plants

MDPI

Article

Exploring the Proteomic Profile of Soybean Bran: Unlocking the Potential for Improving Protein Quality and Quantity

Mayla Daiane Corre Molinari [1,†], Renata Fuganti-Pagliarini [2,†], Yanbao Yu [3], Lilian Hasegawa Florentino [4], Liliane Marcia Mertz-Henning [5], Rayane Nunes Lima [4], Daniela Matias de Carvalho Bittencourt [4], Marcelo Oliveira Freire [6] and Elibio Rech [4,*]

[1] Arthur Bernardes Foundation, Embrapa Soybean, Londrina 86085-981, Brazil; mdcmolinari@gmail.com
[2] Piccola Scientific Consulting, Saskatoon, SK S7P 0B6, Canada; renatafuganti@gmail.com
[3] J. Craig Venter Institute, Rockville, MD 20850, USA; yayu@jcvi.org
[4] Embrapa Genetic Resources and Biotechnology, National Institute of Science and Technology in Synthetic Biology, Distrito Federal 70770-917, Brazil; lilian.florentino@embrapa.br (L.H.F.); rayanenl@gmail.com (R.N.L.); daniela.bittencourt@embrapa.br (D.M.d.C.B.)
[5] Embrapa Soybean, Londrina 86085-981, Brazil; liliane.henning@embrapa.br
[6] J. Craig Venter Institute, La Jolla, CA 92037, USA; mfreire@jcvi.org
* Correspondence: elibio.rech@embrapa.br
† These authors contributed equally to this work.

Abstract: Soybean is a rich source of vegetal protein for both animal and human consumption. Despite the high levels of protein in soybean seeds, industrial processing to obtain soybean bran significantly decreases the final protein content of the byproducts. To overcome this problem, cultivars with higher protein contents must be developed. However, selecting the target proteins is difficult because of the lack of information on the proteome profile of soybean bran. Therefore, this study obtained the comparative proteomic profiles of both natural coatless seeds and defatted bran from an elite tropical-soybean cultivar. Thus, their extracts were characterized using LC–MS/MS and a total of 550 proteins were identified. Among these, 526 proteins were detected in coatless seeds and 319 proteins in defatted bran. Moreover, a total of 139 proteins were identified as presenting different levels of content in coatless seeds and defatted bran. Among them, only 46 were retained after the seed processing. These proteins were clustered in several important metabolic pathways, such as amino-acid biosynthesis, sugar biosynthesis, and antioxidant activity, meaning that they could act as targets for bioactive products or genome editing to improve protein quality and quantity in soybean grains. These findings can enhance our understanding regarding protein robustness for both soybean crops and the commercial bran improvement because target proteins must remain intact after processing and must be bioactive when overexpressed. Overall, the soybean bran proteomic profile was explored for the first time, providing a valuable catalogue of target proteins that can tolerate the industrial process.

Keywords: *Glycine max*; BRS 537; industrial processing; bioactive proteins; grain quality; LC–MS/MS; proteomic profile

Citation: Molinari, M.D.C.; Fuganti-Pagliarini, R.; Yu, Y.; Florentino, L.H.; Mertz-Henning, L.M.; Lima, R.N.; Bittencourt, D.M.d.C.; Freire, M.O.; Rech, E. Exploring the Proteomic Profile of Soybean Bran: Unlocking the Potential for Improving Protein Quality and Quantity. *Plants* **2023**, *12*, 2704. https://doi.org/10.3390/plants12142704

Academic Editor: Fernando Henrique Reboredo

Received: 7 June 2023
Revised: 28 June 2023
Accepted: 13 July 2023
Published: 20 July 2023

1. Introduction

Soybean (*Glycine max*) is one of the most important commodities on the international market. In addition to its multiple uses in animal nutrition and human food consumption, this grain can be used in many industrial processes, as well as in medicine, owing to its health benefits [1–6]. Among the byproducts of soybean, bran is produced through standard soybean milling processes. It is commonly utilized in animal feed, as a fiber source for dairy and beef cattle, in pet food with reduced fat content, and as a supplement in animal feed [7,8].

Overall, soybean seeds have a high concentration of protein (40%) and are abundant in essential amino acids, such as phenylalanine, valine, tryptophan, threonine, lysine, leucine, and isoleucine, and sulfur-containing amino acids, such as methionine and cystine, which are essential for a balanced diet [9,10]. Although natural soybean grains are rich in proteins, most are degraded during soybean processing. This degradation occurs because of many factors, including the mechanical, thermal, and chemical steps involved in processing [11]. These processing factors influence protein properties and can modify their solubility and digestibility, thereby affecting their potential applications in the food industry. Thus, there are two challenges that hamper the development of soybean bran with a higher quantity and quality of proteins. The first one is industrial soybean processing, which involves high heating treatments, resulting in a significant loss of proteins due to denaturation [12]. The second one is that practical breeding progress increases seed protein but decreases yield. Given that seed protein contents have a negative genetic correlation with seed yield, other seed components such as oil and sucrose, along with interactions with environmental effects such as temperature during seed development [12–14]. To reduce these obstacles, the development of soybean cultivars with enhanced protein and amino acid content would further increase the crop's economic value and would contribute to enriching the entire value chain, from farmers to processors to end-users [12].

To mitigate these losses, which decrease the economic value of soybean grains, many biotechnology strategies have focused on improving complex multigene traits, such as the overexpression of genes involved in improving the protein content in the grain [15]. However, to improve the soybean bran protein content, it is necessary to identify target proteins that must remain intact after processing and become bioactive when overexpressed by synthetic biology. If the overexpressed protein is not tolerant to processing, it is lost in the bran.

Currently, the most significant research on soybean seed proteomics includes the differential expression of the cotyledon, mature seed, seed coat, and seeds at different developmental stages, including conventional, transgenic, and mutated soybean cultivars [16–25]. Some of these studies have aimed at the quantitative analysis of proteomic profiles, comparing different stress conditions [6]. Nevertheless, the available literature on the soybean proteome is limited to data regarding unprocessed soybean seeds, and the studies do not represent the protein content available in by-products, such as bran, after industrial processes.

To overcome the lack of information on the protein remaining after soybean processing, we identified and compared the protein profiles of both natural coatless seeds and processed defatted bran from the elite soybean cultivar, BRS 537 (Embrapa, Brazil). The proteomic and biological pathway data described here provide knowledge on the molecular composition of elite tropical soybeans by focusing on proteins tolerant to industrial processing stages, aiming to allow researchers to develop innovative strategies to improve the nutritional quality of soybean products by providing a robust catalogue of target proteins.

In this work, we focused on identifying and comparing the protein profiles of natural coatless seeds and processed defatted bran from the elite soybean cultivar, BRS 537 (Embrapa, Brazil). Our intention is not only to contribute to the existing body of knowledge on the molecular composition of elite tropical soybeans, but also to highlight proteins that display resilience during processing stages. The proteome analysis of soybean bran after processing allowed the detection and quantification of anti-nutritional factors and allergenic proteins, which is vital to ensure food safety and proper labeling of allergens in food products. Lastly, the ultimate aim is to provide a robust catalogue of target proteins that could aid researchers in developing innovative strategies to improve the nutritional quality of soybean products.

2. Results

2.1. Comparative Proteomic Profile

LC–MS/MS was used to analyze and identify the comparative proteomic profile from the elite soybean cultivar, BRS 537 (Embrapa, Brazil). A total of 550 proteins were identified. Among these, 526 proteins were detected in coatless seeds and 319 proteins in defatted bran. Moreover, a total of 139 proteins were identified as presenting different levels of content in coatless seeds and defatted bran (Figure 1). However, 93 proteins were completely lost in the bran after soybean processing and the 46 proteins retained after the seed treatment showed lower protein levels (Figure 1). The chromosomal positions of all the genes that encode the 139 differentially expressed proteins are indicated in Figure 2. Furthermore, the complete list of the 550 identified proteins along with their absolute counts in spectrum counts is shown in the Supplementary File S1.

Regarding the amino acid composition, of the 139 proteins with different contents, 64 were identified as rich in essential amino acids if, in their protein's total amino acid composition, more than 10% comprised either valine, leucine, lysine, or a combination of these essential amino acids. It is important to note that essential amino acids with less than 10% of the total protein composition were not considered. Among those considered, 14 proteins showed more than 10% valine in their composition, 30 proteins contained more than 10% leucine, and 11 proteins featured more than 10% lysine. A combination of more than 10% leucine and valine was present in four proteins, 10% lysine and leucine were observed in three proteins, and 10% lysine and valine were noted in two proteins. In total, 34 proteins (53%) considered rich in essential amino acids were lost in the defatted bran, with 30 proteins (47%) retained after processing (Figure 1).

