JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Application of Two-Dimensional Gel Electrophoresis To Interrogate Alterations in the Proteome of Genetically Modified Crops. 2. Assessing Natural Variability

MARTIN C. RUEBELT,*,§ MARKUS LIPP,§ TRACEY L. REYNOLDS,§ JAMES D. ASTWOOD,§ KARL-HEINZ ENGEL,† AND KLAUS-DIETER JANY‡

Product Safety Center, Monsanto Company, 800 North Lindbergh Boulevard, St. Louis, Missouri 63167; Center for Food and Life Sciences, Chair of General Food Technology, Technical University of Munich, Am Forum 2, 85350 Freising-Weihenstephan, Germany; and Federal Research Centre for Nutrition and Food, Haid-und-Neu-Strasse 9, 76131 Karlsruhe, Germany

Proteomics is currently tested as a complementary tool for the safety assessment of genetically modified (GM) crops. Understanding the natural variability of the proteome is crucial for the interpretation of biological differences between transgenic and nontransgenic parental lines. The natural variation of seed protein profiles among a set of 12 *Arabidopsis thaliana* ecotypes was determined by utilizing two-dimensional electrophoresis (2DE). The total number of different resolved protein spots found among the 12 ecotypes was 931 with a range of 573 (Mt-0) to 653 (Condara) in any one ecotype. Although the ecotypes were grown side-by-side in an environmentally controlled growth chamber, almost half of the resolved spots varied with respect to their presence/absence, and 95% of the spots present in all ecotypes varied in spot quantity (2−53-fold). In the evaluation of unintended effects of genetic modification, it is concluded that the experimental design must account for existing natural variability, which, in the case of the expressed proteome, can be substantial.

KEYWORDS: Two-dimensional gel electrophoresis; *Arabidopsis thaliana*; seed proteome; natural variability

INTRODUCTION

To evaluate the biological significance of protein differences between a transgenic plant and its parental line, it is important to have a comprehensive understanding of the natural variation of protein expression within and among the investigated plant species. It is critical to assess whether a detected difference in protein expression, including expression of a novel protein or the increase/decrease of a specific protein, can be found in nature, that is, under different environmental conditions or in other varieties of the plant species. This is particularly true for the assessment of the safety relevance of detected differences between transgenic and nontransgenic plants. Many international organizations (*1*−*4*) suggest that further assessment (nutritional and toxicological) should be required only if the differences exceed the natural variation present in traditional food crops.

Naturally occurring genetic variation is commonly found in all mammals, microbes, and plants, including *Arabidopsis*. Those random differences in an organism's genome are the basis of the natural selection of a species. Horizontal evolution, transposition, gene rearrangements, fusions, and deletions are considered to be important evolutionary forces. One source of protein variability is caused by point mutations that result in codon changes. Such heritable changes in the DNA directly affect the proteome unless the mutated codon codes for the same amino acid (silent mutation) due to the degeneracy of the genetic code. A point mutation (base-pair substitution) of the DNA, for example, can result in (1) missense mutation (*5*, *6*), (2) nonsense mutation leading to protein elongation or truncation, (3) loss of phosphorylation or glycosylation sites, and (4) alteration of the degradation stability of the resulting protein. Addition or deletion of a base pair in a gene shifts the reading frame by one base and leads to a change in the amino acid sequence of the protein (frameshift mutation). All of these mutations could affect the net charge (isoelectric point) and/or molecular mass of the resulting protein and, therefore, the protein's two-dimensional electrophoresis (2DE) migration behavior (electrophoretic mobility). Quantitative variation may be a result of changes in the amino acid sequence (e.g., change of degradation stability) or mutations of noncoding DNA sequences (e.g., regulatory regions) (*7*).

Genetic variation within a species has been investigated by 2DE for many crops, including barley (*8*−*11*), maize (*12*−*15*), wheat (*16*−*20*), peanut (*21*), and rice (*22*). Although *Arabidopsis* provides an extensive resource for natural genetic variation among ecotypes, only two studies have compared the 2DE profiles of various *Arabidopsis* ecotypes. Marques et al. (*23*)

* Corresponding author (e-mail m.ruebelt@gmail.com).
§ Monsanto Company.
† Technical University of Munich.
‡ Federal Research Centre for Nutrition and Food.

10.1021/jf052357y CCC: $33.50
Published on Web 02/16/2006

included the ecotypes Landsberg *erecta* and Columbia in an interspecies comparison within the Brassicaceae family and compared the 2DE protein profiles of the aerial part of etiolated seedlings. In a recent study, Chevalier et al. (*24*) investigated natural variation in the root proteome among eight *Arabidopsis* ecotypes (Col-0, Col-4, Be-0, Ll-0, Rld-1, Cvi-0, Ws-1, and Ler-1). To date, no data have been published on the differences between seed protein profiles among *Arabidopsis thaliana* (*A. thaliana*) ecotypes. The seed proteome is the focus of this proteomics study as seed crops such as soybean, corn, and canola are important sources for nutrients and, therefore, an essential part of the human diet. Therefore, the impact of the genetic background on the seed proteomes of various *A. thaliana* ecotypes was investigated in this study.

MATERIALS AND METHODS

Plant Material. Seeds for the *A. thaliana* ecotypes were obtained from the Arabidopsis Biological Resource Center in Ohio (http://arabidopsis.org/abrc/) with the exception of seeds for Col-0, which were obtained from Monsanto Co., St. Louis, MO. The stock numbers are C24 (CS906), Condara (CS6175), Cvi-0 (CS6675), Ll-0 (CS6781), Ma-0 (CS6789), Mr-0 (CS6795), Mt-0 (CS6799), Nd-0 (CS6803), Oy-0 (CS6824), Tsu-0 (CS6874), and Ws (CS6891). The 12 *A. thaliana* ecotypes were grown in individual 2.5-in. pots (Metro-Mix 200 soil; Hummert International, Earth City, MO) side-by-side in a growth chamber at 20 °C and 70% relative humidity with 16 h of light (150−200 μeinstein/s/m^2). For each ecotype, 12 replicates were planted and randomly distributed in the growth chamber to limit the influence of environmental factors. The plants were watered and fertilized (100 ppm of Peter's 20:20:20) twice weekly via subirrigation. Seeds were harvested after complete maturity of all seeds on a plant as each ecotype matured at a different time. To avoid seed deterioration, seeds were placed in freezer bags containing desiccant (Drierite Anhydrous Calcium Sulfate) and stored at 4 °C.

