheless, the whole proteome-based classification could observed again by using other metric, such as the Manhattan distance (Fig. 3c), suggesting that the position of Rld-1 could be less stable.

3.3 Comparison of *Arabidopsis* ecotypes from the functions of major proteins

The 25 MPs of each ecotype were picked out from gels and processed for MALDI-TOF-MS analysis. In no case, contradictory identifications were obtained for spots that had been matched by image analysis. On the whole, 32 different protein accessions were identified in 44 from the 49 MP spots, and 5 could not be assigned to any protein accession (Table 2). Simultaneously, one half of the identified MPs was found to be present in 2 or 3 different protein spots. In this latter case, the difference between spots concerned mostly the p*I*, suggesting the occurrence of post-translational modifications or the presence of punctual protein variations between ecotypes. However, this encompassed different situations within ecotypes: (i) a same jasmonate-inducible protein was identified, for instance, in spots 6 and 14 (Fig. 5). These two spots showed a shift of 0.14 pH unit in p*I* and were observed simultaneously for the 8 *Arabidopsis* ecotypes, indicating that the protein existed under different states whatever the genotype was. However, only spot 6 corresponded to a MP in all ecotypes, the spot 14 showing a low abundance in the two Col accessions. Similar situations were observed for several other proteins, like the β-chain of the mitochondrial ATPase, a malate dehydrogenase, and a triose phosphate isomerase, indicating that the most frequent form of a protein could be ecotype-dependent. (ii) On the other hand, a same glutathione *S*-transferase could be detected in spots 25 and 47, differing by 0.17 pH unit in p*I* (Fig. 6). However, these two spots constituted MPs only in two ecotypes (Col-4 and Ler-1), and where either present at low abundance (spot 25) or not detected (spot 47) in other ecotypes. A similar situation was also observed for an alcohol dehydrogenase, suggesting that the occurrence of the modification itself could be ecotype-dependent. (iii) An even more contrasted situation was observed for a same blue-copper binding protein that was identified as MP in two spots

© 2004 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 2. Identified MPs in *Arabidopsis thaliana* ecotypes