The cultivar's genome was sequenced by Embrapa, and it is available at NCBI GenBank under access GCA_012273815.2 assembly (https://www.ncbi.nlm.nih.gov/assembly/GCA_012273815.2/ accessed on 29 May 2022). The novelty of this information is that most of the available soybean-protein databases, such as the Soybean Proteome Database (http://proteome.dc.affrc.go.jp/Soybean/ accessed on 29 May 2022), are limited to data from unprocessed soybean seeds and generally do not describe the protein content available in the processed bran.

2.2. Protein Functional Categories

The KEGG-analysis results identified 32 different pathways represented in the coatless seeds and defatted bran. Of these 32 pathways, 15 were lost after soybean processing, with no proteins remaining (amino acid retrieval, alkaloid biosynthesis, aminoacyl-tRNA biosynthesis, calcium-permeable channel, chitin catalysis, folate biosynthesis, kinase modulator, nucleosome, proteasome compound, protein processing, protein regulation, protein transport, sterol transport, storage iron, and translation initiation). For all the others 17, at least one protein remained in the defatted bran (Figure 1).

Applying the criterion of the identification of three or more proteins in a pathway for it to be considered enriched, the KEGG results pointed to 14 enriched pathways. Thus, 25 proteins were identified as involved in the amino acid biosynthesis pathway; 16 in the ribosome compound; 11 in the modulation of protein folding; 9 in sugar biosynthesis; 8 in defense response; 6 in oxidoreductase, proteasome compound, protein transport, and anti-nutritional factor (2 putative); 5 in antioxidant activity; 4 in translation-elongation factor and acetaldehyde; and 3 in cellulose biosynthesis and the cytoskeleton (Figure 1).

Ptt ID	GENE ID	Annotation	Activity	Unique Peptides	Ptt Size (# of aas)	Mature coatless seed (NSC)	Deffated Bran (NSC)	Log2FC	p-Value	FDR	Essential amino acids (≥10%) in each protein
H2D5S3	Glyma.10G006500	malate dehydrogenase (mdh1)	aa biosynthesis	2	328	658	3	-8	0,0	0,0	10.1% Valine
A0A0R4J2L9	Glyma.02G005500	malate dehydrogenase (mdh1)	aa biosynthesis	2	332	661	3	-8	0,0	0,0	10.2% Valine
I1MB71	Glyma.14G189400	fructose-bisphosphate aldolase, class i (aldo)	aa biosynthesis	8	358	788	7	-7	0,0	0,0	10.3% Leucine
I1JH86	Glyma.02G222400	fructose-bisphosphate aldolase, class i (aldo)	aa biosynthesis	5	358	830	7	-7	0,0	0,0	10.3% Leucine
I1KDM8	Glyma.06G231500	nad-dependent malate dehydrogenase	aa biosynthesis	4	345	219	19	-4	0,1	0,1	10.4% Valine
Q9SPB8	Glyma.12G159300	malate dehydrogenase (mdh2)	aa biosynthesis	4	345	292	19	-4	0,0	0,0	10.4% Valine
A0A0R0KA84	Glyma.04G193500	glyceraldehyde 3-phosphate dehydrogenase	aa biosynthesis	5	338	1643	136	-4	0,0	0,0	10.9% Valine
C6TN36	Glyma.18G009700	glyceraldehyde-3-phosphate dehydrogenase gapc1	aa biosynthesis	4	340	999	28	-6	0,0	0,0	11.8% Valine
A0A368UH36	Glyma.15G262100	phosphoglycerate kinase	aa biosynthesis	2	401	350	24	-5	0,0	0,0	12.2% Leucine
I1LXY1	Glyma.13G028400	mete	aa biosynthesis	7	803	256	4	-6	0,0	0,0	NA
I1JWK3	Glyma.20G055900	mete	aa biosynthesis	8	799	308	4	-7	0,0	0,0	NA
I1LCQ1	Glyma.10G201100	pyruvate kinase	aa biosynthesis	7	526	14	0	-7	0,0	0,0	10.3% Leucine
A0A0R0GFW6	Glyma.15G261900	phosphoglycerate kinase 1	aa biosynthesis	2	483	91	0	-10	0,0	0,0	11.6% Leucine
I1M561	Glyma.13G344500	fructokinase	aa biosynthesis	7	331	41	0	-9	0,0	0,0	12.7% Leucine
A0A0R0G6T3	Glyma.15G038100	triosephosphate isomerase	aa biosynthesis	4	253	985	0	-13	0,0	0,0	14.2% Valine
A0A0R4J2U8	Glyma.03G027500	transketolase	aa biosynthesis	6	731	12	0	-7	0,0	0,0	NA
I1JC43	Glyma.02G036300	atp-citrate synthase	aa biosynthesis	2	423	20	0	-8	0,0	0,0	NA
I1KWM7	Glyma.08G254500	phosphogluconic acid dehydrogenase	aa biosynthesis	14	486	30	0	-8	0,0	0,0	NA
I1M3M4	Glyma.13G291300	ketol-acid reductoisomerase (ilvc)	aa biosynthesis	4	586	42	0	-9	0,0	0,0	NA
I1MT35	Glyma.17G075300	glyceraldehyde-3-phosphate dehydrogenase	aa biosynthesis	8	497	71	0	-10	0,0	0,0	NA
A0A0R4J3C9	Glyma.05G152100	serine hydroxymethyltransferase	aa biosynthesis	3	471	88	0	-10	0,0	0,0	NA
I1MFH9	Glyma.15G108400	glyoxalase i homolog	aa biosynthesis	2	287	90	0	-10	0,0	0,0	NA
A0A0R0EPH6	Glyma.19G186000	triose-phosphate isomeras	aa biosynthesis	4	309	157	0	-11	0,0	0,0	NA
I1JPW5	Glyma.03G190500	enolase	aa biosynthesis	5	444	232	0	-11	0,0	0,0	NA
I1NAI7	Glyma.19G190900	enolase	aa biosynthesis	5	444	246	0	-11	0,0	0,0	NA
I1JGP8	Glyma.02G204000	leucyl aminopeptidase (carp, pepa)	aa retrieval	5	570	56	0	-9	0,0	0,0	10.2% Leucine
I1N7G4	Glyma.19G077800	leucyl aminopeptidase (carp, pepa)	aa retrieval	21	582	89	0	-10	0,0	0,0	11.2% Leucine
