Proteomics analyses of herbicide-tolerant genetically modified, conventionally, and organically farmed soybean seeds

Varunjikar, M.S.; Bohn, T.; Sanden, M.; Belghit, I.; Pineda-Pampliega, J.; Palmblad, M.; ... ; Rasinger, J.D.

Food Control 151 (2023) 109795

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

Proteomics analyses of herbicide-tolerant genetically modified, conventionally, and organically farmed soybean seeds

M.S. Varunjikar [a], T. Bøhn [b], M. Sanden [a], I. Belghit [a], J. Pineda-Pampliega [a], M. Palmblad [c], H. Broll [d], A. Braeuning [d], J.D. Rasinger [a,*]

[a] Institute of Marine Research, Bergen, Norway
[b] Institute of Marin Research, Tromsø, Norway
[c] Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, the Netherlands
[d] German Federal Institute for Risk Assessment, Dept. Food Safety, Max-Dohrn-Str. 8-10, 10589, Berlin, Germany

ARTICLE INFO

Keywords:
Glycine max
Proteomics
Bioinformatics
Mass spectrometry
Transgenic soy
Genetic modification
GMO detection
GMO analysis
Food control
Feed control
Risk assessment

ABSTRACT

The present study compared genetically modified (GM) crops with crops from different farming practices using high-resolution tandem mass spectrometry (HR-MS) and proteomics bioinformatics tools. In a previously published study, a number of significant differences regarding nutritional and elemental composition between a selection of GM, non-GM conventionally farmed, and organic soybeans have been found. In the present study, the proteome-level equivalence of the same samples was assessed using HR-MS. Direct comparison of tandem mass spectra and bottom-up proteomics bioinformatics indicated that proteomes of all samples investigated were very similar overall, with only a few distinct protein expression clusters obtained for GM and organic samples. Standard bottom-up proteome analyses identified 1025 soy proteins; of these 39 were found to be differentially expressed ($p < 0.01$) between GM, non-GM conventionally farmed, and organically farmed soybeans. Subsequent bioinformatics analyses of these proteins highlighted several potentially affected biochemical pathways that could contribute to the compositional differences reported earlier. In addition, protein markers separating conventionally, and organically farmed soybean seeds were found and peptide markers for the detection of GM soy in food and feed samples are described. Taken together, the data presented here shows that HR-MS based proteomics approaches can be used for the detection of transgenic events in food and feed grade soy, the differentiation of organically and conventionally farmed plants, and provide mechanistic explanations of effects observed on the phenotypic level of GM plants. HR-MS and proteomic bioinformatics thus should be considered key tools when developing molecular panel approaches for detection and safety assessments of novel crop varieties destined for use in feed and food.

1. Introduction

Soybean (*Glycine max*) is an essential component of global food and feed value chains (S. Natarajan et al., 2013). To fulfill the rising global demand for soybeans, and to overcome challenges associated with soybean farming, genetically modified (GM) soy is the leading commercial biotech crop with 91.9 million hectares occupying almost half of the global biotech crop area in 2019 (ISAAA, 2019). The challenges associated with soybean farming (including, weeds, drought, diseases, and growth time), can be dealt with using traditional breeding techniques, genetic engineering and more recently, gene editing approaches (Ma et al., 2018). Among GM crops, glyphosate-tolerant GM soybeans are widely cultivated and have dominated the market by accounting for approximately 74% of total global production (ISAAA, 2019). At the

Abbreviations: (GM), Genetically Modified; (GMO), Genetically Modified Organism; (EPSPS), Enolpyruvylshikimate-3-Phosphate Synthase; (HR-MS), High Resolution-Mass Spectrometry; (PCR), polymerase chain reaction; (UHPLC-MS/MS), Ultra-High-Performance Liquid Chromatography coupled to tandem Mass Spectrometry; (ROS), Reactive Oxygen Species; (MS), Mass Spectrometry; (TPP), Trans-Proteomic Pipeline; (MGF), Mascot Generic Format; (mzXML), mass to charge ratio in eXtensible Markup Language.

* Corresponding author.
E-mail addresses: madhushri.shrikant.varunjikar@hi.no (M.S. Varunjikar), thomas.boehn@hi.no (T. Bøhn), monica.sanden@hi.no (M. Sanden), ikram.belghit@hi.no (I. Belghit), javier.pineda.pampliega@hi.no (J. Pineda-Pampliega), n.m.palmblad@lumc.nl (M. Palmblad), Hermann.Broll@bfr.bund.de (H. Broll), Albert.Braeuning@bfr.bund.de (A. Braeuning), josef.rasinger@hi.no (J.D. Rasinger).

https://doi.org/10.1016/j.foodcont.2023.109795
Received 19 January 2023; Received in revised form 29 March 2023; Accepted 14 April 2023
Available online 18 April 2023

time of writing, in the few dominating soy-producing countries (such as, United States, Brazil, Argentina, China and India), herbicide-tolerant GM varieties of soybean are among the preferred crops currently being cultivated (Bøhn et al., 2014; ERS, 2020).

In GM soybeans, herbicide tolerance is still the dominating trait. Within herbicide tolerance, resistance against glyphosate-ammonium is the most widely used application. It can be induced by inserting a transgenic construct into the plant genome which expresses the *Agrobacterium* strain CP4 analog of 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS); an enzyme which is essential for plant survival (Duke & Powles, 2008). This creates a glyphosate tolerant event that facilitates eradication of plant weeds by spraying glyphosate on the crops (Shaner et al., 2012). Previous studies point toward the differential expression of EPSPS in GM vs. non-GM soybean plants but this was considered unlikely to cause toxic effects or allergies and hence did not raise serious safety concerns (Swatkoski & Croley, 2020). Many countries do have regulations in place structuring the safety assessment and authorization of GM plants and animals (Turnbull et al., 2021). Also, the European Food Safety Authority (EFSA) concluded that the GM soybean varieties assessed so far are nutritionally equivalent to their conventional counterparts (EFSA, 2010). In addition, EFSA concluded that the products derived from these GM plants are as safe as conventional soybeans in terms of toxicity and allergenicity (EFSA, 2010). In 2022, more than 20 different GM soybean events resistant to glyphosate are authorized for various purposes in different countries around the world (OECD, 2022). However, in GM plants, along with desired traits, unintended metabolic variations might occur, potentially causing so-called unintended effect (Fernandez & Paoletti, 2018) with respect to plant growth, the environment or human health (Benevenuto et al., 2021; Gould et al., 2022). It was shown that GM soybean can contain glyphosate residues and glyphosate degradation products such as aminomethyl phosphonic acid (AMPA); these might affect or disturb plant metabolism in herbicide-tolerant GM varieties (Bøhn et al., 2014).

To better understand similarities and differences between GM and non-GM crops currently on the market, the monitoring of the biology and physiology using cutting-edge analytical technologies is key (Aguilera et al., 2018). The data generated and the experience gained from the application of these tools, also will aid the required development and application of omics-centered panel approaches for future product-focused safety assessments of crops as championed by (Gould et al., 2022). Several omics-based analytical approaches (including, genetics, proteomics, and metabolomics tools), were found to be suitable for profiling GM and non-GM soybeans (Natarajan et al., 2013). Recently, high-throughput gel-free proteomics techniques have increasingly become available and were implemented in protein analytics (Valletta et al., 2021). Approaches based on Ultra-High-Performance Liquid Chromatography coupled to tandem Mass Spectrometry (UHPLC-MS/MS), also referred to as High Resolution-Mass Spectrometry (HR-MS here onwards), can separate, identify and quantify several thousand low-abundance proteins in a single run (García-Cañas et al., 2010). Using different HR-MS approaches, several studies investigated differential protein expression in GM and non-GM parent soybean seeds and attempts have been made to implement proteomics bioinformatics to assess the safety of GM soybean in relation to the insertion of EPSPS (Benevenuto et al., 2021; de Campos et al., 2020; Jin et al., 2021; Liu et al., 2018, 2020). Overall, all these studies highlighted varying degrees of proteome-level similarities between GM and non-GM soybeans cultivated under different laboratory conditions (Benevenuto et al., 2022). An omics-based assessment of 'ready-to-market' samples obtained from field sampling has yet to be performed.