Spot	Protein		Score	Coverage (%)	p*I*		MW (kDa)	
	Accession No.	Name			calc.	meas.	calc.	meas.
1	JQ1187	Phosphopyruvate hydratase	137	36	5.54	5.76	47.7	51.4
2	Q9C5B0	H+-Transporting ATP synthase β-chain	195	35	6.18	5.59	59.6	50.9
3	T02507	Peroxidase	94	18	5.66	6.29	38.7	35.8
4	T47550	Fructose bisphosphate aldolase-like protein	185	50	6.05	6.42	38.5	32.3
5	B86176	Malate dehydrogenase	110	31	6.11	6.13	35.5	31.6
6	AAK43865	Jasmonate inducible protein isolog	181	39	5.46	5.57	32.1	26.9
7	T51311	Malate dehydrogenase	66	15	8.54	6.43	36.0	27.9
8	T52558	Translation elongation factor eEF1Bα	69	26	4.42	4.47	24.2	25.9
9	T50646	Triose-phosphate isomerase	82	22	5.24	5.63	27.1	22.8
10	B84720	Probable fructokinase	183	42	5.31	5.33	35.3	30.3
11	B86176	Malate dehydrogenase	105	34	6.11	6.38	35.5	32.3
12	AAD00509	Germin-like oxalate oxydase	65	26	5.82	6.08	23.2	21.8
13	DEMUAM	Alcohol dehydrogenase	119	29	5.77	6.18	41.1	36.9
14	AAK43865	Jasmonate inducible protein isolog	123	35	5.46	5.52	32.1	26.7
15	T47550	Fructose bisphosphate aldolase	139	36	6.05	6.25	38.5	32.6
17	Q9C5B0	H+-Transporting ATP synthase β-chain	65	15	6.18	5.48	59.6	51.6
18	S68107	Actin 7	193	37	5.31	5.36	41.7	38.8
20	AAF98403	GSH-dependent dehydroascorbate reductase	78	23	5.56	5.59	23.6	22.1
21	T51862	Malate dehydrogenase	91	22	8.48	6.11	42.4	29.2
22	S18600	Glutamate-ammonia ligase	116	24	6.43	5.26	47.4	38.7
23	T52613	Chaperonin 21 precursor	124	36	8.86	5.47	26.8	22.4
24	BAA97523	1,4-Benzoquinone reductase like	78	40	5.96	6.17	21.8	22.3
25	G86159	Glutathione *S*-transferase	58	16	5.8	6.43	23.5	22.1
26	C96605	Calreticulin	85	21	4.46	4.42	48.7	54.3
27	AAG45246	Adenosine kinase	71	23	5.29	5.25	38.3	35.9
28	AAF02115	Reversibly glycosylated polypeptide	99	24	5.61	5.66	41.1	32.7
29	T50646	Triose-phosphate isomerase	157	54	5.24	5.42	27.1	23.1
30	Q9SS67	Putative peroxidase	87	20	5.04	4.73	34.5	25.2
31	APX1	L-Ascorbate peroxidase	60	19	5.72	5.81	27.4	22.3
32	AAD28242	Peroxyredoxin TPX1	82	25	5.21	5.18	17.4	21.3
33	AAD41430	Disulfide isomerase	203	49	4.64	4.76	55.6	62.9
34	H96686	Probable glutamine synthetase	76	23	5.14	5.10	39.3	33.7
35	AAD56335	60S ribosomal protein	64	24	4.52	5.13	24.3	28.5
36	Q9LSM8	Pyruvate dehydrogenase E1	106	19	5.67	5.30	31.2	30.1
37	JQ1187	Phosphopyruvate hydratase	141	33	5.54	5.66	47.7	51.5
38	T47550	Fructose bisphosphate aldolase	198	49	6.05	6.10	38.5	32.9
39	JQ1187	Phosphopyruvate hydratase	85	19	5.54	5.73	47.7	51.2
40	DEMUAM	Alcohol dehydrogenase	122	22	5.77	6.02	41.8	37.8
41	T13020	Peroxidase ATP 19A	96	28	6.53	6.40	35.6	35.2
42	Q9SS67	Putative peroxidase	87	20	5.04	4.89	34.5	26
43	T52408	Blue copper-binding protein II	60	22	5.47	4.84	20.5	26.1
45	S59519	Tryptophan synthase	82	26	6.77	5.66	33.3	24.7
47	G86159	Glutathione *S*-transferase	110	28	5.80	6.26	23.5	22.2
48	T52408	Blue copper-binding protein II	65	15	5.47	5.18	20.5	25.7

Proteins were identified from their peptide mass fingerprint by searching the NCBInr database (scores greater than 58 are significant at $p < 0.05$).

(spot No. 43 and 48) with a shift of 0.31 pH unit in p*I* (Fig. 5). In this case, only one or the other of the two spots could be detected in a given *Arabidopsis* ecotype, suggesting the occurrence of ecotype-specific isoforms.

In terms of function, most MPs corresponded to proteins involved in energy metabolism and oxydoreduction processes (Table 3), and different functions participating to a same pathway, such as glycolysis (3 functions in 8 MP

© 2004 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

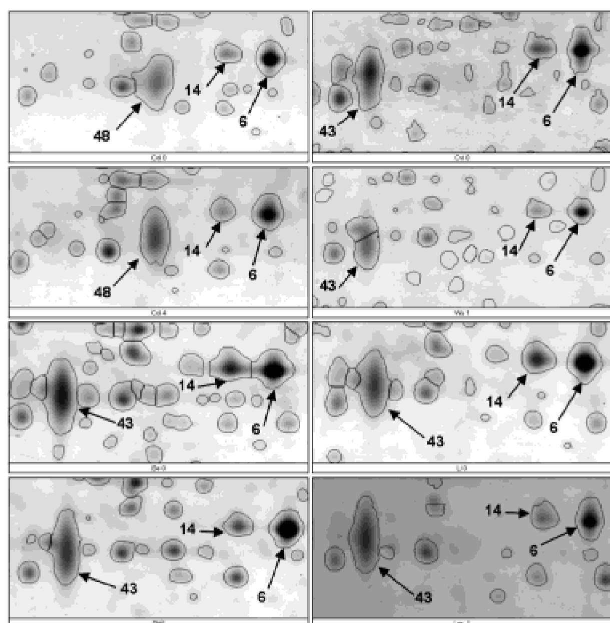

Figure 5. Example of proteins isoforms showing quantitative differences or ecotype-specific expression.