I1MAE6	Glyma.14G121200	alcohol dehydrogenase	acetaldehyde	4	380	461	11	-6	0,0	0,0	10.0% Valine
A0A0R0JL19	Glyma.06G122600	alcohol dehydrogenase	acetaldehyde	12	381	620	11	-6	0,0	0,0	10.2% Valine
A0A0R4J4U4	Glyma.13G035200	alcohol dehydrogenase	acetaldehyde	11	379	3955	24	-8	0,0	0,0	NA
I1N1F9	Glyma.18G137300	phenylacetaldehyde dehydrogenase	acetaldehyde	21	536	53	0	-9	0,0	0,0	NA
A0A0R4J4R5	Glyma.12G059100	tropinone reductase-related	alkaloid biosynthesis	3	267	18	0	-8	0,0	0,0	10.5% Leucine
I1LBD0	Glyma.10G155000	seryl-trna synthetase (sars, sers)	aminoacyl-tRNA biosynthesis	3	447	38	0	-9	0,0	0,0	NA
I1KYW3	Glyma.08G341000	kunitz family trypsin and protease inhibitor protein-related	antinutricional factor	8	203	2455	864	-2	0,0	0,0	NA
C6SWW4	Glyma.01G095000	kunitz family trypsin and protease inhibitor protein-related	antinutricional factor	12	203	13754	12001	-1	0,0	0,0	NA
C6K8D0	Glyma.08G342300	kunitz family trypsin and protease inhibitor protein-related	antinutricional facto	8	238	21	0	-8	0,1	0,1	10.1% Leucine
B1ACD5	Glyma.09G155500	kunitz family /dr4 protein-related	antinutricional factor	5	211	128	0	-10	0,0	0,0	NA
C6TFC1	Glyma.03G040400	protease inhibitor/seed storage/ltp family	antinutricional factor-putative	3	122	545	143	-3	0,0	0,0	NA
I1JL08	Glyma.03G040000	protease inhibitor/seed storage/ltp family	antinutricional factor-putative	5	117	21	0	-8	0,1	0,1	12% Valine
I1LKZ3	Glyma.11G192700	superoxide dismutase cu-zn -related	antioxidant activity	6	204	78	4	-5	0,0	0,0	10.3% Valine
I1L8Q0	Glyma.10G047700	glutathione s-transferase	antioxidant activity	8	237	212	100	-2	0,1	0,1	10.5% Leucine
A0A0R0HQQ2	Glyma.10G078400	peroxiredoxin (alkyl hydroperoxide reductase subunit c)	antioxidant activity	6	258	247	27	-4	0,0	0,0	10.5% Leucine
A0A0R0IAF2	Glyma.09G192800	peroxiredoxin	antioxidant activity	12	162	1384	272	-3	0,0	0,0	NA
Q9FQE8	Glyma.08G175200	glutathione s-transferase u21-related	antioxidant activity	2	219	28	0	-8	0,0	0,0	10.0% Lysine
G3E7M9	Glyma.13G088700	annexin d1-related	calcium-permeable channel	19	316	15	0	-7	0,0	0,0	12.3% Leucine
A0A0R0GZF0	Glyma.13G199800	annexin a7/11 (anxa7_11)	calcium-permeable channel	3	313	156	0	-11	0,0	0,0	NA
I1MXZ6	Glyma.17G254200	thioredoxin h1	cell redox homeostase	7	123	670	437	-2	0,1	0,1	12.2% Valine
I1K7N6	Glyma.06G028600	thioredoxin	cell redox homeostase	2	117	84	0	-10	0,0	0,0	12.8% Lysine/10.3% Valine
I1LWV7	Glyma.13G053000	udp-arabinopyranose mutase	cellulose biosynthesis	3	357	34	0	-9	0,0	0,0	NA
I1JEI0	Glyma.02G120900	udp-arabinopyranose mutase	cellulose biosynthesis	3	368	36	0	-9	0,0	0,0	NA
I1J637	Glyma.01G063000	udp-arabinopyranose mutase	cellulose biosynthesis	4	368	36	0	-9	0,0	0,0	NA
I1MI18	Glyma.15G206800	chitinase	chitin catalysis	16	365	84	0	-10	0,0	0,0	NA
Q9SDY6	Glyma.02G042500	chitinase	chitin catalysis	12	320	134	0	-11	0,0	0,0	NA
C6T0B5	Glyma.13G170900	tubulin/ conserved hypothetical protein	cytoskeleton	9	129	1551	852	-2	0,0	0,0	NA
I1K7J4	Glyma.06G024100	tubulin	cytoskeleton	2	448	54	0	-9	0,0	0,0	NA
I1KRC0	Glyma.08G081100	tubulin beta-4 chain-related	cytoskeleton	2	443	63	0	-9	0,0	0,0	NA
I1M0K3	Glyma.13G189500	cysteine proteinase inhibitor 6	defense response	4	245	354	151	-2	0,0	0,0	NA
I1JFX0	Glyma.02G167700	universal stress protein family (usp)	defense response	6	191	495	14	-6	0,0	0,0	NA
Q9XER5	Glyma.16G031300	late embryogenesis abundant protein (lea_2)	defense response	16	152	3031	2877	-1	0,1	0,1	NA
C6SZN6	Glyma.12G058500	universal stress protein family (usp)	defense response	2	164	209	0	-11	0,0	0,0	11.6% Leucine
C6ZS00	Glyma.06G319700	leucine-rich repeat (lrr) family protein-related	defense response	14	489	45	0	-9	0,0	0,0	15.3% Leucine
K7L949	Glyma.08G268000	gamma-thionin family (gamma-thionin)	defense response	2	127	19	0	-8	0,1	0,1	NA
I1KS58	Glyma.08G108000	selenium-binding protein 1 (selenbp1)	defense response	2	484	136	0	-11	0,0	0,0	NA
C6SVT0	Glyma.10G147700	universal stress protein family (usp)	defense response	9	162	454	0	-12	0,0	0,0	NA
A0A0R0H3L4	Glyma.12G088300	dna ligase/ poly (adp-robose) polymerase 3	DNA repair	16	815	177	1	-8	0,0	0,0	NA
I1KUQ0	Glyma.08G189200	linoleate 9s-lipoxygenase (lox1_5)	fatty acid oxidation	2	860	991	210	-3	0,0	0,0	10.0% Leucine
I1M596	Glyma.13G347500	linoleate 9s-lipoxygenase (lox1_5)	fatty acid oxidation	24	866	2403	283	-4	0,0	0,0	10.2% Leucine
I1LFG6	Glyma.10G293800	formate--tetrahydrofolate ligase	folate biosynthesis	4	636	75	0	-10	0,0	0,0	NA
I1NJ85	Glyma.20G243700	formate--tetrahydrofolate ligase	folate biosynthesis	3	636	82	0	-10	0,0	0,0	NA
I1MCF0	Glyma.15G007300	amp-activated protein kinase	kinase modulator	6	413	65	0	-10	0,0	0,0	10.4% Leucine/ 10.2% Valine