Phenotypic Analysis and Methodology. Four phenotypic traits, that is, first flowering date (FFD), rosette diameter (RD), seed yield, and seed protein content, were assayed. The FFD is the number of days from the date of planting until the opening of the first flower and was assayed by daily inspection of the plants. The RD is the diameter (in centimeters) of the leaf rosette at the time of first flowering. The seed yield is the amount of harvested seeds for one plant. Leaf and stem morphology was visually assessed by the overall shape, length, thickness, and pubescence. The seed protein content was determined using the FlashEA 1112 protein analyzer. The protein content was calculated with Eager 300 software using the protein factor of 6.25. The analysis was performed in replicates of the pooled seed samples, also used for the 2DE analysis.

Sample Preparation. Protein extracts were prepared as described in ref *25*. In brief, seeds harvested from six to eight plants were pooled and ground with a paint shaker-like device. Proteins were extracted with an extraction buffer containing 7 M urea, 2 M thiourea, 0.75% (w/v) CHAPS, 0.75% (v/v) Triton X-100, 100 mM DTT, 1% (v/v) carrier ampholytes stock, 20% (v/v) 2-propanol, and protease inhibitor cocktail Complete (Roche, Mannheim, Germany) for 1 h while shaking on a Nutator (Becton-Dickinson, Sparks, MD) at room temperature. After centrifugation, the supernatant was stored in aliquots at −80 °C until analysis. Protein concentration was estimated using the Bio-Rad Protein Assay with BSA as a standard.

Two-Dimensional Gel Electrophoresis. 2DE and gel staining were performed according to the methods described in ref *25*. Briefly, protein extracts were diluted in rehydration buffer containing 7 M urea, 2 M thiourea, 0.75% (w/v) CHAPS, 0.75% (v/v) Triton X-100, 100 mM DTT, 0.3% (v/v) carrier ampholytes stock, 10% (v/v) 2-propanol, 12.5% (v/v) water saturated isobutanol, protease inhibitor cocktail Complete (Roche), and a trace of bromophenol blue. Nonlinear immobilized pH gradient gel strips (IPG) with nonlinear pH 3−10 gradients (13 cm, GE Healthcare) were rehydrated using 230 μL of diluted sample (150 μg of total protein). The IEF was carried out using a Bio-Rad PROTEAN IEF cell with a controlled cell temperature of 20 °C to a total of 35000 Vh. The IPG strips were equilibrated first for 10 min in 50 mM Tris-HCl (pH 8.8), 6 M urea, 30% (w/v) glycerol, 2.3% (w/v) SDS, 1% (w/v) DTT, and bromophenol blue and then for another 10 min in the same solution except DTT was replaced with 4% (w/v) iodoacetamide. The second dimension was run in a Bio-Rad Criterion Dodeca cell system in 8−16% Tris-HCl linear gradient Criterion gels. The gels were run for the first 15 min at 130 V and then at 180 V until the tracking dye reached the bottom of the gel. The gels were stained with colloidal CBB solution according to the method of Neuhoff et al. (*26*). The CBB staining solution was prepared fresh by mixing 4 parts of 0.1% (w/v) CBB G-250 in 2% (w/v) phosphoric acid and 11% w/v ammonium sulfate with 1 part of methanol. The gels were incubated in this solution for 3 days at room temperature on an orbital rotator. Imaging of the stained proteins was performed at a scan resolution of 36.3 × 36.3 μm using the GS-800 calibrated densitometer (Bio-Rad Laboratories).

Image Analysis. The scanned images of the 2DE gels were processed and analyzed with PDQuest 2-DE Gel Analysis software version 7.1 (Bio-Rad Laboratories). The images were cropped and oriented using the image editing controls of the program. All images were processed with the following software settings for spot detection and background subtraction: sensitivity, 40; size scale, 3; min peak, 400; power mean, 3 × 3; floater, 97; speckles filter. Spots detected by the software program were manually verified. False-positive spots (e.g., artifacts and multiple spots in a cluster) were manually removed; false-negative spots (obviously missed spots with OD > LOD) were added to the images. A spot was considered to be reproducibly present/absent when it was present/absent in all three replicate gels of one extraction. To compare spots across gels, a match set was created from the images of the gels in an experiment. A standard gel (master) was generated out of the image with the greatest number of spots. Spots reproducibly present in a match set member but not present in the image with the most spots were manually added to the standard gel. The automated matching tool of the PDQuest software package was used to match spots across the gels. A few landmarks were manually defined to improve the automated matching results. All spots matched by the software program were manually verified. The spots were quantified by 2D Gaussian modeling. Spot quantities of all gels were normalized to remove non-expression-related variations in spot intensity, so the raw quantity of each spot in a gel was divided by the total quantity of all the spots in that gel that have been included in the standard. Data were exported to Excel and from there to JMP for statistical analysis. All statistical analyses were performed with the statistical software package JMP v. 5 (SAS Institute Inc., Cary, NC).

Natural Variability Study. A phenetic tree was constructed according to the method of Marques et al. (*23*) based on the pairwise comparison of the qualitative (presence/absence of spot) protein profiles of the ecotypes. These 66 pairwise comparisons were done by counting the numbers of spots present in both ecotypes (N_{AB}) and specifically present in one (N_{A0}) or the other (N_{0B}) of the two ecotypes. The Jaccard or dissimilarity index ($D_j = 1 - N_{AB}/(N_{AB} + N_{A0} + N_{0B})$ was used to compute a dissimilarity matrix. From this dissimilarity matrix, the unrooted phenetic tree was calculated with the neighbor-joining algorithm using the Phylip 3.6 software package (*27*).