In addition to assessing the safety of such food and feed products, EU regulation requires correct labeling of GM feed and food to inform consumers if products do contain GM material (EU regulation No 1829/2003). To follow these regulations, the most commonly used analytical method to detect the presence of GM plant ingredients is polymerase chain reaction (PCR) (Abdullah et al., 2006; Mafra et al., 2008). PCR is a highly sensitive method to detect the genetic material specific for a species that is suitable for most samples. Proteomics approaches allow for the detection of the expression of transgenic proteins and can quantify such proteins in mixtures (Swatkoski & Croley, 2020). Detection of EPSPS in GM samples was shown to be feasible; multiple reaction monitoring (MRM) and parallel reaction monitoring (PRM) were previously used to detect low levels of this protein in GM soybean mixtures after enrichment and purification using HR-MS (Devi et al., 2020). In addition, it was demonstrated that using PRM, EPSPS also can be detected in complex processed soy- and corn-based products (e.g., infant formulas, corn-based cereals, tortilla chips and cornmeal) without the need for enrichment of the target protein (Swatkoski & Croley, 2020).

In short, HR-MS-based methods have shown potential for both the targeted detection of transgenic proteins in highly processed feed and food products and the analysis of proteome-level equivalence. For the latter, PCR-based methods are not suitable. To date, most proteomics analyses have been performed exclusively on soybeans cultivated under different laboratory conditions. In the present study, 31 soybean samples which were obtained directly from farmers' fields, compositionally analyzed and described by Bøhn et al. (2014) were subjected to a standard label free HR-MS-based proteomics workflow (Varunjikar et al., 2022) to assess: (i) if protein profiles differ systematically across three categories of soybean (GM, non-GM conventional and organic farmed); (ii) which proteins were differentially expressed across the three categories; (iii) if differentially expressed proteins affected different biochemical pathways in the three soybean categories; and (iv) if the expression of transgenic EPSPS products and peptides can be used for GMO detection/traceability in soybean samples generally.

2. Materials and methods

2.1. Soybean samples and characterization

Samples (n = 31) were obtained as described in Bøhn et al. (2014) from individual sites in Iowa, USA. Information such as seed type, cultivation process whether samples were Roundup Ready GM (n = 10), conventional (n = 10) or organically farmed (n = 11) were noted for all samples along with pesticide and seed treatments used (Bøhn et al., 2014). Detailed information about the samples is given in the Supplementary Table S1.

2.2. Extraction, solubilization and quantification of proteins

50 mg of soybean seeds were weighted into test tubes of the PlusOne GE Healthcare Life Science, USA. Sample grinding kit and solubilized with 1 mL lysis buffer prepared with 4% SDS, 0.1M Tris-HCl, pH 7.6. Samples were homogenized by keeping on ice and 1M Dithiothreitol (DTT) was added to obtain a final concentration of 0.1M. to remove resin and other debris samples were centrifuged for 10 min at 15,000g. The supernatant was collected, heated at 95 °C for 5 min, and centrifuged once again. The supernatants were collected into new tubes and stored at −20 °C until further processing. Protein concentrations were determined using a Pierce 660 assay (Thermo Fisher scientific).

2.3. In-solution digestion of proteins

Peptide extracts were prepared for mass spectrometric analyses using a Filter Aided Sample Preparation digestion protocol (Wiśniewski, 2016). In short, 40 μg of extracted proteins were diluted using 200 μL of urea (8M) solution prepared in Tris-HCl with pH 8.5 (100m) and transferred to Microcon 30, Millipore, ultrafiltration spin column. Proteins were alkylated with 50 mM of iodoacetamide (C_2H_4INO) for 20 min in the dark at room temperature. Subsequently, protein mixtures in the column were washed with 200 μL of 8M urea solution along with

100 μL of 50 mM ammonium bicarbonate (NH_4HCO_3). Trypsin was added to the filters (1:50 enzyme to protein ratio), and tubes were incubated for 16 h at 37 °C. Filters were centrifuged and washed with 40 μL of 50 mM ammonium bicarbonate solution and then eluted with 0.5M NaCl. After centrifugation, peptide concentrations were determined using the Nanodrop system (Thermo Scientific). Lastly, eluates were vacuum dried and stored at −20 °C.

2.4. Mass spectrometry

After peptide extraction, samples were analyzed at the Proteomics Unit at the University of Bergen, Norway (PROBE) as described in detail in Bernhard et al. (2019). In short, dried peptides after FASP extraction were dissolved in 2% acetonitrile and 0.1% formic acid solution. Samples were injected into an Ultimate 3000 RSLC system (Thermo Scientific, Sunnyvale, California, USA), which was connected to a linear quadrupole ion trap orbitrap (LTQ-Orbitrap Elite) mass spectrometer (Thermo Scientific, Bremen, Germany). The LTQ-Orbitrap was equipped with a nano spray Flex ion source (Thermo Scientific). Data was obtained in data-dependent-acquisition (DDA)- mode. Samples were loaded on Acclaim PepMap 100, 2 cm × 75 μm ID nano viper column, packed with 3 μm C18 beads. A flow rate of 5 μL/min for 5 min with 0.1% trifluoracetic acid was used for desalting. Peptides were separated employing a biphasic acetonitrile gradient from two nanoflow UPLC pump (flow rate of 270 nL/min) on a 500 mm Acclaim PepMap analytical column 500 mm × 75 μm ID nanoViper column packed with 3 μm C18 beads. Solvent A comprised 0.1% trifluoracetic acid (v/v) and water and solvent B was 100% acetonitrile. The gradient was 5% B during initial trapping for 5 min followed by 5–7% B over 1 min, 7–21% B for 134 min, 21–34% B for 45 min and 34–80% B for 10 min. Elution of very hydrophobic peptides and conditioning of the column were performed (20 min of isocratic elution using 80% B and 20 min isocratic elution with 5% B, respectively). Peptides eluting from the HPLC-column were ionized by electrospray. Mass spectrometry data were collected in data-dependent-acquisition (DDA) mode automatically switching between full scan MS and MS/MS acquisition. The instrument was controlled using Tune 2.7.0 and Xcalibur 2.2. Survey full scan MS spectra were recorded in the range of 300–2000 *m/z* with a resolution of R = 240,000 at 400 *m/z*. The most intense peptides (n = 12) above an ion threshold value of 3000 counts, and charge states of ≥ 2, were sequentially isolated and fragmented by low-energy collision-induced-dissociation (CID). The accumulation time for CID was 150 ms, the isolation was maintained at 2 Da, activation q = 0.25, and activation time was 10 ms. Fragment ions were scanned in a low-pressure ion-trap at normal scan rate and recorded. Lock-mass internal calibration was disabled and the dynamic exclusion range was 40s for MS2 spectra.

2.5. Direct spectral comparison using compareMS2

For direct spectral comparison of tandem mass spectra using compareMS2 (compareMS2 GUI, 2021; Marissen et al., 2022; Palmblad & Deelder, 2012). mgf files containing the top 10000 most intense tandem mass spectra were created using msConvert (version: 3.0., ProteoWizard (Kessner et al., 2008)). compareMS2 GUI (version 2.0.1) (Marissen et al., 2022) was used for pairwise comparison of MS2 spectra across all samples in the set.

2.6. Proteomics bioinformatics and pathway analyses

Proteomics data (.raw files) were processed using MaxQuant (version 1.6.10.43) following the protocol for label-free quantification described in (Tyanova et al., 2016)). The *Glycine max* reference proteome (UP000008827) and an additional sequence of 5-enolpyruvylshikimate 3-phosphate (EPSP) synthase from *Agrobacterium* sp. CP4 were used for protein identification and quantification using MaxQuant's own inbuilt peptide database search engine (Cox et al., 2011; Cox & Mann, 2008). As parameters, trypsin was selected as a digestive enzyme (allowing for two non-enzymatic termini), carbamidomethylation of cysteine and oxidation of methionine were set as fixed and variable modification, respectively and precursor mass tolerances were set to 4.5 ppm and fragment ion tolerances to 20 ppm for peptide identification. For label-free quantification (LFQ) in MaxQuant unique peptides, shared and razor peptides were used were used for identification and quantification of proteins. The output from the MaxQuant search was post-processed using Proteus (Gierlinski et al., 2018) in R (version 4.0.3) and data of protein intensity were log2 transformed and grouped into the respective categories.