Figure 1. *Cont.*

Ptt ID	GENE ID	Annotation	Activity	Unique Peptides	Ptt Size (# of aas)	Mature coatless seed (NSC)	Deffated Bran (NSC)	Log2FC	p-Value	FDR	Essential amino acids (≥10%) in each protein
E9KNA6	Glyma.18G298300	14-3-3-like protein gf14 lambda	modulation of folding proteins	7	257	150	3	-6	0,0	0,0	10.5% Leucine
A0A0R4J5A5	Glyma.14G176900	14-3-3 protein	modulation of folding proteins	3	261	176	34	-3	0,0	0,1	NA
I1KPN3	Glyma.08G025700	heat shock 70kda protein 5 (hspa5, bip)	modulation of folding proteins	2	667	515	157	-2	0,0	0,1	NA
I1K670	Glyma.05G219400	heat shock 70kda protein 5 (hspa5, bip)	modulation of folding proteins	5	668	584	210	-2	0,0	0,1	NA
A0A0R4J460	Glyma.09G231200	bag family molecular chaperone regulator 7	modulation of folding proteins	9	402	27	0	-8	0,0	0,1	10.2% Lysine
I1MJ28	Glyma.15G250500	chaperonin 60 subunit beta 1	modulation of folding proteins	12	591	69	0	-10	0,0	0,0	10.7% Valine
A0A0R0HF31	Glyma.12G210400	14-3-3 protein	modulation of folding proteins	5	261	23	0	-8	0,0	0,1	NA
I1LR60	Glyma.12G078100	chaperonin 60 subunit alpha 1	modulation of folding proteins	17	584	34	0	-9	0,0	0,0	NA
C6SV97	Glyma.18G062900	peptidyl-prolyl cis-trans isomerase	modulation of folding proteins	5	204	162	0	-11	0,0	0,0	NA
A0A0R0H9T5	Glyma.12G024700	peptidyl-prolyl cis-trans isomerase cyp18-3-related	modulation of folding proteins	7	172	1441	0	-14	0,0	0,0	NA
A0A0R0HEL8	Glyma.11G098700	peptidyl-prolyl cis-trans isomerase cyp19-1	modulation of folding proteins	7	172	1504	0	-14	0,0	0,0	NA
C6T9B1	Glyma.08G084800	urease accessory protein (ureg)	nitrogen metabolism	13	285	297	37	-4	0,0	0,1	NA
I1N2I4	Glyma.18G182600	nitrogen metabolic regulation protein	nitrogen metabolism	5	255	29	0	-8	0,0	0,0	NA
A0A0R0F3E0	Glyma.19G234100	histone h2a	nucleossome	2	144	42	0	-9	0,0	0,0	12.5% Lysine/ 11.1% Leucine
I1NJ59	Glyma.20G240300	glutathione dehydrogenase (ascorbate)	oxidoredutase	4	213	512	215	-2	0,0	0,1	11.7% Leucine/ 10.8% Valine
A0A0R4J681	Glyma.20G233000	glucose and ribitol dehydrogenase homolog 1-related	oxidoredutase	3	293	1387	38	-6	0,0	0,0	NA
I1J6Z7	Glyma.01G102600	aldo-keto reductase family 4 member c10	oxidoredutase	4	313	27	0	-8	0,0	0,0	10.2% leucine
C6TIQ5	Glyma.03G097600	allyl-alcohol dehydrogenase	oxidoredutase	7	343	40	0	-9	0,0	0,0	11.4% Leucine
I1K843	Glyma.06G044600	3-oxo-delta(4,5)-steroid 5-beta-reductase	oxidoredutase	2	388	22	0	-8	0,0	0,2	NA
C6T871	Glyma.01G227900	11-beta-hydroxysteroid dehydrogenase-like 5	oxidoredutase	3	355	41	0	-9	0,0	0,0	NA
A0A0R4J4W3	Glyma.13G130400	26s proteasome regulatory subunit t2 (psmc1, rpt2)	proteasome compound	5	443	14	0	-7	0,0	0,1	10.4% Leucine
I1KDR0	Glyma.06G236100	26s proteasome non-atpase regulatory subunit 5 (psmd5)	proteasome compound	14	525	28	0	-8	0,0	0,0	12.8% Leucine
A0A0R0EW30	Glyma.18G050900	26s proteasome regulatory subunit t3 (psmc4, rpt3)	proteasome compound	6	422	17	0	-8	0,0	0,1	NA
A0A0R4J321	Glyma.04G077400	20s proteasome subunit beta 1 (psmb6)	proteasome compound	7	233	47	0	-9	0,0	0,0	NA
I1MN39	Glyma.16G126200	20s proteasome subunit alpha 4 (psma7)	proteasome compound	7	249	54	0	-9	0,0	0,0	NA
K7MS03	Glyma.18G138900	20s proteasome subunit beta 6 (psmb1)	proteasome compound	6	223	65	0	-10	0,0	0,0	NA
I1JPP3	Glyma.03G182800	transitional endoplasmic reticulum atpase (vcp, cdc48)	Protein processing	6	808	100	0	-10	0,0	0,0	NA
I1JXA0	Glyma.04G186000	transitional endoplasmic reticulum atpase (vcp, cdc48)	Protein processing	3	814	126	0	-10	0,0	0,0	NA
K7KWZ7	Glyma.06G238200	protein phosphatase 2c 39-related	protein regulation	6	282	48	0	-9	0,0	0,0	NA
C6TFP4	Glyma.08G105100	nuclear transport factor 2	protein transport	3	123	20	0	-8	0,1	0,2	10.6% Leucine
I1NH59	Glyma.20G170700	ran-binding protein 1 (ranbp1)	protein transport	4	223	38	0	-9	0,0	0,0	12.1% Lysine
I1L0S5	Glyma.09G037800	mitochondrial inner membrane translocase subunit	protein transport	2	181	108	0	-10	0,0	0,0	12.7% Leucine
I1K5E6	Glyma.05G243800	rack1	protein transport	13	325	15	0	-7	0,0	0,1	NA
I1KSQ0	Glyma.08G126200	rab gdp dissociation inhibitor (gdi1_2)	protein transport	7	444	19	0	-8	0,0	0,1	NA
C6SX05	Glyma.11G150400	nucleoporin-related	protein transport	4	191	154	0	-11	0,0	0,0	NA
A0A0R0KVB4	Glyma.03G239700	aspartyl proteases	proteolysis	10	427	5439	922	-3	0,0	0,0	10.5% Leucine
O64458	Glyma.08G116300	chymopapain / papaya proteinase ii	proteolysis	12	379	4829	2536	-2	0,0	0,0	NA
I1LZ03	Glyma.13G136400	large subunit ribosomal protein l4e (rp-l4e, rpl4)	ribossome compound	4	405	85	2	-6	0,0	0,0	10.6% Lysine
A0A0R0L140	Glyma.02G255500	large subunit ribosomal protein lp0 (rp-lp0, rplp0)	ribossome compound	2	320	84	3	-6	0,0	0,0	10.6% Valine
I1L8R1	Glyma.10G048600	large subunit ribosomal protein l4e (rp-l4e, rpl4)	ribossome compound	4	405	88	2	-6	0,0	0,0	10.9% Lysine
C6T048	Glyma.20G169900	small subunit ribosomal protein s18e (rp-s18e, rps18)	ribossome compound	2	152	692	325	-2	0,0	0,2	11.8% Lysine
C6TLT3	Glyma.11G006600	small subunit ribosomal protein s3ae (rp-s3ae, rps3a)	ribossome compound	3	261	98	6	-4	0,0	0,2	13.0% Lysine
A0A0R4J5K4	Glyma.16G127500	small subunit ribosomal protein s4e (rp-s4e, rps4)	ribossome compound	15	264	88	0	-10	0,0	0,0	10.6% Lysine
A0A0R0HXR3	Glyma.10G224200	small subunit ribosomal protein s18e (rp-s18e, rps18)	ribossome compound	2	152	201	0	-11	0,0	0,0	11.8% Lysine
I1NJ30	Glyma.20G237400	large subunit ribosomal protein l18e (rp-l18e, rpl18)	ribossome compound	2	187	91	0	-10	0,0	0,0	11.8% Lysine/ 10.2% Leucine
C6SVV7	Glyma.13G369700	small subunit ribosomal protein s12e (rp-s12e, rps12)	ribossome compound	6	141	156	0	-11	0,0	0,0	12.8% Valine/ 10.6% Leucine
C6SYZ8	Glyma.09G024000	large subunit ribosomal protein l13ae (rp-l13ae, rpl13a)	ribossome compound	2	206	35	0	-9	0,0	0,0	13.1% Lysine/ 11.2% Valine
C6T4P7	Glyma.07G198600	small subunit ribosomal protein s6e (rp-s6e, rps6)	ribossome compound	2	249	69	0	-10	0,0	0,0	13.7% Lysine
C6SVK2	Glyma.10G018200	large subunit ribosomal protein l22e (rp-l22e, rpl22)	ribossome compound	2	119	82	0	-10	0,0	0,0	15.1% Lysine
I1LIV8	Glyma.11G103700	large subunit ribosomal protein l7ae (rp-l7ae, rpl7a)	ribossome compound	3	259	19	0	-8	0,1	0,2	17.0% Lysine/ 10.8% Leucine
A0A0R0IZE4	Glyma.07G043800	small subunit ribosomal protein sae (rp-sae, rpsa)	ribossome compound	7	310	40	0	-9	0,0	0,0	NA
C6SYE0	Glyma.09G045000	small subunit ribosomal protein s21e (rp-s21e, rps21)	ribossome compound	2	82	91	0	-10	0,0	0,0	NA
I1LGB7	Glyma.11G020100	small subunit ribosomal protein s2e (rp-s2e, rps2)	ribossome compound	3	275	143	0	-11	0,0	0,0	NA
C6T049	Glyma.02G270200	niemann pick type c2 protein npc2-related	sterol transport	2	166	29	0	-8	0,0	0,1	10.8% Leucine
I1J7H3	Glyma.01G124500	ferritin heavy chain (fth1)	storage iron	3	257	29	0	-8	0,0	0,0	10.9% Leucine
A0A0R0IIZ3	Glyma.09G178100	beta-glucosidase 43-related	sugar biosynthesis	5	506	236	13	-5	0,0	0,1	NA
C6TKQ3	Glyma.12G036700	mannose-6-phosphate 6-reductase	sugar biosynthesis	4	309	32	0	-8	0,0	0,0	10.7% Leucine
I1MBR7	Glyma.14G210700	utp--glucose-1-phosphate uridylyltransferasev	sugar biosynthesis	10	469	265	0	-12	0,0	0,0	11.5% Leucine
I1N6A5	Glyma.19G028400	alpha-glucan phosphorylase 1	sugar biosynthesis	3	981	47	0	-9	0,0	0,0	NA
I1LWR4	Glyma.13G057800	alpha-glucan phosphorylase 1	sugar biosynthesis	3	978	48	0	-9	0,0	0,0	NA
I1KQ93	Glyma.08G044100	phosphoglucomutase (pgm)	sugar biosynthesis	18	582	54	0	-9	0,0	0,0	NA
I1LIL5	Glyma.11G095100	glucan endo-1,3-beta-d-glucosidase	sugar biosynthesis	19	340	64	0	-10	0,0	0,0	NA
A0A0R0FT50	Glyma.16G209800	alpha-galactosidase (e3.2.1.22b, gala, rafa)	sugar biosynthesis	3	421	102	0	-10	0,0	0,0	NA
I1L6Y9	Glyma.09G273400	alpha-galactosidase 1	sugar biosynthesis	6	410	138	0	-11	0,0	0,0	NA
A0A0R0JI98	Glyma.06G170900	elongation factor 1-beta	translation elongation activity	6	230	674	241	-2	0,0	0,0	10.0% Leucine/ 10.0% Valine
C6T034	Glyma.02G276600	elongation factor 1-beta (eef1b)	translation elongation activity	2	223	829	429	-2	0,0	0,1	NA
I1KCA8	Glyma.06G176900	elongation factor tu (tuf, tufm)	translation elongation activity	3	479	13	0	-7	0,0	0,1	NA
C6TNT2	Glyma.16G001900	elongation factor 1-gamma (eef1g)	translation elongation activity	2	419	258	0	-11	0,0	0,0	NA
A0A0R4J4Z2	Glyma.13G106200	translation initiation factor 4a (eif4a)	translation initiation	5	413	98	0	-10	0,0	0,0	10.2% Leucine

Figure 1. Differentially expressed proteins identified in coatless soybean seeds and defatted bran. Protein ID or UniProt is provided as well as gene ID, annotation, related activity, number of unique

peptides, protein size (in number of amino acids), normalized spectrum counts, soybean coatless seed and defatted bran Log2FC, *p*-value, false-discovery rate, and percentage of essential amino acids (higher than 10%). Proteins classified as anti-nutritional factors are highlighted in red; lipoxygenases are marked in pink, and proteins present in coatless seed but that were completely lost in defatted bran are highlighted in gray. Legend: Ptt: protein; FDR: false-discovery rate; NA: not applicable; NSC: normalized spectrum counts.