To establish the range of quantitative variation of commonly expressed spots, only ecotype spots with RSD below 55% and mean-spot-quality above 40 were taken into account. The spot quality is a number ranging between 0 and 100 and is calculated by PDQuest on the basis of the spot shape and overlapping effects. The spot quantities of spots of low quality tend to be overestimated by the software program (data not shown).

RESULTS AND DISCUSSION

A. thaliana occurs naturally throughout temperate regions of the world including Europe, East Africa, Asia, Japan, North America, and Australia (*28*). To cover a broad range of geographies, a set of 11 ecotypes from four different continents (10 countries) and one common laboratory line were selected (**Table 1**). The 12 ecotypes, all interfertile, represent a wide

Table 1. Phenotypic Measurements of the Selected *A. thaliana* Ecotypes[a]

name	continent (country)	*N*	FFD[b] (days)	RD[c] (cm)	seed yield (mg)	protein[d] (% fw[e])
Cvi-0	Africa (Cape Verde Island)	10	42 ± 6	10.4 ± 1.7	*222* ± 72	31.0
Mt-0	Africa (Libya)	9	*30* ± 2	8.0 ± 1.0	540 ± 218	25.8
Condara	Asia (Tajikistan)	12	35 ± 3	11.5 ± 1.1	875 ± 300	28.2
Tsu-0	Asia (Japan)	11	38 ± 3	12.3 ± 2.1	1106 ± 309	27.8
Ws	Asia (Russia)	7	*63* ± 8	12.2 ± 0.8	720 ± 344	31.0
Ll-0	Europe (Spain)	8	54 ± 6	*13.5* ± 0.8	*1293* ± 528	28.8
Ma-0	Europe (Germany)	11	*30* ± 2	7.6 ± 1.3	523 ± 71	26.9
Mr-0	Europe (Italy)	8	62 ± 6	11.3 ± 0.7	1137 ± 636	28.9
Nd-0	Europe (Germany)	12	34 ± 6	7.5 ± 0.6	455 ± 144	27.2
Oy-0	Europe (Norway)	8	38 ± 5	9.9 ± 1.4	972 ± 302	26.1
Col-0	North America (U.S.)	8	31 ± 1	*7.4* ± 1.0	722 ± 92	25.7
C24	laboratory line	11	37 ± 3	8.0 ± 0.7	590 ± 132	26.1

[a] Values are means ± SD; extreme values are italicized. [b] Number of days from the date of planting until the opening of the first flower. [c] Rosette diameter at the time of first flowering. [d] Protein content of pooled seed samples of six to eight plants ($N = 2$). [e] Fresh weight.

range of genetic diversity as assessed by restriction fragment length polymorphism (RFLP) (*29*), amplified fragment length polymorphism (AFLP) (*28*, *30*, *31*), and cleaved amplified polymorphic sequence (CAPS) (*32*).

Comparison of Phenotypic Traits. The 12 *A. thaliana* ecotypes were grown side-by-side in an environmentally controlled growth chamber. Four phenotypic traits [i.e., first flowering date (FFD), rosette diameter (RD), seed yield, and seed protein content] were assessed and are summarized in **Table 1**. On the basis of the four measured phenotypic traits, the 12 ecotypes covered a wide spectrum of phenotypic diversity. The average FFD [Boyes' growth stage 6.00 (*33*)] varied between 30 and 63 days for Mt-0/Ma-0 and Ws, respectively. The average RD was found to range from 7.4 cm for Col-0 to 13.5 cm for Ll-0. The chosen ecotypes had average seed yields ranging from as low as 222 mg per plant for Cvi-0 to 1293 mg per plant for Ll-0. To determine the seed protein content, seeds from six to eight plants of one ecotype were pooled, and the pooled sample was analyzed in duplicate. Therefore, the standard deviation does not reflect the natural variation within an ecotype and was not specified in **Table 1**. The average seed protein content (% fw) varied between 25.7% (Col-0) and 31.0% (Cvi-0 and Ws). A correlation between seed yield and FFD, RD, or protein content of seeds was not found, possibly due to the fact that the measured parameters are not considered to be linked physiologically. These measurements have shown that two ecotypes, Mt-0 from Libya (Africa) and Ma-0 from Germany (Europe), are very similar with respect to measured phenotypes and their leaf and stem morphology (visually assessed). All other ecotypes demonstrated a range of values in the measured parameters. Therefore, the ecotypes chosen for this study represent a broad range of morphological and physiological parameter.

Comparison of 2DE Patterns. To ensure that all seeds of a plant reach maturity with minimal environmental perturbations, irrigation was continued until the plant reached complete senescence and all the siliques were yellow [Boyes' growth stage 9.70 (*33*)]. Pooled seed samples were extracted, and the extracts were subjected to an optimized 2DE method (*25*). For the ecotypes 573 (Mt-0) to 653 (Condara) seed proteins were reproducibly resolved with a p*I* range from 4 to 9 and a molecular mass range from 6 to 120 kDa (**Figure 1**). The presence of several abundant, possibly storage, proteins in the 29−36 kDa range make protein resolution difficult as well as horizontal and vertical streaks in the high-pH range. A protein spot was considered to be reproducible when it was present in all three replicate 2DE gels of an ecotype. The verification of

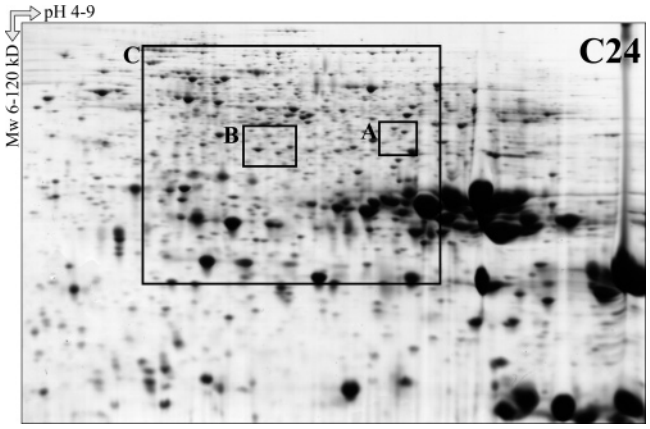

Figure 1. Representative seed proteome pattern (2DE) of *A. thaliana* (150 μg of total seed protein of ecotype C24); boxes correspond to gel regions enlarged in **Figure 3** (boxes A and B) and **Figure 4** (box C).