Data processing and statistical comparison of proteomics samples was performed in Qlucore Omics Explorer V 3.6 (Qlucore AB, Lund, Sweden). Missing values in the dataset were replaced using Qlucore's inbuilt average value-based substitution algorithm. In Supplementary Table 2, information on the degree of sparsity is provided and imputed values are shown in italicized font. Following preprocessing, data were analyzed using ANOVA using the different cultivation regimes as grouping variable. The data was further explored using unsupervised principal component analyses (PCA) based on differentially expressed proteins across three groups and hierarchical cluster analysis (HCA). For all statistical analyses, a p-value <0.01 was used as significance cut-off; in addition, multiple test corrected q-values and log-fold changes are reported.

Differentially expressed proteins ($p < 0.01$) from *Glycine max* were matched against the *Arabidopsis thaliana* reference proteome, using PAW_BLAST db_to_db_blaster.py (https://github.com/pwilmart/PAW_BLAST; Supplementary Table S4). Proteins (with $p < 0.01$ and a log2 fold change of higher than 0.5 or lower than −0.5) were subjected to pathway-level analyses using the web-based tool AgriGO V2.0 (Du et al., 2010) and GO terms for both *Glycine max* and *Arabidopsis thaliana* were retrieved (Supplementary Table S5). GO term reduction was performed using REVIGO (Supek et al., 2011) (Supplementary Table S6). A list of up- and downregulated differentially expressed proteins ($p < 0.01$) obtained from Qlucore HM clusters (Supplementary Table S7) also was submitted to AgriGO V2.0. For said analyses, lists of differentially expressed proteins were submitted to the web-interface for Singular Enrichment Analyses (SEA) against the background *Glycine max*. Gene ontology (GO) terms associated with differentially expressed proteins in GM, conventional or organically farmed soy were retrieved; significantly affected GO terms also were plotted using pathway information from *Arabidopsis thaliana*, a well-studied model organism.

2.7. Peptide marker detection

For GM peptide marker detection, .raw files of soy samples were converted using msconvert version: 3.0 from ProteoWizard (Kessner et al., 2008), to. mgf and. mzML formats and a Trans-Proteomic Pipeline (TPP) (version 5.2.0 (Deutsch et al., 2015)) analysis was performed. Processed mass spectrometry data were deposited in an online repository (MassIVE id: MSV000089618 and ProteomeXchange id: PXD034405). The TPP pipeline-based Comet (Eng et al., 2013) searches were used for identification of peptide specific markers for GM soybean, using the same fasta file as mentioned above. Generated pepXML files were then processed further using R (version 4.0.3) and specific hits from *Agrobacterium* sp. CP4 were quantified in each sample. Only valid hits were filtered from this search and peptide markers were predicted.

3. Results

In the present study, the protein expression profiles of 31 soybean seed samples cultivated under three different cultivation conditions (GM, non-GM conventional and organic) and sampled in individual fields/sites in Iowa, USA (Supplementary Table S1) were studied using HR-MS and proteomics bioinformatics tools (Maier et al., 2010). The

sample set analyzed was described in detail in Bøhn et al. (2014). Raw MS data, as well as proteins and peptides identified in the present study are made available in the attached supplementary material (Supplementary Tables S1–S7 and Supplementary Figs. S2–S9) and publicly accessible proteomics data repositories (MSV000089618 and PXD034405).

3.1. Proteomics profiling of 31 soybean seed samples

Using HR-MS, on average 50,000 high quality MS2 spectra were acquired per sample. A direct comparison of the 10,000 highest intensity MS2 spectra using compareMS2 did not distinguish the samples in the set (Supplementary Fig. 1). A quantitative proteomic analysis (LFQ) of the 31 samples using MaxQuant and *Proteus* mapped over 6300 soy specific peptides and lead to the identification of 1289 proteins in the dataset; of these, 1025 were identified and quantified in at least one sample. The expression levels of 39 identified proteins were found to be differentially regulated ($p < 0.01$; Supplementary Table S2) due to the different cultivation conditions (GM, non-GM conventional and organic) applied (Supplementary Table S1).

A post-hoc analyses of the differentially regulated proteins (Supplementary Table S3) revealed that when compared to conventionally and organically farmed soy, GM soy displayed differential expression in 12 (2 up and 10 down regulated) and 23 (13 up and 10 down regulated) proteins, respectively. As can be seen in Fig. 1A and B, both number of proteins affected and magnitude of fold change were slightly higher when GM soy was compared with organically farmed seed samples. When comparing conventionally farmed soy with organically farmed soy, a total of 17 proteins (all up regulated) were found to be differentially expressed (Fig. 1C). Of note, the transgenic protein EPSPS only was observed in sample comparisons in which GM soy was included (Fig. 1A and B). Across all conditions tested, a total of 16 proteins overlapping responses were detected (Fig. 1D).

To further investigate the proteome level equivalence across the 31 samples cultivated under three different environmental conditions, principal component analyses (PCA) were used. The overall percentage of variance explained by the first three principal components (PCs) was 52%. A clear separation of conventionally (GM and non-GM) and organically farmed samples was observed along PC1; across PC2, a separation between conventionally farmed GM and non-GM soy was observed (Fig. 2A) indicating that differential protein expression was observed as a result of different farming practices. Hierarchical cluster analyses (HCA) of differentially expressed proteins also showed that a change in proteomic profiles of soybean seed samples can be observed depending on the cultivation regimes (Fig. 2B); according to the heatmap in Fig. 2, on the horizontal axis, a clear separation between organic and conventional and GM soy was observed.

3.2. Pathway analyses

To assess which biological functions are likely to be affected by the altered protein expression levels detected across GM, non-GM conventionally farmed, and organic soybean samples, significantly ($p < 0.01$) differentially expressed proteins displaying differential regulation with a log2 fold change of higher than 0.5 or lower than −0.5 (highlighted in blue color in the scatter plots of Fig. 2A–C) were analyzed further using AgriGO V2. The analysis yielded no results as the number of proteins passing the set significance and expression level thresholds (Supplementary Table S3) was below the minimum number of entries required by AgriGO V2 to successfully map entries to their corresponding GO terms. Instead of a GO term enrichment analysis, AgriGO V2 was then used to assess if individual GO terms associated with the selected subset

