Benevenuto *et al.*
Environmental Sciences Europe (2023) 35:14
https://doi.org/10.1186/s12302-023-00715-6

Environmental Sciences Europe

RESEARCH

Open Access

Integration of omics analyses into GMO risk assessment in Europe: a case study from soybean field trials

Rafael Fonseca Benevenuto[1], Caroline Bedin Zanatta[1], Friedrich Waßmann[2], Michael F. Eckerstorfer[3] and Sarah Zanon Agapito-Tenfen[4*]

Abstract

In Europe, genetically modified organisms (GMOs) are subject to an authorization process including a mandatory risk assessment. According to the respective guidance by the European Food Safety Authority (EFSA), one of the pillars of this GMO risk assessment is a comparative analysis of the compositional and agronomic characteristics. This targeted approach has been criticized for its limitations, as it only considers pre-determined compounds, being insufficient to assess a comprehensive range of relevant compounds, including toxins and anti-nutrients, on a case-specific basis. Strategies based on advanced untargeted omics technologies have been proposed as a potential broader approach to be implemented into the initial step of the risk assessment framework. Here, we provide an example of a step-by-step omics analysis based on systems biology approach to fit into the context of European GMO regulation. We have performed field trial experiments with genetically modified (GM) Intacta™ Roundup Ready™ 2 Pro soybean containing both cry1Ac and cp4epsps transgenic inserts and analyzed its proteomic profile against the non-GM counterpart and reference varieties. Based on EFSA's comparative endpoint-by-endpoint approach, the proteomics analysis revealed six proteins from the GMO outside the 99% tolerance intervals of reference varieties (RVs) in the equivalence test. Interestingly, from the near-isogenic (non-GM) comparator we found as many as ten proteins to be outside of the said RVs' equivalence limits. According to EFSA's statistical guidelines, differences found in metabolite abundance between a GMO and its non-GM comparator would not be considered biologically relevant as all compounds of concern remained within the equivalence limits of commercial RVs. By assessing the proteomic and metabolomic data through our proposed systems biology approach, we found 70 proteins, and the metabolite xylobiose as differentially expressed between the GMO and its non-GM comparator. Biological relevance of such results was revealed through a functional biological network analysis, where we found alterations in several metabolic pathways related to protein synthesis and protein processing. Moreover, the allergenicity analysis identified 43 proteins with allergenic potential being differentially expressed in the GM soybean variety. Our results demonstrate that implementation of advanced untargeted omics technologies in the risk assessment of GMOs will enable early and holistic assessment of possible adverse effects. The proposed approach can provide a better understanding of the specific unintended effects of the genetic modification on the plant's metabolism, the involved biological networks, and their interactions, and allows to formulate and investigate dedicated risk hypotheses in the first place. We draw conclusions on a detailed comparison with the comparative assessment according to EFSA and provide scientific arguments and examples on how the current comparative approach is not fit for purpose.

*Correspondence:
Sarah Zanon Agapito-Tenfen
saag@norceresearch.no
Full list of author information is available at the end of the article

Keywords Genetically modified organism, Safety, Biosafety, Biotechnology, Environmental risk assessment, Substantial equivalence, Proteomics, Metabolomics

Introduction

Genetically modified organisms (GMOs) intended for marketing as food and feed and derived products or for cultivation and release in the environment are subject to an authorization process in Europe in accordance with Directive 2001/18/EC and Regulation (EC) No 1829/2003. Commission Implementing Regulation (EU) No 503/2013 specifies the content of applications for authorization of genetically modified food and feed including the information required to be submitted. In order to advice this process, the European Food Safety Authority (EFSA) has issued several guidance documents with regard to the preparation and presentation of applications and, most importantly, the scientific information within risk assessment.

EFSA guidance documents are based on four pillars of GMO risk assessment: (i) a molecular characterization, which is an assessment of the molecular structure of the intended modification as well as any other unintended changes in the GMO; (ii) the comparative analysis, which is focused on compositional, nutritional and agronomic characteristics; (iii) an evaluation of potential toxicity and allergenicity; and (iv) an evaluation of the potential environmental impact of the GMO [1].

According to the EFSA guidances, possible alterations in the phenotype are identified through a comparative analysis of growth performance, yield, chemical composition, and more. A targeted approach (i.e., measurements of a limited number of individual compounds such as macronutrients, micronutrients, and certain crop-specific secondary metabolites) is used for the detection of compositional and nutritional differences between the GMO and its near-isogenic non-GM counterpart. This comparative approach applies the concept of substantial equivalence as food and feed derived from GMOs are compared to an appropriate comparator, defined by Regulation (EC) No 1829/2003 as "*a similar food or feed produced without the help of genetic modification and for which there is a well-established history of safe use*" [1, 2]. In general terms, the concept of substantial equivalence is based on the notion that existing organisms, such as those used as food sources, can be used as comparators when assessing the safety of the genetically modified organism. According to EFSA, the compositional analysis is not considered an endpoint analysis itself, but as a starting point of the case-specific risk assessment as it serves the purpose of identifying intended and unintended differences and/or lack of equivalences between GM plants and derived food and feed of their comparator(s) [1]. Since the application of substantial equivalence principle by EFSA, it became clear that the underlying criteria left scope for different interpretation by various risk assessors and academics who described the principle as unfit for purpose [3–9]. In addition, EFSA's comparative approach has been long and frequently criticized for its limitations with respect to a restricted and 'biased' selection of compounds that can be analyzed, as the detection of unknown toxins or anti-nutrients is not possible using this method [10–13].

There is a long-lasting and ongoing debate concerning the potential value of much broader scale, such as the use of unbiased molecular profiling approaches in risk assessment [14]. Such untargeted approaches, through the quantity of the data they generate, may help to: (a) identify effects which could trigger additional risk assessment hypotheses to be tested and (b) reduce the level of uncertainty that unintended changes have been overlooked [15]. Strategies based on advanced massive analysis of molecular data have been developed and successfully applied to screen genetically modified plant varieties for altered transcriptomic, proteomic or metabolomic profiles when compared to their non-GM counterparts [16–23], revised in [24]). The application of such molecular profiling analyses has been also suggested by the expert group on risk assessment and management serving the Cartagena Protocol on Biosafety serving the United Nations Convention on Biological Diversity [25]; these should be employed in those comparison studies where the scientifically most justifiable near-isogenic and conventional comparator would not grow under the relevant stress condition, or not grow as well, e.g., after herbicide application. In addition to these previous debates, the current ongoing discussion of the GMO regulations in Europe will certainly trigger revisions of its technical guidance, including EFSA guidance documents, to accommodate the risk assessment of organisms derived from New Genomic Techniques (NGTs)[26].

Several ways to apply or implement omics analysis into the risk assessment framework have been proposed [11, 13, 15, 24, 27–29]. More recently, EFSA has explored opportunities for integration of datasets produced via specific omics tools within risk assessment approaches in several fields, including

GMO risk assessment. In their report, the authority suggested the use of case examples that could be tested to enhance confidence in the use of omics datasets in risk assessment [30]. Similar to EFSA, the U.S. National Academies of Sciences, Engineering, and Medicine also acknowledges the usefulness of omics technologies to enable an examination of a plant's DNA sequence, gene expression, and molecular composition, as these techniques are expected to improve the efficiency of development of both non-GM and GM crops and could likewise be used to analyze new GMOs and test for unintended changes caused by the genetic engineering process [31]. However, none of these studies provide a clear implementation pathway for the GMO risk assessment in Europe. In this paper, we produced empirical data to test implementation, and we provide a clear pathway for omics analysis integration in the context of the European GMO regulation and EFSA's guidance documents.

Material and methods

Plant material

A total of seven soybean cultivars were selected for the field experiment and subsequent omics analysis: the stacked GM event BRS1001 Intacta™ Roundup Ready™ 2 Pro soybean (IPRO; unique identifier MON-877Ø1-2 × MON-89788-1) from Embrapa Brazil, containing transgenic elements from event MON89788, conferring glyphosate-tolerance (i.a. CP4-epsps behind the chimeric promoter P-FMV/Tsf1 and a chloroplast transit peptide sequence), and from event MON87701, conferring Lepidoptera-resistance (i.a. Cry1Ac from *Bacillus thuringiensis* behind the *A. thaliana* rbcS promoter and transit peptide); the near-isogenic non-GM variety BRS284 from Embrapa; as well as five non-GM commercial reference varieties (BRS 232; BRS 283; BRS 257; BRS 511 all from Embrapa and CD 216 from Codetec Brazil). Seeds were supplied by the local representatives and not tested in house for other GM events as they are certified seeds and follow Brazilian seed quality regulations.

Field conditions

A field experiment was conducted in the state of Santa Catarina, southern part of Brazil. The area is situated in 27°25' S and 51°31' W, and it is dedicated to agriculture land use only. The exact location of the field can be provided upon request to authors to avoid publication of a private area location. Soil is classified as Red Latosol dystrophic with clay texture [32]. To comply with the practices used in the region, soybean seeds were planted in no-tillage system. The area was previously treated with systemic glyphosate-based herbicide (GBH). Prior to sowing, potassium (150 kg/ha) and phosphorus (400 kg/ha)-based fertilizers were applied and incorporated in each planting line. Soybean seeds were subjected to treatments with insecticides and fungicides (active ingredients used: pyraclostrobin, thiophanate-methyl, and fipronil), as well as with *Bradyrhizobium japonicum*-based inoculant (1.2 ml per kg of seed, 6 × 10^9 cfu/ml). Seeds were manually planted on November 18th (2017) in a density of 200,000 plants/ha, with a distance of 0.10 m between plants and 0.50 m between lines. The following pesticides were used during the growing season following agricultural praxis in the region: Thiamethoxam, Lambda Cyhalothrin, Lufenuron, Trifloxystrobin, Prothioconazole, Diflubenzuron Mancozeb, Azoxystrobin, Chlorantraniliprole, Difenoconazole, Cyproconazole, Bentazon, Fomesafen, Clethodim. Other adjuvants and chemosynthetics were also used (Nimbus, Áureo, Triunfo, and methyl ester-based adjuvants). All soybean varieties were treated equally. No glyphosate-based herbicides were applied during the growing season. Leaf samples were taken at phenological stage V5 and V6 before flowering. Samples were composed from material from the third upper leaf taken from four plants from inner lines. Samples were immediately placed in 3.8-ml cryogenic tubes, frozen in liquid nitrogen, and kept in a -80° C freezer until protein and metabolites were extracted.

This field experiment followed the EFSA guidelines for statistical analysis for the safety of genetically modified organisms [33]. Briefly, the plot area was replicated at seven plots, each one was defined as an area of 40 m × 7.5 m (L x W), divided into four randomized blocks of 20 m^2 (named blocks) (Additional File 1).

Proteomics analysis

Total protein was extracted from soybean leaf samples (86 samples corresponding to the different varieties, 4 blocks, 8 plots) according to Carpentier et al. [34] with modifications. Briefly, 100 mg of leaf tissue was weighed and mixed with the extraction buffer (EDTA 5 mM, KCl 100 mM, sucrose 30%, TRIS HCl pH 8.5 50 Mm and protease inhibitor following the concentration provided by the manufacturer (cOmplete™, EDTA-free Protease Inhibitor Cocktail from Roche). Samples were ground in a Precellys 24 automatic homogenizer. Proteins were extracted using a phenol-based solution and precipitated in ammonium acetate (100 mM) in methanol. After precipitation, proteins were washed twice with 20% DTT in acetone. The pellet was resuspended in 1.5% urea and 100 mM TEAB. Finally, protein extracts were quantified in a spectrophotometer (reading at 562 nm) with the BCA protein kit (Novagen working reagent) and adjusted to a concentration of 2 µg/µl total. Samples

were then analyzed using the mass spectrometer with labeling TMT 11-plex, LC–MS/MS (Arctic University of Tromsø, Proteomic Platform). First, protein samples were placed on a nanoLC, before sequential injection into an Orbitrap (Q-Exactive) instrument with high mass accuracy. Then, the peptides were fragmented in an order of ten times in MS by high-energy collisional dissociation (HCD). The mass spectra of peptides and fragmented peptides were used for the identification of proteins and post-translational modification (PTM), as well as their quantification. For protein identification, Proteome Discoverer (Thermo Fisher Scientific) was used, and quantification was carried out using MaxQuant software with Perseus.

Metabolomic analysis

Metabolites were extracted from the collected leaf samples and sent to the Swedish Metabolic Center of Sweden (University of Umeå, Sweden) for subsequent GC and LC–MS analysis. Sample preparation and metabolite extraction were performed as described by Jiye et al. [35]. In a 20 µl sample, 1000 µl of extraction buffer (60/20/20 v/v methanol:chloroform:water) were added together with a tungsten granule. Additionally, quality control (QC) (metabolite extract grouped) as well as the extraction blanks was analyzed and processed. Samples were shaken in a mixing mill and then centrifuged at 4° C and 14,000 rpm for 10 min. The supernatants were transferred to LC and GC microvials, respectively, and the solvents were evaporated. The resulting samples were divided into three aliquots for analysis on three platforms of MS instruments. These included two UPLC/MS platforms, one optimized for positive ionization and a second optimized for negative ionization UHPLC Agilent 1290 Infinity (Agilent, Waldbronn, Germany). The third was derivatized and analyzed by GC / MS. The UPLC–MS/MS platform included a Waters ACQUITY ultra-performance liquid chromatography (UPLC). The compounds were detected with an Agilent 6550 Q-TOF mass spectrometer equipped with a jet electrospray ion source operating without negative ion mode. A reference interface has been connected for specific mass accuracy. Compounds were quantified via peak area of the total ion mass. The identification of chemical compounds was based on comparisons with entries from the metabolic library of purified standards.