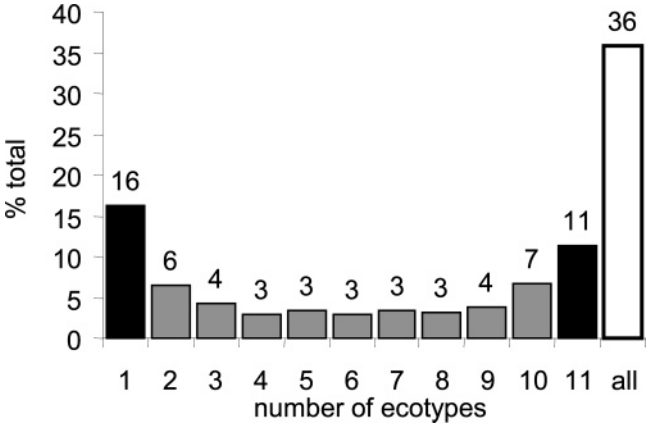

Figure 2. Distribution of 931 distinct spots detected among the 12 ecotypes according to the number of ecotypes where they were reproducibly detected.

a spot's presence or absence and the accurate matching of spots between ecotypes were constrained by large differences in their protein patterns and were often ambiguous.

The total number of different protein spots found in the 12 ecotypes was 931. **Figure 2** shows the entire spot distribution according to the number of ecotypes where they were detected. Among these 931 spots, 334 spots (36%; white bar, **Figure 2**) were present in all of the ecotypes and 597 spots (64%) were variable, that is, absent in at least one ecotype. Twenty-seven percent of all protein spots appeared to be either specifically

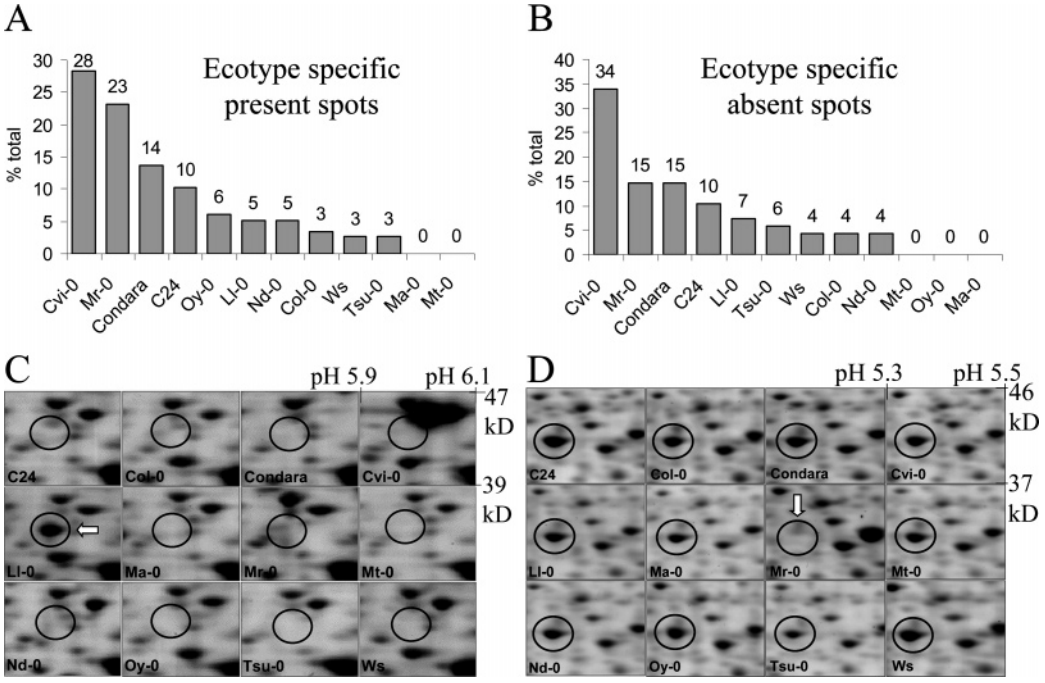

Figure 3. Ecotype-specific spots: (**A**, **B**) distribution of ecotype-specific present spots (**A**) and absent spots (**B**) according to the ecotypes; (**C**, **D**) examples of an ecotype-specific present (**C**) and absent (**D**) protein spot. Enlarged gel regions A and B from **Figure 1**.

present or absent for one ecotype. Each ecotype-specific protein spot (black bars, **Figure 2**) was visually inspected and verified. Of the 150 spots (16%) present in only one ecotype (ecotype-specific), 33 spots were ambiguously absent in at least one other ecotype; that is, the spot may be obscured by another spot or detected in less than three replicate gels due to edge or focusing effects and, therefore, not included in the data set. Of the 106 spots (11%) absent in only one ecotype (specific ecotype-absent), 38 spots were ambiguous.

The distribution of 117 unambiguous ecotype-specific spots (**A**) and 68 unambiguous specific ecotype-absent spots (**B**) was analyzed according to the ecotype in which they were detected (**Figure 3**). Two examples of ecotype-specific present or absent protein spots are shown in **Figure 3C**,**D**. In example **C**, the marked protein spot is present only in the ecotype Ll-0 and not present in any of the other ecotypes. In example **D**, the marked spot is present in all ecotypes but ecotype Mr-0. Both examples also demonstrate the high resolution power and great reproducibility of the 2DE method. The ecotypes Cvi-0, Mr-0, Condara, and C24 accounted for 75% of these specifically present spots (**Figure 3A**) and 74% of the specifically absent spots (**Figure 3B**). For Ma-0 or Mt-0, no uniquely present or absent spots were identified. Although there is a large portion of common protein spots (36%), many vary and are specific for one or a few ecotypes.

At this time, very little information is available about the proteome of *Arabidopsis* ecotypes. More information is available for crop plant species, such as barley (*8*−*11*), maize (*12*−*15*), wheat (*16*−*20*), peanut (*21*), rice (*22*), and potato (*34*), for which large variability in the proteomes has been demonstrated. Chevalier et al. (*24*) were the first group to assess the natural variation in the proteome of *Arabidopsis* ecotypes. They investigated the natural variation in the root proteome among eight *Arabidopsis* ecotypes and resolved an average of 250 spots for each ecotype. The variability of the root proteome is similar to the variability seen in the seed proteome. The numbers of variable spots and specific spots were 75 and 26% for the root proteome and 64% and 26% for the seed proteome, respectively.

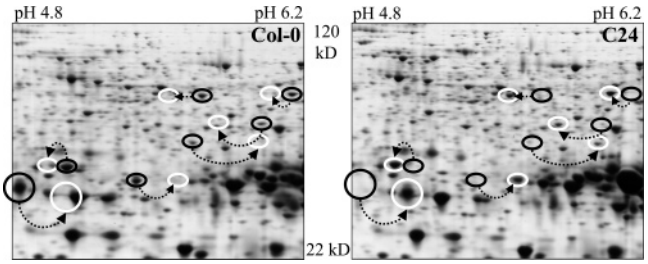

Figure 4. Comparison of 2DE gels from Col-0 and C24. The marked protein spots are examples of hypothetical p*I* shifts. Solid circles mark spot position of seed proteins found in Col-0; open circles mark spot position of seed proteins found in C24; arrows imply the direction of the hypothetical p*I* shifts. Enlarged gel region D from **Figure 1**.

The nature of the protein pattern variability was not investigated here and could be hypothesized to rely on post-translational modifications and allelic variations for proteins identifying the same ecotype. Upon inspection of the ecotype-specific spots, apparent position shifts (PS) of proteins were observed. As shown in **Figure 4**, a comparison of Col-0 and C24 gels revealed many hypothetical pI shifts. These shifts are in both directions and range from small (<0.1 p*I*) to large (0.6 p*I*) differences. Position shifts or allelic variations of a protein can be suspected when two spots, differing by their p*I* and/or molecular mass, are mutually exclusive in different ecotypes. Such allelic variation of a protein may be the result of a point mutation, frameshift, deletion, addition, or post-translational modification as discussed previously. For example, Jungblut et al. (*35*) compared the proteomes of various *Helicobacter pylori* strains and demonstrated that a single amino acid change caused a change of p*I* of 0.05 unit, which resulted in a clearly detectable shift in the 2DE pattern. Finnie et al. (*11*, *36*) compared the protein patterns of a series of barley cultivars with different malting properties and demonstrated by mass spectroscopy that a single amino acid substitution is sufficient to explain the 0.1-pH-unit difference between two β-amylase spots. Also, Schlesier et al. (*37*) demonstrated that a single amino acid substitution

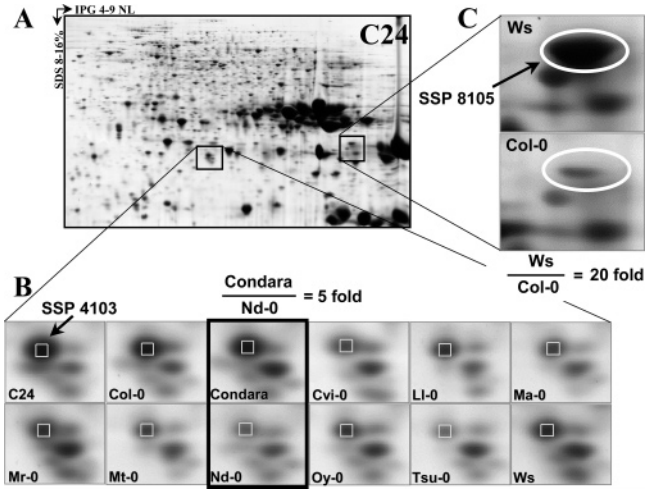

Figure 5. Quantitative comparison of two specific protein spots in different ecotypes: (**A**) seed proteome pattern (2DE) of ecotype C24; (**B**) enlarged gel region with protein spot SSP 4103 for all 12 ecotypes; (**C**) enlarged gel region with protein spot SSP 8105 for the ecotypes Ws and Col-0.

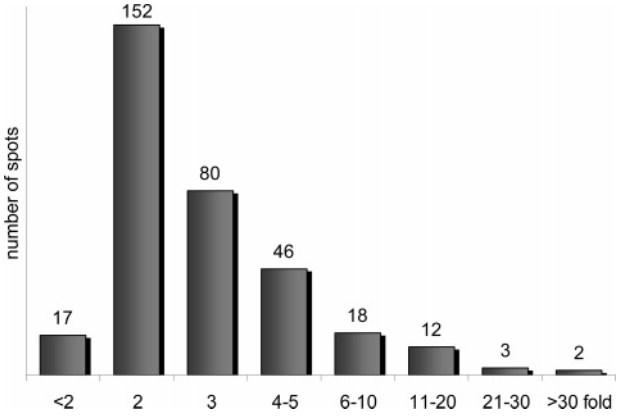

Figure 6. Natural variation of protein spot quantities among the 12 ecotypes considering the 330 spots detected in all ecotypes.

resulted in a 0.45 pI shift of a germin-like protein found in the leaf proteome of two *Arabidopsis* ecotypes (Col-0 vs Ws-2). This finding highlights the genetic basis for proteome differences and the power of 2DE to detect such differences. Anderson et al. (*38*) analyzed various wheat lines and found that charge modifications often occur in the major storage proteins of wheat and that mass modifications occur less frequently. Similarly, in this study, more horizontal position shifts were observed than vertical position shifts. A possible explanation for this observation may be that mutations leading to charge changes are more frequent or that horizontal position shifts are easier to detect than vertical position shifts due to a higher resolution power in this direction.

The natural variability of qualitative seed protein profiles due to genetic backgrounds is extensive among the 12 selected *A. thaliana* ecotypes. Novel (ecotype specific) protein spots and changes in the electrophoretic mobility (p*I* and/or molecular mass changes) of proteins occur within a species.

Quantitative Comparison of 2DE Patterns. To assess the natural variability of spot quantities, only proteins expressed in all ecotypes were considered. Two examples for the natural variability in spot quantity will be discussed (**Figure 5**). For protein spot SSP 4103, there was considerable variation in the quantity of this spot among the 12 ecotypes. Most significant is the 5-fold difference in SSP 4103 between the two ecotypes Condara and Nd-0 (black box, **Figure 5B**). However, the extreme spot quantities for this spot are not unique. Ecotypes C24 and Col-0 are like Condara and Tsu-0 is like Nd-0 with respect to the spot quantities of SSP 4103. A larger difference (20-fold) is seen between the ecotypes Ws and Col-0 for the spot SSP 8105 (**Figure 5C**). The Ws ecotype is unique in having such a high quantity of SSP 8105. The Oy-0 ecotype, with the second highest spot quantity for SSP 8105, has 60% less spot quantity than the spot of WS. These two examples illustrate the range of protein levels a commonly expressed protein can exhibit across the 12 *Arabidopsis* ecotypes.

To avoid overestimation of natural variation, only spots with CVs below 55% and a preset spot quality to compensate for streaking, overlap, etc., were utilized for the assessment of protein spot quantity variation. Four of the 334 spots had CVs and/or spot qualities below the threshold and were excluded from the data set. The range (ratio of the highest to the lowest spot quantity) computed for the remaining 330 protein spots varied from 1- to 53-fold, with ranges equal to or higher than 2-fold for ~95% of the spots (**Figure 6**). Only 5% of the common protein spots are less variable across the 12 ecotypes with quantity ranges of <2-fold. It is presumed that the quantitative variation is a result of changes in the amino acid sequence of the structural gene (e.g., change of degradation stability) and/or mutations of noncoding DNA sequences (e.g., regulatory sequences) (*7*). Proteins represent also polygenic traits. Damerval et al. (*39*) showed in maize that the quantity of a single protein can depend on several chromosomal loci. There are no data published regarding genetic variability in protein quantity for *Arabidopsis*. However, Bustin et al. (*13*), analyzing 21 maize (*Zea mays* L.) inbred lines by 2DE, found that the ratio of the highest to the lowest intensity in 21 lines ranged from 1.4 to 26 for 190 quantified spots. This range is similar to the range found in this study, even though Bustin et al. (*13*) used silver staining, which may not allow for large dynamic ranges. However, large spot quantity variation of commonly expressed proteins seems to be common within a species. Therefore, a difference in spot quantity observed between a GM sample and its non-GM counterpart should be evaluated in the context of the natural variation of this spot in order to evaluate its biological or safety relevance.

Quantification of the Natural Variation. The distribution charts of the ecotype-specific protein spots (**Figure 3A**,**B**) suggest that four ecotypes (Cvi-0, Mr-0, Condara, and C24) have the most unique protein profiles compared to the other ecotypes but do not allow a conclusion about the overall relationships between the ecotypes. To quantify and visualize the relationship (calculate the distance) between the ecotypes, an unrooted phenetic tree was constructed according to the methodof Marques et al. (*23*) based on the pairwise comparison of the qualitative (presence/absence of spot) protein profiles of the ecotypes. The Jaccard index was used to compute a dissimilarity matrix. From this dissimilarity matrix, an unrooted phenetic tree (**Figure 7**) was calculated with the neighbor-joining algorithm using the Phylip 3.6 software package (*27*). An unrooted phenetic tree specifies the relationships among ecotypes and does not define the evolutionary path.

The length of the branch (**Figure 7**) is proportional to the number of differences in the seed proteomes. The distance between two ecotypes is the sum of the length of all branches connecting them. The greatest distance was found between Cvi-0 and C24, with a total of 271 different spots and a Jaccard or dissimilarity index of 0.356. Ecotypes Ma-0 and Mt-0 appeared to be very closely related, with 30 different spots and a Jaccard

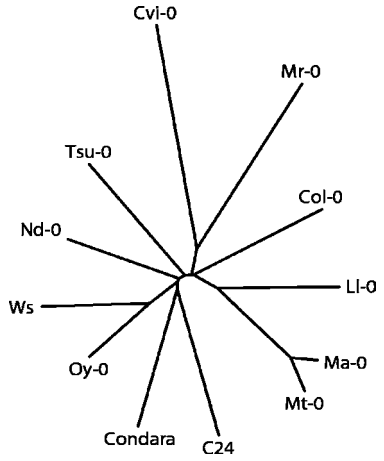

Figure 7. Unrooted phenetic tree built from the distance matrix calculated according to the Jaccard index on all of the spots of the 12 ecotypes using the neighbor-joining algorithm.

index of 0.050. The next group of closely related ecotypes is Ws and Oy-0, with 102 different spots and a Jaccard index of 0.147. This type of analysis agrees with data for the two very closely related ecotypes Ma-0 and Mt-0 that had displayed very similar phenotypes (FFD, RD, seed yield, and protein in **Table 1**). Erschadi et al. (*30*), using 15 AFLP primer combinations, grouped 20 *Arabidopsis* ecotypes, and Ma-0 and Mt-0 were also clustered into one group. Kliebenstein et al. (*40*) analyzed the glucosinolate profiles in the leaves and seeds of 39 *Arabidopsis* ecotypes. Ma-0 and Mt-0 showed a very similar glucosinolate profile. Therefore, the phenetic tree built with the proteome data generated in these experiments provides ecotype relationships similar to AFLP primer combinations and metabolic profiles regarding the ecotypes Ma-0 and Mt-0. Due to differences in selected ecotypes and methods of data analysis, it is difficult to compare the present data with population genetics publications based on RFLP (*29*), AFLP (*28*, *30*, *31*), and CAPS (*32*). However, it is known that *Arabidopsis* ecotypes are not easily grouped, that is, do not conform to a bifurcating pattern of evolution, and there is no "ecotype phylogeny" (*28*). The analysis of genetic relationship by AFLP among ecotypes revealed a star- or bushlike dendrogram (*28*, *31*). The phenetic tree based on variation in the protein patterns illustrates the large genetic variability among the 12 ecotypes. The fact that most of the ecotype branches rise from the center of the phenetic tree suggests that the selected 12 ecotypes cover a large range of natural variability.

The understanding of the natural variability of the proteome is crucial for the interpretation of biological and safety-relevant differences between transgenic and nontransgenic parental lines. The natural variability of seed protein profiles resulting from different genetic backgrounds was found to be extensive among a set of 12 *A. thaliana* ecotypes. Almost half of the resolved spots varied with respect to their presence/absence, and 95% of the spots, present in all ecotypes, varied in spot quantity (2−53-fold). These data will be used as a baseline for the head-to-head comparison of transgenic versus parental *Arabidopsis* lines in order to assess differences in the context of natural variability.

ACKNOWLEDGMENT

We thank Daniel Free for assistance with plant cultivation and Margaret A. Nemeth for statistical expertise.

LITERATURE CITED

(1) *EC Scientific Steering Committee Guidance Document for the Risk Assessment of Genetically Modified Plants and Derived Food and Feed* (March 6−7, 2003); European Commission: Brussels, Belgium, 2003; http://europa.eu.int/comm/food/fs/sc/ssc/out327_en.pdf.

(2) *European Food Safety Authority Draft Guidance Document for the Risk Assessment of Genetically Modified Plants and Derived Food and Feed*; European Food Safety Authority (EFSA): 2004; http://www.efsa.eu.int/consultation/372/consultation_guidance_gmo_01_en1.pdf.

(3) FAO/WHO. *Safety Aspects of Genetically Modified Foods of Plant Origin. Joint FAO/WHO Consultation on Foods Derived from Biotechnology*, Geneva, Switzerland, May 29−June 2, 2000; Food and Agriculture Organization of the United Nations and World Health Organization: Rome, Italy, 2000.

(4) Chassy, B.; Hlywka, J. J.; Kleter, G. A.; Kok, E. J.; Kuiper, H. A.; McGloughlin, M.; Munro, I. C.; Phipps, R. H.; Reid, J. E. Nutritional and safety assessments of foods and feeds nutritionally improved through biotechnology: an executive summary. *Compr. Rev. Food Sci. Food Saf.* **2004**, *3*, 38−104.

(5) El-Din El-Assal, S.; Alonso-Blanco, C.; Peeters, A. J.; Raz, V.; Koornneef, M. A QTL for flowering time in *Arabidopsis* reveals a novel allele of CRY2. *Nat. Genet.* **2001**, *29*, 435−440.

(6) Koornneef, M.; Alonso-Blanco, C.; Vreugdenhil, D. Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annu. Rev. Plant Biol.* **2004**, *55*, 141−172.

(7) Klose, J.; Nock, C.; Herrmann, M.; Stuhler, K.; Marcus, K.; Bluggel, M.; Krause, E.; Schalkwyk, L. C.; Rastan, S.; Brown, S. D.; Bussow, K.; Himmelbauer, H.; Lehrach, H. Genetic analysis of the mouse brain proteome. *Nat. Genet.* **2002**, *30*, 385−393.

(8) Gorg, A.; Postel, W.; Domscheit, A.; Gunther, S. Two-dimensional electrophoresis with immobilized pH gradients of leaf proteins from barley (*Hordeum vulgare*): method, reproducibility and genetic aspects. *Electrophoresis* **1988**, *9*, 681−692.

(9) Gorg, A.; Postel, W.; Baumer, M.; Weiss, W. Two-dimensional polyacrylamide gel electrophoresis, with immobilized pH gradients in the first dimension, of barley seed proteins: discrimination of cultivars with different malting grades. *Electrophoresis* **1992**, *13*, 192−203.

(10) Ostergaard, O.; Melchior, S.; Roepstorff, P.; Svensson, B. Initial proteome analysis of mature barley seeds and malt. *Proteomics* **2002**, *2*, 733−739.

(11) Finnie, C.; Melchior, S.; Roepstorff, P.; Svensson, B. Proteome analysis of grain filling and seed maturation in barley. *Plant Physiol.* **2002**, *129*, 1308−1319.

(12) Higginbotham, J. W.; Smith, J. S.; Smith, O. S. Quantitative analysis of two-dimensional protein profiles of inbred lines of maize (*Zea mays* L.). *Electrophoresis* **1991**, *12*, 425−431.

(13) Burstin, J.; Devienne, D.; Dubreuil, P.; Damerval, C. Molecular markers and protein quantities as genetic descriptors in maize. 1. Genetic diversity among 21 inbred lines. *Theor. Appl. Genet.* **1994**, *89*, 943−950.

(14) de Vienne, D.; Bost, B.; Fievet, J.; Zivy, M.; Dillmann, C. Genetic variability of proteome expression and metabolic control. *Plant Physiol. Biochem.* **2001**, *39*, 271−283.

(15) Consoli, L.; Damerval, C. Quantification of individual zein isoforms resolved by two-dimensional electrophoresis: genetic variability in 45 maize inbred lines. *Electrophoresis* **2001**, *22*, 2983−2989.

(16) Zivy, M.; Thiellement, H.; de Vienne, D.; Hofmann, J.-P. Study on nuclear and cytoplasmic genom expression in wheat by two-dimensional gel electrophoresis. 1. First results on 18 alloplasmic lines. *Theor. Appl. Genet.* **1983**, *66*, 1−7.

(17) Marlow, G. C.; Wurst, D. E.; Loschke, D. C. The use of ultrathin-layer polyacrylamide gel isoelectric focusing in two-dimensional analysis of plant and fungal proteins. *Electrophoresis* **1988**, *9*, 693−704.

(18) Weiss, W.; Huber, G.; Engel, K. H.; Pethran, A.; Dunn, M. J.; Gooley, A. A.; Gorg, A. Identification and characterization of wheat grain albumin/globulin allergens. *Electrophoresis* **1997**, *18*, 826−833.

(19) David, J. L.; Zivy, M.; Cardin, M. L.; Brabant, P. Protein evolution in dynamically managed populations of wheat: adaptive responses to macro-environmental conditions. *Theor. Appl. Genet.* **1997**, *95*, 932−941.

(20) Jacobsen, S.; Nesic, L.; Petersen, M.; Sondergaard, I. Classification of wheat varieties: use of two-dimensional gel electrophoresis for varieties that can not be classified by matrix assisted laser desorpiton/ionization−time of flight−mass spectrometry and an artificial neural network. *Electrophoresis* **2001**, *22*, 1242−1245.

(21) Basha, S. M. M. Identification of cultivar differences in seed polypeptide composition of peanuts (*Arachis hypogaea* L.) by two-dimensional polyacrylamide gel electrophoresis. *Plant Physiol.* **1979**, *63*, 301−306.

(22) Abe, T.; Gusti, R. S.; Ono, M.; Sasahara, T. Variations in glutelin and high molecular weight endosperm proteins among subspecies of rice (*Oryza sativa* L.) detected by two-dimensional gel electrophoresis. *Gen. Genet. Syst.* **1996**, *71*, 63−68.

(23) Marques, K.; Sarazin, B.; Chane-Favre, L.; Zivy, M.; Thiellement, H. Comparative proteomics to establish genetic relationships in the Brassicaceae family. *Proteomics* **2001**, *1*, 1457−1462.

(24) Chevalier, F.; Martin, O.; Rofidal, V.; Devauchelle, A. D.; Barteau, S.; Sommerer, N.; Rossignol, M. Proteomic investigation of natural variation between *Arabidopsis* ecotypes. *Proteomics* **2004**, *4*, 1372−1381.

(25) Ruebelt, M. C.; Leimgruber, N. K.; Lipp, M.; Reynolds, T. L.; Nemeth, M. A.; Astwood, J. D.; Engel, K.-H.; Jany, K.-D. Application of two-dimensional gel electrophoresis to interrogate alterations in the proteome of genetically modified crops. 1. Assessing analytical validation. *J. Agric. Food Chem.* **2006**, *54*, 2154−2161.

(26) Neuhoff, V.; Arold, N.; Taube, D.; Ehrhardt, W. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* **1988**, *9*, 255−262.

(27) Felsenstein, J. Phylip−Phylogeny Inference Package, version 3.6 (alpha3), July 2002; http://evolution.genetics.washington.edu/phylip.html.

(28) Sharbel, T. F.; Haubold, B.; Mitchell-Olds, T. Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. *Mol. Ecol.* **2000**, *9*, 2109−2118.

(29) King, G.; Nienhuis, J.; Hussey, C. Genetic similarity among ecotypes of *Arabidopsis thaliana* estimated by analysis of restriction fragment length polymorphism. *Theor. Appl. Genet.* **1993**, *86*, 1028−1032.

(30) Erschadi, S.; Haberer, G.; Schoniger, M.; Torres-Ruiz, R. Estimating genetic diversity of *Arabidopsis thaliana* ecotypes with amplified fragment length polymorphisms (AFLP). *Theor. Appl. Genet.* **2000**, *100*, 633−640.

(31) Miyashita, N. T.; Kawabe, A.; Innan, H. DNA variation in the wild plant *Arabidopsis thaliana* revealed by amplified fragment length polymorphism analysis. *Genetics* **1999**, *152*, 1723−1731.

(32) Baumbusch, L.; Thorstensen, T.; Krauss, V.; Fischer, A.; Naumann, K.; Assalkhou, R.; Schulz, I.; Reuter, G.; Aalen, R. The *Arabidopsis thaliana* genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. *Nucleic Acids Res.* **2001**, *29*, 4319−4333.

(33) Boyes, D. C.; Zayed, A. M.; Ascenzi, R.; McCaskill, A. J.; Hoffman, N. E.; Davis, K. R.; Gorlach, J. Growth stage-based phenotypic analysis of *Arabidopsis*: a model for high throughput functional genomics in plants. *Plant Cell* **2001**, *13*, 1499−1510.

(34) Lehesranta, S. J.; Davies, H. V.; Shepherd, L. V.; Nunan, N.; McNicol, J. W.; Auriola, S.; Koistinen, K. M.; Suomalainen, S.; Kokko, H. I.; Karenlampi, S. O. Comparison of tuber proteomes of potato varieties, landraces, and genetically modified lines. *Plant Physiol.* **2005**, *138*, 1690−1699.

(35) Jungblut, P. R.; Bumann, D.; Haas, G.; Zimny-Arndt, U.; Holland, P.; Lamer, S.; Siejak, F.; Aebischer, A.; Meyer, T. F. Comparative proteome analysis of *Helicobacter pylori*. *Mol. Microbiol.* **2000**, *36*, 710−725.

(36) Finnie, C.; Maeda, K.; O, O. S.; Bak-Jensen, K. S.; Larsen, J.; Svensson, B. Aspects of the barley seed proteome during development and germination. *Biochem. Soc. Trans.* **2004**, *32*, 517−519.

(37) Schlesier, B.; Berna, A.; Bernier, F.; Mock, H.-P. Proteome analysis differentiates between two highly homologues germin-like proteins in *Arabidopsis thaliana* ecotypes Col-0 and Ws-2. *Phytochemistry* **2004**, *65*, 1565−1574.

(38) Anderson, N. G.; Tollaksen, S. L.; Pascoe, F., H.; Anderson, L. Two-dimensional electrophoretic analysis of wheat seed proteins. *Crop Sci.* **1985**, *25*, 667−674.

(39) Damerval, C. Quantification of silver-stained proteins resolved by two-dimensional electrophoresis: genetic variability as related to abundance and solubility in two maize lines. *Electrophoresis* **1994**, *15*, 1573−1579.

(40) Kliebenstein, D. J.; Kroymann, J.; Brown, P.; Figuth, A.; Pedersen, D.; Gershenzon, J.; Mitchell-Olds, T. Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. *Plant Physiol.* **2001**, *126*, 811−825.

Received for review September 23, 2005. Revised manuscript received January 19, 2006. Accepted January 20, 2006.

JF052357Y