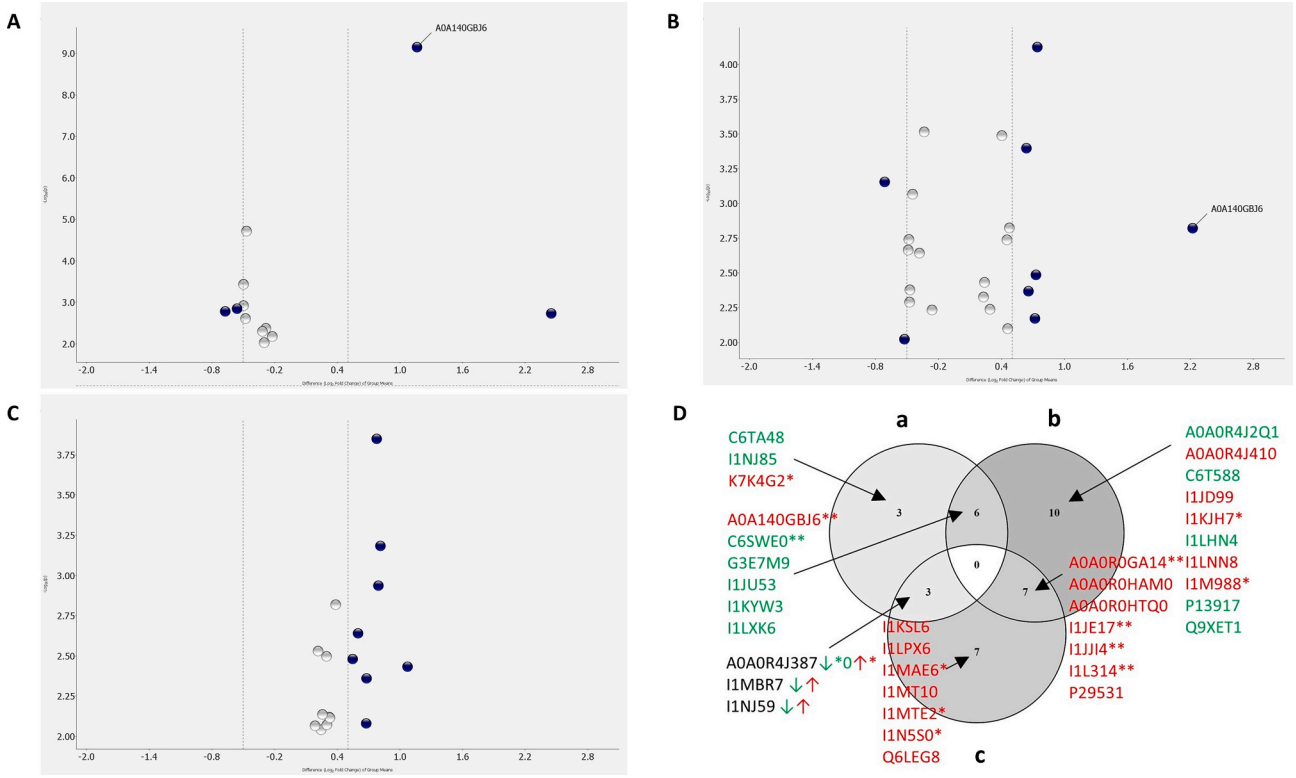

Fig. 1. (A) Differentially expressed proteins; comparison between GM vs Conventional, differences (Log_2 Fold Change) of Group Means; (B) Differentially expressed proteins; comparison between GM vs Organic, differences (Log_2 Fold Change) of Group Means; (C) Differentially expressed proteins; comparison between Conventional vs Organic, differences (Log_2 Fold Change) of Group Means. (D) Venn diagram comparing differentially expressed proteins in three groups GM, Conventional and organic a. GM vs Conventional differentially expressed proteins, b. GM vs Organic differentially expressed proteins, c. Conventional vs Organic differentially expressed proteins. Further information can be found in Supplementary Table S2.

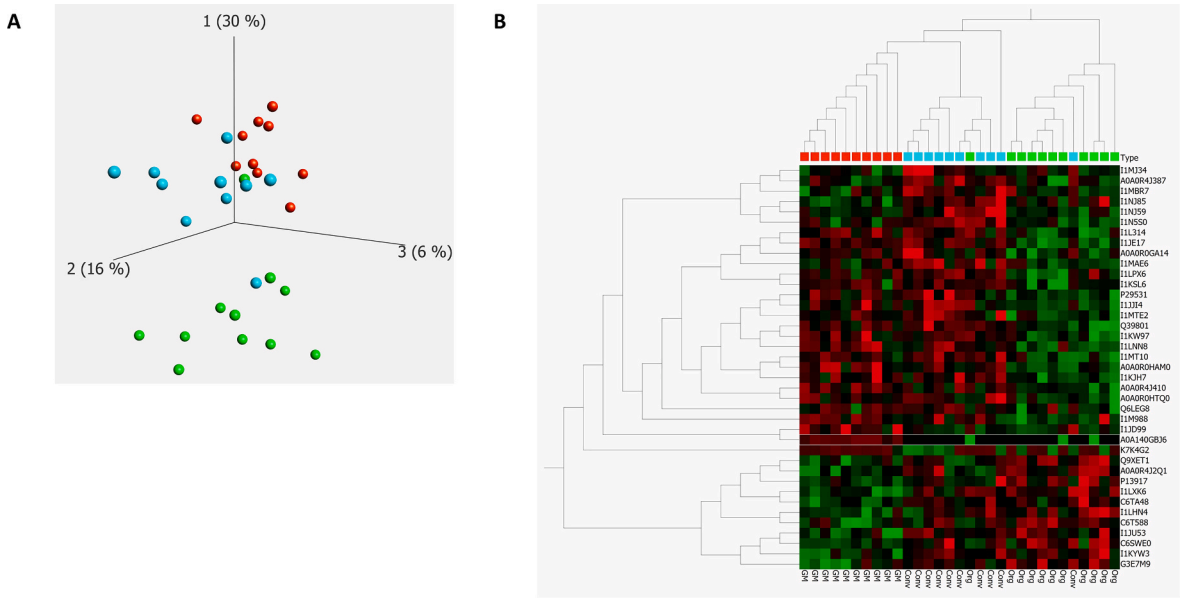

Fig. 2. (A) Principal component analysis (PCA) plot of differentially expressed proteins in three different categories (organic represented in green, GM represented in orange, and conventional represented in blue) of soybean seeds Glycine max depending on the cultivation regimes. GM soybean samples were separated from conventionally and organically farmed non-GM soybean seeds (p values as given in Supplementary Table S2). (B) Hierarchical clustering (HC) of soybean samples and differentially expressed proteins of three different categories (organic represented in green, GM represented in orange, and conventional represented in blue) of soybean seeds Glycine max depending on the cultivation regimes. Differential analyses (DA) and multigroup comparison were performed using the Qlucore Omics Explorer V 3.6. The heatmap represents the up and downregulated proteins within each measured sample where red bar represents upregulation of proteins and green bar represents downregulations of the same from the scale of −2 to 2 where 1 unit is equal to variance of 1 from the mean.

of *Glycine max* proteins could be extracted. Only one protein (A0A0R4J387) was mapped to two GO terms listed in AgriGO V2 namely, GO:0045430 (*chalcone isomerase activity*) and GO:0009813 (*flavonoid biosynthetic process*). To mine for additional GO terms, PW_BLAST was run, and soy proteins were matched with their respective orthologs from the model organism *Arabidopsis thaliana* (Supplementary Table S4). Re-running the AgriGO V2 analysis on the successfully mapped *Arabidopsis thaliana* protein identifiers, yielded several GO terms matches which in a subsequent REVIGO analysis were linked to different biological processes, molecular functions, and cellular components including, *flavonoid biosynthetic process, response to stress, proteolysis involved in cellular protein catabolic process, and monolayer-surrounded lipid storage body*. All successfully mapped *Glycine max* and *Arabidopsis thaliana* proteins and associated GO functions are listed in Supplementary Tables S5 and S6.

In addition to the conservative analysis of the 13 differentially expressed proteins ($p < 0.01$ and a log2 fold change of higher than 0.5 or lower than −0.5) described above, we subjected all 39 differentially expressed proteins ($p < 0.01$) with no log-fold change cut off, and an additional subset of differentially expressed 23 proteins ($p < 0.01$, no log-fold change cut off), which according to HCA analysis formed distinct clusters in the heatmap of Fig. 2B (Supplementary Table S7), to GO term enrichment analyses in AgriGO V2. These analyses also revealed significantly affected GO terms ($p < 0.05$) which were visualized and further analyzed using scatter plots and pathway information from the model organism *Arabidopsis thaliana*. GO terms including *seed oil body biogenesis*, *response to freezing*, *response to cold*, *lipid storage*, and *cellular modified amino acid metabolic process* were found to be affected in GM and conventionally farmed soybean samples when compared to organically farmed samples (Fig. 3).