spots), were identified as MPs. However, large differences in the functional classification were observed among ecotypes. For the majority of functional classes, these relative differences were in a twofold to fourfold range. For nitrogen assimilation, however, qualitative differences were also observed, no corresponding enzyme being found as MPs for half ecotypes. Accordingly, this suggested that MPs distinguished ecotypes not only in terms of protein isoforms but also on a function basis. Therefore, ecotypes

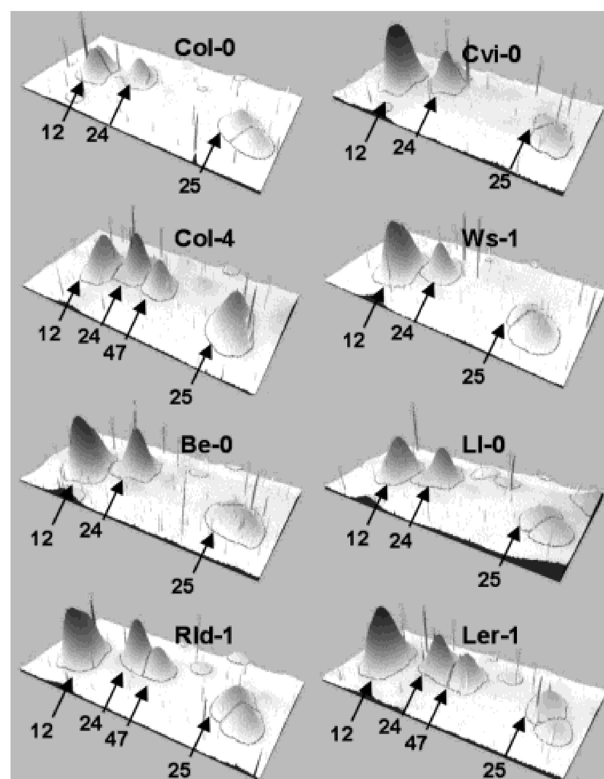

Figure 6. Example of proteins isoforms showing quantitative and ecotype-dependent expression.

were further classified according to the major functions identified. Figure 3d shows that, when using Euclidian distances to calculate a similarity matrix from the data of Table 3, and the Ward's criterion for linkage, three main

Table 3. Functional distribution of MPs in *Arabidopsis thaliana* ecotypes

Function	Col-0	Col-4	Be-0	Ll-0	Rld-1	Cvi-0	Ws-1	Ler-1
	(% total amount)							
Energy and *C*-metabolism								
glycolysis	16.6	13.3	11.0	11.2	8.0	12.8	9.7	8.8
TCA cycle	8.0	8.7	5.8	6.5	5.7	10.1	12.0	9.0
mitochondrial respiration	3.2	3.7	6.2	3.4	4.6	5.7	5.1	3.9
other	4.4	2.8	3.6	5.1	3.4	1.0		2.6
Electron transfer	3.3	5.0	2.7	1.7	4.1	4.3	2.7	6.3
Oxidative processes	2.5	4.5	2.4	2.8	2.7	10.8	7.5	11.6
Hormone response/stress	4.3	3.5	4.4	3.9	6.1	5.1	4.6	6.8
N-Metabolism	1.5		0.8	1.2	2.1			
Protein expression	1.4	3.6	1.0	2.9	3.7	2.5	4.3	1.5
Other	2.6	2.1	2.3	2.1	4.6	4.0	2.9	7.7
Unidentified proteins	3.4	2.3	3.9	2.4		0.9	2.8	1.4

Lacking values correspond to proteins either not ranked as MPs or not detected.

© 2004 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

clusters could be obtained. Furthermore, this tree was almost identical to those obtained when classifying ecotypes on the basis of their whole proteome (Fig. 3a) or according to the anonymous MP subset (Figs. 3b, c).