Statistical analyses

Two statistical approaches were performed in this study, hereafter referred to as 'statistical analysis #i' and 'statistical analysis #ii'. The first analysis followed the statistical guidelines proposed by EFSA for the comparative assessment of compositional data from the statistical considerations for the safety evaluation of GMOs (2010) as discussed in the EFSA Omics Colloquium [30]. A second statistical analysis is presented here as an alternative approach as part of the molecular characterization in risk assessment. This new approach is based on a comprehensive untargeted metabolic and physiological assessment for the identification of unintended changes in the plant as whole.

Prior to both statistical analyses, metabolomics and proteomics data were normalized to the median distribution, auto-scaled, and log transformed aiming to facilitate statistical comparisons. For interpretation of the numerical values, means and differences of means on the logarithmic scale have been back-transformed to geometric means and ratios of geometric means on the original scale. These data treatments follow consensus standards for these analyses [36–39].

Statistical analysis #i

Identified proteins and metabolites were statistically analyzed based on 'endpoint-by-endpoint' comparative analysis as suggested by the EFSA statistical guidance for compositional analysis. This assessment is composed of two sets of comparative tests: first, a difference test to demonstrate whether the GMO is different from its near-isogenic control comparator (i.e., GMO vs. non-GM); followed by an equivalence test of the GMO compared to a range of conventional varieties to demonstrate whether it is equivalent to commercial reference varieties with a 'history-of-safe-use' (i.e., GMO vs. reference varieties (RVs) range) [33]. Statistical significance of the difference test was defined as $p < 0.05$ as determined using the two-tailed Student's t-test. Values for significance were adjusted for the false discovery rate (FDR) with the Bonferroni–Holm method (p-adj FDR < 0.05) [40]. Plots were analyzed both separately (within plots) and combined (across all plots). For the equivalence test, the range of observed values from the reference varieties was determined for each analytical component and used to calculate tolerance intervals. A tolerance interval is a range of values, with a specified degree of confidence, which contains at least a specified proportion, of an entire sampled population for the parameter measured. For each significant protein or metabolite in the difference test (GMO vs. non-GM), a 99% tolerance interval representing the equivalence limit was calculated that is expected to contain, with 95% confidence, 99% of the quantities expressed in the population of commercial conventional varieties. The tolerance intervals estimate was based on a total of 20 observations from 6 reference varieties from all plots. Finally, each mean value and standard deviation (mean ± SD) from statistically

different compound in the GMO (p-adj < 0.05) was compared to the 99% tolerance interval for the equivalence test. Non-equivalence is determined when the statistically different mean ± SD from the GMO sample falls outside the 99% RVs tolerance interval.

Statistical analysis #ii

The proposed alternative statistical model for omics data analysis was performed for both single plots and analysis across all plots. Exploratory multiple co-inertia analysis (MCIA) and principal component analysis (PCA) were conducted to investigate and geometrically projects the main sources of variation present in the proteomic and metabolomic data sets (Lever, Krzywinski & Altman 2017). Then, a comparative statistical analysis of whole-proteome and metabolome data for GM vs. non-GM plants was performed searching for potential metabolic alterations. Therefore, the focus was not on the endpoint-by-endpoint analysis but rather on the relationships and metabolic functions of proteins and metabolites as a whole. Scaling of the data, PCA, MCIA, volcano plots, and tolerance intervals were produced using MetaboAnalyst 5.0 (https://www.metaboanalyst.ca), as well as *ggplot2*, *msmsTests*, *omicade4*, and *tolerance* packages in R environment. Functional enrichment analysis and interaction network of differential proteins and metabolites were performed using Stitch 5.0 (http://stitch.embl.de) and String 11.0 (https://string-db.org).

The statistical significance of PCA was defined as $p < 0.05$, as determined using the two-tailed Student's t-test and false discovery ratio correction with the Bonferroni–Holm method [40]. Fold changes are also presented in logarithm base 2 (Log2FC), a widely used transformation for a continuous spectrum of values to represent up- (positive) and down-regulated (negative) compound values in a reader friendly fashion. Functional annotation and identification of enriched metabolic pathways were performed with UniProt database (https://www.uniprot.org) and KEGG pathway enrichment analysis (Kyoto Encyclopedia of Genes and Genomes) using the differentially expressed proteins and metabolites as input. Pathways with p-adj FDR < 0.05 were considered as significantly enriched. Additionally, Stitch and String databases were used to generate biological networks of protein-metabolite interactions aiming to facilitate data interpretation. A cut-off score for the confidence of interaction ≥ 0.4 was used for a more reliable biological network.

A second statistical analysis was performed in order to search for allergenic proteins in statistically different proteins using Allergen Online v.20 database (http://www.allergenonline.org). Allergens were searched using Full Fasta 36 algorithm method with E-value cutoff = 1. Our search parameters followed the guidelines of Codex ([41]/2005) for the evaluation of the potential allergenicity of novel proteins, which suggests that matches of at least 35% identity may indicate the possibility of cross-reactivity. The presented E-value statistical score is calculated based on the overall length sequence alignments and the quality (% identity and similarity) of the overlap amino acids. The size of the E-value is inversely related to similarity of two proteins, meaning a low E-value indicates a high degree of similarity between the query sequence and the matching sequence from the database, while a value close to 1 indicates the proteins are not likely to be related in evolution, or structure.

Results

Statistical analysis #i—omics data integrated into current comparative analysis of GMO

The comparative analysis of GM vs. non-GM proteomic datasets showed 15 differentially expressed proteins (eight down-regulated and seven up-regulated in the GMO) in plot 1; four proteins (one up-regulated and three down-regulated) in plot 3; 70 proteins (17 up-regulated and 53 down-regulated) in plot 4; 14 proteins (seven up-regulated and seven down-regulated) in plot 5; four proteins (one up-regulated; and three down-regulated) in plot 6; two proteins (one up-regulated and one down-regulated) in plot 7; and six proteins (five up-regulated and one down-regulated) in plot 9. Combined plot analysis showed only two differentially expressed proteins, both down-regulated in the GMO (Additional file 2).

Equivalence tests showed that the majority of the differentially expressed proteins fell within the 99% tolerance interval representing the equivalence limits established from reference varieties. However, two proteins (I1KG57; I1LI58) from plot 4 and four proteins (I1KXW8; A0A0R0HT35; A0A0R0EPX0; I1MFX5) from plot 5 from the GMO fell outside the equivalence limits (Table 1). In fact, these proteins were not detected in the reference varieties. According to the EFSA statistical guidance, these particular results are considered statistically different from the near-isogenic non-GM variety, as well as non-equivalent to the commercial reference varieties available in the market.

Most surprisingly, ten proteins from the non-GM variety also fell outside the equivalence limits calculated based on RVs tolerance interval (Table 1). Also, a RVs tolerance interval could not be calculated in two cases (plot 4 analysis: I1MPE8; A0A0R0KAT4) in which the proteins were not detected in more than two biological replicates of the reference varieties samples.

Table 1 List of statistically different proteins among the GMO and non-GM soybean varieties single-plot analyses which fall outside the reference varieties equivalence limits

Protein ID	GMO (± SD)	Non-GM (± SD)	p-adj	RVs tolerance interval
Proteins in the GMO falling outside RVs equivalence limits				
Plot 4				
I1KG57	1.2083 (± 0.04)	1.6010 (± 0.02)	0.025	[1.275, 1.874]
I1LI58	0.5967 (± 0.02)	0.9899 (± 0.02)	0.011	[0.696, 1.293]
Plot 5				
I1KXW8	0	0.6891 (± 0.04)	< 0.001	[0.304, 1.669]
A0A0R0HT35	0	0.9271 (± 0.14)	0.014	[0.400,1.598]
A0A0R0EPX0	0	1.1651 (± 0.05)	< 0.001	[0.346, 1.664]
I1MFX5	0	1.0678 (± 0.08)	0.002	[0.181, 1.593]
Proteins in the non-GM plants falling outside RVs equivalence limits				
Plot 1				
A0A368UGC3	0.8394 (± 0.02)	0	0.004	[0.792, 1.325]
I1K6K2	1.8341 (± 0.04)	0	0.009	[1.507, 2.522]
Q42447	1.6765 (± 0.09)	0	0.022	[0.535, 2.907]
I1L0T3	2.0508 (± 0.07)	0	0.014	[0.057, 5.402]
A0A0R0HP07	1.1579 (± 0.03)	0	0.004	[0.696, 1.684]
Plot 4				
I1JTZ7	1.4583 (± 0.06)	1.8714 (± 0.02)	0.031	[1.001, 1.598]
I1JBW7	− 0.5662 (± 0.16)	2.0108 (± 0.24)	0.025	[− 2.662, 1.681]
I1LF43	1.1062 (± 0.23)	0	0.025	[0.819, 2.081]
Plot 6				
I1NAY9	0.6706 (± 0.04)	0	< 0.001	[0.106, 1.382]
Plot 7				
A0A0R0HJQ2	2.1241 (± 0.05)	− 0.3230 (± 0.03)	0.022	[0.999, 2.656]

This table shows the protein UniProt ID; mean ± standard deviation expression values for the GMO and non-GM variety; false discovery rate adjusted p-value; and tolerance interval with 95% confidence, representing the equivalence limits of 99% of the values expressed in the population of reference varieties

Metabolomics data analysis showed three metabolites (glycine; tyrosine; melibiose) with statistically significant differences between the GMO and non-GM from plot 3; and one metabolite (xylobiose) in plot 4. Glycine (log2FC = 0.24; p-adj = 0.008), melibiose (log2FC = 0.79; p-adj = 0.023), and xylobiose (log2FC = 0.58; p-adj = 0.023) showed higher concentrations in GM samples, while tyrosine (log2FC = − 1.65; p-adj = 0.023) showed significantly lower amounts compared to non-GM samples. The analysis of combined data from all plots did not show any statistical differences between both varieties. According to equivalence testing, all metabolites fell within the 99% tolerance interval calculated based on the values observed in the reference varieties. This leads us to assume that, despite the significant differences between samples derived from GM and non-GM plants, such differential metabolites are equivalent to the range of commercial reference varieties observed and, therefore, these differences are not considered biologically relevant based on the current EFSA guidelines for comparative risk assessment of food and feed from GM plants.

Statistical analysis #ii—omics data integrated into the molecular characterization and allergenicity assessment

Multiple co-inertia analysis (MCIA) was performed in order to explore the experimental quality and the main sources of variation, including environmental variation, in the proteomic and metabolomic datasets from all seven plots (1, 3, 4, 5, 6, 7, 9) and three genotypes (GMO, non-GM, and RV) simultaneously. MCIA has been recognized as an excellent tool for integrating the results of different omics techniques. It is an exploratory data analysis method that is able to provide a simple graphical representation that identifies the concordance between these multiple datasets [42].

First, we performed a MCIA with datasets from all experimental plots aiming to evaluate how the variation in all data obtained behave. The coordinates of each plot for each treatment are connected by lines, the lengths of which indicate the divergence (the shorter the line, the higher the level of concordance) between the metabolites and protein abundance levels for a particular plot. In the

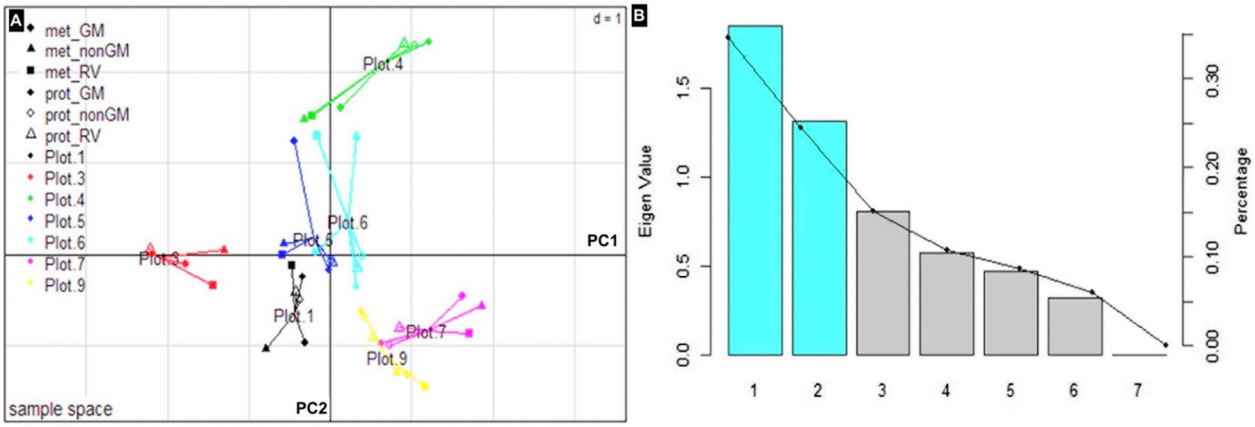

Fig. 1 MCIA projection plot. **A** MCIA projection plot representing the proteomics and metabolomics datasets from seven experimental plots: PC1 = 34.7% and PC2 = 24.6%. PC1 is represented by the first axis (horizontal), and PC2 is represented by the second axis (vertical). Different symbols represent the respective treatments and omics analyses and are connected by lines where the length is proportional to the divergence between the data from the same replicate. Lines are joined by a common point, representing the reference structure, which maximizes covariance derived from the MCIA synthetic analysis. Colors represent the different field plots. **B** Eigenvalue and percentage graphics show the amount of variation in the dataset corresponding to each PC

principal component 1 (PC1), there is a clear grouping of plots 1, 3, and 5, and another grouping of plots 4, 6, 7, and 9, accounting for 34.7% of the total variation in the datasets. On the other hand, PC2 shows separation of plots 1, 3, 7, and 9, from plots 4, 5, and 6, accounting for 24.6% of the total variation (Fig. 1). Such results are generally in accordance with visual observations of agronomic characteristics made during the field experiment, such as differences in the development of plants from the lowland plots (1, 3, 4) compared to the other plots, probably due to the substantial variation in environmental conditions, which includes the sun light incidence and soil moisture accumulation.