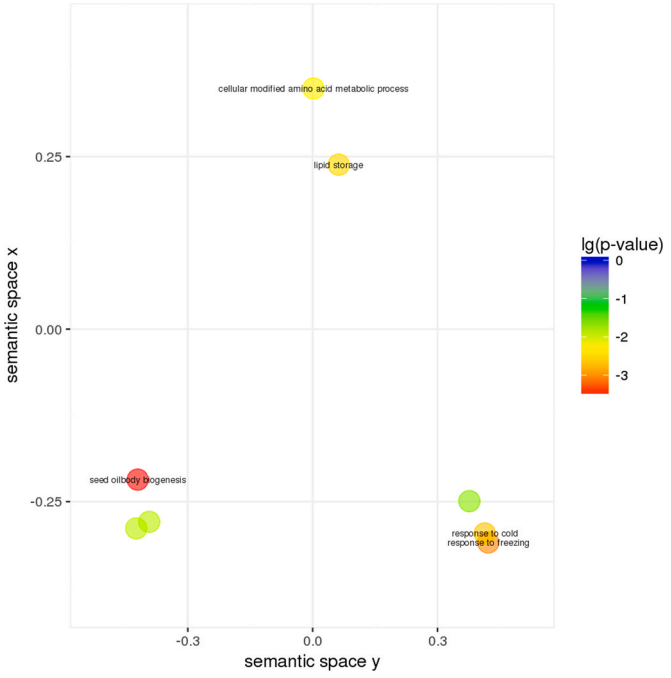

Fig. 3. AgriGO v2 analyses of differentially expressed proteins (GM and Conv) vs Org) according to heatmap clusters upregulated in GM and conventional (Supplementary Table S7); Scatter plot analyses of biological process similarity for the GO term with the other GO terms from model organism Arabidopsis (recommended for plant).

3.3. Peptide markers for EPSPS from Agrobacterium sp. CP4

As can been seen in Fig. 1 A, B and Fig. 2B, the MaxQuant analysis revealed that the transgenic protein EPSPS from bacteria "A0A140GBJ6" was found expressed in GM samples. It appeared to be absent in all other samples. For a more thorough identification of peptides specific to EPSPS from *Agrobacterium* sp. CP4, a TPP analysis was performed; three peptides specific to EPSPS were detected consistently across all GM soybean samples (Table 1). As the three peptides listed in Table 1 were absent in both conventional and organic soybean, three of

Table 1
List of CP4 EPSPS peptides and respective calculated neutral peptides mass, retention time, *m/z* ratios, xcorr, deltacn, expected value, ions, and detection status.

peptide	calc_neutral pep_mass	Retention time sec	mzratio	xcorr	deltacn	expect	ions	Previously detected
R.LAGGEDVADLR.V	1114.5619	3867.3	558.29	2.729	0.550	0.000766	18/20	Yes (Devi et al., 2020)
R.ITGLLEGEDVINTGK.A	1557.8250	7324.6	779.92	2.852	0.515	6.85e-006	17/28	Yes (Devi et al., 2020)
K.SAVLLAGLNTPGITTVIEPIMTR.D	2382.3192	10486.5	1192.2	3.208	0.520	0.00169	29/88	Yes (Swatkoski & Croley, 2020)

these could potentially be used as additional markers to detect and separate GM soybean from non-GM counterparts in pure and mixed food and feed samples. As shown in Table 1, the number of the detected peptides varies across the different GM soybean varieties; however, all these four peptides were detected in the GM samples exclusively.

4. Discussion

Direct comparison of tandem mass spectra, as implemented in compareMS2 (Marissen et al., 2022; Palmblad & Deelder, 2012), have been used earlier for food and feed authentication (Belghit et al., 2021; Ohana et al., 2016; Rasinger et al., 2016; Varunjikar et al., 2022; Wulff et al., 2013). Here, we evaluated if the tool can be used for differentiating field samples of the same species but with different genetic makeup (GM and non-GM) and grown under different cultivation regimes (conventional and organic farming), respectively.

Following data analysis with compareMS2, which did not reveal any specific patterns in the samples, mass spectrometry data were analyzed applying a standard bottom-up proteomics approach to further assess and describe differences between GM, non-GM conventionally farmed, and non-GM organic soybeans. A total of 1025 proteins were consistently identified across the 31 samples and subjected to relative quantification using LFQ (Supplementary Table S2). Compared to the total number of quantified proteins in the set, the proportion of differentially expressed proteins was rather low, namely 39 proteins (with a p-value cut off <0.01) and 20, when in addition to the statistical cut off, a log2 fold change cut-off of above and below 0.5 and -0.5, respectively was applied. This finding indicated that on proteome-level overall, the 31 soybean field samples were very similar irrespective of genetic modification or cultivation regime applied.

When comparing the total number of proteins identified across all samples with recent studies in literature (Jin et al., 2021; Liu et al., 2018, 2020), it was found that these were comparable in magnitudes. For example, using isobaric tags for relative and absolute quantitation (iTRAQ), Liu et al. (2018) identified approximately 1600 *Glycine max* proteins of which over 900 proteins were detected consistently across three biological replicates. When comparing the number of differentially regulated proteins detected in the present study with literature, depending on the cut-offs chosen, comparable or slightly higher total numbers were reported when different non-GM and GM soybean seeds were compared. For example, using a p-value cut off of $p < 0.05$ and fold change cut offs similar to the ones used in the present study (1.5 for upregulation and 0.67 for downregulation), Liu et al. (2018) detected 13 to 67 differentially expressed proteins, and Liu et al. (2020) reported on the detection of 35–38 differentially expressed proteins.

Among the 39 differentially expressed proteins described in the previous section (Fig. 1D), Supplementary Table S3), also the transgenic protein EPSPS from bacteria "A0A140GBJ6" was detected; it was exclusively found in GM soy samples (Fig. 2A and B). This transgenic protein was also found to be differentially expressed in several other studies investigating protein expression differences in transgenic soybean seeds (Jin et al., 2021; Liu et al., 2018, 2020). Transgenic EPSPS is considered non-toxic and non-allergenic and no safety concerns have been raised due to its incorporation in GM soy analyzed in this study (S. S. Natarajan, 2010; Swatkoski & Croley, 2020).

Further analysis of the data obtained in the present study using a TPP-based workflow confirmed the findings obtained using MaxQuant; peptides specific to EPSPS *Agrobacterium* sp. CP4 were detected only in GM-soybean samples. This finding was based on three EPSPS-specific peptides of which two already were described in literature previously (Devi et al., 2020) and two are described here for the first time. To the best of our knowledge, the three GM-specific peptides described in the present work (Table 1), were detected for the first time in soybean seeds using an untargeted proteomics workflow without any prior enrichment steps. For the screening and detection of transgenic proteins in processed food products, targeted mass spectrometry assays were previously developed and were tested on processed soybean infant formulation, corn cereals and tortilla chips (Swatkoski & Croley, 2020). It was also shown earlier that multiple-reaction monitoring can be used to successfully identify low concentrations of EPSPS in soybean mixtures if enrichment and purification steps are applied before MS analysis (Devi et al., 2020). The peptides detected here (Table 1) could thus be tested in future studies to assess if additional targeted mass spectrometry assays not relying on sample enrichment procedures could be developed to complement the ones already published in literature (Devi et al., 2020; Swatkoski & Croley, 2020). In addition, it could be assessed if spectra library focused approaches similar to the ones recently described for the HR-MS-based detection of non-permitted ingredients in insect feed (Belghit et al., 2021) could be adapted for use for GM detection. Both targeted and SL based mass spectrometry-based assays have been shown earlier to be effective in regulatory settings including for example, the detection and identification of smaller quantities of non-permitted PAP in highly processed food and feed samples (Belghit et al., 2021; Lecrenier et al., 2021) and the detection of insects in feed and food (Stobernack et al., 2022; Varunjikar et al., 2022). Ultimately, the development of additional assays will increase both specificity and sensitivity of GM-soybean detection in complex food and feed matrices which might be required when moving towards a product-based regulation of crops (Gould et al., 2022).

In addition to EPSPS, several other differentially expressed proteins were detected in the present study (Fig. 1D) whose expression level differences were separating conventionally farmed and GM soybeans from organically farmed samples (Fig. 2B). The transgenic protein EPSPS regulates shikimate and branched pathways in GM soybean samples and was in a recent study found to cause changes in expression levels of proteins involved in the maintenance of the metabolic balance of the GM plant under normal growth and stress conditions (Jin et al., 2021). To assess which biological and cellular functions were affected across three categories of soybeans (GM, non-GM conventional and organic farmed); differentially expressed proteins were analyzed using AgriGO V2.0; a pathway analysis tool of relevance to the agricultural community (Du et al., 2010) and used earlier for GO analyses of transgenic soybean seeds (Jin et al., 2021; S. Natarajan et al., 2020). An initial GO term mapping based on conservative threshold settings for differential expression revealed that as reported earlier, also in the present work proteins linked to biological processes including response to stress (GO:0006950) were detected (Table 1). A follow-up analysis using less stringent settings showed that in addition to proteins related to stress response, several molecular functions were affected in GM, conventionally, and organically farmed soybean (Supplementary Figs. 2–9) including a number of metabolic processes (*seed oil body biogenesis, lipid storage, cellular modified amino acid metabolic response*) possibly related to differences in composition.