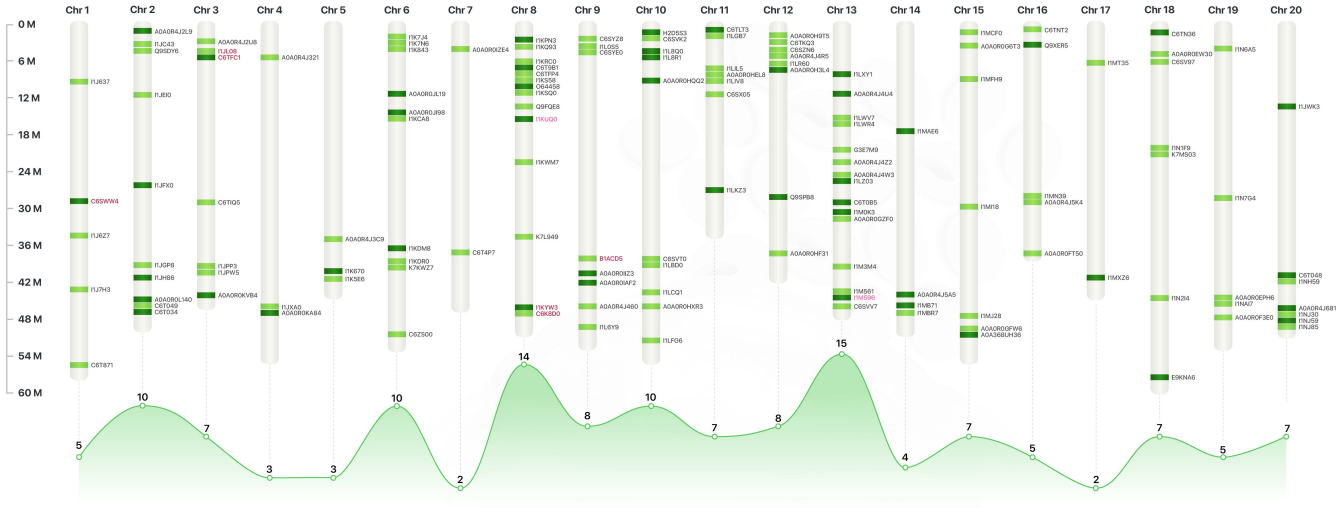

Figure 2. Chromosomal positions of all the genes that encode the 139 differentially expressed proteins. Dark green represents proteins completely retained in the defatted bran after soybean processing. Light green represents proteins lost in defatted bran. Proteins marked in red are anti-nutritional factors, and proteins marked in pink are lipoxygenases (LOX).

Two proteins were identified in the pathways involved in amino acid retrieval, the calcium-permeable channel, cell-redox homeostasis, chitin catalysis, fatty-acid oxidation, folate biosynthesis, nitrogen metabolism, protein processing, and proteolysis. Only one protein was identified in the following pathways: alkaloid biosynthesis, aminoacyl-RNA biosynthesis, DNA repair, kinase modulation, protein regulation, nucleosome, sterol transport, iron storage, and translation initiation (Figure 1). One important group of proteins for the human food industry identified in both the coatless seeds and the defatted bran was the lipoxygenases, despite the group's undesirable flavor (2).

To illustrate the complexity of the protein biosynthesis pathway, all 25 proteins identified either in the coatless seeds or in the defatted bran are combined in Figure 3. It is important to highlight that 15 of these 25 proteins were identified as rich in essential amino acids. In addition, 14 of the 25 proteins were lost in the defatted bran after industrial processing, and 6 of them were classified as rich in essential amino acids.

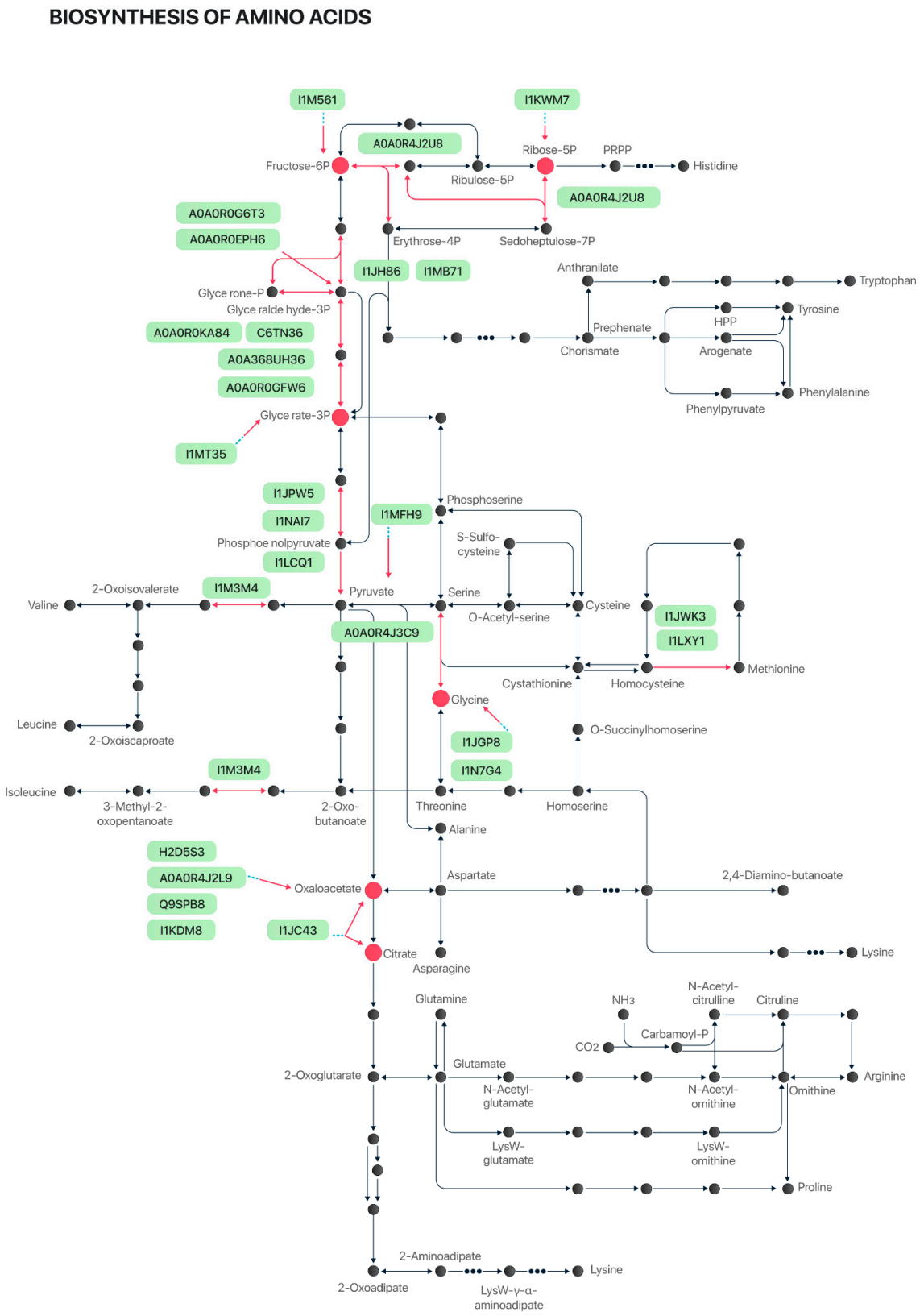

Figure 3. Amino-acid biosynthesis pathway. The 25 proteins differentially expressed are highlighted in light green and their respective metabolic-activity positions are indicated by red arrows and red dots.

3. Discussion

Soybeans are an important driving force in the industrial economies of producing countries. Some data suggest that the international market for soybean-protein ingredients was valued at around USD 9.7 billion in 2018 and it is projected to reach USD 16.6 billion by 2026, at a compound annual growth rate (CAGR) of 6.9% [26]. These numbers explain why the industry is focused on developing cultivars with more sources of added value, such as higher protein content. However, the use of industrial processes to obtain soybean by-products can decrease protein and essential amino acid contents. Soybean bran is the by-product that remains after the outer layer of the soybean seed has been removed and the oil has been extracted. It is widely recognized for its high protein content and significant levels of dietary fiber, making it a promising ingredient for diverse food applications and a potential source of bioactive compounds. Therefore, analyzing the proteomic profile of soybean bran after processing provides valuable information about its nutritional and functional properties. This knowledge can help us find value-added applications for this agricultural by-product, thereby reducing waste and increasing the overall profitability of the soybean industry. Thus, researchers have focused on identifying key proteins and developing genetic engineering and synthetic-biology strategies to overcome the loss of protein content and nutritional quality during industrial food processes.