4 Discussion

To date, little effort was devoted to investigate the differences eventually observed between proteomes of ecotypes from a same plant species. Even in the case of the model plant *Arabidopsis thaliana*, the only available data concern a pair of ecotypes that was included into an inter-species comparison within the Brassicaceae family, with no attempt to identify proteins [3]. Therefore, presently available data do not allow to decipher, for instance, whether eventual differences between ecotypes originate from the expression of different ecotype-specific isoforms of proteins showing the same function or from the differential expression of proteins with different functions.

4.1 Different *Arabidopsis* ecotypes display contrasted proteomes

In the present work, two main series of evidences were obtained arguing for the existence of large differences between the proteomes of the eight *Arabidopsis* ecotypes analyzed.

(i) In terms of protein patterns, gel comparison demonstrated first that at least one-third of the spots was not detected in half the accessions. On the other hand, ranking the spots in terms of MPs showed that only one-quarter of these constituted MPs in all the ecotypes, several MPs for one ecotype displaying lower abundance or being not detected in others. Further identification of MPs gave examples for different types of variations, to which quantitative differences were also superimposed: simultaneous expression of the different spots in all the ecotypes, expression of one or two forms of the protein in all ecotypes, whereas another form was ecotype-dependent, or exclusive expression of one form of the protein in an ecotype-specific manner. Therefore, collectively, these results demonstrate the occurrence of a qualitative and quantitative plasticity of the protein profiles between ecotypes which concerns both low-abundance and high-abundance proteins. The nature of these differences was not investigated here and could be hypothesized to rely as well on post-translational modifications as on allelic variations for those proteins identifying the same accession. Overall, very little information is presently available concerning the variations in proteome among *Arabidopsis* ecotypes, more extensive information being available for some other plant species. Using metrics based on the presence or absence of spots, the two *Arabidopsis* ecotypes Landsberg and Columbia were shown to be first merged together when compared to other species from the Brassicaceae family [5]. For maritime pine [4], plants from seven origins were compared, and less than 20% of spots was observed simultaneously in all patterns. However, in none of these works, a specific analysis according to the protein abundance was made. Accordingly, our results both constitute a novel example of the proteome plasticity among plant ecotypes and give the first evidences that the major proteins could contribute strongly to the variations observed. Colloidal Coomassie blue staining of 2-DE gels is known to favor the visualization of abundant proteins. Therefore, as the proteins ranked here as MPs constituted half of the amount of this subset, the accumulation of MPs is likely to have a major cost for the cell. Accordingly, it can be speculated whether, beside the selection of isoforms or allelic forms conferring a superiority with respect to some functions, the variations demonstrated in the proteomes of ecotypes reveal also physiological re-equilibrations.

(ii) In terms of functions associated to MPs, large variations were also observed, mostly as quantitative differences, showing that not the same abundant proteins are accumulated in all ecotypes. Actually, whenever more than one-quarter of MPs was not detected in some ecotypes, for most of them the same accession was identified in other spots indicating the occurrence of isoforms or protein modifications. On the other hand, in several cases, given functions were highly expressed in some ecotypes, but minor in others. A striking example concerns the alcohol dehydrogenase function, that constitutes a major function in two MP spots for Ll-0 ecotype (where it cumulates 3.7% of the total protein amount), but a minor function in Ws-1 ecotype (0.1% of the total). Such amplitude suggests the occurrence of strong functional variations between ecotypes. Moreover, the fact that several of the major functions participate to the same pathway support the hypothesis that the plasticity of the proteome reveals different physiological equilibriums between ecotypes. For instance, in terms of MPs, Cvi-0 ecotype appears to dedicate 13% of its protein amount to glycolysis and 27-fold less to nitrogen assimilation, whereas this ratio is 7 times lower for Rld-1 ecotype. In terms of experimental strategy, it should be emphasized that, although a small proportion of proteins was characterized here, the conclusions derived can be assumed to be robust as it is unlikely that such ratio could be strongly modified by the identification of minor proteins. However, it should be also noticed that the five spots that could not be identified from their peptide mass fingerprint contribute substantially to the variation between ecotypes (3.9% of the total protein amount in

© 2004 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Be-0 ecotype, not detected in Rld-1 ecotype). Beside the important question of the function of these MPs, this leaves open the possibility that they participate to some of the pathways identified, thus altering their contribution in the present analysis.