We have conducted MCIA for both omics datasets within each experimental plot aiming to evaluate the convergence and divergence of proteomic and metabolomic data from the GMO, non-GM and RV varieties inside the plots. We found that four (plots 3, 4, 7, and 9) out of seven plots presented similar trends in the proteome and metabolic profiles, in which PC1 showed clear distinct separation between GM and non-GM plants accounting for 49–67% of the total variation in the dataset. However, there was no pattern in the distribution of the RV group in the MCIA analysis across all plots. Therefore, we have run MCIA with only GM and non-GM groups for the same plots. Running MCIA without the RV samples results in more distinct clustering of the GM and non-GM groups, with two PCs accounting for more than 80% of the total variance (Additional file 3). Experimental plots 1, 5 and 6 did not show any clear pattern in the MCIA distribution of datasets. There was a high divergence between proteomic and metabolomic profiles depicted by the length of coordinates for each biological replicate when compared to other plots. Such result might be attributed to the variation in environmental conditions found in the respective plots located in a specific area of the field experiments (i.e., differences in the micro-climate between plots, differences in forest shading, differences in fertility and water drainage due to slope difference).

In order to test our alternative statistical approach, we selected the experimental plot with lower environmental variation. MCIA of plot 4 showed that PC1 clearly separated proteomics and metabolomics data from GM and non-GM groups, which accounted for 70.58% of the total variation in the data (Fig. 2A). We found similar results from both omics data sets separately by PCA. Clear differences between GM and non-GM groups, as well as within-group clustering of biological replicates were demonstrated by PCA for both omics data, with a total of 67.9% (metabolomics) and 69.4% (proteomics) of the variance accounted for in 2 PCs (Fig. 2B, C).

Analysis of unintended changes in proteomic and metabolomic profiles

We conducted an overall comparative analysis of the proteome and metabolome profiling aiming at searching for unintended metabolic changes in plants. We first applied a comparative assessment conducting pairwise *t*-tests ($p < 0.05$) between the proteomic and metabolomic profiles of the GMO vs. non-GM varieties within plot 4 as an example. Among the total of 74 analyzed metabolites, only xylobiose was abundant in significantly higher concentration (fold change = 1.5; Log2FC = 0.58;

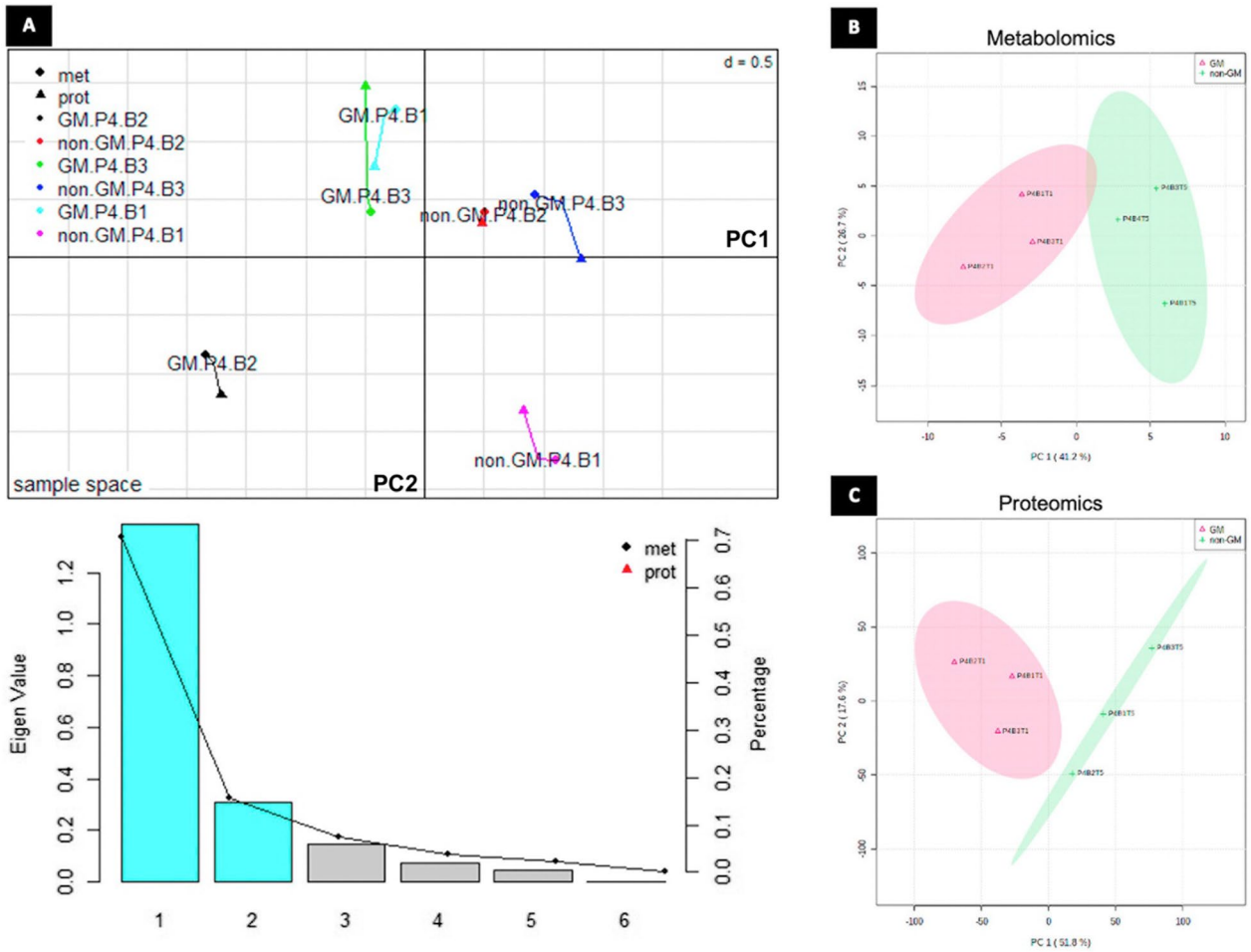

Fig. 2 Exploratory analysis of omics data from plot 4. **A** MCIA projection plot representing the proteomics and metabolomics datasets from experimental plot 4: PC1 = 70.58% and PC2 = 15.68%. Different symbols represent the different omics analyses, and colors represent the biological replicates. **B** PCA projection plot for metabolomics data of plot 4: PC1 = 41.2% and PC2 = 26.7%. **C** PCA projection plot for proteomics data of plot 4: PC1 = 51.8% and PC2 = 17.6%

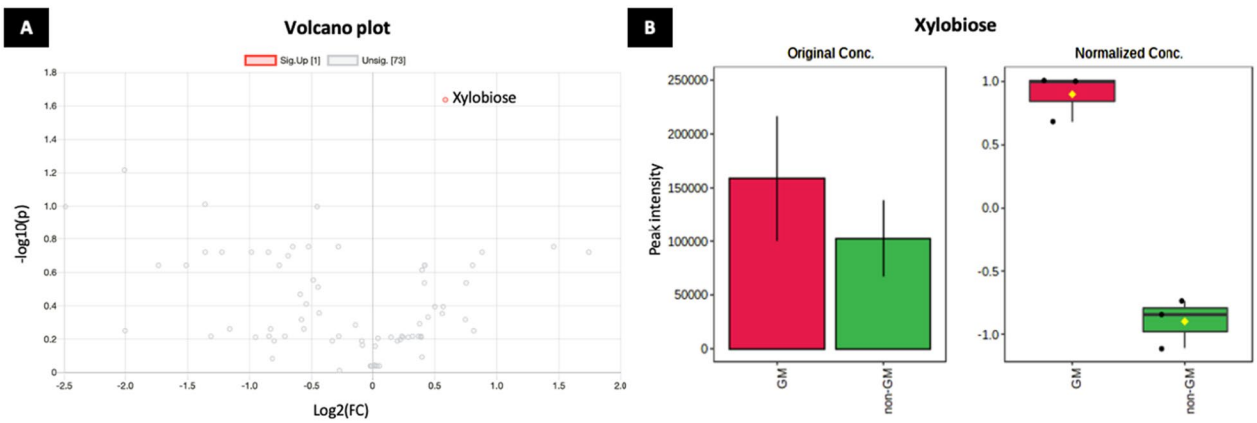

Fig. 3 Metabolomics comparative analysis. **A** Volcano plot displaying the distribution of 74 analyzed features of GM vs. non-GM groups separated by magnitude (x-axis, log2-fold change) and statistical significance (y-axis, p-adj FDR threshold = 0.05) in signal intensity. The only significant metabolite xylobiose (up-regulated for GM) is highlighted. **B** Average concentration for xylobiose in the GMO and non-GM samples (Student's *t*-test; $^{*}p < 0.05$). The bar plots on the left show the original values (mean ± SD). The box and whisker plots on the right summarize the normalized values

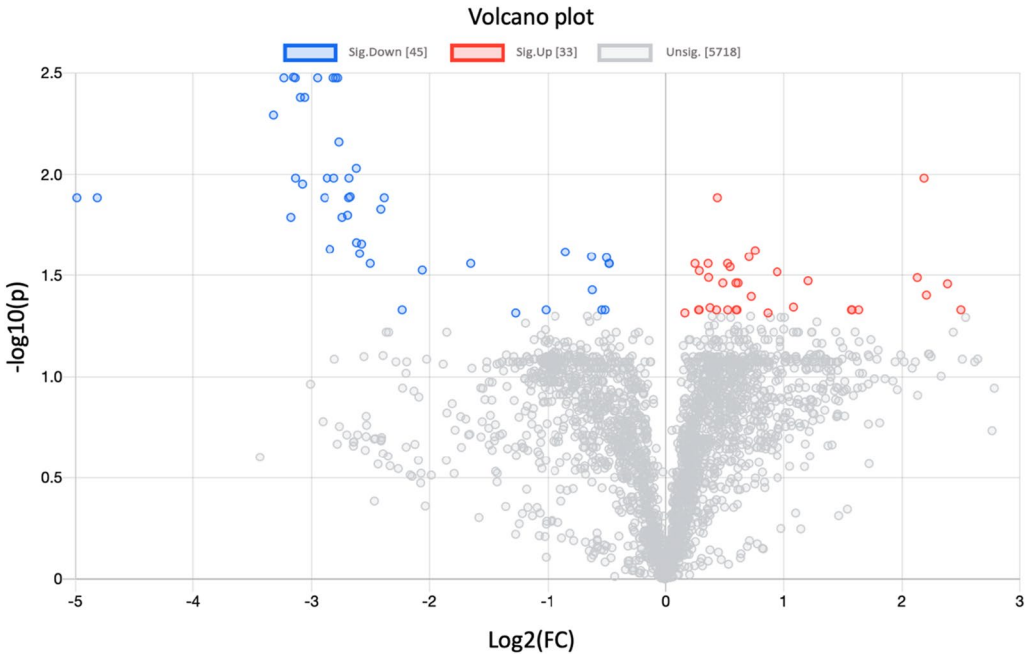

Fig. 4 Volcano plot of proteomics comparative analysis. The plot shows the distribution of 5718 analyzed proteins of GM vs. non-GM groups separated by magnitude (x-axis, log2-fold change) and statistical significance (y-axis, p-adj FDR threshold = 0.05) in signal intensity. 78 significant differentially expressed proteins being up (33) or down-regulated (45) in the GMO are highlighted

p-adj = 0.023) in the GMO compared to the non-GM variety samples (Fig. 3). Fifteen other chemicals showed variation in concentrations between both varieties, but did not present statistically significant differences according to the t-test with FDR correction (Additional file 4). The volcano plot distribution of between-group differences based on fold change and statistical significance results shows all compounds analyzed (Fig. 3A, B).