Soybean bran has emerged as an abundant and cost-effective source of essential amino acids. It comprises all nine essential amino acids in significant amounts, making it a valuable component in a balanced diet. The essential amino acids found in soybean bran include methionine, valine, threonine, phenylalanine, isoleucine, leucine, histidine, tryptophan, and lysine. In this study, more than 50% (26 out of 46) of the proteins maintained in the defatted bran were considered rich in essential amino acids, since more than 10% of their protein composition comprised valine, leucine, lysine, or a combination of these. The loss of these proteins in the defatted bran after processing is an industry concern, as these essential amino acids are directly related to the quality and quantity of the total protein content in soybean by-products. It is worth noting that even if not fully intact, proteins rich in essential amino acids enhance the quality of soy-based products. This is an important feature for industry animal feed as poor bran quality interferes with the maintenance, growth, and reproduction of animals, as well as the production of final products, such as meat, eggs, and milk [27].

In this study, although some of the proteins identified in coatless seeds and defatted bran did not present a composition rich in essential amino acids, they can play important roles in the metabolic pathways that are involved in the synthesis of proteins and essential amino acids, or even act on their targets (bioactive). This would be of interest in genetic engineering and the development of bioproducts (Figure 1).

In addition to the biosynthesis pathways of essential amino acids, other pathways are also shown in Figure 3. The metabolism of sulfur-containing amino acids, such as methionine and cysteine, has been connected to several crucial aspects of human and animal health and nutrition [27,28]. Methionine is an essential amino acid, and cysteine is considered a "conditional" essential amino acid since mammals can produce cysteine from methionine. However, mammals are unable to synthesize essential amino acids on their own and must obtain them from their diet [27], highlighting the importance of cultivating soybean varieties with higher protein content for animal feed and human supplementation.

In the animal feed industry, when properly processed for specific purposes, soybean grains and by-products can be used to feed all types of animals, including companion animals, domestic animals, poultry, swine, and aquatic life. Soybean by-products are rich in amino acids such as lysine, tryptophan, threonine, isoleucine, and valine, which are often lacking in cereal grains, such as corn and sorghum, both commonly used in poultry and swine feed. These amino acids are essential for ruminants and monogastric animals. However, soybeans also contain anti-nutritional factors, such as lectins, hemagglutinins, isoflavones, phytic acid, trypsin, and protease inhibitors, and this must be addressed to increase their nutritional value for the industry [29]. Soybean anti-nutritional factors can

have significant implications for the nutritional value and utilization of soybeans. The impact of anti-nutritional factors in soybean consumers can vary depending on the species, age, and physiological status of the animals, as well as the level and duration of exposure to these compounds. Further research is needed to understand the complex interactions between anti-nutritional factors in soybeans and their effects on animal and human health. Efforts should focus on developing novel techniques to effectively reduce or eliminate anti-nutritional factors while preserving the nutritional integrity of soybean-based products. In this study, six proteins were identified as anti-nutritional factors, all of which are related to the Kunitz family of trypsin and protease inhibitors (Figure 1). The effect of soybean trypsin inhibitors on ruminants and monogastric animals has been extensively studied. Most of these reports show that the nutritional value of soybeans for ruminants and monogastric animals is limited by these anti-nutritional factors, which can interfere with feed intake and nutrient metabolism [30–33].

Soybean possessing high levels of protease inhibitors, particularly trypsin inhibitors, can negatively affect protein digestibility and amino acid availability. While heat processing can inactivate these protease inhibitors, excessive heat can also destroy other proteins, nutrients, and essential amino acids [29]. In this context, the identification of key proteins from these anti-nutritional families and the use of biotechnological tools to develop cultivars with lower contents of anti-nutritional factors are desired, as these would allow for better and more efficient use of the protein intake in the diet, making them promising approaches for the industry. It is important to emphasize that the expression of anti-nutritional factors is influenced by diverse elements. One previous study evaluated soybean genotypes and the expression of five anti-nutritional factors and two Kunitz trypsin inhibitors, demonstrating that the expression of these proteins varied according to the genotype, growing location, season, and sowing time, with early cultivation showing lower levels of anti-nutritional factors [34]. However, more studies are required to explore the potential health benefits and risks associated with these factors.

Another aspect of studying the proteomic profile of soybean bran is to assess the presence of allergenic proteins. When considering the use of soybean in human food, soy is one of the main food allergens and the group of proteins to be considered comprises the inflammation-resolution enzymes, lipoxygenases (LOX) (Figure 2). While LOX plays crucial roles in the biosynthesis of bioactive compounds and defense responses in soybean plants, their activity can also lead to the development of off-flavors in soybean oil and affect the functional properties of soy proteins [35–37]. In plants, these enzymes play an important physiological role because the hydroperoxidation of linoleic acid is the first step in the biosynthesis of substances that regulate the growth, the factors involved in wound healing, and the control of chronic-disease-related inflammation [35]. Lipoxygenases have numerous potential applications in the food industry, but the use of soybean seeds as food ingredients has sometimes been limited, particularly in Western cultures, due to their "grassy/beany" flavor [36], resulting from the enzymatic oxidation of linoleic acid and linolenic acid by lipoxygenases. Some consumers also prefer a more neutral flavor in soybean products [37]. Although our results show only two lipoxygenases in the coatless seeds and defatted bran, with lower levels in the latter, these proteins are still a focus for the industry, which continuously searches for ways to reduce or completely remove the amount of these proteins in soy-derived products to improve taste and acceptance and expand the consumer market [38]. Nevertheless, in the context of human food LOX can be a target to nutraceutical industry.

Besides that, in a previous study with these biological materials, among the 139 proteins, 19 presented parts with allergenic potential, but only seven (I1LXY1; I1JWK3; I1KDM8; C6TFC1; I1KPN3; O64458; C6SWW4) remained in the defatted bran after processing [39]. These proteins possess different amino acid sequences, molecular weights, and tertiary structures, contributing to their distinctive allergenic properties.

Global efforts have been made to develop hypoallergenic soybean varieties with diminished allergenic potential. Reducing the level of these proteins could be an interesting

strategy to both reduce the allergenicity content and increase the scope of the use of the soybean cultivar BRS 537, and other genotypes in the human diet. Genetic engineering techniques, including gene silencing and protein modification, aim to decrease the expression or alter the structure of allergenic proteins. Thus, these approaches hold promise for the production of safer soy-based products. Moreover, continued research is needed to investigate the soybean allergenic proteins present in bran, as well as the impact of processing techniques on allergenicity.

4. Materials and Methods

4.1. Biological Material

The soybean cultivar BRS 537 was maintained in a greenhouse within optimal cultivation conditions until seed production. BRS 537 is an early conventional tropical cultivar with high yield potential and stability and was launched in 2020 [40]. To extract the protein, samples of mature seed without their seed coat and industrially processed defatted bran were collected in three biological repetitions. The biological materials were freshly collected and pulverized in a mortar with a pestle in liquid nitrogen before analysis.

The simulation of the industrial processing of soybeans to obtain further defatted bran was performed by the Food Science and Quality Center of ITAL (Institute of Food Technology of the Government of Sao Paulo). The process consisted of removing the tegument from mature soybean seeds, crushing the particles to homogenize them, and removing the oil using a solvent. The defatted bran was purified to remove solvent residues, and then underwent treatment with humid steam for 30 min and dry heat (60 °C) for 1 h.

To extract proteins, soybeans from three biological replicates were first ground into fine powder in a chilled mortar and pestle; about 3 g were then transferred to a 15-milliliter Falcon tube, followed by incubation with SDS-based lysis buffer (4% SDS, 50 mM DTT, 0.1% Tween 20, 100 mM Tris-HCl, pH 8.0) on a shaker for about 60 min. The samples were centrifuged at 4000× *g* for 20 min, and the supernatants were transferred to clean tubes. Acetone precipitation was performed by adding 6 × volumes of cold acetone followed by overnight incubation at −20 °C. The precipitates were collected by centrifugation at 16,000× *g* for 20 min at 4 °C and washed twice with ice-cold acetone. The resulting pellets were dissolved in 50 mM ammonium bicarbonate and subjected to overnight tryptic digestion at 37 °C. The digests were desalted using C18-based StageTips, as described previously [41], dried with a SpeedVac, and stored at −80 °C until further use.

4.2. LC–MS/MS Analysis

The LC–MS/MS analysis was performed using an Ultimate 3000 nanoLC coupled with a Q Exactive mass spectrometer (Thermo Fisher, Waltham, MA, USA) with minor modifications [42]. In brief, the peptides were resuspended in LC buffer A (0.1% formic acid in water) and loaded onto a trap column (PepMap C18, 2 cm × 100 µm, Thermo Fisher, USA) followed by separation on an in-house packed column (C18 ReproSil, 3.0 µm, 17 cm × 75 µm; Dr. Maisch, Ammerbuch, Germany). Survey scans (MS1) were acquired in a data-dependent-top-10 method with a resolution of 70,000, a scan range of 350–1700 Da, maximum injection time of 20 ms, and AGC target of 1×10^6. The MS/MS scans were obtained via higher-energy collisional dissociation (HCD) fragmentation with a resolution of 17,500, target value of 5×10^5, and maximum injection time of 100 ms.

A soybean-protein database (85,142 sequences) downloaded from UniProt Knowledge base was used for the database search. The search parameters included trypsin as the enzyme, with a maximum of two missed cleavage sites allowed; mass tolerances of 10 ppm and 20 ppm for precursor and fragments, respectively; protein N-terminal acetylation and methionine oxidation as variable modifications; cysteine-carbamide methylation as a fixed modification; peptide length of at least seven amino acids; and a false-discovery rate (FDR) of 1% for proteins.

4.3. Identification of Differentially Expressed Proteins

The output of the LC–MS/MS analysis was normalized and quantified based on Scaffold Software instructions, including protein size and sample depth as variables [43]. The differential expression in Log2 fold change was performed using Benjamin–Hochberg correction of *p*-values and FDR values using EdgeR software v.3.32.136 [44]. Only proteins with ≥2 unique peptides using peptide spectrum matches (PSMs) were retrieved [45].

The annotation of proteins was performed using Phytozome v13 software (https://phytozome-next.jgi.doe.gov/ accessed on 29 May 2022) and the Persephone genome browser (https://web.persephonesoft.com/ accessed on 29 May 2022). The annotation of the biological function of proteins was analyzed using the Wm82.a2 and Wm82.a4 reference genomes of soybean.

4.4. Systems Biology

The enrichment was performed using differentially expressed protein datasets via KEGG pathways and Phytozome databases (https://www.genome.jp/kegg/pathway.html accessed on 10 June 2022). Pathways with 3 or more representative proteins were considered enriched according to ShinyGO software (http://bioinformatics.sdstate.edu/go/ accessed on 10 June 2022). The amino acid composition was examined using ProtParam software (https://web.expasy.org/protparam/ accessed on 15 June 2022). A threshold of more than 10% essential amino acids in the protein composition was applied.

5. Conclusions

To our knowledge, this is the first work to describe and provide the global protein profiles of coatless seeds and processed soybean defatted bran from the elite tropical cultivar, BRS 537 (Embrapa, Brazil). The global proteomics analysis allowed a better understanding of the composition and function of this elite soybean. In addition, the protein panel presented here will allow researchers to develop new strategies to improve the nutritional quality of soybean products or improve other soybean genotypes. In addition, through a variety of biotechnological tools that are currently available, it is possible to remove unwanted metabolites and increase the number of those that are beneficial, adding value to the use of soybeans in human food, as well as in the animal-feed market. Moreover, these proteins could be targets for several synthetic biology strategies aiming to develop high quality protein contents, such as providing insights for the development of synthetic mimetic plant organelles for protein storage. Thus, by combining comparative proteomics with other omics tools and genome editing, researchers can identify specific genes that may be targeted for cultivar improvement.

For the first time, the soybean–bran proteome was explored and, as a result, a catalogue of proteins that tolerate industrial processes was generated. Soybean bran is often considered a waste product in soy processing. Understanding the proteomic changes that occur in soybean bran after industrial processing provides insights about the utilization strategies for this by-product. By identifying proteins that are resistant to degradation or with unique functional properties, researchers can explore new methods for using soy bran in various food formulations. Thus, understanding the protein profile in soybean bran can orient the development of sustainable and value-added industrial application strategies for soybean bran.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/plants12142704/s1. The Supplementary File S1 can be accessed at https://github.com/Rech-PBSyn/Soybean-Proteome/ accessed on 14 July 2023. The raw data are available on the Mass Spectrometry Interactive Virtual Environment-MassIVE proteomic server (MSV000086419; https://massive.ucsd.edu/ProteoSAFe/static/massive-quant.jsp/ accessed on 29 May 2022).

Author Contributions: Conceptualization, E.R., M.O.F., M.D.C.M. and R.F.-P.; methodology, M.D.C.M., R.F.-P., Y.Y., L.H.F., L.M.M.-H. and M.O.F.; validation, M.D.C.M., R.F.-P., Y.Y., L.H.F., L.M.M.-H. and M.O.F.; formal analysis, M.D.C.M., R.F.-P., Y.Y., L.H.F., L.M.M.-H., E.R. and M.O.F.; investigation, M.D.C.M., R.F.-P., Y.Y., L.H.F., L.M.M.-H., E.R. and M.O.F.; data curation, M.D.C.M., R.F.-P., Y.Y., L.H.F., L.M.M.-H., R.N.L., D.M.d.C.B., M.O.F. and E.R.; writing—original draft preparation, M.D.C.M., R.F.-P., Y.Y., L.M.M.-H. and R.N.L.; funding acquisition, E.R. and M.O.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Embrapa Genetic Resources and Biotechnology/National Institute of Science and Technology in Synthetic Biology, the National Council for Scientific and Technological Development (465603/2014-9), the Research Support Foundation of the Federal District (0193.001.262/2017), and the Coordination for the Improvement of Higher Education Personnel.

Data Availability Statement: Data are available on request from the authors.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Pathan, M.S.; Sleper, D.A. Advances in Soybean Breeding. In *Genetics and Genomics of Soybean*; Stacey, G., Ed.; Springer: New York, NY, USA, 2008; pp. 113–133.
2. Hartman, G.L.; West, E.D.; Herman, T.K. Crops that feed the World 2. Soybean—Worldwide production, use, and constraints caused by pathogens and pests. *Food Secur.* **2011**, *3*, 5–17. [CrossRef]
3. Pettersson, D.; Pontoppidan, K. Soybean Meal and The Potential for Upgrading Its Feeding Value by Enzyme Supplementation. In *Soybean*; Hany, A.E.-S., Ed.; IntechOpen: Rijeka, Italy, 2013; Chapter 13.
4. Phommalath, S.; Teraishi, M.; Yoshikawa, T.; Saito, H.; Tsukiyama, T.; Nakazaki, T.; Tanisaka, T.; Okumoto, Y. Wide genetic variation in phenolic compound content of seed coats among black soybean cultivars. *Breed. Sci.* **2014**, *64*, 409–415. [CrossRef] [PubMed]
5. Seo, J.-H.; Kim, K.-S.; Ko, J.-M.; Choi, M.-S.; Kang, B.-K.; Kwon, S.-W.; Jun, T.-H. Quantitative trait locus analysis for soybean (*Glycine max*) seed protein and oil concentrations using selected breeding populations. *Plant Breed.* **2019**, *138*, 95–104. [CrossRef]
6. Lilian Hasegawa, F.; Rayane Nunes, L.; Mayla, D.C.M. Soybean Functional Proteins and the Synthetic Biology. In *Soybean*; Takuji, O., Yoshihiko, T., Norikuni, O., Takashi, S., Sayuri, T., Eds.; IntechOpen: Rijeka, Italy, 2022; Chapter 11.
7. Sessa, D.J. Processing of soybean hulls to enhance the distribution and extraction of value-added proteins. *J. Sci. Food Agric.* **2004**, *84*, 75–82. [CrossRef]
8. Tkaczewska, J.; Zając, M.; Jamróz, E.; Derbew, H. Utilising waste from soybean processing as raw materials for the production of preparations with antioxidant properties, serving as natural food preservatives—A pilot study. *LWT* **2022**, *160*, 113282. [CrossRef]
9. Krishnan, H.B. Engineering Soybean for Enhanced Sulfur Amino Acid Content. *Crop Sci.* **2005**, *45*, 454–461. [CrossRef]
10. Basson, A.R.; Ahmed, S.; Almutairi, R.; Seo, B.; Cominelli, F. Regulation of Intestinal Inflammation by Soybean and Soy-Derived Compounds. *Foods* **2021**, *10*, 774. [CrossRef] [PubMed]
11. Erickson, D.R. Chapter 5—Overview of Modern Soybean Processing and Links Between Processes. In *Practical Handbook of Soybean Processing and Utilization*; Erickson, D.R., Ed.; AOCS Press: Urbana, IL, USA, 1995; pp. 56–64.
12. Patil, G.; Mian, R.; Vuong, T.; Pantalone, V.; Song, Q.; Chen, P.; Shannon, G.J.; Carter, T.C.; Nguyen, H.T. Molecular mapping and genomics of soybean seed protein: A review and perspective for the future. *TAG. Theor. Appl. Genet. Theor. Und Angew. Genet.* **2017**, *130*, 1975–1991. [CrossRef]
13. Patil, G.; Vuong, T.D.; Kale, S.; Valliyodan, B.; Deshmukh, R.; Zhu, C.; Wu, X.; Bai, Y.; Yungbluth, D.; Lu, F.; et al. Dissecting genomic hotspots underlying seed protein, oil, and sucrose content in an interspecific mapping population of soybean using high-density linkage mapping. *Plant Biotechnol. J.* **2018**, *16*, 1939–1953. [CrossRef]
14. Liu, S.; Liu, Z.; Hou, X.; Li, X. Genetic mapping and functional genomics of soybean seed protein. *Mol. Breed. New Strateg. Plant Improv.* **2023**, *43*, 29. [CrossRef]
15. Roell, M.S.; Zurbriggen, M.D. The impact of synthetic biology for future agriculture and nutrition. *Curr. Opin. Biotechnol.* **2020**, *61*, 102–109. [CrossRef] [PubMed]
16. Min, C.W.; Hyeon, H.; Gupta, R.; Park, J.; Cheon, Y.E.; Lee, G.H.; Jang, J.W.; Ryu, H.W.; Lee, B.W.; Park, S.U.; et al. Integrated Proteomics and Metabolomics Analysis Highlights Correlative Metabolite-Protein Networks in Soybean Seeds Subjected to Warm-Water Soaking. *J. Agric. Food Chem.* **2020**, *68*, 8057–8067. [CrossRef] [PubMed]
17. Islam, N.; Krishnan, H.B.; Natarajan, S. Proteomic Profiling of Fast Neutron-Induced Soybean Mutant Unveiled Pathways Associated with Increased Seed Protein Content. *J. Proteome Res.* **2020**, *19*, 3936–3944. [CrossRef] [PubMed]

18. Liu, W.; Xu, W.; Li, L.; Dong, M.; Wan, Y.; He, X.; Huang, K.; Jin, W. iTRAQ-based quantitative tissue proteomic analysis of differentially expressed proteins (DEPs) in non-transgenic and transgenic soybean seeds. *Sci. Rep.* **2018**, *8*, 17681. [CrossRef]
19. Kim, H.T.; Choi, U.K.; Ryu, H.S.; Lee, S.J.; Kwon, O.S. Mobilization of storage proteins in soybean seed (*Glycine max* L.) during germination and seedling growth. *Biochim. Et Biophys. Acta* **2011**, *1814*, 1178–1187. [CrossRef] [PubMed]
20. Meyer, L.J.; Gao, J.; Xu, D.; Thelen, J.J. Phosphoproteomic analysis of seed maturation in Arabidopsis, rapeseed, and soybean. *Plant Physiol.* **2012**, *159*, 517–528. [CrossRef]
21. Yu, X.; Jin, H.; Fu, X.; Yang, Q.; Yuan, F. Quantitative proteomic analyses of two soybean low phytic acid mutants to identify the genes associated with seed field emergence. *BMC Plant Biol.* **2019**, *19*, 569. [CrossRef]
22. Murad, A.M.; Rech, E.L. NanoUPLC-MSE proteomic data assessment of soybean seeds using the Uniprot database. *BMC Biotechnol.* **2012**, *12*, 82. [CrossRef]
23. Barbosa, H.S.; Arruda, S.C.; Azevedo, R.A.; Arruda, M.A. New insights on proteomics of transgenic soybean seeds: Evaluation of differential expressions of enzymes and proteins. *Anal. Bioanal. Chem.* **2012**, *402*, 299–314. [CrossRef]
24. Yin, Y.; Ren, Z.; Zhang, L.; Qin, L.; Chen, L.; Liu, L.; Jia, R.; Xue, K.; Liu, B.; Wang, X. In Situ Proteomic Analysis of Herbicide-Resistant Soybean and Hybrid Seeds via Matrix-Assisted Laser Desorption/Ionization-Mass Spectrometry Imaging. *J. Agric. Food Chem.* **2023**, *71*, 7140–7151. [CrossRef]
25. Min, C.W.; Park, J.; Bae, J.W.; Agrawal, G.K.; Rakwal, R.; Kim, Y.; Yang, P.; Kim, S.T.; Gupta, R. In-Depth Investigation of Low-Abundance Proteins in Matured and Filling Stages Seeds of Glycine max Employing a Combination of Protamine Sulfate Precipitation and TMT-Based Quantitative Proteomic Analysis. *Cells* **2020**, *9*, 1517. [CrossRef]
26. Soy Protein Ingredients Market To Reach USD 16.6 Billion By 2026, Reports and Data. Available online: https://www.globenewswire.com/news-release/2019/09/24/1920183/0/en/Soy-Protein-Ingredients-Market-To-Reach-USD-16-6-Billion-By-2026-Reports-And-Data.html (accessed on 22 May 2022).
27. Krishnan, H.B.; Jez, J.M. Review: The promise and limits for enhancing sulfur-containing amino acid content of soybean seed. *Plant Sci. Int. J. Exp. Plant Biol.* **2018**, *272*, 14–21. [CrossRef] [PubMed]
28. Shoveller, A.K.; Stoll, B.; Ball, R.O.; Burrin, D.G. Nutritional and functional importance of intestinal sulfur amino acid metabolism. *J. Nutr.* **2005**, *135*, 1609–1612. [CrossRef] [PubMed]
29. Samuel, N.N.; Agnes, K.K.-N. Advances in Soybean and Soybean By-Products in Monogastric Nutrition and Health. In *Soybean and Nutrition*; Hany, E.-S., Ed.; IntechOpen: Rijeka, Italy, 2011; Chapter 7.
30. Dilger, R.N.; Sands, J.S.; Ragland, D.; Adeola, O. Digestibility of nitrogen and amino acids in soybean meal with added soyhulls. *J. Anim. Sci.* **2004**, *82*, 715–724. [CrossRef]
31. Opapeju, F.O.; Golian, A.; Nyachoti, C.M.; Campbell, L.D. Amino acid digestibility in dry extruded-expelled soybean meal fed to pigs and poultry1. *J. Anim. Sci.* **2006**, *84*, 1130–1137. [CrossRef]
32. Ansia, I.; Stein, H.H.; Brøkner, C.; Hayes, C.A.; Drackley, J.K. Nutrient digestibility and endogenous protein losses in the foregut and small intestine of weaned dairy calves fed calf starters with conventional or enzyme-treated soybean meal. *J. Dairy Sci.* **2021**, *104*, 2979–2995. [CrossRef]
33. Kertz, A.F.; Hill, T.M.; Quigley, J.D., 3rd; Heinrichs, A.J.; Linn, J.G.; Drackley, J.K. A 100-Year Review: Calf nutrition and management. *J. Dairy Sci.* **2017**, *100*, 10151–10172. [CrossRef]
34. Maria John, K.M.; Khan, F.; Luthria, D.L.; Garrett, W.; Natarajan, S. Proteomic analysis of anti-nutritional factors (ANF's) in soybean seeds as affected by environmental and genetic factors. *Food Chem.* **2017**, *218*, 321–329. [CrossRef]
35. Freire, M.O.; Van Dyke, T.E. Natural resolution of inflammation. *Periodontology 2000* **2013**, *63*, 149–164. [CrossRef]
36. Krishna, A.; Singh, G.; Kumar, D.; Agrawal, K. Physico-chemical characteristics of some new varieties of soybean. *J. Food Sci. Technol.* **2003**, *40*, 490–492.
37. Shi, Y.; Mandal, R.; Singh, A.; Pratap Singh, A. Legume lipoxygenase: Strategies for application in food industry. *Legume Sci.* **2020**, *2*, e44. [CrossRef]
38. Yang, A.; Smyth, H.; Chaliha, M.; James, A. Sensory quality of soymilk and tofu from soybeans lacking lipoxygenases. *Food Sci. Nutr.* **2016**, *4*, 207–215. [CrossRef] [PubMed]
39. Molinari, M.D.C.; Fuganti-Pagliarini, R.; Barbosa, D.A.; Kafer, J.M.; Marin, S.R.R.; Mertz-Henning, L.M.; Nepomuceno, A.L.; Rech, E. Screening de proteínas bioativas com potencial antioxidante em proteoma de soja cv. BRS 537. In Proceedings of the IX Congresso de Soja 2022 Mercosoja, Foz do Iguaçu, Brazil, 19 May 2022.
40. Embrapa. Soja-BRS 537. Available online: https://www.embrapa.br/en/busca-de-solucoes-tecnologicas/-/produto-servico/8340/soja-brs-537 (accessed on 29 May 2022).
41. Yu, Y.; Pieper, R. Using Proteomics to Identify Inflammation During Urinary Tract Infection. *Methods Mol. Biol.* **2019**, *2021*, 259–272. [CrossRef] [PubMed]
42. Lin, Y.H.; Platt, M.P.; Fu, H.; Gui, Y.; Wang, Y.; Gonzalez-Juarbe, N.; Zhou, D.; Yu, Y. Global Proteome and Phosphoproteome Characterization of Sepsis-induced Kidney Injury. *Mol. Cell. Proteom. MCP* **2020**, *19*, 2030–2047. [CrossRef] [PubMed]
43. Searle, B.C. Scaffold: A bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics* **2010**, *10*, 1265–1269. [CrossRef]

44. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. [CrossRef]
45. Bern, M.W.; Kil, Y.J. Two-dimensional target decoy strategy for shotgun proteomics. *J. Proteome Res.* **2011**, *10*, 5296–5301. [CrossRef]