A large part of identified proteins corresponds to enzymes involved in energy metabolism. This is not unexpected owing to their functional importance, although little reference data is presently available. Very recently, a first large-scale quantification of genes expressed in *Arabidopsis* root was made using the SAGE approach [21]. None of the 25 most frequent tags (> 12 700 different tags) corresponds to any of the 25 MPs identified here in one or the other of the 8 ecotypes, although the product of some of these genes, like a glyceraldehyde-3-phosphate dehydrogenase, falls clearly in the p*I* and MW ranges investigated. Differences in physiological state of roots between these experiments are likely. Nevertheless, this suggests that the major functions expressed could be not deducible from a transcript analysis. As these functions appear to discriminate ecotypes and possibly to reveal different physiological status, this would confer to the proteome analysis an unique advantage in characterizing the plant biodiversity.

Taken together, the differences quoted in terms of protein pattern and of protein function appear to encompass both a genetic and a physiological information. The former, which corresponds to approaches already well established in plants [2], was not analyzed here, although some of the observed examples (position shifts, presence/absence of spots) suggest the possible occurrence of allelic variations. The latter was not addressed up to now in ecotypes. On this basis, proteomic analysis of naturally occurring variations can be speculated to constitute a new powerful way to get insights into functional traits.

4.2 Proteomes as signatures for ecotype classification

Hierarchical clustering was used to get insights into the structure of the differences in the proteome that were quoted among ecotypes. This resulted in a classification into three groups: the two ecotypes known to be very close according to their origin, the three ecotypes introduced for their alternative root growth responses, and the three last currently used ecotypes. The same structure holds as well for the whole and anonymous protein pattern, or when restricting the focus on the 25 anonymous MPs accounting for half the protein amount in each ecotype, as for the main functions deduced from the identification of MPs. It should be emphasized that these classifications rely on both quantitative (protein abundance), qualitative (protein presence or absence), and functional features (protein function class). Therefore, it can be speculated whether they reveal characteristic proteomic signatures. This was not investigated into more details, and no similar analysis was published to date for plant ecotypes. In the case of *Arabidopsis* mutants, however, the classification of anonymous protein patterns was previously shown to group the *cri* mutant with the corresponding wild-type ecotype when grown in the presence of cytokinins, and this mutant was demonstrated thereafter to overproduce cytokinins [22]. Although such examples are rare, they suggest that different but close genotypes can be classified from their proteome in a way reflecting the physiological status of plants. Therefore, owing to the important differences quoted here among proteomes in terms of major functions, both these data and ours support the working hypothesis that the proteomic signature includes a physiological information.

One striking result is that the three ecotypes that were introduced in the study for their alternative response to environmental stress were grouped in a same cluster. According to the discussion above, this would suggest that Be-0, Ll-0, and Rld-1 ecotypes would share major physiological features. To date, however, the only data available about these ecotypes concern their atypical response to phosphate starvation. On the other hand, a close link between phosphate and energy metabolisms is well known [23], and the present proteome analysis was performed with plants grown under normal conditions. Therefore, one tentative hypothesis is that the three ecotypes would display specific and constitutive features in their energy metabolism that would be detectable at the proteome level. Under this hypothesis, their original response to phosphate starvation would result from these specific features.

A noticeable point in this work is that only a small part of the total proteomic information was used. According to the experimental conditions, the analysis concerned a defined window in terms of p*I* and MW and was restricted to soluble proteins. The classification obtained can address therefore only the functions associated to the corresponding subset of abundant and soluble proteins, illustrated here by enzymes involved in energy metabolism. Accordingly, a different classification would be not unexpected when looking at other classes of proteins, such as, for instance, transporters which include numerous basic and hydrophobic proteins. On the other hand, both anonymous and function-based classifications of MPs led to the same classification that was obtained when exploiting all the protein pattern. This suggests that MPs

© 2004 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

would be responsible for a main part of variations between ecotypes, and that an analysis focusing on a limited number of proteins bearing major functions may have the capacity to reveal specific features of the proteomes. No attempt was made here to assess these points, for instance by changing the number of MPs selected and the proportion of the total protein amount taken into account to classify the proteomes. Further work remains therefore required to rationalize a general strategy suitable for large-scale classification of ecotypes from a limited but detailed characterization of proteomes.