GM and conventionally farmed non-GM soybean samples analyzed in the present work were collected directly from farmers' fields where these have been treated with different herbicides and pesticides

(Supplementary Table S1), which was reported earlier to affect biological processes on the phenotypic level (Bøhn et al., 2014). Upregulated proteins in GM and conventionally farmed non-GM soybeans (I1NJ59, I1MJ34) were associated with *glutathione transferase activity* (GO:0004364; Supplementary Fig. 8). Higher activity of *glutathione transferase* shows an increased stress response in seeds exposed to herbicides and pesticides (Benevenuto et al., 2021; Cummins et al., 1997). As samples for the present study were obtained from different agricultural fields within the same region (for details see Bøhn et al., 2014), the observed protein-level stress response could have been caused either by differing environmental and geographical conditions (micro-climate, soil quality, etc.), which could not be controlled for in the present study or be due to the different genetic backgrounds of the GM soybean varieties. However, the findings obtained when comparing GM with organic soybean samples were consistent with previous findings where ROS and stress-related enzymes also were found to be elevated in transgenic soy samples grown under controlled laboratory conditions (Benevenuto et al., 2021; Komatsu et al., 2017).

SEA enrichment analyses indicated that *seed oil body biogenesis*, *response to stress*, and *lipid storage* were affected in GM and conventionally farmed non-GM samples when compared to organically farmed soy (Fig. 3 and Supplementary Fig. S1). We found that *accumulation of lipid particles* and *lipid storage bodies* varied specifically in GM and conventionally farmed samples when compared to organically farmed soy (Fig. 3 and Supplementary Fig. S2). The study by Bøhn et al. (2014) reported significant variations in fatty acid components in the three soybean categories; specifically, changes were observed for linoleic and palmitic acid (Bøhn et al., 2014). The pathway-level variations found in the present study may provide a molecular explanation for the observed effects on phenotypic level i.e., the altered fatty acid lipid profiles of the seed samples reported by Bøhn et al. (2014).

Molecular functions associated with the *cellular modified amino acid metabolic process*, which is part of protein metabolism, were affected in GM and conventionally farmed soy. As shown in the previous study using the same sample material (Table 2, Bøhn et al., 2014), the total protein content of GM and conventionally farmed soybean was significantly lower than in organic soy samples. Varying protein content indicated a *modified amino acid metabolic process* (Supplementary Figs. 2 and 7) which again may influence the amino acid profile of these soybean sample groups. These findings overlap with a previous study by Natarajan et al. (2020), wherein proteomic profiling of soybean mutants showed enrichment of endoplasmic reticular-based proteins and several proteins associated with export metabolism (S. Natarajan et al., 2020). In the present study, in organically farmed soybeans also differentially expressed proteins related to *peptidase activity* (Supplementary Fig. 5) were detected. The previous study by Bøhn et al., 2014, indicated that the organic soybean material contained higher amounts of indispensable amino acids (IAAs) when compared to conventional and GM soybeans. The higher level of IAAs in the organically farmed soybean samples described by Bøhn et al., 2014 might be a result of the observed pathway level differences in amino acid metabolism observed here (Supplementary Figs. 2, 7, 8 and Fig. 3).

Protein level and abundance differences have been detected and described earlier in several studies which investigated proteome-level differences between GM and non-GM soybean samples (Jin et al., 2021; Liu et al., 2018, 2020). Previous studies have also related differentially expressed proteins to changes in GM *Glycine max* metabolic pathways, stress response and ROS (Jin et al., 2021). However, to the best of our knowledge, no studies exist to date which report on the comparison of HR-MS derived proteomic data obtained from field samples of organic, conventionally farmed and GM soybeans. The present study links molecular level changes of the proteome to phenotypic data of the same samples reported earlier (Bøhn et al., 2014) and revealed GM-soy-specific peptide markers not yet reported in the literature (Table 1). Due to the limited information available on the environmental conditions where the farm-grown soy samples were collected, we cannot exclude that the difference we observed were due to variation in local environmental conditions (e.g., soil type or microclimate) rather than genetic modifications and/or farming practices. Additional meta-data or further experiments performed in controlled laboratory conditions would be required to rule out confounding factors due to variation in environmental conditions.

Altogether our findings indicate that proteomic analyses of *Glycine max* HR-MS data can contribute to and support the rapid advance in the development of omics tools for the next generation risk assessment of plants (EFSA, 2022) and support the safety assessments of crop varieties, as suggested by Gould et al. (2022). To tackle omics data scarcity in non-model species (Rasinger et al., 2016) and in line with a recent call for the application of FAIR (findable, accessible, interoperable and reproducible) principles (Pineda-Pampliega et al., 2022) for ecological and evolutionary studies in plants (Manzano & Julier, 2021), all data presented and discussed here was made publicly available for the scientific community (MassIVE id: MSV000089618 and ProteomeXchange id: PXD034405).

5. Conclusions

This study aimed to evaluate and compare the proteomes of GM, non-GM, conventionally farmed, and non-GM organically farmed soybean, respectively. A total of 31 GM and non-GM soybean seed field samples cultivated under different agricultural practices were analyzed using HR-MS and several state-of-the-art proteomics bioinformatics analyses approaches. Three peptides specific to EPSPS from *Agrobacterium* sp. CP4 were detected exclusively in GM soybean seeds and several differentially expressed proteins were found that allowed for a separation of conventionally and organically farmed soybeans. In addition, a number of biochemical pathways were highlighted that possibly could contribute to the differences in the composition of these samples reported by Bøhn et al. (2014). Taken together, the data presented here shows that proteomics approaches can be used for the detection and tracing of transgenic events in food and feed grade soy, the differentiation of organically and conventionally farmed plants, and provide mechanistic explanations of effects observed on the phenotypic level of GM plants. Thus, HR-MS and proteomic bioinformatics analysis tools should be considered when developing panel approaches for product-focused safety assessments of crop varieties destined for use in feed and food.

CRediT authorship contribution statement

M.S. Varunjikar: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **T. Bøhn:** Investigation, Writing – review & editing. **M. Sanden:** Investigation, Writing – review & editing. **I. Belghit:** Writing – original draft, Writing – review & editing. **J. Pineda-Pampliega:** Writing – original draft, Writing – review & editing. **M. Palmblad:** Conceptualization, Data curation, Investigation, Methodology. **H. Broll:** Writing – original draft, Writing – review & editing. **A. Braeuning:** Writing – original draft, Writing – review & editing. **J.D. Rasinger:** Conceptualization, Supervision, Data curation, Investigation, Methodology, Project administration, Software, Writing – review & editing, Writing – original draft.

Declaration of competing interest

The authors declare no conflict of interest.

Data availability

The data have been deposited to MassIVE (with dataset identifier MSV000089618) and ProteomeXchange (with dataset identifier PXD034405).

Acknowledgments

Funding was provided by the Institute of Marine Research (Programledelse Fiskeernæring, MultiOmicsTools project 15470) and by the German Federal Institute for Risk Assessment (project 51–004).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodcont.2023.109795.

References

compareMS2 GUI. (2021). https://github.com/524D/compareMS2. https://github.com/524D/compareMS2.

Abdullah, T., Radu, S., Hassan, Z., & Hashim, J. K. (2006). Detection of genetically modified soy in processed foods sold commercially in Malaysia by PCR-based method. *Food Chemistry, 98*(3), 575–579. https://doi.org/10.1016/j.foodchem.2005.07.035

Aguilera, J., Aguilera-Gomez, M., Barrucci, F., Cocconcelli, P. S., Davies, H., Denslow, N., Lou Dorne, J., Grohmann, L., Herman, L., Hogstrand, C., Kass, G. E. N., Kille, P., Kleter, G., Nogué, F., Plant, N. J., Ramon, M., Schoonjans, R., Waigmann, E., Wright, M. C., & EFSA. (2018). EFSA scientific colloquium 24 – 'omics in risk assessment: State of the art and next steps. *EFSA Supporting Publications, 15*(11). https://doi.org/10.2903/sp.efsa.2018.EN-1512

Belghit, I., Varunjikar, M., Lecrenier, M.-C. C., Steinhilber, A. E., Niedzwiecka, A., Wang, Y. V. V., Dieu, M., Azzollini, D., Lie, K., Lock, E.-J. J., Berntssen, M. H. G. H. G., Renard, P., Zagon, J., Fumière, O., van Loon, J. J. A. J. A., Larsen, T., Poetz, O., Braeuning, A., Palmblad, M., & Rasinger, J. D. D. (2021). Future feed control – tracing banned bovine material in insect meal. *Food Control, 128*(April), Article 108183. https://doi.org/10.1016/j.foodcont.2021.108183

Benevenuto, R. F., Venter, H. J., Zanatta, C. B., Nodari, R. O., & Agapito-Tenfen, S. Z. (2022). Alterations in genetically modified crops assessed by omics studies: Systematic review and meta-analysis. *Trends in Food Science & Technology, 120*, 325–337. https://doi.org/10.1016/J.TIFS.2022.01.002

Benevenuto, R. F., Zanatta, C. B., Guerra, M. P., Nodari, R. O., & Agapito-Tenfen, S. Z. (2021). Proteomic profile of glyphosate-resistant soybean under combined herbicide and drought stress conditions. *Plants, 10*(11), 2381. https://doi.org/10.3390/plants10112381

Bernhard, A., Rasinger, J. D., Betancor, M. B., Caballero, M. J., Berntssen, M. H. G., Lundebye, A. K., & Ørnsrud, R. (2019). Tolerance and dose-response assessment of subchronic dietary ethoxyquin exposure in Atlantic salmon (Salmo salar L.). *PLoS One, 14*(Issue 1). https://doi.org/10.1371/journal.pone.0211128

Bøhn, T., Cuhra, M., Traavik, T., Sanden, M., Fagan, J., & Primicerio, R. (2014). Compositional differences in soybeans on the market: Glyphosate accumulates in Roundup Ready GM soybeans. *Food Chemistry, 153*, 207–215. https://doi.org/10.1016/j.foodchem.2013.12.054

de Campos, B. K., Galazzi, R. M., Dos Santos, B. M., Balbuena, T. S., dos Santos, F. N., Mokochinski, J. B., Eberlin, M. N., Arruda, M. A. Z. Z., Dos Santos, N., Ao, J., Mokochinski, B., Eberlin, M. N., & Arruda, M. A. Z. Z. (2020). Comparison of generational effect on proteins and metabolites in non-transgenic and transgenic soybean seeds through the insertion of the cp4-EPSPS gene assessed by omics-based platforms. *Ecotoxicology and Environmental Safety, 202*, Article 110918. https://doi.org/10.1016/j.ecoenv.2020.110918

Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology, 26*(12), 1367–1372. https://doi.org/10.1038/nbt.1511

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., & Mann, M. (2011). Andromeda: A peptide search engine integrated into the MaxQuant environment. *Journal of Proteome Research, 10*(4), 1794–1805. https://doi.org/10.1021/PR101065J/SUPPL_FILE/PR101065J_SI_002.ZIP

Cummins, I., Moss, S., Cole, D. J., & Edwards, R. (1997). Glutathione transferases in herbicide-resistant and herbicide-susceptible black-grass (Alopecurus myosuroides). *Pesticide Science, 51*(3), 244–250. https://doi.org/10.1002/(SICI)1096-9063(199711)51:3<244::AID-PS643>3.0.CO;2-2

Deutsch, E. W., Mendoza, L., Shteynberg, D., Slagel, J., Sun, Z., & Moritz, R. L. (2015). Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. *Proteomics - Clinical Applications, 9*(7–8), 745–754. https://doi.org/10.1002/prca.201400164

Devi, S., Chu, P.-Y., Wu, B.-H., & Ho, Y.-P. (2020). Mass spectrometry combined with affinity probes for the identification of CP4 EPSPS in genetically modified soybeans. *Journal of Mass Spectrometry, 55*(2), Article e4371. https://doi.org/10.1002/jms.4371

Duke, S. O., & Powles, S. B. (2008). Glyphosate: A once-in-a-century herbicide. *Pest Management Science, 64*(4), 319–325. https://doi.org/10.1002/ps.1518

Du, Z., Zhou, X., Ling, Y., Zhang, Z., & Su, Z. (2010). agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Research, 38*(suppl_2), W64–W70. https://doi.org/10.1093/nar/gkq310

EFSA. (2010). Scientific Opinion on applications (EFSA-GMO-RX-40-3-2[8-1a/20-1a], EFSA-GMO-RX-40-3-2) for renewal of authorisation for the continued marketing of (1) food containing, consisting of, or produced from genetically modified soybean 40-3-2; (2) feed containi. *EFSA Journal, 8*(12). https://doi.org/10.2903/j.efsa.2010.1908

EFSA. (2022). Theme (concept) paper - application of OMICS and BIOINFORMATICS approaches: Towards next generation risk assessment, 19(5). In *Iacono, Giovanni Guerra, Beatriz Kass, Georges Paraskevopoulos, Konstantinos Kleiner, Juliane Heppner, Claudia Hugas, MartaEFSA Supporting Publications*. https://doi.org/10.2903/sp.efsa.2022.e200506.

Eng, J. K., Jahan, T. A., & Hoopmann, M. R. (2013). Comet: An open-source MS/MS sequence database search tool. *Proteomics, 13*(1), 22–24. https://doi.org/10.1002/pmic.201200439

ERS. (2020). *Ers USA*. https://Www.Ers.Usda.Gov/Topics/Crops/Soybeans-Oil-Crops/.

EU regulation No 1829/2003. (1829). *EU regulation*, 1–5. https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:2006R1881:20100701:EN:PDF%0Ahttps://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A01985L0374-19990604&qid=1604918047856.

Fernandez, A., & Paoletti, C. (2018). Unintended effects in genetically modified food/feed safety: A way forward. *Trends in Biotechnology, 36*(9), 872–875. https://doi.org/10.1016/J.TIBTECH.2018.03.005

García-Cañas, V., Simó, C., León, C., Ibáñez, E., & Cifuentes, A. (2010). *MS-BASED analytical methodologies to characterize genetically modified crops*. https://doi.org/10.1002/mas.20286

Gierlinski, M., Gastaldello, F., Cole, C., & Barton, G. (2018). Proteus : an R package for downstream analysis of MaxQuant output. *bioRxiv*. , Article 416511. https://doi.org/10.1101/416511

Gould, F., Amasino, R. M., Brossard, D., Buell, C. R., Dixon, R. A., Falck-Zepeda, J. B., Gallo, M. A., Giller, K. E., Glenna, L. L., Griffin, T., Magraw, D., Mallory-Smith, C., Pixley, K. V., Ransom, E. P., Stelly, D. M., & Stewart, C. N. (2022). Toward product-based regulation of crops. *Science, 377*(6610), 1051–1053. https://doi.org/10.1126/science.abo3034

ISAAA. (2019). *Global status of commercialized biotech/GM crops in 2019: Biotech crops drive SocioEconomic development and sustainable environment in the new frontier* (Vols. 3–13). Ithaca, NY: ISAAA. https://www.isaaa.org/resources/publications/briefs/55/executivesummary/pdf/B55-ExecSum-English.pdf.

Jin, L., Wang, D., Mu, Y., Guo, Y., Lin, Y., Qiu, L., & Pan, Y. (2021). Proteomics analysis reveals that foreign cp4-epsps gene regulates the levels of shikimate and branched pathways in genetically modified soybean line H06-698. *GM Crops & Food, 12*(1), 497–508. https://doi.org/10.1080/21645698.2021.2000320

Kessner, D., Chambers, M., Burke, R., Agus, D., & Mallick, P. (2008). ProteoWizard: Open source software for rapid proteomics tools development. *Bioinformatics, 24*(21), 2534–2536. https://doi.org/10.1093/bioinformatics/btn323

Komatsu, S., Wang, X., Yin, X., Nanjo, Y., Ohyanagi, H., & Sakata, K. (2017). Integration of gel-based and gel-free proteomic data for functional analysis of proteins through Soybean Proteome Database. *Journal of Proteomics, 163*, 52–66. https://doi.org/10.1016/j.jprot.2017.05.009

Lecrenier, M.-C., Marien, A., Veys, P., Belghit, I., Dieu, M., Gillard, N., Henrottin, J., Herfurth, U. M., Marchis, D., Morello, S., Oveland, E., Poetz, O., Rasinger, J. D., Steinhilber, A., Baeten, V., Berben, G., & Fumière, O. (2021). Inter-laboratory study on the detection of bovine processed animal protein in feed by LC-MS/MS-based proteomics. *Food Control, 125*(November 2020), Article 107944. https://doi.org/10.1016/j.foodcont.2021.107944

Liu, W., Xu, W., Li, L., Dong, M., Wan, Y., He, X., Huang, K., & Jin, W. (2018). iTRAQ-based quantitative tissue proteomic analysis of differentially expressed proteins (DEPs) in non-transgenic and transgenic soybean seeds. *Scientific Reports, 8*(1), 1–10. https://doi.org/10.1038/s41598-018-35996-y

Liu, W., Zhang, Z., Liu, X., & Jin, W. (2020). iTRAQ-based quantitative proteomic analysis of two transgenic soybean lines and the corresponding non-genetically modified isogenic variety. *Journal of Biochemistry, 167*(1), 67–78. https://doi.org/10.1093/jb/mvz081

Mafra, I., Ferreira, I., & Oliveira, M. B. P. P. (2008). Food authentication by PCR-based methods. *European Food Research and Technology, 227*(3), 649–665. https://doi.org/10.1007/s00217-007-0782-x

Maier, T. S., Kuhn, J., & Müller, C. (2010). Proposal for field sampling of plants and processing in the lab for environmental metabolic fingerprinting. *Plant Methods, 6*(1), 1–14. https://doi.org/10.1186/1746-4811-6-6

Ma, X., Mau, M., & Sharbel, T. F. (2018). Genome editing for global food security. *Trends in Biotechnology, 36*(2), 123–127. https://doi.org/10.1016/J.TIBTECH.2017.08.004

Manzano, S., & Julier, A. C. M. (2021). *Biological science practices how FAIR are plant sciences in the twenty-first century? The pressing need for reproducibility in plant ecology and evolution*. https://doi.org/10.1098/rspb.2020.2597

Marissen, R., Varunjikar, M. S., Laros, J. F. J., Rasinger, J. D., Neely, B. A., & Palmblad, M. (2022). compareMS2 2.0: An improved software for comparing tandem mass spectrometry datasets. *Journal of Proteome Research*. https://doi.org/10.1021/acs.jproteome.2c00457, 0–5.

Natarajan, S. S. (2010). Natural variability in abundance of prevalent soybean proteins. *Regulatory Toxicology and Pharmacology, 58*(3), S26–S29. https://doi.org/10.1016/j.yrtph.2010.08.005

Natarajan, S., Islam, N., & Krishnan, H. B. (2020). Proteomic profiling of fast neutron-induced soybean mutant unveiled pathways associated with increased seed protein content. *Journal of Proteome Research, 19*(10), 3936–3944. https://doi.org/10.1021/acs.jproteome.0c00160

Natarajan, S., Luthria, D., Bae, H., Lakshman, D., & Mitra, A. (2013). Transgenic soybeans and soybean protein analysis: An overview. *Journal of Agricultural and Food Chemistry, 61*(48), 11736–11743. https://doi.org/10.1021/jf402148e

OECD. (2022). *No title*. https://biotrackproductdatabase.oecd.org/byOrganism.aspx.

Ohana, D., Dalebout, H., Marissen, R. J., Wulff, T., Bergquist, J., Deelder, A. M., & Palmblad, M. (2016). Identification of meat products by shotgun spectral matching. *Food Chemistry, 203*, 28–34. https://doi.org/10.1016/j.foodchem.2016.01.138

Palmblad, M., & Deelder, A. M. (2012). Molecular phylogenetics by direct comparison of tandem mass spectra. *Rapid Communications in Mass Spectrometry, 26*(7), 728–732. https://doi.org/10.1002/rcm.6162

Pineda-Pampliega, J., Bernhard, A., Hannisdal, R., Ørnsrud, R., Mathisen, G. H., Solstad, G., & Rasinger, J. D. (2022). Developing a framework for open and FAIR data management practices for next generation risk- and benefit assessment of fish and seafood. *EFSA Journal, 20*(S2). https://doi.org/10.2903/J.EFSA.2022.E200917

Rasinger, J. D., Marbaix, H., Dieu, M., Fumière, O., Mauro, S., Palmblad, M., Raes, M., & Berntssen, M. H. G. (2016). Species and tissues specific differentiation of processed animal proteins in aquafeeds using proteomics tools. *Journal of Proteomics, 147*, 125–131. https://doi.org/10.1016/j.jprot.2016.05.036

Shaner, D. L., Lindenmeyer, R. B., & Ostlie, M. H. (2012). What have the mechanisms of resistance to glyphosate taught us? *Pest Management Science, 68*(1), 3–9. https://doi.org/10.1002/ps.2261

Stobernack, T., Zarske, M., Niedzwiecka, A., Zagon, J., Steinhilber, A., Poetz, O., & Herfurth, U. M. (2022). LC-MS-based detection of silkworm pupae in feed with and without prior immunoaffinity enrichment. *Journal of Insects as Food and Feed*. https://doi.org/10.3920/JIFF2022.0098, 1–12.

Supek, F., Bošnjak, M., Škunca, S., & Šmuc, S. (2011). Summarizes and visualizes long lists of gene ontology terms. *PLoS One, 6*(7), Article 21800. https://doi.org/10.1371/journal.pone.0021800

Swatkoski, S. J., & Croley, T. R. (2020). Screening of processed foods for transgenic proteins from genetically engineered plants using targeted mass spectrometry. *Analytical Chemistry, 92*(4), 3455–3462. https://doi.org/10.1021/acs.analchem.9b05577

Turnbull, C., Lillemo, M., & Hvoslef-Eide, T. A. K. (2021). Global regulation of genetically modified crops amid the gene edited crop boom – a review. *Frontiers of Plant Science, 12*(February), 1–19. https://doi.org/10.3389/fpls.2021.630396

Tyanova, S., Temu, T., & Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature Protocols, 11*(12), 2301–2319. https://doi.org/10.1038/nprot.2016.136

Valletta, M., Ragucci, S., Landi, N., Di Maro, A., Pedone, P. V., Russo, R., & Chambery, A. (2021). Mass spectrometry-based protein and peptide profiling for food frauds, traceability and authenticity assessment. *Food Chemistry, 365*, Article 130456. https://doi.org/10.1016/J.FOODCHEM.2021.130456

Varunjikar, M. S., Belghit, I., Gjerde, J., Palmblad, M., Oveland, E., & Rasinger, J. D. (2022). Shotgun proteomics approaches for authentication, biological analyses, and allergen detection in feed and food-grade insect species. *Food Control, 137*(December 2021), Article 108888. https://doi.org/10.1016/J.FOODCONT.2022.108888

Wiśniewski, J. R. (2016). Quantitative evaluation of filter aided sample preparation (FASP) and multienzyme digestion FASP protocols. *Analytical Chemistry, 88*(10), 5438–5443. https://doi.org/10.1021/acs.analchem.6b00859

Wulff, T., Nielsen, M. E., Deelder, A. M., Jessen, F., & Palmblad, M. (2013). Authentication of fish products by large-scale comparison of tandem mass spectra. *Journal of Proteome Research, 12*(11), 5253–5259. https://doi.org/10.1021/pr4006525