By an advanced analytical approach based on the TMT-11 × plex technique, a total of 5718 proteins were detected in samples from plot 4. In the comparative analysis, a total of 78 statistically different proteins were found between the GMO and the non-GM variety (p-adj FDR < 0.05). Volcano plot distribution of differentially expressed proteins displayed 33 (42.1%) proteins significantly up-regulated and 45 (57.7%) down-regulated in GM plants (Fig. 4). Table 2 shows the protein ID and name according to the Uniprot database, as well as the functional annotation and the fold change variation of the altered proteins significantly different in the GMO.

Metabolic pathway and interaction network analysis

We performed a functional enrichment analysis in order to rank associations between differentially regulated metabolites and proteins representing metabolic networks and the respective statistical probability. The association of chemicals and proteins in the biological network provides hints to their metabolic functions. Also, this analysis allows us to identify the relevant results of potentially altered pathways in the genetically modified plant. By conducting a KEGG pathway enrichment analysis in the generated biological network, we found ribosome, spliceosome, and protein export pathways as the most enriched, followed by biosynthesis of secondary metabolites, carbon fixation in photosynthetic organisms, carbon metabolism, and biosynthesis of amino acids (Table 3).

Network analysis using String database revealed key modules likely playing a role in the metabolism of GM plants (Fig. 5). This interaction network was divided into three main functional modules which correspond to the significantly altered pathways. Module 1 includes the KEGG altered pathways of ribosome and protein export. This module interacts with Module 2, which includes six altered proteins in the spliceosome pathway, via protein–protein interaction between a hydrolase uncharacterized protein (GLYMA09G05810.1) and 40S ribosomal proteins S12 (GLYMA13G44690.1; GLYMA15G00610.1). Module 3 is related to protein processing in endoplasmic reticulum with the protein disulfide isomerase S-2 (GLYMA19G41690.1) and an uncharacterized protein with a putative function assigned to retrograde protein transport from endoplasmic reticulum to cytosol (GLYMA03G33990.1), among other proteins.

Module 1 connects to Modules 2 and 3 by sharing strong protein interactions with stromal 70 kDa heat

Table 2 List of differentially expressed proteins among the GMO and the non-GM soybean variety in statistical analysis #ii

	Protein ID	Protein name (Uniprot)	FC	log2(FC)	p.adjusted	Gene ontology (biological process)	Gene ontology (molecular function)
1	A0A0R0I9D1;I1MQJ9	eRF1_1 domain-containing protein	0.11235	− 3.154	0.0033152	Cytoplasmic translational termination [GO:0002184]; regulation of growth [GO:0040008]	Sequence-specific mRNA binding [GO:1990825]; translation release factor activity, codon specific [GO:0016149]
2	I1K5D3;I1KQJ6	Eukaryotic translation initiation factor 3 subunit C (eIF3c) (Eukaryotic translation initiation factor 3 subunit 8) (eIF3 p110)	0.10624	− 3.2346	0.0033406	Formation of cytoplasmic translation initiation complex [GO:0001732]; translational initiation [GO:0006413]	Translation initiation factor activity [GO:0003743]; translation initiation factor binding [GO:0031369]
3	A0A0R0EBE6	Carboxypeptidase (EC 3.4.16.-)	0.11348	− 3.1395	0.0033406		Serine-type carboxypeptidase activity [GO:0004185]
4	I1M5W5	Matrin-type domain-containing protein	0.12941	− 2.95	0.0033406	Spliceosomal complex assembly [GO:0000245]	nucleic acid binding [GO:0003676]; zinc ion binding [GO:0008270]
5	K7N4K5;I1LCG9	Uncharacterized protein	0.14227	− 2.8133	0.0033406	mRNA splicing, via spliceosome [GO:0000398]	mRNA binding [GO:0003729]
6	I1LJF7	MFS domain-containing protein	0.14433	− 2.7926	0.0033406		Carbohydrate:proton symporter activity [GO:0005351]
7	C6TML4	Hydrolase_4 domain-containing protein	0.14609	− 2.7751	0.0033406		lipase activity [GO:0016298]
8	A0A0R0F137	Ribosomal_L18e/L15P domain-containing protein	0.11698	− 3.0956	0.0041719	Translation [GO:0006412]	mRNA binding [GO:0003729]; RNA binding [GO:0003723]; structural constituent of ribosome [GO:0003735]
9	K7N502;A0A0R0ENA0;K7LJV1	TRAM domain-containing protein	0.11998	− 3.0592	0.0041719	RNA processing [GO:0006396]	RNA binding [GO:0003723]
10	C6TMK2;I1MI73;I1MI72	2,4-Dienoyl-CoA reductase ((3E)-enoyl-CoA-producing) (EC 1.3.1.124)	0.099955	− 3.3226	0.005094	fatty acid catabolic process [GO:0009062]	2,4-Dienoyl-CoA reductase (NADPH) activity [GO:0008670]
11	I1MWX3;A0A0R0GNM9	Uncharacterized protein	0.14721	− 2.764	0.0069169		
12	I1LRP6;K7KYB8;A0A0R0JXB5	PDZ domain-containing protein	0.16292	− 2.6178	0.0093213		Metalloendopeptidase activity [GO:0004222]

Table 2 (continued)

	Protein ID	Protein name (Uniprot)	FC	log2(FC)	p.adjusted	Gene ontology (biological process)	Gene ontology (molecular function)
13	I1JE93	Ribosomal RNA-processing protein 40	0.11372	− 3.364	0.010422	CUT catabolic process [GO:0071034]; exonucleolytic catabolism of deadenylated mRNA [GO:0043928]; exonucleolytic trimming to generate mature 3′-end of 5.8S rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA) [GO:0000467]; nuclear polyadenylation-dependent rRNA catabolic process [GO:0071035]; nuclear polyadenylation-dependent tRNA catabolic process [GO:0071038]; nuclear retention of pre-mRNA with aberrant 3′-ends at the site of transcription [GO:0071049]; nuclear-transcribed mRNA catabolic process, exonucleolytic, 3′-5′ [GO:0034427]; polyadenylation-dependent snoRNA 3′-end processing [GO:0071051]; U4 snRNA 3′-end processing [GO:0034475]	RNA binding [GO:0003723]
14	A0A0R0EYQ2;K7MR27;K7MR26;K7MR28;A0A0R0I6E6;I1L2F5;A0A0R0IFH3;A0A0R0I6G1;I1L2F4;A0A0R0ID51	FHA domain-containing protein	0.13694	− 2.8684	0.010422		
15	I1KG57	O-fucosyltransferase family protein	0.14274	− 2.8085	0.010422	Fucose metabolic process [GO:0006004]	transferase activity [GO:0016740]
16	K7MP06;K7M8L9;I1MYC5;A0A0R0ETU0	RNA helicase (EC 3.6.4.13)	0.15609	− 2.6796	0.010422	Spliceosomal complex disassembly [GO:0000390]	ATP binding [GO:0005524]; ATP hydrolysis activity [GO:0016887]; RNA binding [GO:0003723]; RNA helicase activity [GO:0003724]
17	A0A0R0JFR0	Chal_sti_synt_N domain-containing protein	4.5469	2.1849	0.010422	Polyketide biosynthetic process [GO:0030639]	Acyltransferase activity, transferring groups other than amino-acyl groups [GO:0016747]
18	I1NIQ4;K7N519	Uncharacterized protein	0.11847	− 3.0774	0.011153	Intracellular protein transport [GO:0006886]; protein transport by the Tat complex [GO:0043953]	
19	A0A0R0H3E6;A0A0R0H303;I1LR08	V-type proton ATPase subunit a	0.15718	− 2.6695	0.01287		Proton transmembrane transporter activity [GO:0015078]
20	I1JBW7	ZnMc domain-containing protein	0.031519	− 4.9876	0.013019	Collagen catabolic process [GO:0030574]; extracellular matrix organization [GO:0030198]	Metalloendopeptidase activity [GO:0004222]; zinc ion binding [GO:0008270]
21	A0A368UIA9	Uncharacterized protein	0.035507	− 4.8157	0.013019		Serine-type endopeptidase activity [GO:0004252]

Table 2 (continued)

	Protein ID	Protein name (Uniprot)	FC	log2(FC)	p.adjusted	Gene ontology (biological process)	Gene ontology (molecular function)
22	A0A0R0FVM5	Aspartate kinase (EC 2.7.2.4)	0.13504	− 2.8886	0.013019	Homoserine biosynthetic process [GO:0009090]; lysine biosynthetic process via diaminopimelate [GO:0009089]; threonine biosynthetic process [GO:0009088]	aspartate kinase activity [GO:0004072]; ATP binding [GO:0005524]
23	I1JTZ7	Uncharacterized protein	0.15577	− 2.6825	0.013019	Lipid biosynthetic process [GO:0008610]	
24	H9TN50	Expansin B protein	0.19189	− 2.3816	0.013019	Sexual reproduction [GO:0019953]; syncytium formation [GO:0006949]	
25	I1N0Z5;C6THR1;I1KXZ9;A0A0R0J3C7	ECH_2 domain-containing protein	1.3518	0.43483	0.013019		3-Hydroxyisobutyryl-CoA hydrolase activity [GO:0003860]
26	I1M9X6	BURP domain-containing protein	0.18809	− 2.4105	0.014827		
27	C6TCJ7	Uncharacterized protein	0.15478	− 2.6917	0.015882	SRP-dependent cotranslational protein targeting to membrane, signal sequence recognition [GO:0006617]	7S RNA binding [GO:0008312]
28	K7LKW9;K7N3W6;I1NH16	Deubiquitinating enzyme MINDY-3 (EC 3.4.19.12)	0.11065	− 3.1759	0.016262		Lys48-specific deubiquitinase activity [GO:1990380]; thiol-dependent deubiquitinase [GO:0004843]
29	K7MXN7;I1ML93	Uncharacterized protein	0.14987	− 2.7382	0.016262		
30	A0A0R0EKI3	Uncharacterized protein	0.16313	− 2.6159	0.021712		
31	I1J4Z0;I1L4T1	Exostosin domain-containing protein	0.168	− 2.5734	0.022036	Protein glycosylation [GO:0006486]	Glycosyltransferase activity [GO:0016757]
32	I1KQH7;I1KQH9;I1KQI0	Uncharacterized protein	0.13934	− 2.8434	0.023391	mRNA processing [GO:0006397]; RNA splicing [GO:0008380]	
33	Q84ZV1	Disease-resistance protein (R 9 protein)	1.6868	0.75431	0.023795	Signal transduction [GO:0007165]	Calcium ion binding [GO:0005509]; NAD + nucleosidase activity [GO:0003953]
34	I1KJJ3;A0A0R0J9U8	Patatin (EC 3.1.1.-)	0.5554	− 0.8484	0.024191	Lipid catabolic process [GO:0016042]	Acylglycerol lipase activity [GO:0047372]; phospholipase activity [GO:0004620]
35	A0A0R0GB30	Uncharacterized protein	0.16619	− 2.5891	0.024617	Proteolysis involved in cellular protein catabolic process [GO:0051603]	Cysteine-type endopeptidase activity [GO:0004197]
36	I1LNY2;I1L5L1;K7MDE9;A0A0R0I109	Ribosomal protein	0.6477	− 0.6266	0.025515	Maturation of LSU-rRNA [GO:0000470]; translation [GO:0006412]	RNA binding [GO:0003723]; rRNA binding [GO:0019843]; structural constituent of ribosome [GO:0003735]
37	I1KL63	Glycosyltransferase (EC 2.4.1.-)	1.6272	0.70238	0.025542		UDP-glycosyltransferase activity [GO:0008194]

Table 2 (continued)

	Protein ID	Protein name (Uniprot)	FC	log2(FC)	p.adjusted	Gene ontology (biological process)	Gene ontology (molecular function)
38	I1LQ64;I1LQ65;A0A0R0H152	Uncharacterized protein	0.70714	− 0.49993	0.025804	Dimethylallyl diphosphate biosynthetic process [GO:0050992]; isopentenyl diphosphate biosynthetic process, methylerythritol 4-phosphate pathway [GO:0019288]	4-Hydroxy-3-methylbut-2-en-1-yl diphosphate reductase activity [GO:0051745]; 4 iron, 4 sulfur cluster binding [GO:0051539]; metal ion binding [GO:0046872]
39	K7KLL1;I1KBC8	Uncharacterized protein	0.17673	− 2.5004	0.027621		
40	I1KIY3	40S ribosomal protein S7	0.3182	− 1.652	0.027621	Ribosomal small subunit biogenesis [GO:0042274]; rRNA processing [GO:0006364]; translation [GO:0006412]	Structural constituent of ribosome [GO:0003735]
41	I1KA18;A0A0R0JF06	Acetyl-CoA carboxylase (EC 6.4.1.2)	1.4351	0.52118	0.027621	Fatty acid biosynthetic process [GO:0006633]; malonyl-CoA biosynthetic process [GO:2001295]	Acetyl-CoA carboxylase activity [GO:0003989]; ATP binding [GO:0005524]; metal ion binding [GO:0046872]
42	C6TFY0	Uncharacterized protein	0.71744	− 0.47908	0.027621	Translation [GO:0006412]	RNA binding [GO:0003723]; structural constituent of ribosome [GO:0003735]
43	Q2PMP8	50S ribosomal protein L16, chloroplastic	0.71861	− 0.47673	0.027621	Mitochondrial translation [GO:0032543]	rRNA binding [GO:0019843]; structural constituent of ribosome [GO:0003735]
44	I1JPP3;K7MRG3;A0A0R0ITW3	Uncharacterized protein	1.2806	0.35681	0.027621	Autophagosome maturation [GO:0097352]; ER-associated misfolded protein catabolic process [GO:0071712]; mitotic spindle disassembly [GO:0051228]; retrograde protein transport, ER to cytosol [GO:0030970]; ubiquitin-dependent ERAD pathway [GO:0030433]	ATP binding [GO:0005524]; ATP hydrolysis activity [GO:0016887]; polyubiquitin modification-dependent protein binding [GO:0031593]
45	I1KVP9;A0A0R0FGA3	Uncharacterized protein	1.1859	0.24594	0.027621		
46	A0A0R0I9M4;I1L417;I1L416;A0A0R0IHL7;K7KRD7	Annexin	1.4558	0.54185	0.02869	Response to stress [GO:0006950]	Calcium-dependent phospholipid binding [GO:0005544]; calcium ion binding [GO:0005509]
47	A0A368UJV7;I1JXI1;A0A368UID7;A0A368UIA6	Uncharacterized protein	0.23964	− 2.061	0.029791		
48	I1L171	Uncharacterized protein	1.2162	0.28237	0.030006		ATP binding [GO:0005524]; hydrolase activity [GO:0016787]; mRNA binding [GO:0003729]; RNA binding [GO:0003723]; RNA helicase activity [GO:0003724]

Table 2 (continued)

	Protein ID	Protein name (Uniprot)	FC	log2(FC)	p.adjusted	Gene ontology (biological process)	Gene ontology (molecular function)
49	I1KWF3;A0A0R0IX71	Endo-1,3(4)-beta-glucanase (EC 3.2.1.6)	1.9221	0.94266	0.030422	Metabolic process [GO:0008152]	Beta-glucosidase activity [GO:0008422]; glucan endo-1,3-beta-glucanase activity, C-3 substituted reducing group [GO:0052861]; glucan endo-1,4-beta-glucanase activity, C-3 substituted reducing group [GO:0052862]
50	E3W9C1;A0A0R0ERR6	Protein disulfide-isomerase (EC 5.3.4.1)	1.2844	0.36108	0.032384		protein disulfide isomerase activity [GO:0003756]
51	C6TA28;K7N1I9;A0A0R0GQL8	Nucleoside diphosphate kinase (EC 2.7.4.6)	4.3713	2.1281	0.032453	CTP biosynthetic process [GO:0006241]; GTP biosynthetic process [GO:0006183]; UTP biosynthetic process [GO:0006228]	ATP binding [GO:0005524]; nucleoside diphosphate kinase activity [GO:0004550]
52	A0A0R0HG00;A0A0R0HHN9;A0A0R0HLL3;A0A0R0H9A9;A0A0R0H9D4;I1M2X3	Protein kinase domain-containing protein	2.3078	1.2065	0.033657		ATP binding [GO:0005524]; protein serine/threonine/tyrosine kinase activity [GO:0004712]; protein serine/threonine kinase activity [GO:0004674]; protein serine kinase activity [GO:0106310]
53	I1KSR4;A0A0R0IL12;A0A0R0K4U1;I1K4A1;A0A0R0JWR6;A0A0R0JWY2;A0A0R0K2I7;A0A0R0JWR2;K7KQP4	Uncharacterized protein	1.5259	0.60968	0.034514	ER-associated misfolded protein catabolic process [GO:0071712]; protein N-linked glycosylation via asparagine [GO:0018279]	UDP-glucose:glycoprotein glucosyltransferase activity [GO:0003980]; unfolded protein binding [GO:0051082]
54	I1JUD5	Uncharacterized protein	1.507	0.59165	0.034514		ATP binding [GO:0005524]; hydrolase activity [GO:0016787]; RNA binding [GO:0003723]; RNA helicase activity [GO:0003724]
55	I1KNN1;I1MQW2	Uncharacterized protein	1.3958	0.48107	0.034514	Ribosomal small subunit assembly [GO:0000028]; translation [GO:0006412]	RNA binding [GO:0003723]; structural constituent of ribosome [GO:0003735]
56	I1JN31;A0A0R0KQC7;K7MY41	Carboxypeptidase (EC 3.4.16.-)	5.2164	2.3831	0.034854		Serine-type carboxypeptidase activity [GO:0004185]
57	I1JNS6	Probable aspartic proteinase GIP1 (EC 3.4.23.-) (Glucanase inhibitor protein 1) (GmGIP1)	0.65033	-0.62076	0.037281	Defense response [GO:0006952]	Aspartic-type endopeptidase activity [GO:0004190]
58	I1L1U5;I1L1U4;I1L1U2;I1MHJ8;I1MHJ6	Sucrose synthase (EC 2.4.1.13)	4.6113	2.2052	0.039612	Sucrose metabolic process [GO:0005985]	Sucrose synthase activity [GO:0016157]
59	I1NA94;A0A0R0EPL0;K7MZ15	Uncharacterized protein	1.6485	0.72119	0.040167	Lysine biosynthetic process via diaminopimelate [GO:0009089]	Diaminopimelate decarboxylase activity [GO:0008836]
60	C6TB56	PKS_ER domain-containing protein	2.1187	1.0832	0.045368		Oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor [GO:0016616]; zinc ion binding [GO:0008270]

Table 2 (continued)

	Protein ID	Protein name (Uniprot)	FC	log2(FC)	p.adjusted	Gene ontology (biological process)	Gene ontology (molecular function)
61	I1KUN7;I1LTF6;I1KBS6;I1JXV4;Q43466;I1K552;I1MV85;I1MV84;I1K0W9;I1KSY9;K7KQW3;A0A0R0JX61;A0A0R0K6Z1;Q43465;I1J9Q1	Protein kinase domain-containing protein	1.2955	0.37348	0.045682	Intracellular signal transduction [GO:0035556]; protein phosphorylation [GO:0006468]	ATP binding [GO:0005524]; protein serine/threonine kinase activity [GO:0004674]
62	C6T8R3;I1L5M3;I1L5M4;I1L5M2	Uncharacterized protein	5.641	2.496	0.046811		
63	A0A0R0H4Q8;K7M167;A0A0R0H1N2;I1M1L8;K7MAF1	Uncharacterized protein	0.213	− 2.2311	0.046811	Response to symbiotic fungus [GO:0009610]	Serine-type endopeptidase activity [GO:0004252]
64	I1L7M8;A0A0R0KXZ7;K7K5V4;K7KDZ7	Long-chain fatty acid CoA ligase (EC 6.2.1.3)	3.1029	1.6336	0.046811	cutin biosynthetic process [GO:0010143]; wax biosynthetic process [GO:0010025]	Long-chain fatty acid CoA ligase activity [GO:0004467]
65	I1KV03;I1KGH4;I1KGH3	Uncharacterized protein	29.835	1.577	0.046811		Hydrolase activity, acting on ester bonds [GO:0016788]
66	I1JBW8;A0A0R0KY07	ZnMc domain-containing protein	2.9738	1.5723	0.046811	Collagen catabolic process [GO:0030574]; extracellular matrix organization [GO:0030198]	metalloendopeptidase activity [GO:0004222]; zinc ion binding [GO:0008270]
67	I1M3I8;I1LUR0	Uncharacterized protein	0.49454	− 1.0158	0.046811		
68	I1LTG9;A0A0R0GXL3;I1L2Q0;I1LKD9	Protein transport protein Sec61 subunit beta	1.5167	0.6009	0.046811	Posttranslational protein targeting to membrane, translocation [GO:0031204]; SRP-dependent cotranslational protein targeting to membrane, translocation [GO:0006616]	
69	I1JPL2;A0A0R0JTD7	J domain-containing protein	1.5082	0.59287	0.046811		
70	I1NAV4	Uncharacterized protein	0.68786	− 0.53981	0.046811	Chloroplast rRNA processing [GO:1901259]	mRNA binding [GO:0003729]
71	I1KYT6;I1N048	RuvB-like helicase (EC 3.6.4.12)	1.4362	0.52223	0.046811	Box C/D snoRNP assembly [GO:0000492]; chromatin remodeling [GO:0006338]; histone acetylation [GO:0016573]; regulation of transcription by RNA polymerase II [GO:0006357]	5′-3′ DNA helicase activity [GO:0043139]; ATP binding [GO:0005524]; ATP hydrolysis activity [GO:0016887]; DNA helicase activity [GO:0003678]
72	I1MJU7;A0A0R0HS21	Uncharacterized protein	0.701	− 0.51252	0.046811	Protein folding [GO:0006457]	ATP binding [GO:0005524]; ATP hydrolysis activity [GO:0016887]; unfolded protein binding [GO:0051082]
73	C6TJD3	Uncharacterized protein	1.3455	0.42817	0.046811	Protein phosphorylation [GO:0001934]; rescue of stalled ribosome [GO:0072344]	Protein kinase C binding [GO:0005080]; ribosome binding [GO:0043022]
74	I1MW40	Uncharacterized protein	1.2156	0.28167	0.046811	Mitochondrial proton-transporting ATP synthase complex assembly [GO:0033615]	

Table 2 (continued)

	Protein ID	Protein name (Uniprot)	FC	log2(FC)	p.adjusted	Gene ontology (biological process)	Gene ontology (molecular function)
75	I1L1K7	Uncharacterized protein	1.2113	0.27653	0.046811	Cytoplasmic translational initiation [GO:0002183]	ATP binding [GO:0005524]; hydrolase activity [GO:0016787]; RNA helicase activity [GO:0003724]; translation initiation factor activity [GO:0003743]
76	I1N806;I1MLA2;I1N807;I1MLA3	Uncharacterized protein	0.41395	− 1.2725	0.048495		
77	I1MCW6;C6ZRU0;A0A0R0G434;K7M940;I1MCX0	Protein kinase domain-containing protein	1.818	0.86234	0.048495		ATP binding [GO:0005524]; protein serine/threonine kinase activity [GO:0004674]
78	I1LQ02;I1LQ03	WBb225L1.16 protein	1.1179	0.16078	0.048495	Ribosome biogenesis [GO:0042254]	snoRNA binding [GO:0030515]

The table shows protein ID; protein name via UniProt accession; fold change (FC); log2 of fold change (Log2FC); false discovery rate adjusted p-value (p-adj); and annotation of biological process and/or molecular function according to the Gene ontology categories. Blank cells indicate that the information was not present in the databases for that protein

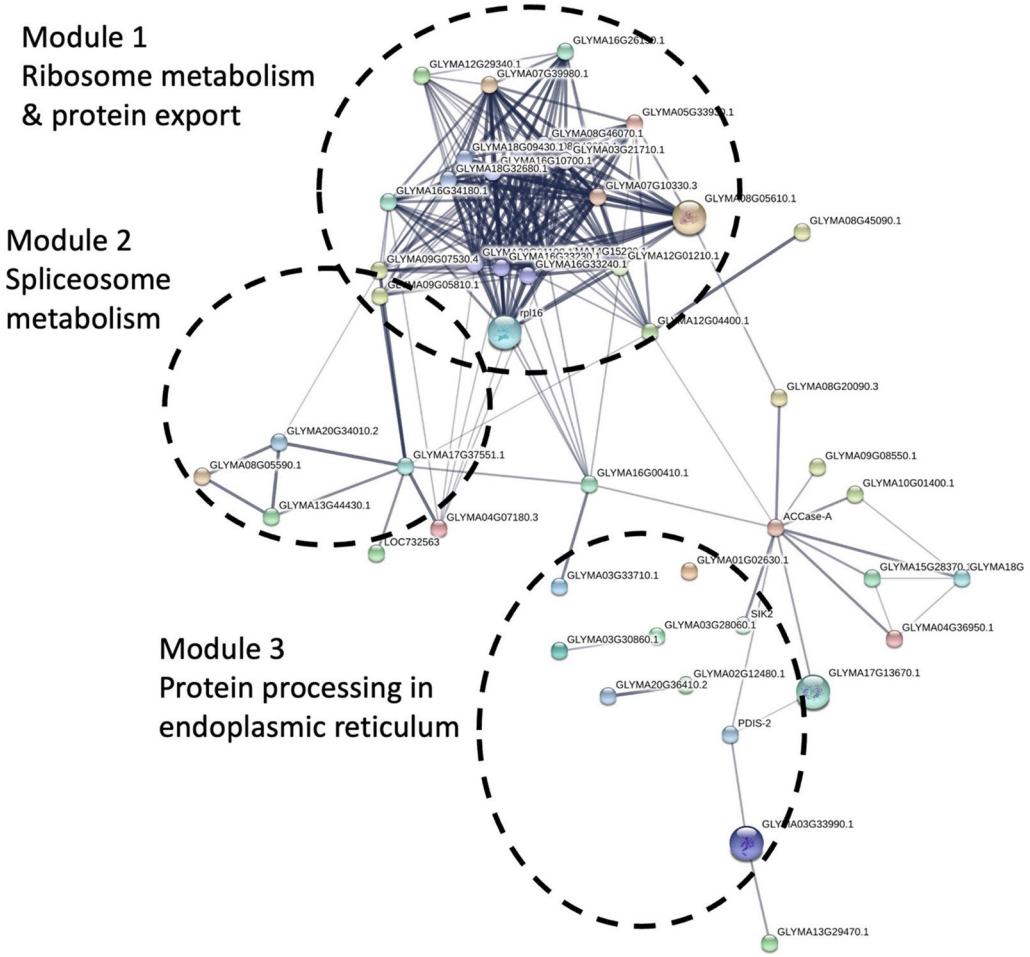

Fig. 5 Interaction network of proteins and metabolites statistically different in GM soybean plants. The visual network was built using Stitch and String databases. Three distinct functional modules are highlighted in black dotted circles. Protein names for accessions are present in Table 2. Stronger associations are represented by thicker lines. Protein–protein interactions are shown in grey, chemical–protein interactions in green and interactions between chemicals in red

Table 3 List of altered metabolic pathways in the GM soybean plants

Pathway ID	Pathway description	Observed gene count	False discovery rate
3010	Ribosome	11	7.89e-09
3040	Spliceosome	5	0.0036
3060	Protein export	3	0.0069
4141	Protein processing in endoplasmic reticulum	5	0.0069

Table shows KEGG pathway ID; metabolic pathway description; observed gene count; and false discovery rate adjusted p-value

shock-related protein (GLYMA16G00410.1) further connected to acetyl-Coa carboxylase enzymes which are involved in ATP and RNA binding of spliceosome pathway. ACCase-A then connects to a serine/threonine-protein kinase srk2a isoform × 1 (GLYMA08G20090.3), thus linking back Modules 1 and 3.

Untargeted allergenicity analysis

We have searched for potential allergens among the statistically different proteins using the peer-reviewed allergen database Allergen Online v.20 (http://www.allergenonline.org/) intended for the identification of proteins that may present a potential risk of allergenic cross-reactivity. Among the 78 proteins with statistically significantly different expression levels in the GMO vs. non-GM variety comparison (p-adj FDR < 0.05), 43 were identified to have allergenic potential (Table 4). These proteins show at least 35% identity with overlapping amino acid sequences with known allergens according to the database search algorithm. Three proteins are related to pollen allergens and have been identified in different

Table 4 List of differentially expressed proteins with allergenic potential in the GMO variety

Protein ID	Allergen group ID	Allergen name	E-value	% id	% sim	Alignment length
I1MJU7	gid|2708	Heat shock cognate 70 [Aedes aegypti]	1.90E−96	0.505	0.778	626
A0A368UIA9	gid|466	Pre-pro-cucumisin [Cucumis melo]	2.40E−75	0.519	0.740	655
A0A0R0GB30	gid|18	Actinidain [Actinidia deliciosa]	1.20E−50	0.516	0.779	217
I1JN31	gid|775	Serine carboxypep [Triticum aestivum]	6.20E−43	0.401	0.661	454
H9TN50	gid|80	Beta-expansin 1 [Zea mays]	2.80E−37	0.448	0.714	241
I1KJJ3	gid|193	Patatin [Solanum tuberosum]	3.90E−35	0.386	0.675	378
C6TB56	gid|2582	Alcohol dehydrogenase [Cochliobolus lunatus]	1.30E−33	0.292	0.625	360
I1KV03	gid|585	ENSP-like protein [Hevea brasiliensis]	1.10E−27	0.318	0.592	368
A0A0R0H4Q8	gid|466	Pre-pro-cucumisin [Cucumis melo]	4.70E−26	0.391	0.664	763
C6TMK2	gid|2371	Seed maturation-like protein [Sesamum indicum]	1.60E−21	0.322	0.625	267
A0A0R0EBE6	gid|775	Serine carboxypep [Triticum aestivum]	2.10E−20	0.321	0.596	240
E3W9C1	gid|63	Protein disulfide-isomerase (PDI) [Triticum aestivum]	1.90E−11	0.333	0.612	183
I1MCW6	gid|773	Putative leucine-rich repeat protein [Triticum aestivum]	1.40E−07	0.299	0.584	137
A0A0R0HG00	gid|773	Putative leucine-rich repeat pro [Triticum aestivum]	5.00E−06	0.388	0.675	80
K7MP06	|gid|1067	Conglutin beta [Lupinus angustifolius]	5.10E−06	0.231	0.592	277
I1JNS6	gid|1747	Pollen allergen CPA63 [Cryptomeria japonica]	2.00E−05	0.225	0.527	395
K7N4K5	gid|1248	Eukaryotic translation initiation factor [Forcipomyia taiwana]	8.60E−05	0.199	0.497	306
I1KNN1	gid|626	Putative allergen [Lepidoglyphus destructor]	2.00E−02	0.284	0.556	81
I1JBW8	gid|151	Gliadin [Triticum aestivum]	2.00E−02	0.354	0.523	65
K7N502	|gid|359	Chain A, Nmr Solution Structure [Blomia tropicalis]	1.10E−01	0.313	0.582	67
I1M5W5	gid|150	Omega-gliadin [Triticum aestivum]	1.40E−01	0.254	0.465	256
I1JBW7	gid|1338	Ragweed homologue of Art v 1 precursor [Ambrosia artemisiifolia]	1.80E−01	0.263	0.545	99
C6TJD3	gid|1248	Eukaryotic translation initiation factor [Forcipomyia taiwana]	1.90E−01	0.216	0.538	320
I1KQH7	gid|90	Tropomyosin [Allergen Ani s 3]	2.20E−01	0.234	0.558	154
A0A0R0EKI3	gid|1067	Conglutin beta [Lupinus angustifolius]	2.40E−01	0.273	0.557	88
C6TCJ7	gid|2215	Glutathione S-tran [Ascaris suum]	3.20E−01	0.692	0.846	13
I1MWX3	gid|1542	Cys peroxiredox [Triticum aestivum]	3.30E−01	0.270	0.549	122
I1KUN7	gid|1044	Tropomyosin [Balanus rostratus]	3.40E−01	0.311	0.581	74
K7LKW9	gid|1743	Troponin C [Crangon crangon]	3.80E−01	0.327	0.714	49
I1L1U5	gid|471	Cyn d 1 [Cynodon dactylon]	5.70E−01	0.235	0.513	115
I1L171	gid|1910	Non-specific lip [Lycium barbarum]	7.00E−01	0.387	0.645	31
K7KLL1	gid|1191	Lit v 1 tropomyosin [Litopenaeus vannamei]	7.50E−01	0.232	0.568	241
A0A0R0JFR0	gid|74	Pollen allergen Amb t 5 precursor [Ambrosia trifida]	8.10E−01	0.474	0.737	19
I1JE93	gid|1560	Salivary antigen 5 precursor [Glossina morsitans morsitans]	8.70E−01	0.213	0.521	169
I1LQ02	gid|594	Latex allergen [Hevea brasiliensis]	1.20E+01	0.256	0.556	117
I1K5D3	gid|1828	Tropomyosin [Onchocerca volvulus]	1.50E+01	0.238	0.531	277
I1JPL2	gid|1338	Ragweed homologue of Art v 1 precursor [Ambrosia artemisiifolia]	1.60E+01	0.521	0.563	48
I1NAV4	gid|150	Omega-5 gliadin [Triticum aestivum]	2.10E+01	0.253	0.578	83
I1LTG9	gid|1338	Ragweed homologue of Art v 1 precursor [Ambrosia artemisiifolia]	2.70E+01	0.432	0.514	37
I1N806	gid|2725	Putative galactose oxidase [Artemisia argyi]	4.50E+01	0.265	0.554	83
C6T8R3	gid|2371	Seed maturation-like protein precursor [Sesamum indicum]	4.90E+01	0.248	0.524	145
C6TFY0	gid|150	Omega-gliadin, partial [Triticum aestivum]	6.90E+01	0.284	0.612	67
Q84ZV1	gid|698	Calcium-binding protein [Olea europaea]	9.10E+01	0.484	0.766	64

The table shows differential protein UniProt ID; allergen GenBank ID and name; E-value statistical score; percent identity of the overlapping alignment; percent similarity; and amino acids alignment length in query protein to the aligned allergen

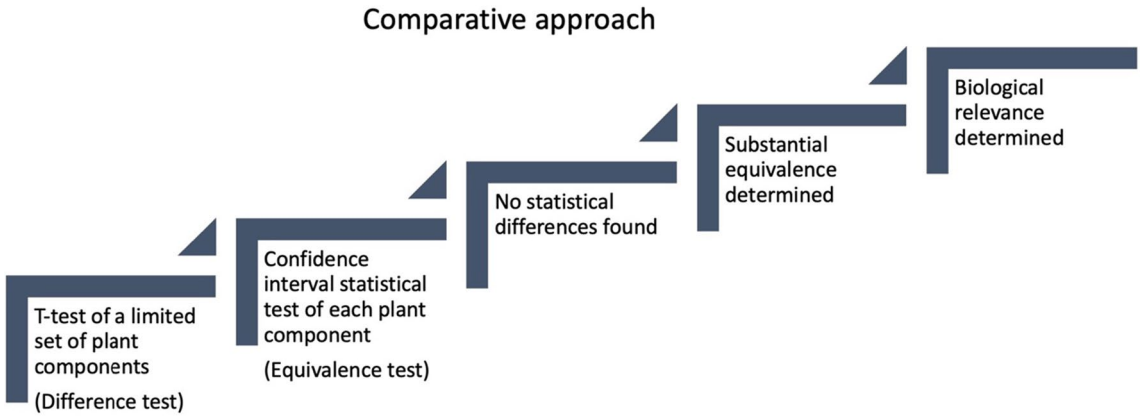

Fig. 6 Graphical representation of the steps of the comparative approach according to EFSA and decision-making framework for GMO risk assessment in the EU

plant species. Two identified proteins showed significant statistical score and high similarity with the gliadin allergen protein, which is a component of gluten, present in wheat. Protein matches with the highest E-value score were heat shock cognate 70 (I1MJU7) which showed 51% identity to the allergen identified in *Aedes aegypti*; the pre-pro-cucumisin allergen (A0A368UIA9) with 53% identity to the full sequence identified in *Cucumis melo*; and actinidain allergen (A0A0R0GB30) with 52% sequence identity to the protein found in kiwi (*Actinidia deliciosa*). The results suggest further experimental studies with some of the identified potential allergens or allergenic epitopes sharing identities lower than 50% and having E-scores larger than 1,00E-4 regarding immunoglobulin E (IgE) binding and clinical reactivity. In addition, the assessment of literature would then contribute to the design of appropriately allergic study subjects.

Discussion

The biological relevance of substantial equivalence

According to EFSA, biological relevance is based on the following three aspects: (i) the outcomes of the difference test; (ii) the outcomes of the equivalence test, as well as (iii) expert judgement regarding the implications of the changes for food and feed safety of a particular GMO [43]. The difference and equivalence tests are the basis for the analysis of substantial equivalence which in reality is an endpoint-by-endpoint comparison of a limited set of components between the GMO, the non-GM near-isogenic counterpart and several reference varieties [33]. The list of such components is determined by species and derives from external sources, like the list of analytes outlined in the OECD consensus documents for soybean composition [44].

In general, the comparative analysis is a Student's *t*-test to verify the null hypothesis which is "no difference between the GMO and its conventional counterpart" against the alternative hypothesis: "difference between the GMO and its conventional counterpart" [33]. In the case of a GMO safety assessment, measured changes are considered to be of no biological relevance if compositional data fall within the range observed (99% tolerance interval) in traditionally cultivated crops that are considered to have a history of safe use for consumption by humans and/or domesticated animals [43]. In practice, if all tested components are found within the interval of equivalence limits the organism is determined as substantially equivalent with no threshold level for equivalence. In other words, GM soy could be determined equivalent to common bean or maize if they were included in the analysis (Fig. 6).

On the other hand, when statistically significant differences are found, usually additional data in support of the substantial equivalence are provided by the applicant. In the case of the data submitted by the applicant for Intacta soybean (MON 87701 × MON 89788), the comparative compositional analysis revealed 11 components (out of 53) with significant differences ($p \leq 0.05$) between the GMO and the conventional control (Monsanto [45]. However, when data of soybean component levels from published scientific literature and ILSI's[1] Crop Composition Database were included in the comparative evaluation (Table 18 in Monsanto [45]), the statistically significant different data points now fall inside the equivalence limits. EFSA statistical guidelines require inclusion of reference varieties conducted with

[1] International Life Sciences Institute.

a fully randomized plot layout. External datasets should only be used when a strong justification can be given why the first option was impossible. However, as also observed for other applications, the applicant did not provide reasoning on why these databases were used in addition to the 20 soybean reference varieties which were grown side-by-side in the field trials. Including those data enlarged equivalence limits leading to the product meeting substantial equivalence. In this case, field trial data for reference varieties were available, which would meet EFSA's guidelines. The EFSA scientific opinion on Intacta soybean considered that "the information available for soybean MON 87701 × MON 89788 addressed the scientific issues indicated by the Guidance document of the EFSA GMO Panel" and that "the soybean MON 87701 × MON 89788 is as safe as its comparator with respect to potential effects on human and animal health or the environment in the context of its intended uses" [46].

In this case, EFSA assumes that the list of components tested in the trial is sufficient to establish equivalence when no differences are found, but does not consider it sufficient to attest non-equivalence when differences are observed. This unbalanced interpretation of the same set of components is a weakness of this comparative framework and lacks scientific justification (Bohn et al., 2014; Millstone et al., 2020).

When we assessed our proteomics data as indicated by EFSA's Guidance document (statistical analysis #i), i.e., in the same way compositional data are presented in dossiers, we were challenged with two different statistical results in which we could not assess biological relevance and which are not addressed in EFSA´s statistical guideline (Fig. 1) [33]. In the first case, we found 10 proteins from the non-GM near-isogenic comparator which fell outside the equivalence limits. In our second case, a 99% tolerance interval of equivalent limits could not be calculated for two proteins from the reference varieties, because they could not be detected in three or more replicates of the samples. In case of the assessment of Intacta soybean EFSA took note of such statistical outcome and wrote in its opinion under chapter 4.1.3 that constituents at levels below the limit of quantification for more than 50% of samples were omitted from the analysis [46]. In summary, EFSA's requirements for statistical analysis were met in our dataset analysis and we can only conclude that the GMO is not equivalent to the non-GM counterpart.

A multi-omics approach based on systems biology

Systems biology is generally understood as the study of biological systems "whose behavior cannot be reduced to the linear sum of their parts' functions" [47]. In practice, it has been also described as a "computational modeling of molecular systems and the integrative interpretation of ever larger postgenomic datasets are accepted as useful, and perhaps even necessary, components of biological research" [48].

Our newly proposed approach described in 'Statistical analysis 2' follows a systems biology approach. The idea is to establish a holistic perspective of the genetically modified organism, in which the genetic modification is perceived as causing a perturbation of a system (i.e., the near-isogenic non-GM counterpart). The strategy is then to monitor the responses, integrate the data and perform a computational analysis, based on bioinformatics, to describe the modified system. This strategy is not new, and it has been routinely used to understand complex traits in all fields of biology, including the study of human diseases [49].

In this proposed implementation approach, omics datasets are used to investigate gene-by-gene interactions by network modeling; and even the flow of genetic information when multiplex omics are applied. In this way, many biological processes and metabolisms can be tested by the identification of the biochemical functions from a large network of molecular interactions, including interactions among molecules of the same type, for example, protein–protein interactions, or among molecules of different types, for example, protein–RNA, or protein–metabolite interactions [50].

The organism's adaptation to changing conditions of the receiving environment depends on their capacity to change their molecular constitution, which can be achieved by modulation of the quantitative composition and the diversity of the cell's molecular repertoire. Molecular diversification is particularly pronounced on the proteome level, at which multiple proteoforms derived from the same gene can in turn combinatorially form different protein complexes, thus expanding the repertoire of functional modules in the cell" [50]. The understanding of the plant protein–protein interaction network and interactome provides crucial insights into the regulation of plant developmental, physiological, and pathological processes [51]. Thus, data extracted from biochemical networks are more informative than the analysis of each single molecule alone, like the endpoint-by-endpoint analysis performed according to EFSA's guidance.

There are several advantages in performing a systems biology approach as compared to the comparative assessment: (1) the untargeted and unsupervised analysis of molecules provides additional chance of detecting unintended and unexpected changes, such as new toxins and allergens; (2) the analysis of compounds or molecules that are relevant for each GMO event as opposing to a

pre-determined list of compounds per species; (3) the list of altered compounds can be used for a network analysis based on their biological functions and their participation in certain metabolic pathways; (4) the range of molecules to be analyzed is only dependent on the state of the art of the analytical and technological development and will not be restricted to a pre-determined consensus list, hence, allowing to keep pace with increasingly complex metabolic changes and technological progress. Finally, the identification of potential metabolic disturbances due to the genetic modification will inform the testing of dedicated risk hypotheses, for example, if stress related metabolism is altered in the GM variety, then acute stress-response assays are recommended. The specific testing will be applicable to the GM event on a case-by-case basis, and such analyses will complement the animal feeding studies and the molecular characterization in the hazard identification step of a risk assessment [27]. In contrast, the current approach by EFSA requires submission of particular data sets to conclude on the "comparative safety" of a GMO, however the interpretation of these data has not been guided by specific test hypotheses.

A relevant aspect of any new analytical and statistical approach for implementation is standardization. This article does not provide a full pathway towards standardization but rather outlines the first steps for future validation. It is important to highlight that several initiatives have already accomplished a great deal of work over the past two decades towards standardization. The HUPO Proteomics Standards Initiative (PSI; www.psidev.info) defines community standards for data representation in proteomics to facilitate data comparison, exchange and verification and a list of scientific publications with standards can be found in their webpage.

Intacta soybean with altered metabolism

In this study, we identified metabolic disturbance at major pathways: the ribosome, spliceosome, protein processing and protein export metabolism. Alteration in protein-related metabolism can be related to the heterologous expression of the stacked cassette. Whereas several carbon metabolism-related proteins were present in the enriched networks, these pathways were not statistically significant. However, any metabolic imbalance in the plant can be expected to also have an impact on the carbon metabolism. The current strategy for transgenic expression is based on strong constitutive promoters (e.g., the viral P35S) which can be problematic as transgenes are overexpressed at all developmental stages and tissues, leading to competition for energy and building blocks for synthesis of proteins, RNA and metabolites that are required for plant growth (Singhal et al. 2015). In addition, genomic insertions and disruptions caused by transgenesis may lead to pleiotropic effects. These effects have to be investigated as to whether they are associated with risks (risk pathway). Whereas a growth penalty might only have an agronomic impact, increased sensitivity to stress as a pleiotropic effect can lead to the production of certain secondary metabolites, such as toxins or allergens, in the plant. In our allergenicity analysis, we have identified 43 potential allergenic proteins which should be further assessed. Our results contradict the data which Monsanto presented when requesting placing on the EU market for Intacta soybean in 2009 and which was assessed by EFSA in 2012. In their dossier, Monsanto researchers listed 11 compounds which were statistically different from the near-isogenic counterpart in their limited comparative analysis (53 compounds). However, the differences were considered to be within the equivalence limits of the reference varieties and further analysis of their functions was not performed (Monsanto 2009).

It is not yet clear how the metabolic disturbances identified in our study would affect the performance and the safety of Intacta soybean in the field, as additional analyses are necessary. However, there are few studies showing unintended effects which seemed to be caused in response to changes in other plant traits and compounds rather than the heterologous Bt protein per se. In 2014, Monsanto scientists have published results that "should be viewed as an alert that *S. eridania* [*Spodoptera eridania*] populations may increase in Bt soybeans [Intacta soybean]" [52]. Their results showed that Intacta soybean reduced larval development by 2 days and increased adult male longevity by 3 days, which indicates that the effect of Intacta soybean MON 87701 × MON 89788 on *S. eridania* development and reproduction can be favorable to pest development. In addition, the effect of GM Bt maize and BT proteins on non-target organisms (e.g., Neuroptera insects) has been extensively observed over the past two decades ([53], [54]). There have also been reports on phytotoxicity in Intacta soybean in response to glyphosate applications which could not yet be explained [55, 56]. Thus, understanding the underlying effects of transgene expression and mechanisms on plant molecular biology, biochemistry and physiology is crucial for predicting the effects on plant fitness and altered substances which may lead to potential risks [57, 58].

Conclusion

Taken together, our results show that a science-based, risk-related approach based on omics techniques can be implemented for risk assessment of GMOs according to the EU legislation. We demonstrated that a systems biology approach based on a holistic perspective can be more informative in risk assessment than the currently employed endpoint-by-endpoint analysis for the assessment of potential unintended effects in a GM plant. We show that current tolerance and equivalence interval analyses based on data from reference varieties creates a quantitative noise with a high threshold level due to genotypic variability. In contrast, the approach proposed in this paper offers several advantages for the risk assessment procedure. Untargeted omics techniques allow for monitoring case-by-case responses. It also opens the possibility for the integration of large datasets by generating metabolic networks. The proposed analysis pipeline addresses the existing gap between animal feeding studies and molecular characterization in the hazard identification step of a risk assessment.

In this study, we provide a concrete case explaining how this analysis can be included in risk assessment, the outcome of the analysis and how to further investigate risk-related hypotheses. The proposed systems biology-based approach identified alterations in protein and energy related metabolism of the Intacta soybean variety, which is different from the conclusions based on the current EFSA risk assessment approach. Based on the results generated by the approach proposed in our study, we conclude that the comparative assessment according to the current EFSA guidance is not fit for purpose and needs to be improved.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12302-023-00715-6.

Additional file 1. Image from the field location and the identification of the experimental plots in the field. Zórtea, Brazil.

Additional file 2. Detailed statistical analysis #i.

Additional file 3. MCIA statistical analysis #ii (within experimental plots)

Additional file 4. Detailed statistical analysis #ii (metabolomics data)

Acknowledgements

The authors would like to thank Leticia Dambroz Filipini, Rodrigo Arnt Sant'Ana, Giulia Bortolon, Francieli Marian and Dylan Thomas Telles for their assistance during the field experiment; Prof Rubens Nodari for his assistance in the CropScience Department at UFSC. We would also like to thank Wolfram Reichenbecher, Anke Hein and Samson Simon for their constructive comments and revisions to earlier versions of this manuscript.

Author contributions

Conceptualization, SZA-T, FW; methodology, RFB, CBZ, MFE, FW and SZA-T; software, RFB and CBZ; validation, RFB; formal analysis, RFB, CBZ, MFE, FW, and SZA-T; data curation, RFB, CBZ, FW and SZA-T; writing—original draft preparation, RFB, CBZ and SZA-T; writing—review and editing, FW, MFE, and SZA-T; supervision, FW and SZA-T; project administration, FW and SZA-T. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the German Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection and commissioned by the Federal Agency for Nature Conservation (Bundesamt für Naturschutz, BfN) under grant number FKZ 3517841000. F.W. is employed by BfN who commissioned the study. Scholarships for C.B.Z and R.O.N. have been granted by National Council for Scientific and Technological Development (CNPq) in Brazil.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]CropScience Department, Federal University of Santa Catarina, Florianópolis, Brazil. [2]Department of GMO Regulation and Biosafety, Federal Agency for Nature Conservation, Bonn, Germany. [3]Department Landuse & Biosafety, Environment Agency Austria, Vienna, Austria. [4]NORCE Norwegian Research Centre AS, Climate & Environment Department, Siva Innovasjonssenter, Sykehusvn 21, 9019 Tromsø, Norway.

Received: 19 September 2022 Accepted: 22 January 2023
Published online: 21 February 2023

References

1. EFSA (2011) Panel on genetically modified organisms: guidance for risk assessment of food and feed from genetically modified plants. EFSA J 9(5):2150
2. ADAS (2013) Review of the strategies for the comprehensive food and feed safety and nutritional assessment of GM plants per se. EFSA supporting publication 2013:EN-480, 115 pp.
3. Hilbeck A, Meyer H, Wynne B et al (2020) GMO regulations and their interpretation: how EFSA's guidance on risk assessments of GMOs is bound to fail. Environ Sci Eur 32:54. https://doi.org/10.1186/s12302-020-00325-6
4. Jiang C, Meng C, Schapaugh A (2019) Comparative analysis of genetically-modified crops: Part 1. Conditional difference testing with a given genetic background. PLoS ONE 14(1):e0210747. https://doi.org/10.1371/journal.pone.0210747
5. Kok EJ, Kuiper HA (2003) Comparative safety assessment for biotech crops. Trends Biotechnol 21:439–444
6. Hilbeck A and Otto M (2015). Specificity and Combinatorial Effects of Bacillus Thuringiensis Cry Toxins in the Context of GMO Environmental Risk Assessment. Front. Environ. Sci, 09 November Sec. Environmental health and Exposome. 2015;(3). https://doi.org/10.3389/fenvs.2015.00071
7. Levidow L et al (2007) Recasting 'substantial equivalence': transatlantic governance of GM food. Sci Technol Hum Values 32:26–64
8. Millstone E, Brunner E, Mayer S (1999) Beyond 'substantial equivalence.' Nature 401:525–526. https://doi.org/10.1038/44006
9. Millstone E (2002) The limitations and potential utility of substantial equivalence. In: Evaluating substantial equivalence: a step towards improving the risk/safety evaluation of GMOs, Austrian Federal Environment Agency, Conference papers, vol 32. Vienna, pp 32–41.

10. Catchpole GS, et al (2005) Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops. Proc Natl Acad Sci USA 107, 6198–6203 102, 14458–14462.
11. Christ B, Pluskal T, Aubry S, Weng JK (2018) Contribution of untargeted metabolomics for future assessment of biotech crops. Trends Plant Sci 23(12):1047–1056. https://doi.org/10.1016/j.tplants.2018.09.011. (**Epub 2018 Oct 22 PMID: 30361071**)
12. García-Cañas V, Simó C, León C, Ibáñez E, Cifuentes A (2011) MS-based analytical methodologies to characterize genetically modified crops. Mass Spectrom Rev 30(3):396–416. https://doi.org/10.1002/mas.20286. (**Epub 2010 Dec 23. PMID: 21500243**)
13. Kuiper HA, Kok EJ, Engel KH (2003) Exploitation of molecular profiling techniques for GM food safety assessment. Curr Opin Biotechnol 14(2):238–243. https://doi.org/10.1016/s0958-1669(03)00021-1. (**PMID: 12732328**)
14. Weckwerth W (2004) Biochemical profiling in plant systems: phenotypes beyond chemi-cal equivalence. In: Nap JPH, Atanassov A, Stickema WJ (eds) Genomics for biosafety in plant biotechnology. IOS Press
15. Davis H (2010) A role for "omics" technologies in food safety assessment. Food Control 21(12):1601–1610. https://doi.org/10.1016/j.foodcont.2009.03.002
16. Agapito-Tenfen SZ, Guerra MP, Wikmark OG et al (2013) Comparative proteomic analysis of genetically modified maize grown under different agroecosystems conditions in Brazil. Proteome Sci. https://doi.org/10.1186/1477-5956-11-46
17. Brandão AR, Barbosa HS, Arruda MA (2010) Image analysis of two-dimensional gel electrophoresis for comparative proteomics of transgenic and non-transgenic soybean seeds. J Proteomics 73(8):1433–1440. https://doi.org/10.1016/j.jprot.2010.01.009. (**Epub 2010 Feb 1 PMID: 20123049**)
18. Coll A, Nadal A, Rossignol M, Puigdomènech P, Pla M (2011) Proteomic analysis of MON810 and comparable non-GM maize varieties grown in agricultural fields. Transgenic Res 20(4):939–949 (**Epub 2010 Oct 23**)
19. Hao W, Li F, Yan W et al (2017) Comparative metabolic profiling of four transgenic maize lines and two non-transgenic maize lines using high-performance liquid chromatography mass spectrometry. Acta Physiol Plant 39:167. https://doi.org/10.1007/s11738-017-2468-8
20. 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. Sci Rep 8(1):17681. https://doi.org/10.1038/s41598-018-35996-y
21. Vidal N, Barbosa H, Jacob S, Arruda M (2015) Comparative study of transgenic and non-transgenic maize (Zea mays) flours commercialized in Brazil, focussing on proteomic analyses. Food Chem 1(180):288–294. https://doi.org/10.1016/j.foodchem.2015.02.051. (**Epub 2015 Feb 16 PMID: 25766830**)
22. Zhao Y, Zhang L, Zhao C et al (2015) Metabolic responses of rice leaves and seeds under transgenic backcross breeding and pesticide stress by pseudotargeted metabolomics. Metabolomics 11:1802–1814. https://doi.org/10.1007/s11306-015-0834-3
23. Zolla L, Rinalducci S, Antonioli P, Righetti PG (2008) Proteomics as a complementary tool for identifying unintended side effects occurring in transgenic maize seeds as a result of genetic modifications. J Proteome Res 7(5):1850–1861. https://doi.org/10.1021/pr0705082
24. Benevenuto RF, Venter HJ, Zanatta CB, Nodari RO, Agapito-Tenfen SZ (2022) Alterations in genetically modified crops assessed by omics studies: Systematic review and meta-analysis. Trends Food Sci Technol 120:325–337. https://doi.org/10.1016/j.tifs.2022.01.002
25. Ad Hoc Technical Expert Group (AHTEG) (2016) Final report of the ad hoc technical expert group on risk assessment and risk management under the Cartagena protocol on biosafety. In Proceedings of the United Nations Environment Programme Convention for Biodiversity, Mexico City, 25–29 July 2016. Available online: http://www.cbd.int/doc/meetings/bs/bsrarm-ahteg-2016-01/official/bsrarm-ahteg-2016-01-06-en.pdf. Accessed 8 Apr 2021.
26. Agapito-Tenfen SZ, Okoli AS, Bernstein MJ, Wikmark OG, Myhr AI (2018) Revisiting risk governance of GM plants: the need to consider new and emerging gene-editing techniques. Front Plant Sci 21(9):1874. https://doi.org/10.3389/fpls.2018.01874.PMID:30622546;PMCID:PMC6308909
27. Heinemann JA, Kurenbach B, Quist D (2011) Molecular profiling–a tool for addressing emerging gaps in the comparative risk assessment of GMOs. Environ Int 37(7):1285–1293. https://doi.org/10.1016/j.envint.2011.05.006. (**Epub 2011 May 31 PMID: 21624662**)
28. Eckerstorfer M, Narendja F, Roschko R, Heissenberger A, Gaugitsch H (2012) Gutachten zum Stand und zu den Möglichkeiten des Einsatzes von Omics-Technologien in der GVO-Bewertung. BfN-Skripten 313, Bonn, Germany. Doi: https://doi.org/10.13140/RG.2.1.4319.9128.
29. Fontes S, Eliana MG, Carmen SS, Edison PR, and Antônio R, Panizzi (2002). The Environmental Effects of Genetically Modified Crops Resistant to Insects. Neotropical Entomology 31(4):497–513. https://doi.org/10.1590/S1519-566X2002000400001
30. EFSA (2018) EFSA Scientific Colloquium 24: omics in risk assessment: state of the art and next steps. Paper presented at the EFSA Supporting Publications.
31. National Academies of Science, Engineering and Medicine (NASEM) (2016). Committee on Genetically Engineered Crops. Genetically Engineered Crops: Experiences and Prospects. Washington DC. National Academies Press. http://www.nap.edu/catalog/23395/genetically-engineered-crops-experiences-and-prospects
32. dos Santos HG, Jacomine PKT, Dos Anjos L, De Oliveira V, Lumbreras JF, Coelho MR, et al (2018) Sistema brasileiro de classificação de solos: Brasília, DF: Embrapa.
33. EFSA (2010) Panel on Genetically Modified Organisms: statistical considerations for the safety evaluation of GMOs. EFSA J 8(2):1250
34. Carpentier SC, Witters E, Laukens K, Deckers P, Swennen R, Panis B (2005) Preparation of protein extracts from recalcitrant plant tissues: an evaluation of different methods for two-dimensional gel electrophoresis analysis. Proteomics 5(10):2497–2507
35. Jiye A, Trygg J, Gullberg J, Johansson AI, Jonsson PR, Antti H et al (2005) Extraction and GC/MS analysis of the human blood plasma metabolome. Anal Chem 77(24):8086–8094
36. Chong J, Wishart DS, Xia J (2019) Using MetaboAnalyst 4.0 for comprehensive and integrative metabolomics data analysis. Curr Protoc Bioinform 68(1):86
37. Gardinassi LG, Xia J, Safo SE, Li S (2017) Bioinformatics tools for the interpretation of metabolomics data. Curr Pharmacol Rep 3(6):374–383
38. Plubell DL, Wilmarth PA, Zhao Y, Fenton AM, Minnier J, Reddy AP et al (2017) Extended multiplexing of tandem mass tags (TMT) labeling reveals age and high fat diet specific proteome changes in mouse epididymal adipose tissue. Mol Cell Proteom 16(5):873–890
39. Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11(3):1–9
40. Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Stat 65–70.
41. Codex JFWAC (2003) Codex alimentarius: food hygiene, basic texts. Food & Agriculture Org.
42. Meng C, Kuster B, Culhane AC, Gholami AM (2014) A multivariate approach to the integration of multi-omics datasets. BMC Bioinform 15(1):1–13
43. EFSA Scientific Committee, Hardy A, Benford D, Halldorsson T, Jeger MJ, Knutsen HK, More S, Naegeli H, Noteborn H, Ockleford C, Ricci A, Rychen G, Schlatter JR, Silano V, Solecki R, Turck D, Younes M, Bresson J-L, Griffin J, Hougaard Benekou S, van Loveren H, Luttik R, Messean A, Penninks A, Ru G, Stegeman JA, van der Werf W, Westendorf J, Woutersen RA, Barizzone F, Bottex B, Lanzoni A, Georgiadis N, Alexander J (2017) Guidance on the assessment of the biological relevance of data in scientific assessments. EFSA J 15(8):4970, 73 pp. Doi: 10. 2903/j.efsa.2017.4970
44. Organization for Economic Cooperation and Development (OECD) (2012) Revised Consensus Document on compositional considerations for new varieties of soybean [*Glycine max* (l.) merr.]: key food and feed nutrients, anti- nutrients, toxicants and allergens. ENV/JM/MONO(2012)24. http://www.oecd.org/officialdocuments/displaydocument/?cote=ENV/JM/MONO(2012)24&doclanguage=en.
45. Monsanto Company (2009) Application for authorization to place on the market MON 87701 × MON 89788 soybean in the European Union, according to Regulation (EC) No 1829/2003 on genetically modified food and feed.
46. EFSA Panel on Genetically Modified Organisms (GMO) (2012) Scientific Opinion on application (EFSA-GMO-NL-2009- 73) for the placing on the market of insect resistant and herbicide tolerant genetically modified soybean MON 87701 × MON 89788 for food and feed uses, import and processing under Regulation (EC) No 1829/2003 from Monsanto. EFSA J

10(2):2560. [34 pp.]. doi:https://doi.org/10.2903/j.efsa.2012.2560. Available online: www.efsa.europa.eu/efsajournal.
47. Nature Editorial (2021) Systems biology portifolio. https://www.nature.com/subjects/systems-biology. Acessed 28 Apr 2021.
48. Breitling R (2010) What is systems biology? Front Physiol 1:9. https://doi.org/10.3389/fphys.2010.00009
49. Civelek M, Lusis AJ (2014) Systems genetics approaches to understand complex traits. Nat Rev Genet 15(1):34–48. https://doi.org/10.1038/nrg3575. (**Epub 2013 Dec 3. PMID: 24296534; PMCID: PMC3934510**)
50. Bludau I, Aebersold R (2020) Proteomic and interactomic insights into the molecular basis of cell functional diversity. Nat Rev Mol Cell Biol 21(6):327–340. https://doi.org/10.1038/s41580-020-0231-2. (**Epub 2020 Mar 31. Erratum in: Nat Rev Mol Cell Biol. 2020 Apr 17;: PMID: 32235894**)
51. Zhang Y, Gao P, Yuan JS (2010) Plant protein-protein interaction network and interactome. Curr Genomics 11(1):40–46. https://doi.org/10.2174/138920210790218016
52. Bortolotto OC, Silva GV, de Freitas BA, Pomari AF, Martinelli S, Head GP, Carvalho RA, Barbosa GC (2014) Development and reproduction of Spodoptera eridania (Lepidoptera: Noctuidae) and its egg parasitoid Telenomus remus (Hymenoptera: Platygastridae) on the genetically modified soybean (Bt) MON 87701 × MON 89788. Bull Entomol Res 104(6):724–730. https://doi.org/10.1017/S0007485314000546. (**Epub 2014 Sep 24. Erratum in: Bull Entomol Res. 2015 Apr;105(2):259-60. PMID: 25248849**)
53. Hilbeck A, Baumgartner M, Padruot M, Bigler F (1998). Effects of transgenic Bacillus thuringiensis corn-fed prey on mortality and development time of immature Chrysoperla cornea (Neuroptera: Chrysopidae), Environmental Entomology. 27(2):1:480–487. https://doi.org/10.1093/ee/27.2.480
54. Lang A, Lee M, Dolek M. et al. (2019). Laboratory tests with Lepidoptera to assess non-target effects of Bt maize pollen: analysis of current studies and recommendations for a standardised design. Environ Sci Eur 2019;31:(39). https://doi.org/10.1186/s12302-019-0220-2
55. Cesco VJS et al (2018) Agronomic performance of Intacta RR2 soybean submitted to doses of glyphosate. Planta Daninha 2018(v36):e018168727. https://doi.org/10.1590/S0100-83582018360100109
56. Zanatta CB, Benevenuto RF, Nodari RO et al (2020) Stacked genetically modified soybean harboring herbicide resistance and insecticide rCry1Ac shows strong defense and redox homeostasis disturbance after glyphosate-based herbicide application. Environ Sci Eur 32:104. https://doi.org/10.1186/s12302-020-00379-6
57. Chvátalová V (2019) A critical evaluation of EFSA's environmental risk assessment of genetically modified maize MON810 for honeybees and earthworms. Environ Sci Eur 31:52. https://doi.org/10.1186/s12302-019-0238-5
58. Lohn AF, Trtikova M, Chapela I, Van den Berg J, du Plessis H, Hilbeck A (2020) Transgene behavior in *Zea mays L.* crosses across different genetic backgrounds: Segregation patterns, cry1Ab transgene expression, insecticidal protein concentration and bioactivity against insect pests. PLoS ONE 15(9):e0238523. https://doi.org/10.1371/journal.pone.0238523

Publisher's Note