In conclusion, we report here a first use of proteomics to investigate variations between *Arabidopsis* ecotypes. The result obtained demonstrate that the differences in the proteomes originate both from the expression of different ecotype-specific isoforms of proteins showing the same function and from the differential expression of proteins with different functions. Furthermore, the approach appears to have the capacity to discriminate ecotypes and to generate a classification according to their physiological status, directly from the features of their major proteins. Owing to the throughput of the present proteomic technology and to the hundreds of ecotypes available for *Arabidopsis*, it can be speculated that such proteomic approach will constitute a robust tool for mining the natural biodiversity in plants.

This work was supported by the EU research program Quality of Life and Management of Living Resources (contract QLK5-CT-2001-01871) and by the Proteome Platform of the Montpellier-LR Génopole.

5 References

[1] De Vienne, D., Burstin, J., Gerber, S., Leonardi, A. *et al.*, *Heredity* 1996, *76*, 166–177.

[2] Thiellement, H., Bahrman, N., Damerval, C., Plomion, C. *et al.*, *Electrophoresis* 1999, *20*, 2013–2326.

[3] Burstin, J., de Vienne, D., Dubreuil, P., Damerval, C., *Theor. Appl. Genet.* 1994, *89*, 943–950.

[4] Bahrman, N., Zivy, M., Damerval, C., Baradat, P., *Theor. Appl. Genet.* 1994, *88*, 407–411.

[5] Marquès, K., Sarazin, B., Chané-Favre, L., Zivy, M., Thiellement, H., *Proteomics* 2001, *11*, 1457–1462.

[6] Jacobsen, S., Nesic, L., Petersen, M., Sondergaard, I., *Electrophoresis* 2001, *22*, 1242–1245.

[7] Lum, J. H., Fung, K. L., Cheung, P. Y., Wong, M., S. *et al.*, *Proteomics* 2002, *2*, 1123–1130.

[8] Fullaondo, A., Vicario, A., Aguirre, A., Barrena, I., Salazar, A., *Heredity* 2001, *87*, 266–272.

[9] Mosquera, E., López, J. L., Alvarez, G., *Heredity* 2003, *90*, 432–442.

[10] Pigliucci, M., *Trends Plant. Sci.* 1998, *3*, 485–489.

[11] Alonso-Blanco, C., Koornneef, M., *Trends Plant. Sci.* 2000, *5*, 22–29.

[12] Mitchell-Olds, T., *Trends Ecol. Evol.* 2001, *16*, 693–700.

[13] Maloof, J. N., Borevitz, J. O., Dabi, T., Lutes, J. *et al.*, *Nat. Genet.* 2001, *29*, 441–446.

[14] Alonso-Blanco, C., Blankestijn-de Vries, H., Hanhart, C. J., Koornneef, M., *Proc. Natl. Acad. Sci. USA* 1999, *96*, 4710–4717.

[15] Pepper, A. E., Corbett, R. W., Kang, N., *Plant Cell. Environ.* 2002, *25*, 591–600.

[16] Beemster, G. T., De Vusser, K., De Tavernier, E., De Bock, K., Inze, D., *Plant Physiol.* 2002, *129*, 854–864.

[17] Chevalier, F., Pata, M., Nacry, P., Doumas, P., Rossignol, M., *Plant Cell. Environ.* 2003, *26*, 1839–1850.

[18] Lejay, L., Tillard, P., Lepetit, M., Olive, F. D. *et al.*, *Plant J.* 1999, *18*, 509–519.

[19] Schele, C., Lamer, S., Pan, Z., Li, X. P. *et al.*, *Electrophoresis* 1998, *19*, 918–927.

[20] Hartigan, J. A., *Clustering Algorithms*, Wiley, New York 1975.

[21] Eckman, D. R., Lorenz, W. W., Przybyla, A. E., Wolfe, N. L., Dean, J. F. D., *Plant Physiol.* 2003, *133*, 1397–1406.

[22] Santoni, V., Delarue, M., Caboche, M., Bellini, C., *Planta* 1997, *202*, 62–69.

[23] Plaxton, W. C., in: Lynch, J. P., Deikman, J. (Eds.), *Phosphorus in Plant Biology*, ASPP, Rockville, USA 1998, 229–241.

© 2004 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim