

RESEARCH ARTICLE

Shotgun proteomics of *Brassica rapa* seed proteins identifies vicilin as a major seed storage protein in the mature seed

Mahmudur Rahman¹, Qi Guo¹, Abdul Baten^{1,2}, Ramil Mauleon¹, Amina Khatun¹, Lei Liu¹, Bronwyn J. Barkla^{1*}

1 Southern Cross Plant Science, Faculty of Science and Engineering, Southern Cross University, Lismore, New South Wales, Australia, **2** Institute of Precision Medicine & Bioinformatics, Sydney Local Health District, Royal Prince Alfred Hospital, Camperdown, New South Wales, Australia

* bronwyn.barkla@scu.edu.au



OPEN ACCESS

Citation: Rahman M, Guo Q, Baten A, Mauleon R, Khatun A, Liu L, et al. (2021) Shotgun proteomics of *Brassica rapa* seed proteins identifies vicilin as a major seed storage protein in the mature seed.

PLoS ONE 16(7): e0253384. <https://doi.org/10.1371/journal.pone.0253384>

Editor: Thierry Chardot, INRA, FRANCE

Received: April 5, 2021

Accepted: June 4, 2021

Published: July 9, 2021

Copyright: © 2021 Rahman et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Raw spectral data is available at DOI: [10.25918/data.133](https://doi.org/10.25918/data.133). All other relevant data are within the manuscript and its [Supporting information files](#).

Funding: MR received an International Postgraduate Research Scholarship (IPRS) and Australian Postgraduate Award Scholarship (APA) from the Australian Government. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Proteins make up a large percentage of the Brassica seed and are second only to the oil in economic importance with uses for both animal and human nutrition. The most abundant proteins reported in the seeds of Brassica are the seed storage proteins cruciferin and napin, belonging to the 12S globulin and 2S albumin families of proteins, respectively. To gain insight into the *Brassica rapa* seed proteome and to confirm the presence and relative quantity of proteins encoded by candidate seed storage genes in the mature seed, shotgun proteomics was carried out on protein extracts from seeds of *B. rapa* inbred line R-o-18. Following liquid chromatography tandem mass spectrometry, a total of 34016 spectra were mapped to 323 proteins, where 233 proteins were identified in 3 out of 4 biological replicates by at least 2 unique peptides. 2S albumin like napin seed storage proteins (SSPs), 11/12S globulin like cruciferin SSPs and 7S globulin like vicilin SSPs were identified in the samples, along with other notable proteins including oil body proteins, namely ten oleosins and two oil body-associated proteins. The identification of vicilin like proteins in the mature *B. rapa* seed represents the first account of these proteins in the Brassicaceae and analysis indicates high conservation of sequence motifs to other 7S vicilin-like allergenic proteins as well as conservation of major allergenic epitopes in the proteins. This study enriches our existing knowledge on rapeseed seed proteins and provides a robust foundation and rational basis for plant bioengineering of seed storage proteins.

Introduction

It is estimated that up to a billion people worldwide, from both developing and developed nations, do not have an adequate intake of protein in their diet which leads to impaired growth and increased risk of disease [1, 2]. To overcome this deficiency there is a push to identify sustainable supplies of both animal and plant protein for human consumption in order to lower greenhouse gas emissions and improve the protein supply chain in agriculture.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: 1-D-SDS-PAGE, One-Dimensional Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis; ¹H NMR, Proton Nuclear Magnetic Resonance; 3D-structure, Three Dimensional Structure; BLAST, Basic Local Alignment Search Tool; BP, Biological Process; CC, Cellular Component; cDNA, Complementary Deoxyribonucleic Acid; COV, Sequence Coverage; CRISPR/Cas9, Clustered Regularly Interspaced Short Palindromic Repeats-Associated Protein-9 Nuclease; DNA, Deoxyribonucleic Acid; ESI, Electron Spray Ionization; EST, Expressed Sequence Tag; FASTA, Fast Adaptive Shrinkage Threshold Algorithm; FDR, False Discovery Rate; g, Relative Centrifugal Force, Gravity; GM, Genetically Modified; GMQE, Global Model Quality Estimation; GO, Gene Ontology; HPLC, High Performance Liquid Chromatography; IDA, Information Dependent Acquisition; IgE, Immunoglobulin E; kDa, Kilo Dalton; LC-MS/MS, Liquid Chromatography Coupled to Tandem Mass Spectrometry; MF, Molecular Function; min, Minute; mL, Milli Litre; mm, Milli Metre; MS, Mass Spectrometry, Mass Spectrometer; MSA, Multiple Sequence Alignment; NCBI, National Center for Biotechnology Information; PID, Percent Sequence Identity, Pairwise Identity; QMEAN, Qualitative Model Energy Analysis; RNA, Ribonucleic Acid; SDS-PAGE, Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis; SSP, SSPs, Seed Storage Protein, Seed Storage Proteins; TCA, Trichloroacetic Acid; TOF, Time-of-Flight; wiff, File Extension of Applied Biosystems Raw Data Files; µg, Micro Gram; µL, Micro Litre.

Rapeseed is the second-most abundantly cultivated oilseed crop globally [3–5]. After extraction of oil, about 1.2 million tons of defatted rapeseed meal is produced globally on an annual basis. This meal contains about 35–45% protein (dry weight) bringing it close to soybean for protein value [6–8]. This protein rich by-product has a nutritional content comparable to milk casein [9], egg protein [10] and animal proteins [7], with a balanced amino acid profile high in methionine and cysteine [10], providing important sulphur containing amino acids. However, the meal is primarily used to boost the nutritional value of animal or aquaculture feed or as a soil supplementation product [11, 12], rather than as a source of proteins for human consumption. This is primarily due to the presence of proteins that may cause allergenic responses as they pass through the human digestive system [13–15].

B. rapa is the highest oil-bearing species and therefore one of the economically most important species in the tribe Brassiceae belonging to the crucifer family [16]. In contrast to allopolyploid *B. napus* (AC genome), *B. juncea* (AB genome) and *B. carinata* (BC genome), *B. rapa* is a diploid species and contains only the Brassica A genome [17, 18]. In addition to oilseed crops (subsp. *Trilocularis* and *oleifera*), *B. rapa* encompasses edible leafy vegetables like bok choy or pak choy (subsp. *chinensis*), bomdong (var. *glabra*), Chinese cabbage (subsp. *pekinensis*), choy sum (subsp. *parachinensis*), komatsuna (subsp. *perviridis*), napa cabbage (subsp. *pekinensis*), rapini or broccoli rabe (var. *ruvo*) and tatosi, spoon mustard or tat choy (subsp. *narinosa*); root and tuber crops like Japanese vegetable turnip (subsp. *rapa*) [19–22]. Although some seed storage proteins have been partially characterised and allergenicity studies have been carried out on a few proteins from *B. napus* and *S. alba*, seed storage proteins and their potential allergenicity have not been well explored for *B. rapa* [23–31].

The Brassica seed storage proteins napin and cruciferin, two of the most abundant proteins in the seed, have been identified as the most likely candidates for triggering allergies, with key epitopes identified in their sequences [15, 32–35], however evidence shows other seed proteins present in other species outside the Brassica family, namely 7S globulin type vicilins, and oil-body proteins like oleosins and non-specific lipid transfer proteins may also elicit an allergenic response [36, 37].

More information is needed on the protein sequence, structure and biochemistry of *Brassica rapa* allergens. This would allow for biotechnological approaches which could manipulate the quantities of these proteins in the seed and lower the amount of allergenic proteins while not compromising the total amount of protein. This approach has recently been carried out using CRISPR/Cas9 (clustered regularly interspaced short palindromic repeat-associated protein-9 nuclease) gene editing technology in the Brassica, *Camelina sativa* whereby deletions were generated in three of the 12 genes encoding cruciferin to successfully improve seed protein composition [38].

Recent sequencing of the genomes of *Brassica rapa* inbred lines Chifu-401 v3 [39] and R-o-18 v2.2 [40], two of the most widely cultivated genotypes of rapeseed and Canola, has provided important information on the gene families for the seed storage proteins and their similarity to known allergenic proteins [12]. In this study a proteomics approach using LC-MS/MS was carried out on single seeds of *B. rapa* R-o-18 to validate the presence of predicted seed storage protein-coding genes, and to catalogue the proteins in the seed which will further understanding of their functions and identify regulatory proteins that may be targeted to alter the amount of proteins of interest including those involved in oil accumulation, disease and pest sensitivity/resistance and those considered allergenic. Using this approach, we were able to identify the presence of the seed storage protein vicilin for the first time within the Brassica family.

Materials and methods

Seed materials

Non-genetically modified seeds of *Brassica rapa* inbred line R-o-18, were collected from the Southern Cross Plant Science seed repository. In this study, four pooled populations, each consisting of seeds from 4 different plants were used.

Extraction of protein from seeds

Protein was extracted from seeds according to the method described in [41] with some modifications. In summary, seeds were placed in 2.0 mL PCR grade Safe Lock microcentrifuge tubes (Eppendorf AG, Germany, Lot: D15893R) with a single 5 mm stainless steel bead (Qiagen GmbH, D-40724, Hilden, Cat. No. 69989). Methanol (200 μ L of 100%) was added as solvent and seeds were macerated using a ball grinder/homogeniser (TissueLyser Qiagen, Retsch GmbH, Germany) by shaking in a precooled adapter for 20 cycles per second for two min with a one min rest interval. Extracts were then centrifuged at 4°C in a Sigma 4K15 Laboratory centrifuge (Sigma Laboratory Centrifuge 4K15, Germany), at 10,000g for 20 min, followed by drying in a speed vacuum dryer (Eppendorf concentrator plus 5301 Eppendorf, USA) at 30°C under vacuum.

TCA precipitation

Trichloroacetic acid (TCA) precipitation of the dried samples was carried out according to the protocol of Barkla et al. [42] with minor modification. In summary, 500 μ L MilliQ water (Advantage A 10, Millipore, USA) was added to the dried samples and vortexed lightly. This was followed by addition of 100 μ L of 10X TE (standard Tris-EDTA buffer) [43] and 100 μ L of 0.3% NaDoC (Sodium deoxycholate). Finally, 100 μ L of 72% TCA was added, and the samples were vortexed and incubated on ice for one hour. The samples were then centrifuged at 4°C for 20 min at 11000g. The supernatant was carefully aspirated, and the pellet was resuspended in 500 μ L of 90% methanol. The sample was incubated in a -20°C freezer overnight. The samples were then centrifuged at 4°C for 20 min at 11000g and the supernatant aspirated. The pellets were air dried and then stored at -80°C.

NanoHPLC, mass spectrometry and protein identification

Mass spectroscopic analysis was carried out on the protein samples at the Institute for Molecular Biosciences proteomics facility at the University of Queensland, Brisbane. The extracts were analysed by Shimadzu Prominence NanoHPLC- MS/MS on an Eksigent, Ekspert nano LC400 uHPLC (SCIEX, Canada) coupled to a Triple Time of Flight (TOF) 6600 mass spectrometer (SCIEX, Canada) equipped with a PicoView nanoflow (New Objective, USA) ion source. Five μ L of each trypsin digested extract was injected onto a 75 μ m x 150mm ChromXP C18 CL 3 μ m column (SCIEX, Canada) at 400 nL/min and a column temperature of 45°C. Mobile phase contained two solvents, solvent A and solvent B. Solvent A consisted of 0.1% formic acid in water and solvent B contained 0.1% formic acid in acetonitrile. Linear gradients of 5–30% solvent B were run over 120 min at 400 nL/minute flow rate, followed by a steeper gradient of 30% to 90% solvent B for 3 min, then 90% solvent B for 17 min, for peptide elution. The gradient was then returned to 5% solvent B for equilibration prior to the next sample injection. The ion spray voltage was set to 2600V, declustering potential (DP) 80V, curtain gas flow 25 psi, nebuliser gas 1 (GS1) 30 psi, interface heater at 150°C. The mass spectrometer acquired 100ms full scan TOF-MS data followed by up to fifty 50ms full scan product ion data in an Information Dependent Acquisition (IDA) mode. Full scan TOF-MS data was acquired

over the mass range 350–1500 m/z and for product ion MS/MS, 100–1500 m/z . Ions observed in the TOF-MS scan exceeding a threshold of 100 counts and a charge state of +2 to +5 were set to trigger the acquisition of product ion MS/MS spectra of the resultant 50 most intense ions followed by data acquisition. Raw spectral data was submitted to the Southern Cross University research portal (<https://researchportal.scu.edu.au/>) and is available under 10.25918/data.133.

Analysis of proteomic data

Spectral data from the mass spectrometer was acquired and processed using Analyst TF 1.7 software (SCIEX, Canada). A combined database was generated from the *B. rapa* R-o-18 v2.2 database annotated in house at Southern Cross Plant Science, Southern Cross University [43, 245 proteins] [40], merged with previously reported sequences publicly available in the Uniprot and NCBI protein databases including (a) 1.7/2S albumin or napin (b) 11/12S globulin or cruciferin (c) 7S globulin or vicilin (d) oleosin, and (e) putative napin sequences characterised in *B. rapa* [15], (S1 and S2 Tables in S1 File) (f) putative napin, cruciferin and vicilin sequences characterised in *B. rapa* determined by performing BLASTP analysis with the corresponding cruciferin and vicilin sequences from *A. thaliana* [15]. ProteinPilot 5.0.2 (SCIEX, Canada) was used to search spectra against the combined database, visualize fragmentation evidence of identified peptides, process qualitative and quantitative proteomics data and encode the output mzIdentML file (Fig 1) [44]. To interpret, organize, validate and visualize the information, Scaffold 4.8.7 (Proteome Science) was used [45, 46].

Gene ontology enrichment analysis

Proteins identified in at least 3 out of 4 biological replicates with at least 2 unique peptides were selected for downstream analysis. Sequences of identified proteins were compared against the UniprotKB *Arabidopsis thaliana* database using the Basic Local Alignment Search Tool (BLASTP). Uniprot IDs of matched *Arabidopsis* homologs were then submitted to DAVID v6.8 (Functional Annotation Bioinformatics Microarray Analysis database <https://david.ncifcrf.gov/>, 20.04.2020) [47] for functional annotation of identified proteins and Gene

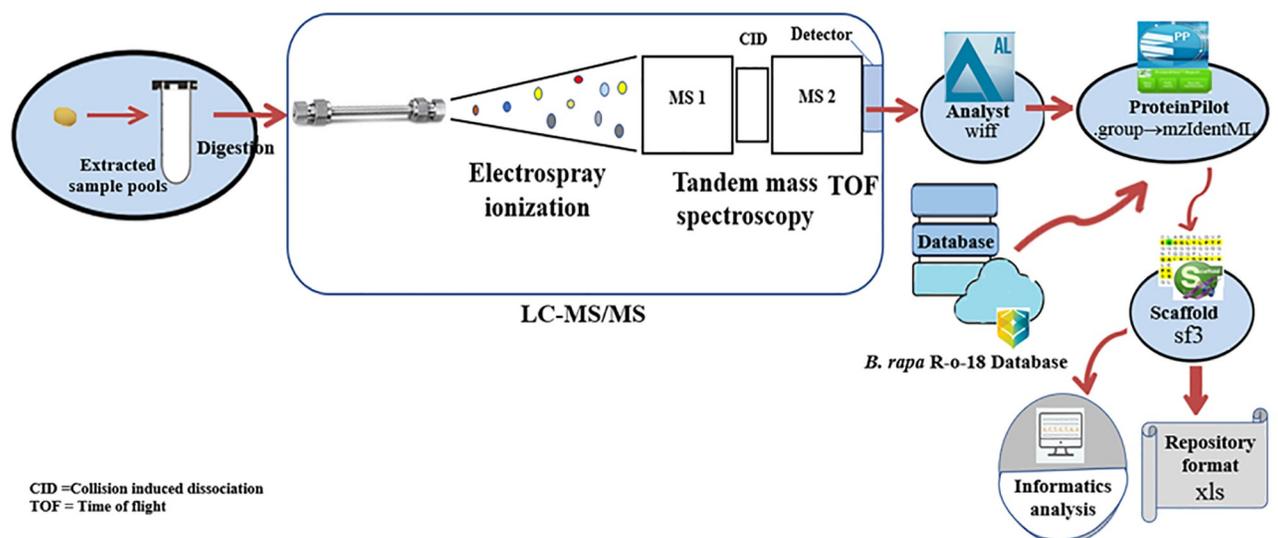


Fig 1. Overview graph of proteomics pipeline. The LC-MS/MS proteomics workflow is illustrated using ovals to represent the key steps within the workflow and the arrows connecting them.

<https://doi.org/10.1371/journal.pone.0253384.g001>

Ontology (GO) enrichment analysis using the hypergeometric method with Benjamini false discovery rate (FDR) correction.

Multiple sequence alignment and phylogenetic analysis

Multiple sequence alignment was performed using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) (S1 Table in [S1 File](#)) [48]. Aligned sequences were formatted using MView [49]. Evolutionary relationships were determined by phylogenetic analysis using the online NGPhylogeny.fr web resource (<https://ngphylogeny.fr/workflows/oneclick/>, 30/05/2020).

Sequence analysis and epitope mapping and of vicilin proteins

To obtain more insight into the allergenicity of vicilin like proteins, epitope mapping was performed according to the method described in [15, 50]. In summary, previously documented vicilin proteins from diverse species representing known allergens reported in the protein databases (S1 Table in [S1 File](#)) were mapped to the sequences identified in this study (S5 Table in [S1 File](#)), to determine conservation of IgE-binding epitopes [50–52].

Results

Mass spectrometric data was searched against a database made up of the *Brassica rapa* R-o-18 genome annotation version 2.2 comprising 43,245 protein entries [53] which was combined with putative and published sequences of seed storage and oil body proteins [15] (S1 and S2 Tables in [S1 File](#)). The data obtained were analysed with ProteinPilot 5.0.2 software (SCIEX), then visualized and validated using Scaffold 4.8.7 (Proteome Science) employing a minimum protein threshold of 99.9% and peptide threshold of 95% (S3 Table in [S1 File](#)).

A total of 34016 spectra were mapped to 323 proteins (S3 Table in [S1 File](#)). These proteins ranged in theoretical molecular weight from 7 kDa (Bra000692) to 175 kDa (Bra026064) (S4 Table in [S1 File](#)). A Venn diagram was constructed ([Fig 2](#)) to show how the proteins were distributed between the biological replicates (Pool 1 to 4). A total of 233 proteins were identified as present in at least 3 out of 4 biological replicates with at least 2 unique peptides (S4 Table in [S1 File](#)) as indicated by the numbers within the red outline ([Fig 2](#)). Of these 164 were identified in all four biological replicates as indicated by the middle square inside the red outline ([Fig 2](#)).

Gene ontology (GO) terms for the identified proteins were initially obtained by BLASTP search against protein sequences in the Arabidopsis database in NCBI. This was refined using manual curation taking into account both experimental evidence available in the literature for protein homologs, detailed sequence analysis, as well as verified information from the Uniprot database [54]. The characterised proteins were then grouped according to their gene ontology (GO) terms within the categories of cellular component and molecular function. Ten different cellular component categories were mapped to the identified proteins ([Fig 3](#)). The majority of the proteins identified were associated to the cytoplasm (50.6%), while 8.1% were associated to lipid/oil bodies and 6.4% to vacuoles. Fourteen of the identified proteins were unable to be assigned to any cellular component.

Identified proteins were then grouped according to their molecular function [55], and the percent contribution of proteins to each functional group compared in [Fig 4](#). The majority of these were directly linked to cellular metabolic processes, including carbohydrate and lipid metabolic processes, as well as cellular homeostasis. There was also a considerable grouping of proteins in the category of response to stress and translation. Proteins grouped in the nutrient reservoir category were all classified as seed storage proteins (SSP), representing 4.72% of all non-redundant proteins identified. This group was made up of 2S albumin like napins, 11/12S

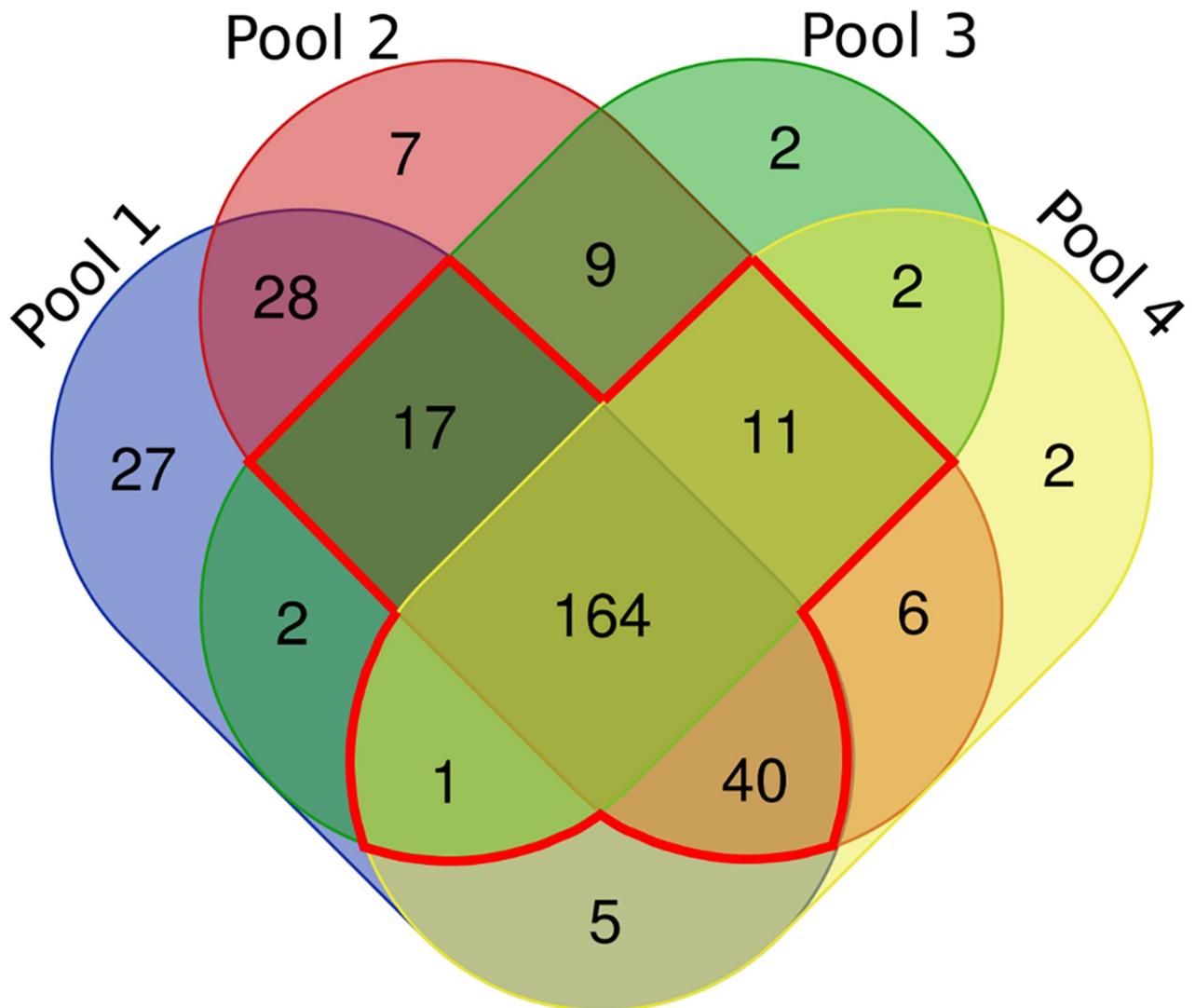


Fig 2. Venn diagram showing the distribution of unique and common non-redundant proteins among the four biological replicates. The overlapping regions show numbers demonstrating the proteins that were expressed in at least three biological replicates with 2 unique peptides. Proteins within the red outline are those present in all biological replicates (Pool 1 to 4). The identities of the expressed proteins are detailed in S4 Table in [S1 File](#).

<https://doi.org/10.1371/journal.pone.0253384.g002>

globulin-like cruciferins and 7S globulin like vicilins [56]. In addition to the napins Bra041165, BraA03000889 and BraA01001883, all previously predicted *Brassica rapa* napin gene products [15] were found to have high percent identity and coverage with the napins identified by LC-MS/MS at the protein level, and the percent identity and coverage for these are reported in S8 Table in [S1 File](#). In total, seed storage proteins made up on average 20.3% of the total spectra identified, with 14.8% of the spectra associated to cruciferins, and 2.6% associated with napins. Another 2.8% of the seed storage protein spectra were associated with 7S-globulin like vicilins, suggesting that these proteins are also a major seed storage protein in the seeds of *Brassica rapa*.

As expected for an oil seed crop, a large number of oil/lipid body proteins, namely oleosins and oil-body associated proteins as well as lipid transfer proteins were identified ([Fig 3](#) and [S5](#)

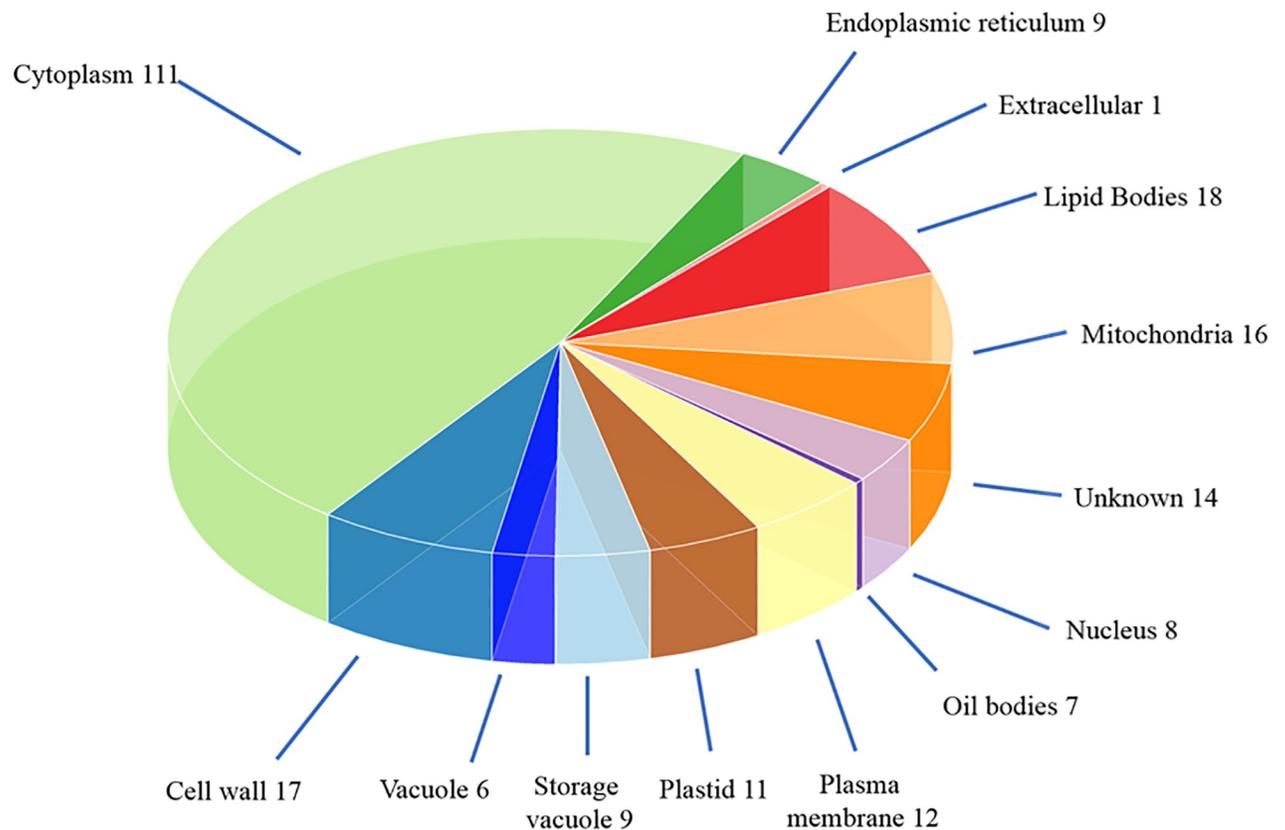


Fig 3. Distribution of proteins identified from *Brassica rapa* R-o-18 seed extract according to GO cellular component annotations. The proteins identified from *Brassica rapa* R-o-18 seed extract using LC-MS/MS were categorized based on Gene Ontology (GO) annotation as described in Material and Methods. Ten 'cellular component' categories were assigned to 219 proteins with 14 unknown proteins remaining uncategorized.

<https://doi.org/10.1371/journal.pone.0253384.g003>

Table in [S1 File](#) [57–59]. Oleosins and other oil body associated proteins accounted for 19.8 and 1.3%, respectively of the total spectra in the seed.

Gene Ontology (GO) term enrichment analysis using the DAVID functional annotation tool, was undertaken to test the over-representation of GO terms within the list of identified proteins compared to the natural abundance in the reference background Arabidopsis dataset to gain insight into their biological significance in the seed proteome. A total of 181, 189, and 157 proteins from the identified proteome were assigned to a biological process, cellular component, or molecular function GO annotation term, respectively. GO enrichment of molecular function identified 14 categories showing enrichment ([Fig 5](#)) with the most significantly enriched in our dataset being structural constituent of ribosome (p -value: $6.4E-17$), copper ion binding (p -value: $2.2E-14$), nutrient reservoir activity (p -value: 0.000000011), mRNA binding (p -value: 0.000000061) and glutathione transferase activity (p -value: 0.00019). The enriched GO terms associated with biological processes totaled 46 ([Fig 5B](#)), with those relating to stress responses showing the most significant enrichment compared to the background dataset, including response to cadmium ion (p -value: $1.10E-24$), response to salt stress (p -value: $2.1E-13$), response to heat (p -value: $3E-13$), response to cytokinin (p -value: $3.3E-10$), response to abscisic acid (p -value: $1.8E-09$); and response to cold (p -value: $5.9E-09$). Translation and oxidation-reduction processes were also enriched significantly (p -value: 0.0000013 and p -value: 0.019 , respectively). Cellular component GO terms significantly enriched totaled 32 ([Fig 5](#)), and included the most enriched related terms of cytosol (p -value: $2.2E-46$), and cytoplasm

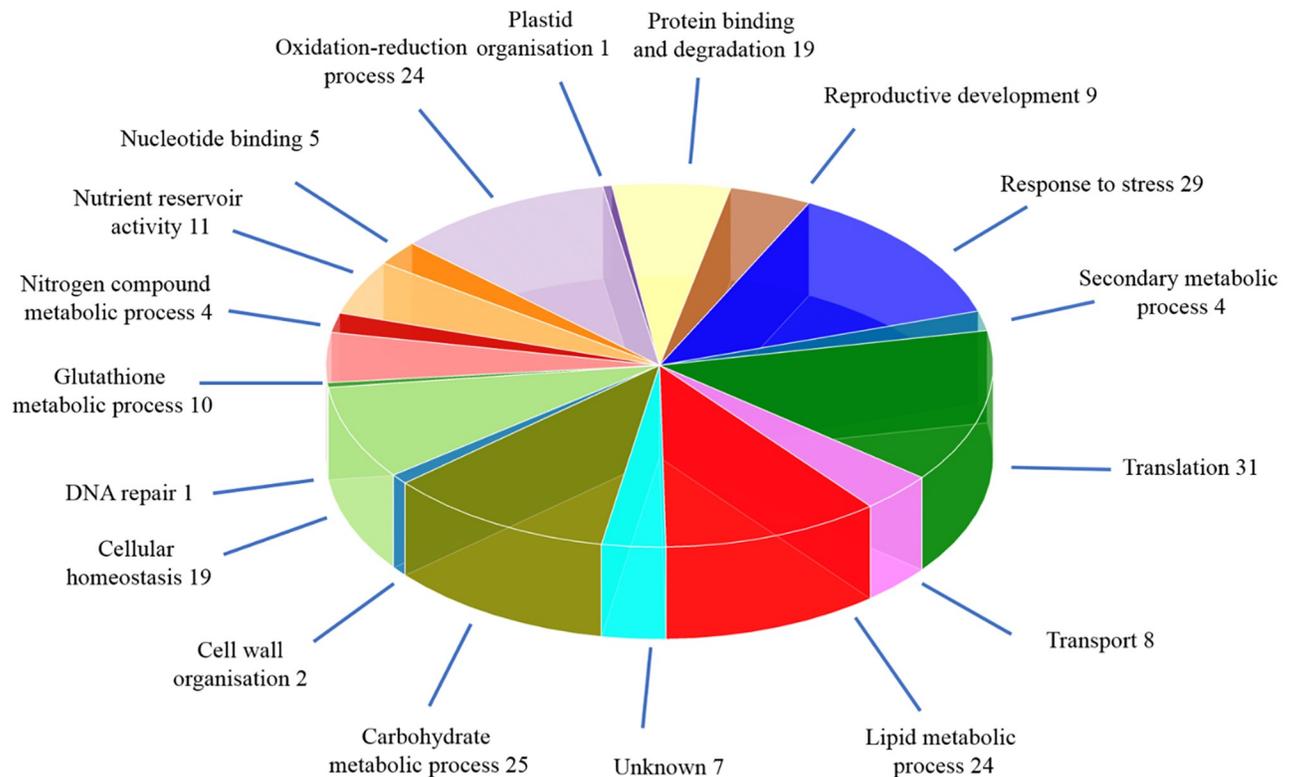


Fig 4. Distribution of proteins identified from *Brassica rapa* R-o-18 seed extract according to GO molecular function annotations. The proteins identified from *B. rapa* R-o-18 seed extract using LC-MS/MS were categorized based on Gene Ontology (GO) annotation as described in Material and Methods. Seventeen 'molecular function' categories were assigned to 226 proteins with 7 unknown proteins remaining uncategorized.

<https://doi.org/10.1371/journal.pone.0253384.g004>

(p -value: $1.5E-22$) as well as membrane and plasma membrane (p -value: $1.6E-12$ and $6.6E-10$, respectively).

The identification of five 7S globulin-like vicilin proteins provides the first experimental evidence of these proteins in Brassica (S5 Table in S1 File), fragmentation evidence for all high confidence peptides (>95%) that matched the vicilin sequences in all four pooled populations are shown in S1 Fig. Previously vicilin proteins in the *B. rapa* R-o-18 genome had only been inferred from partial complementary DNA (cDNA) sequences, RNAseq data or expressed sequence tags (ESTs), and based on sequence similarities to known legume vicilins [59, 60]. Multiple sequence alignment of the five identified *B. rapa* vicilins with vicilin and vicilin-like protein sequences from *Arabidopsis thaliana*, *Cannabis sativa*, cashew (*Anacardium occidentale*), narrow-leaf blue lupine (*Lupinus angustifolius*), white lupine (*Lupinus albus*), sesame (*Sesamum indicum*), pistachio (*Pistacia vera*), European hazel (*Corylus avellana*) and peanut (*Arachis hypogaea*) can be seen in (Fig 6). The alignment indicates significant sequence identity of the five identified *B. rapa* vicilins with other vicilin and vicilin like proteins [51, 61, 62]. To confirm this, a pairwise sequence comparison analyses for percent sequence identities (PIDs) was performed comparing the known vicilin sequences in S1 Table in S1 File with the identified 7S globulin like vicilins using CLC Genomics Workbench v3.6.5 (QIAGEN Bioinformatics). A distance matrix, expressing the nearest neighbour distance was created from a multiple sequence alignment with the sequences of similar length using the distmat server (<https://www.bioinformatics.nl/cgi-bin/emboss/distmat>, 3, 08, 2020). The pairwise comparison shows that there are high sequence similarities among the identified *B. rapa* 7S globulin

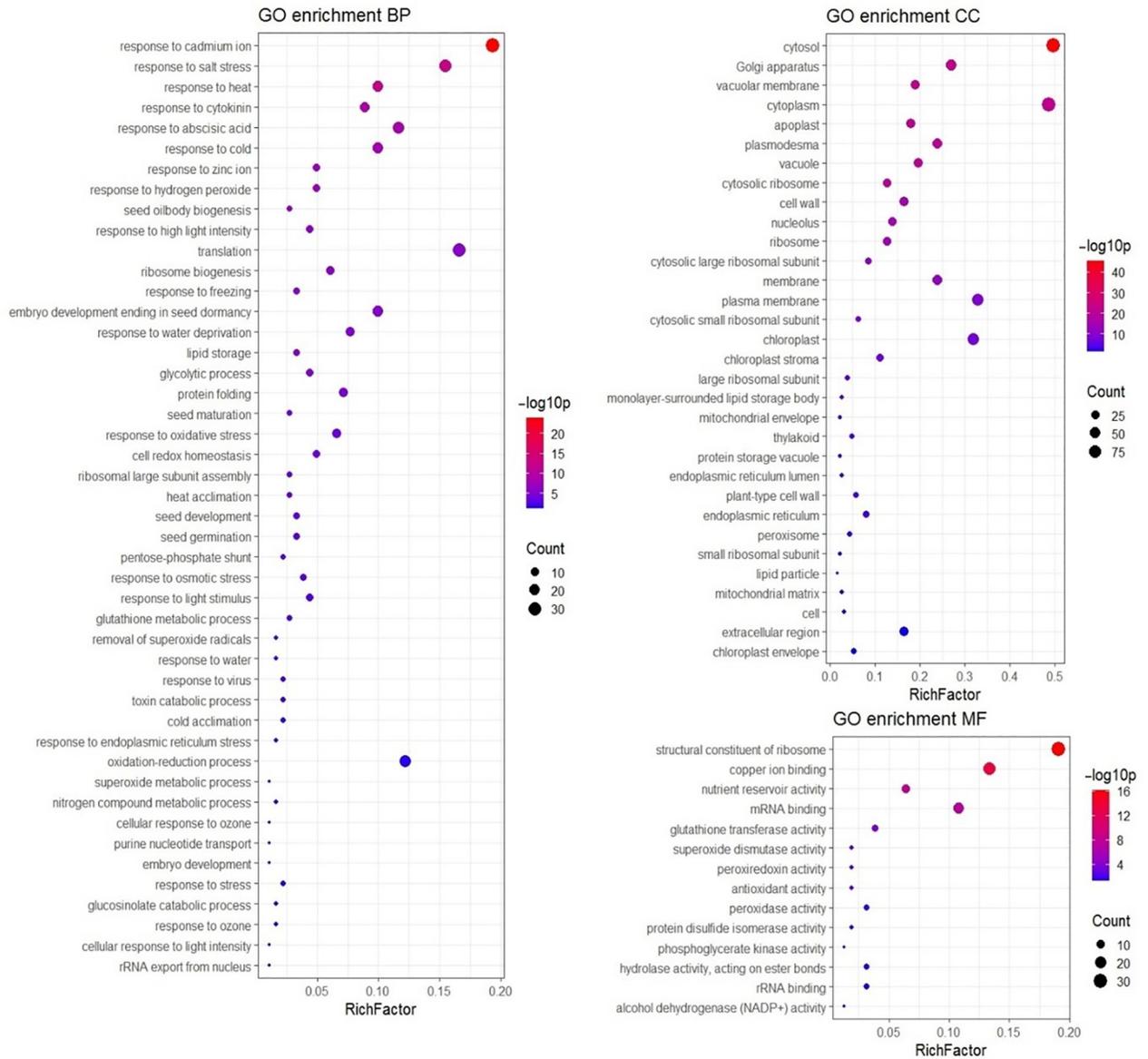


Fig 5. Gene Ontology (GO) term enrichment analysis of identified *Brassica rapa* seed proteins by DAVID. Gene Ontology (GO) term enrichment analysis was carried out using the hypergeometric method with Benjamini false discovery rate (FDR) correction. MF, CC and BP represents molecular functions, cellular components and biological processes, respectively. Rich factor is the ratio of the number of identified proteins annotated in the given GO term pathway to the number of all proteins from the database annotated in the pathway.

<https://doi.org/10.1371/journal.pone.0253384.g005>

like vicilins and previously identified vicilin sequences suggesting there are close evolutionary relationships between members of this cupin superfamily [15, 63, 64]. A distance matrix was constructed based on the score of the evolutionary distance between each pair of sequences (S2 and S3 Figs) [65, 66].

Crystallographic studies suggest that all vicilin proteins are composed of two domains, consisting of 11 β -sheets within the N-terminal region, and 3 to 4 α -helices at the C-terminus of the protein [51, 61, 67]. To confirm the presence of these domains in the *B. rapa*, vicilin proteins identified in this study, the full-length sequences of identified vicilins were analysed using SWISS-MODEL (<https://swissmodel.expasy.org/>, 20.06.20) to perform tertiary structure

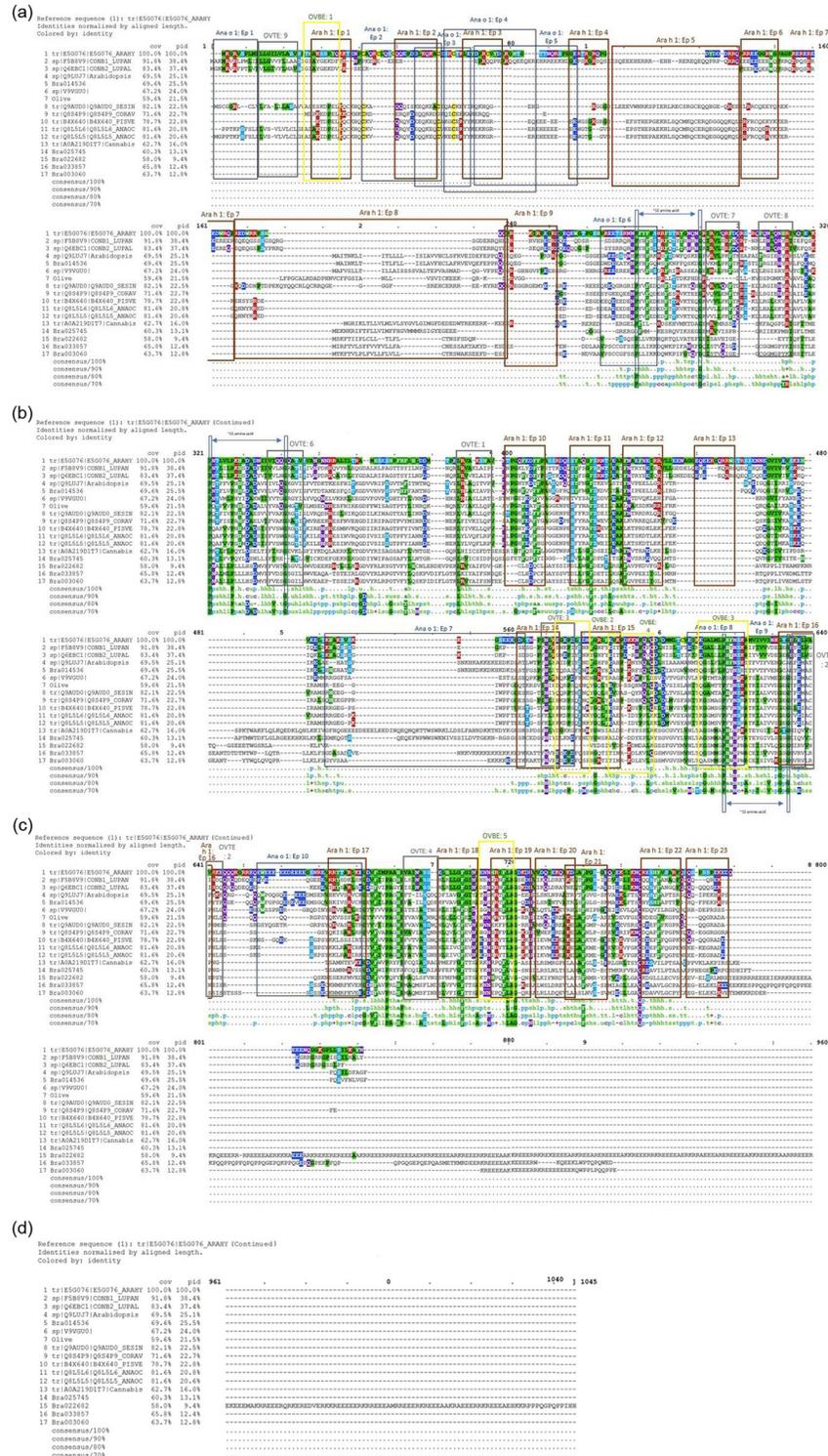


Fig 6. Multiple sequence alignment of the 7S globulin-like vicilin protein sequences identified in *Brassica rapa*. 7S globulin-like vicilin protein sequences from *B. rapa* R-0-18 seed proteome were aligned with sequences from *ArabisIdopsis thaliana*, *Cannabis sativa*, cashew (*Anacardium occidentale*), narrow-leaved blue lupine (*Lupinus angustifolius*), white lupine (*Lupinus albus*), sesame (*Sesamum indicum*), pistachio (*Pistacia vera*), European hazel (*Corylus avellana*) and peanut (*Arachis hypogaea*) available in public domain databases. Ep = epitopes. The orange boxes delineate known Ara h 1 epitopes [50], dark blue boxes outline known Ana o 1 epitopes [52], grey boxes

delineate known olive vicilin T-cell epitopes abbreviated as OVTE and yellow boxes bound the known olive vicilin B-cell epitopes abbreviated as OVBE [51]. All the epitopes are listed in Table 1 and the corresponding sequences identified in the vicilin-like proteins in *B. rapa*. pid = percent sequence identity. Blue boxes show the location of the conserved proline and glycine residues.

<https://doi.org/10.1371/journal.pone.0253384.g006>

prediction. High-resolution three-dimensional structure models of the identified vicilin proteins were drawn without ligands, as homo-trimers, based on the available crystal structures of vicilins (<https://www.rcsb.org/>) as template. Details of the 3D models with the quality scores and global model quality estimate (GMQE) and qualitative model energy analysis (QMEAN) values are presented in S6 Table in S1 File [15, 68]. The best models from the tertiary structure analysis based on GMQE and QMEAN values confirmed that the identified vicilin proteins in the *B. rapa* R-o-18 seed proteome possess this uniquely shared characteristic property of vicilins, with β -sheets in the N-terminal region and α -helices at the C-terminus (Fig 7) [51].

Within these two domains vicilin proteins show conserved glycine and proline residues located 15 amino acid residues apart [67, 69]. This motif is conserved in the *B. rapa* vicilins (Fig 6, blue boxes). Additionally, as noted in other vicilins, there was a lack of cysteine residues and a high abundance of glycine, lysine, glutamic and aspartic acid residues [70–72].

Phylogenetic analysis carried out with well characterized full length vicilin proteins [50–52] showed the evolutionary relationship of the five identified vicilin proteins from *B. rapa* to vicilins from other species (Fig 8). *B. rapa* vicilin Bra014536 was observed to cluster with the *A. thaliana* vicilin (At3g22640), whereas vicilin Bra 025745 was more closely related to the *Can-nabis sativa* 7S vicilin-like protein (Cs7S, A0A219D1T7) (Fig 8, S1 Table in S1 File). The other

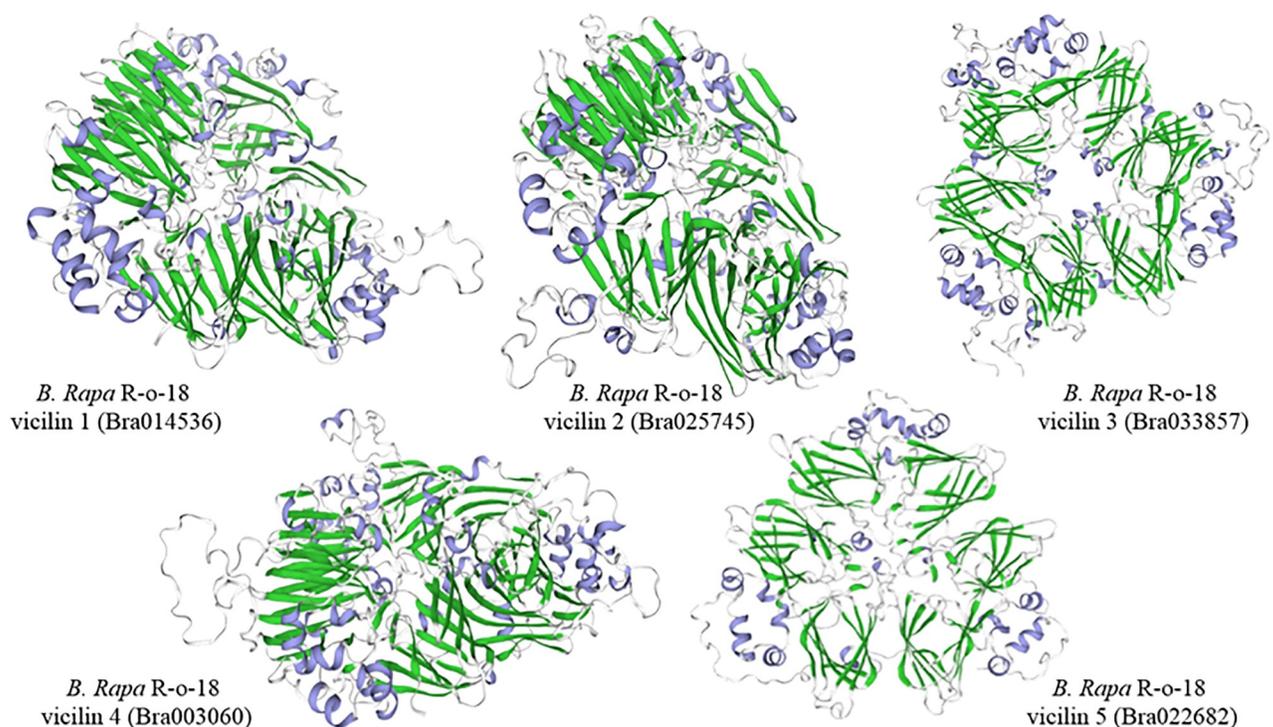


Fig 7. Three-dimensional structure of the vicilin like seed storage proteins identified in *Brassica rapa*. Structural modelling was carried out using the SWISS-MODEL program and was based on the available crystal structures of vicilin (S6 Table in S1 File). The green shading indicates the location of β -sheets in the N-terminal region and the blue shading indicates α -helices at the C-terminus.

<https://doi.org/10.1371/journal.pone.0253384.g007>

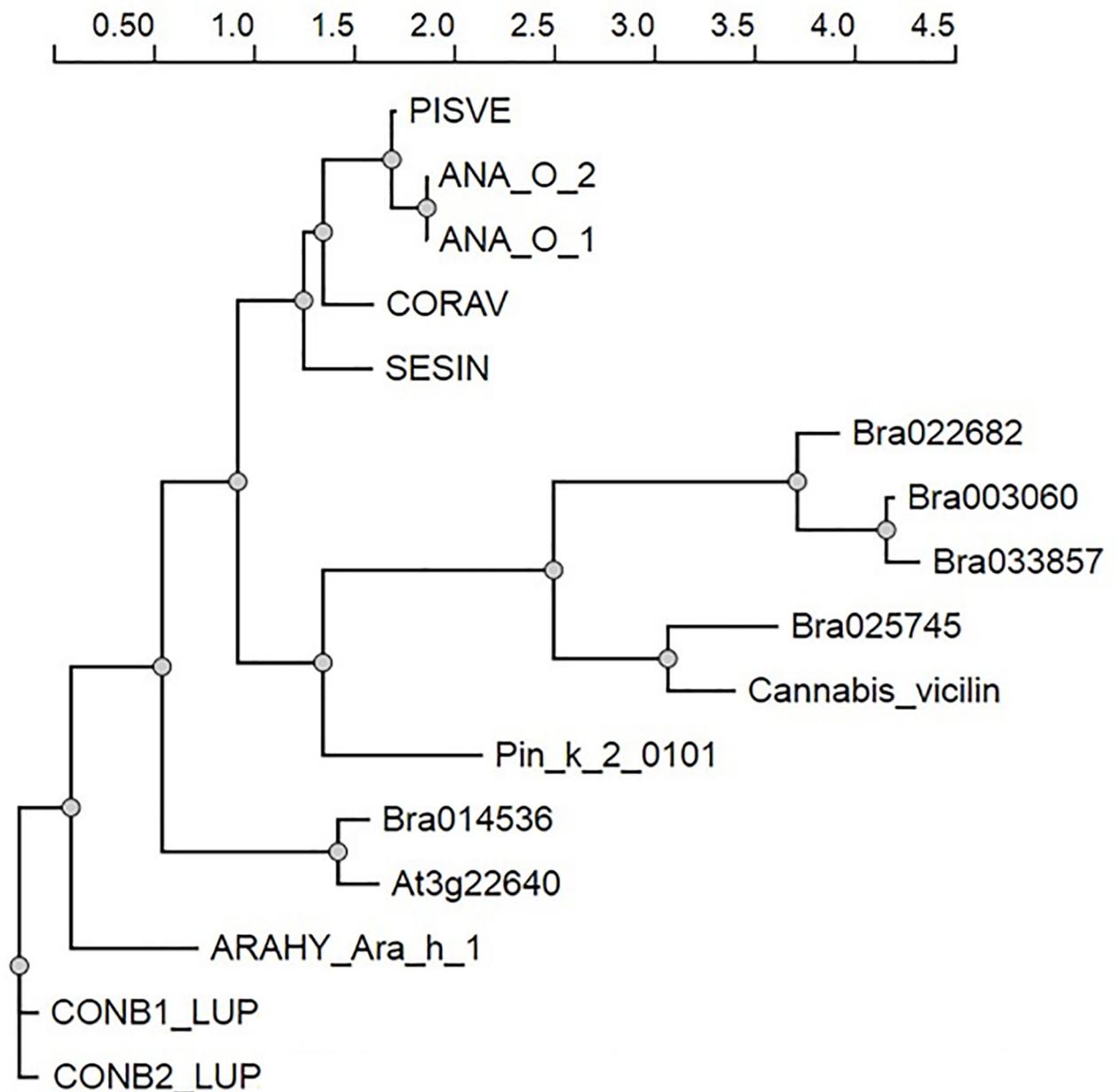


Fig 8. Phylogenetic tree of vicilin-like proteins. The phylogenetic tree was constructed using the maximum likelihood method in NGPhylogeny.fr web resource (<https://ngphylogeny.fr/workflows/oneclick/>, 30/05/2020). The tree was constructed using the five 7S globulin-like vicilin protein sequences identified in *Brassica rapa* R-o-18, as well as 7S globulin like vicilin protein sequences reported in different plant species including *Arabidopsis thaliana* (At3g22640), *Cannabis sativa* (Cannabis_vicilin), cashew (*Anacardium occidentale*) (ANA_O-1 and ANA_O_2), narrow-leaved blue lupine (*Lupinus angustifolius*) (CONB1_LUP), white lupine (*Lupinus albus*) (CONB2_LUP), sesame (*Sesamum indicum*) (SESIN), Korean pine (*Pinus koraiensis*) (Pin_k_2_0101), pistachio (*Pistacia vera*) (PISVE), European hazel (*Corylus avellana*) (CORAV) and peanut (*Arachis hypogaea*) (ARAHY_Ara_h_1), all available in public domain databases. The full-length sequences of identified vicilins are presented in S5 Table in S1 File, while the other sequences are presented in S1 Table in S1 File. The scale above the tree represents the in-built distance matrix employed in NGPhylogeny.fr web resource and indicates the distance based on the sequence similarity of their features.

<https://doi.org/10.1371/journal.pone.0253384.g008>

B. rapa vicilin proteins identified (Bra 033857, Bra 003060 and Bra022682), clustered together in a branch distinct from vicilins from other species.

Vicilins have been shown to induce allergic responses in humans [36, 50, 52, 73, 74]. To investigate potential allergenicity of the *B. rapa* vicilin proteins, epitope mapping was performed using the sequences of *B. rapa* vicilins and comparing them to the twenty-three reported epitopes of the well-studied peanut vicilin allergen Ara h 1 [50], eleven epitopes of a cashew allergen Ana o 1 [52] and T- and B- cell epitopes characterized in olive 7S vicilins corresponding to lupine Lup 1, sesame Ses i 3, pistachio Pis v 3 and European hazel Cor a 11 [51] (Table 1, Fig 6, deep blue, orange, grey and yellow boxes, respectively).

Epitope mapping to compare the known epitope sequences of Ara h 1, Ana o 1, Olive 7S vicilin T and B-cell epitopes and the linear epitopes in lupine and sesame with corresponding sequence motifs in the *B. rapa* vicilin proteins indicates that a large number (13 out of 23) of the peanut vicilin epitopes are highly conserved, while others, including cashew and olive vicilin B-cell epitopes are partially conserved based on the degree of amino acid similarity and identity (Fig 6, S5 Table in S1 File). Most of the highly conserved Ara h 1 epitopes are found in the second half of the proteins comprising epitopes 10 to 23. This corresponds to the region of the protein that contains the α -helices at the C-terminus [50, 51, 67]. Sequence analysis of the identified 7S globulin-like vicilins is presented in S9 Table in S1 File.

Discussion

A shotgun proteomic approach was carried out on extracts from *B. rapa* mature seeds to profile seed proteins with a specific focus to identify and catalogue SSP. From a total of 34016 spectra, 323 proteins were identified (Fig 2). This number is similar to protein numbers reported for shotgun seed proteomics of soybean and quinoa seeds, (243 and 352 respectively) [75, 76], although less than the number of proteins identified in seeds from the tree *Camellia oleifera* (1691 proteins) [77], or from rice and barley grains (822 and 1168 proteins, respectively) [78, 79].

Previous seed focused proteomic studies in Brassica oil-seed crops have primarily employed gel-based approaches to identify proteins showing changes in abundance through seed development or under changing conditions [31, 80–86]. While these studies have advanced understanding of specific biological processes important for the seed, they have not provided a general overview of the proteins in the seed that can be obtained by a bottom-up type approach.

In this study five cruciferin (Bra022801, Bra035434, Bra002906, BraA09006896 and P15456), and three napin proteins (Bra041165, BraA03000889 and BraA01001883) were identified in the proteome of the *B. rapa* R-o-18 seed extracts (S4 Table in S1 File). Molecular genetic analysis indicated the presence of five genes encoding napin proteins in the *B. rapa* R-o-18 genome and 10 genes encoding napin proteins in the Chiifu genome [15] (S8 Table in S1 File). Direct identification of three of these in this study provides protein-level evidence of gene expression which can help to refine gene models.

Previous studies have reported that in Brassica, 12S globulin type cruciferins and 2S albumin type napins together constitute approximately 85%–90% of the total protein in the seed [59, 87]. Cruciferin is the most abundant of these, roughly accounting for 60% of the total protein in the seed. Napin is reported as the second most abundant SSP, representing 20–30% of the total protein [88]. Generally, the ratio of cruciferin to napin, although variable, is in the range of 0.6 to 0.2 [58, 59]. In this study, based on spectral counting of identified proteins, cruciferin accounted for 14.8% of the seed proteome and napin accounted for an additional 2.6%.

Table 1. Epitope mapping of *Brassica rapa* vicilin sequences with the twenty-three reported epitopes of the well-studied peanut vicilin allergen Ara h 1.

Epitope number	A.A. sequences	A.A. position	Identified <i>B. rapa</i> vicilins				
			Bra 025745	Bra 022682	Bra 033857	Bra 003060	Bra 014536
Ara h 1 epitopes (peanut, <i>Arachis hypogaea</i>)							
1	AKSSPYQKKT	25–34	x	x	x	x	x
2	QEPDDLKQKA	48–57	x	x	x	x	x
3	LEYDPRCVYD	65–74	x	x	x	x	x
4	GERTRGRQPG	89–98	x	x	x	x	x
5	PGDYDDRRQ	97–105	x	x	x	x	x
6	PRREEGGRWG	107–116	x	x	x	x	x
7	REREEDWRQP	123–132	x	x	x	x	x
8	EDWRRPSHQQ	134–143	PC	PC	PC	PC	PC
9	QPRKIRPEGR	143–152	PC	PC	PC	PC	PC
10	TPGQFEDFFP	294–303	HC	HC	HC	HC	HC
11	SYLQEFNRNT	311–320	HC	HC	HC	HC	HC
12	FNAEFNEIRR	325–334	HC	HC	HC	HC	HC
13	EQEERGQRRW	344–353	x	x	x	x	x
14	DITNPINLRE	393–402	HC	x	HC	HC	HC
15	DNFGRLFEVK	409–418	HC	x	HC	HC	HC
16	GTGNLELVAV	461–470	HC	HC	HC	HC	HC
17	RRYTARLKEG	498–507	HC	HC	HC	HC	HC
18	ELHLLGFGIN	525–534	HC	HC	HC	HC	HC
19	HRIFLAGDKD	539–548	HC	HC	HC	HC	HC
20	VDQIEKOAKD	551–560	HC	HC	HC	HC	HC
21	KDLAFPGSGE	559–568	HC	HC	HC	HC	HC
22	RESHFVSARP	578–587	HC	PC	HC	HC	HC
23	QSPSSPEKED	597–606	HC	PC	HC	HC	HC
Ana o 1 epitopes (cashew, <i>Anacardium occidentale</i>)							
1	MGPPTKFSFSLFL	1–15	x	x	x	x	x
2	CKVQRQYDEQQKEQC	41–55	x	x	x	x	x
3	EQQKEQCVKECEKYY	49–53	x	x	x	x	x
4	KECEKYYKEKKGRER	57–71	x	x	x	x	x
5	EKKGREREHEEEEE	65–79	x	x	x	x	x
6	DEAEEEDENPYVFED	145–159	x	x	x	x	x
7	RRGEGPKIWPFTTES	337–351	x	x	x	x	x
8	NITKGGMSVPPFYNSR	393–407	PC	PC	PC	PC	PC
9	TKIAIVVSGEGCVEI	409–423	PC	PC	PC	PC	PC
10	SSHPSYKCLRIRK	433–447	x	x	x	x	x
11	EEFFQGPWEWRKEKE	521–535	PC	PC	PC	PC	PC
Olive 7S vicilin T-cell epitopes (OVTE)							
T1, T5	LVIAKLLQP	140–148	x	x	x	x	x
T2	FEMACPHLS	297–305	x	x	x	x	x
T3	INIHDQRPS	231–239	x	x	x	x	x
T4	YVAVASNNQ	336–345	x	x	x	x	x
T6	YVAQGMGTV	194–102	x	x	x	x	x
T7	VVLLPKFTQ		x	x	x	x	x
T8	DSPGMKYRV	63–71	x	x	x	x	x
T9	LVSVLVLCL (in AANOC)	13–22					
Olive 7S vicilin B-cell epitopes (OBVE)							

(Continued)

Table 1. (Continued)

Epitope number	A.A. sequences	A.A. position	Identified <i>B. rapa</i> vicilins				
			Bra 025745	Bra 022682	Bra 033857	Bra 003060	Bra 014536
B1	KHQGEHGRGGDIL	Sesame	x	x	x	x	x
B2	DQRPSQFNQ	Olive	x	x	x	x	x
B3	QGAMTPYYNSKA	Olive	PC	PC	PC	PC	PC
B4	EITPDRNPQVQ	Lupine	x	x	x	x	PC
B5	KNNKRYPLA	Olive	PC	PC	PC	PC	PC

The sequences of Ara h 1 and corresponding sequences identified as vicilin-like proteins in *B. rapa* are taken from Fig 6. X = not conserved, PC = partially conserved ($\geq 50\%$ identity in corresponding sequence alignment), HC = highly conserved ($>90\%$ identity in corresponding sequence alignment) depending on the degree of amino acid similarity and identity among the Ara h 1 and corresponding *Brassica rapa* vicilin sequences. The italicized epitopes were the most commonly recognised peptides in Ara h 1 sequence by IgE in at least 80% of the sera from sensitive individuals [50]. Peptides reported as immunodominant epitopes are shown in bold [52].

<https://doi.org/10.1371/journal.pone.0253384.t001>

This represented a napin to cruciferin ratio of 1 to 5 and agrees with the range previously reported.

Identification of vicilin SSP in the *B. rapa* seed provides evidence for an important role for this family of SSP in Brassica. Vicilins have been predominantly mentioned as SSP in legumes [70] and tree nuts [36, 52, 89], with only partial gene sequences corresponding to incomplete clones identified in Arabidopsis [59].

Previously it was assumed that cruciferins and napins were the main seed storage proteins in the Brassica seed [59]. However, based on spectral counts, vicilins appear to be close in abundance to napins in the seeds of *B. rapa*. The molecular weight of identified vicilins ranges from 52–83 kDa (S5 Table in S1 File), and analysis of seed extract protein profiles from 1D-SDS-PAGE identified a strong band within this range (S4 Fig). Western blotting with a vicilin specific antibody would help to confirm this.

The identification of vicilins in the seed extract of *Brassica rapa* is an important finding, as vicilin-like seed storage proteins are one of the major allergens in tree nuts, including almond (*Prunus dulcis*) [73, 74], cashew (*Anacardium occidentale*) [52], *Pinus koraiensis* (Pinaceae) [90, 91] and walnut (*Juglans regia*) [36]. Vicilins from pulses and legumes, including peanut (*Arachis hypogaea*) [50], pea [92], chickpea [93], and lentils [62] have also been shown to elicit strong allergenic responses in humans. Sequence analysis revealed that sequence identity of the *B. rapa* vicilins to those of other species ranged from 9.4% to 25.5%, with the highest value obtained for Bra014536 with reference to peanut vicilin Ara h 1 (Fig 6) [51]. Sequence comparisons reported among other vicilins also show quite low sequence identity, for example, cashew vicilin and the peanut vicilin have 27% sequence identity although both are predominant allergens [61]. Similarly, the identity between lentil Len c 1.0101 and sesame Ses i 3 is only 31% [62]. Despite low overall sequence identity, a number of epitopes identified in the peanut Ara h 1 vicilin protein were highly conserved in the vicilins from *B. rapa*, including epitopes 10, 11, 12 and 14–24, inclusive (Table 1) [50], with two of these, 17 and 22 corresponding to peptides in the Ara h 1 sequence recognised by IgE antibodies in at least 80% of the sera from sensitive individuals [50]. The epitope modelling shows that identified *B. rapa* R-o-18 vicilin like proteins have higher epitope conservation with the Ara h 1 epitopes than that of Ano a 1 and olive vicilin and its homologues. Further clinical testing would need to be carried out to determine the *B. rapa* R-o-18 vicilin epitopes which elicit the strongest immune response.

Phylogenetic analysis showed the evolutionary relationship among the identified *B. rapa* R-o-18 vicilins to known vicilins and vicilin like proteins (Fig 8). The *B. rapa* R-o-18 vicilins which have demonstrated higher sequence conservation of allergenic epitopes with Ara h 1

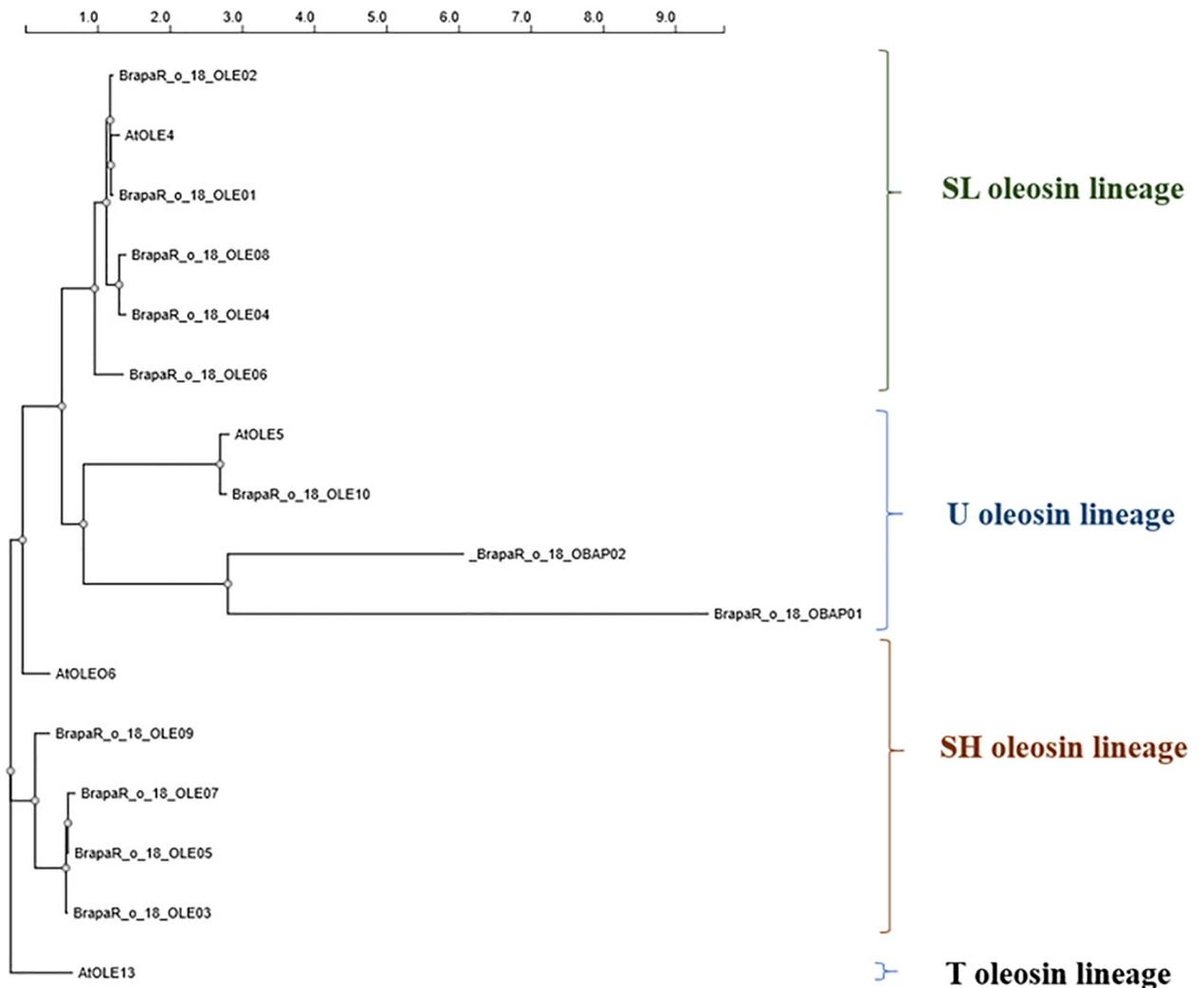


Fig 9. Phylogenetic analysis of identified *Brassica rapa* oleosin and oil body associated proteins. The linear phylogram was built using PRESTO-Phylogenetic tReE viSualisaTiOn server (<https://ngphylogeny.fr/data/displaytree/>, 29/09/2020) and shows the relationship to the Arabidopsis oleosin genes (S7 Table in [S1 File](#)). The branch length scale at the top of the figure indicates the distance between the sequences based on the sequence similarity of their features. The identified oleosin proteins Bra019493, Bra000167, A0A078GHK4, M4E9X1, A0A397KW15, A0A397ZZI8, M4DBK6, M4EI43, Bra035756, Bra032113 and oil-body associated proteins Bra036039 and Bra036039 are designated as OLE01, OLE02, OLE03, OLE04, OLE05, OLE06, OLE07, OLE08, OLE09, OLE10, OBAP 01 and OBAP 02 respectively.

<https://doi.org/10.1371/journal.pone.0253384.g009>

([Table 1](#)), were found to be rooted from the same evolutionary ancestor and distant from the Anoa 1 vicilin which shared fewer epitopes ([Fig 8](#)).

A number of proteins associated with lipid biosynthesis and oil body structure were also identified in the *B. rapa* proteome (S5 Table in [S1 File](#)) including ten oleosins and two oil body associated proteins. Previous studies have identified up to 30 oleosin genes in *B. rapa* after analysis of the draft genome sequence of inbred line Chiifu-401-42 [94], and in this study we were able to identify 12 proteins corresponding to these genes in the seeds. Identified proteins can be placed into three out of the five oleosin lineages described by Huang and Huang, 2015 ([Fig 9](#)) [95]. These described lineages include primitive (in green algae, mosses and ferns); universal (U) encompassing all land plants; seed low-molecular weight (SL) from seed plants; seed high molecular weight (SH) from angiosperms and tapetum (T) from Brassicaceae.

B. rapa oil body associated protein Bra0360391, 2 and Bra036099, designated as OBAP01 and OBAP02, respectively, and oleosin Bra032113 (designated as OLE10) originate from U lineage, *B. rapa* oleosin Bra0194931, Bra0001672, M4E9X1, A0A397ZZI8, M4EI43, (designated as OLE01, OLE02, OLE04, OLE06 and OLE08, respectively) from SL lineage, *B. rapa* oleosin A0A078GHK4, A0A397KW15, M4DBK6 and Bra035756 (designated as OLE03, OLE05, OLE07 and OLE09, respectively) from SH lineage (Fig 8, S6 Table in S1 File) [94]. No oleosin from the T lineage was detected in *B. rapa*. Oleosins have been shown to have important roles in determining oil body size, seed size and seed weight which in turn would likely influence oil production [96].

Many of the proteins identified relate to well-studied seed physiological processes. Brassicaceae crops have been reported to synthesize many unique S-containing secondary metabolites which play an important role in oxidative stress signalling and resistance to herbivores [97, 98], the most notable of these is glutathione. In this study six glutathione S-transferases were identified in the seed extracts along with proteins involved in glutathione-linked responses including but not limited to glutathione peroxidase, thioredoxin, and S-formylglutathione hydrolase as well as a sulphur transporter. Several enzymes involved in the cysteine/methionine metabolism were also identified (O-acetylserine (thiol) lyase, cysteine synthase, adenosylhomocysteinase, glutaredoxin, sulfotransferase, sulfate transporter, 5-methyltetrahydropteroyltriglutamate—homocysteine methyltransferase, glutathione S-transferase, adenylate kinase).

There was also a large number of proteins identified in the seed associated with translation (Fig 4). These proteins would likely play a key role in translation of stored mRNA's during the initial stages of seed imbibition, and evidence has shown that stored translational machinery is already effective during the very first hours upon imbibition [99]. There were also numerous proteins related to desiccation tolerance identified in the seed, including late embryogenesis related proteins (LEA), dehydrins, seed maturation protein (SMP), as well as embryonic maturation protein (EM). These proteins, as well as being important for seed desiccation, play a key role in the early stages of seed imbibition leading to germination [99]. We also identified the repair protein l-isoaspartyl methyltransferase (PIMT) which has been shown to be involved in isoAsp protein repair functions in the seed which functions to increase seed longevity and germination vigour [100, 101].

Conclusion

This study has catalogued the proteins in *B. rapa* R-o-18 seeds which can serve as a reference for future rapeseed protein analyses and provides information that can help to gain insight into the regulation, molecular mechanism and subcellular localization of the identified proteins. Major seed storage proteins were identified, including for the first time in *B. rapa*, proteins belonging to the 7S vicilin family of SSPs. Analysis of the sequences of the identified vicilin proteins provided evidence for their potential allergenicity due to high conservation of sequences that have shown to be major allergenic epitopes in vicilins of other species. This epitope identification could be used in the development of recombinant proteins for diagnosis and therapeutic purposes against rapeseed allergy [50, 102].

This study enriches our existing knowledge on rapeseed seed proteins and because of the close evolutionary relationship of *B. rapa* with other Brassicas, the knowledge obtained from this study should be useful to further understanding of the proteins of its other progenitors [103, 104], as well as construct a robust foundation and rational basis for plant bioengineering of seed storage proteins to allow improvement of seed protein which will allow the utilization of such an abundant and low cost rapeseed by-product for human consumption.

Supporting information

S1 Fig. Fragmentation evidence of high confidence peptides matching vicilin sequences in all four pooled populations.

(DOCX)

S2 Fig. Pairwise analysis of full length 7S vicilin sequences.

(DOCX)

S3 Fig. Distance matrix of the 18 full length vicilin sequences aligned in Supplementary Tables 1 and 5 in S1 File.

(DOCX)

S4 Fig. SDS-PAGE of proteins extracted from single seeds to demonstrate quality of protein used for MS analysis.

(DOCX)

S1 Raw image.

(PPTX)

S1 File. Supplementary tables.

(DOCX)

Acknowledgments

Alun Jones, University of Queensland is acknowledged for providing access to the mass spectroscopic facilities at the University of Queensland, Brisbane and technical support. The authors thank Professor Graham King, Southern Cross Plant Science, Southern Cross University for generously providing the non-GM seeds from *Brassica rapa* inbred line R-o-18. The authors also thank Dr Priyakshee Borpatra Gohain, Southern Cross Plant Science, Southern Cross University for providing information on the Brassica genomic database.

Author Contributions

Conceptualization: Mahmudur Rahman, Bronwyn J. Barkla.

Formal analysis: Mahmudur Rahman, Qi Guo, Abdul Baten, Ramil Mauleon.

Funding acquisition: Mahmudur Rahman.

Investigation: Mahmudur Rahman, Amina Khatun.

Methodology: Mahmudur Rahman, Bronwyn J. Barkla.

Project administration: Bronwyn J. Barkla.

Resources: Bronwyn J. Barkla.

Supervision: Lei Liu, Bronwyn J. Barkla.

Writing – original draft: Mahmudur Rahman.

Writing – review & editing: Qi Guo, Lei Liu, Bronwyn J. Barkla.

References

1. Kabir SMR, Rahman M., Khatun A, Saha S, Roy A, Rashid AHMA, et al. Total flavonoids content and reducing power assay of twelve common Bangladeshi leafy vegetables. *PharmacologyOnline*. 2016; 2016(3):6–14.

2. Wu G, Fanzo J, Miller DD, Pingali P, Post M, Steiner JL, et al. Production and supply of high-quality food protein for human consumption: sustainability, challenges, and innovations. *Annals of the New York Academy of Sciences*. 2014; 1321(1):1–19.
3. Basunanda P, Radoev M, Ecke W, Friedt W, Becker H, Snowdon R. Comparative mapping of quantitative trait loci involved in heterosis for seedling and yield traits in oilseed rape (*Brassica napus* L.). *Theoretical and Applied Genetics*. 2010; 120(2):271–81. <https://doi.org/10.1007/s00122-009-1133-z> PMID: 19707740
4. Joehnke MS. Effects of processing on the *in vitro* digestibility of napin proteins from rapeseed (*Brassica napus* L.): Department of Food Science, Faculty of Science, University of Copenhagen; 2017.
5. Zhu Y, Ye J, Zhan J, Zheng X, Zhang J, Shi J, et al. Validation and characterization of a seed number per silique quantitative trait locus qSN.A7 in Rapeseed (*Brassica napus* L.). *Frontiers in Plant Science*. 2020; 11:11 pages.
6. Schatzki J, Ecke W, Becker HC, Möllers C. Mapping of QTL for the seed storage proteins cruciferin and napin in a winter oilseed rape doubled haploid population and their inheritance in relation to other seed traits. *Theoretical and Applied Genetics*. 2014; 127(5):1213–22. <https://doi.org/10.1007/s00122-014-2292-0> PMID: 24595811
7. Xie H, Wang Y, Zhang J, Chen J, Wu D, Wang L. Study of the fermentation conditions and the antiproliferative activity of rapeseed peptides by bacterial and enzymatic cooperation. *International Journal of Food Science & Technology*. 2015; 50(3):619–25.
8. Linnemann AR, Dijkstra DS. Toward sustainable production of protein-rich foods: appraisal of eight crops for Western Europe. Part I. Analysis of the primary links of the production chain. *Critical Reviews in Food Science and Nutrition*. 2002; 42(4):377–401. <https://doi.org/10.1080/20024091054193> PMID: 12180778
9. Bos C, Airinei G, Mariotti F, Benamouzig R, Bérot S, Evrard J, et al. The poor digestibility of rapeseed protein is balanced by its very high metabolic utilization in humans. *The Journal of Nutrition*. 2007; 137(3):594–600. <https://doi.org/10.1093/jn/137.3.594> PMID: 17311946
10. Yoshie-Stark Y, Wada Y, Wäsche A. Chemical composition, functional properties, and bioactivities of rapeseed protein isolates. *Food Chemistry*. 2008; 107(1):32–9.
11. Akbari A, Wu J. An integrated method of isolating napin and cruciferin from defatted canola meal. *LWT-Food Science and Technology*. 2015; 64(1):308–15.
12. Rahman M, Khatun A, Liu L, Barkla BJ. Brassicaceae mustards: Traditional and agronomic uses in Australia and New Zealand. *Molecules*. 2018; 23(231):18 pages. <https://doi.org/10.3390/molecules23010231> PMID: 29361740
13. Kasprzak M, Houdijk J, Liddell S, Davis K, Olukosi O, Kightley S, et al. Rapeseed napin and cruciferin are readily digested by poultry. *Journal of Animal Physiology and Animal Nutrition*. 2016; 101(4):558–666. <https://doi.org/10.1111/jpn.12576> PMID: 27562881
14. Puumalainen TJ, Puustinen A, Poikonen S, Turjanmaa K, Palosuo T, Vaali K. Proteomic identification of allergenic seed proteins, napin and cruciferin, from cold-pressed rapeseed oils. *Food Chemistry*. 2015; 175:381–5. <https://doi.org/10.1016/j.foodchem.2014.11.084> PMID: 25577095
15. Rahman M, Baten A, Mauleon R, King GJ, Liu L, Barkla BJ. Identification, characterization and epitope mapping of proteins encoded by putative allergenic napin genes from *Brassica rapa*. *Clinical and Experimental Allergy*. 2020; 50(7):848–68. <https://doi.org/10.1111/cea.13612> PMID: 32306538
16. Gómez-Campo C. Morphology and morpho-taxonomy of the tribe Brassiceae. *Brassica crops and wild allies*[]. 1980:3–31.
17. Xue J-Y, Wang Y, Chen M, Dong S, Shao Z-Q, Liu Y. Maternal Inheritance of U's Triangle and Evolutionary Process of Brassica Mitochondrial Genomes. *Frontiers in Plant Science*. 2020; 11(805). <https://doi.org/10.3389/fpls.2020.00805> PMID: 32595682
18. Cheung F, Trick M, Drou N, Lim YP, Park J-Y, Kwon S-J, et al. Comparative analysis between homologous genome segments of *Brassica napus* and its progenitor species reveals extensive sequence-level divergence. *The Plant Cell*. 2009; 21(7):1912–28. <https://doi.org/10.1105/tpc.108.060376> PMID: 19602626
19. Lin K, Zhang N, Severing EI, Nijveen H, Cheng F, Visser RG, et al. Beyond genomic variation-comparison and functional annotation of three *Brassica rapa* genomes: a turnip, a rapid cycling and a Chinese cabbage. *BMC Genomics*. 2014; 15(1):1–17. <https://doi.org/10.1186/1471-2164-15-250> PMID: 24684742
20. Porcher MH. Multilingual multiscrypt plant name database. University of Melbourne w ww plantnames.unimelb.edu.au/Sorting/Mushrooms_Intro.html. 2005.
21. Thomas P. Canola Growers Manual (Canola Council of Canada). 2003.
22. Wikipedia. Brassica rapa: Wikipedia; 2021 [https://en.wikipedia.org/wiki/Brassica_rapa].

23. Abeysekera S. Solubility, *in vitro* digestibility and allergenicity of *Brassica juncea*, *Brassica napus* and *Sinapis alba* proteins. Saskatoon, SK, Canada,; University of Saskatchewan; 2011.
24. Fischer W, Schopfer P. Isolation and characterization of mustard (*Sinapis alba* L.) seed storage proteins. *Botanica Acta*. 1988; 101(1):48–56.
25. L'Hocine L, Pitre M, Achouri A. Detection and identification of allergens from Canadian mustard varieties of *Sinapis alba* and *Brassica juncea*. *Biomolecules*. 2019; 9(489):25 pages. <https://doi.org/10.3390/biom9090489> PMID: 31540036
26. Manchester a-f. Allergy information for: Mustard (*Brassica nigra*, *Brassica juncea*, *Brassica hirta*, *Sinapis alba*). 2006.
27. Menéndez-Arias L, Moneo I, Domínguez J, Rodríguez R. Primary structure of the major allergen of yellow mustard (*Sinapis alba* L.) seed, Sin a I. *European Journal of Biochemistry*. 1988; 177(1):159–66. PMID: 3181153
28. Ruoppolo M, Amoresano A, Pucci P, Pascarella S, Polticelli F, Trovato M, et al. Characterization of five new low-molecular-mass trypsin inhibitors from white mustard (*Sinapis alba* L.) seed. *European Journal of Biochemistry*. 2000; 267(21):6486–92. <https://doi.org/10.1046/j.1432-1327.2000.01741.x> PMID: 11029593
29. Shim Y-Y, Wanasundara JP. Quantitative detection of allergenic protein Sin a 1 from yellow mustard (*Sinapis alba* L.) seeds using enzyme-linked immunosorbent assay. *Journal of agricultural and food chemistry*. 2008; 56(4):1184–92. <https://doi.org/10.1021/jf072660u> PMID: 18217709
30. Jolivet P, Boulard C, Bellamy A, Larre C, Barre M, Rogniaux H, et al. Protein composition of oil bodies from mature *Brassica napus* seeds. *Proteomics*. 2009; 9(12):3268–84. <https://doi.org/10.1002/pmic.200800449> PMID: 19562800
31. Jolivet P, Boulard C, Bellamy A, Valot B, d'Andréa S, Zivy M, et al. Oil body proteins sequentially accumulate throughout seed development in *Brassica napus*. *Journal of Plant Physiology*. 2011; 168(17):2015–20. <https://doi.org/10.1016/j.jplph.2011.06.007> PMID: 21803444
32. Menendez-Arias L, Dominguez J, Moneo I, Rodriguez R. Epitope mapping of the major allergen from yellow mustard seeds, Sin a I. *Molecular Immunology*. 1990; 27(2):143–50. [https://doi.org/10.1016/0161-5890\(90\)90109-d](https://doi.org/10.1016/0161-5890(90)90109-d) PMID: 1690853
33. Monsalve R, de la Pena MG, Menendez-Arias L, Lopez-Otin C, Villalba M, Rodriguez R. Characterization of a new oriental-mustard (*Brassica juncea*) allergen, Bra j IE: detection of an allergenic epitope. *Biochemical Journal*. 1993; 293(3):625–32.
34. Rance F, Dutau G, Abbal M M. Mustard allergy in children. *Allergy*. 2000; 55(5):496–500. <https://doi.org/10.1034/j.1398-9995.2000.00383.x> PMID: 10843432
35. Rancé F, Micheau P, Janka D, Abbal M, Söderström L. Specific IgE level and prediction of tolerance in mustard allergy. *Journal of Allergy and Clinical Immunology*. 2004; 113(2):S146.
36. Dubiela P, Kabasser S, Smargiasso N, Geiselhart S, Bublín M, Hafner C, et al. Jug r 6 is the allergenic vicilin present in walnut responsible for IgE cross-reactivities to other tree nuts and seeds. *Scientific Reports*. 2018; 8(1):11366-. <https://doi.org/10.1038/s41598-018-29656-4> PMID: 30054513
37. Jappe U, Schwager C, Schromm AB, González Roldán N, Stein K, Heine H, et al. Lipophilic allergens, different modes of allergen-lipid interaction and their impact on asthma and allergy. *Frontiers in Immunology*. 2019; 10(122). <https://doi.org/10.3389/fimmu.2019.00122> PMID: 30837983
38. Lyzenga WJ, Harrington M, Bekkaoui D, Wigness M, Hegedus DD, Rozwadowski KL. CRISPR/Cas9 editing of three CRUCIFERIN C homoeologues alters the seed protein profile in *Camelina sativa*. *BMC plant biology*. 2019; 19(1):292-. <https://doi.org/10.1186/s12870-019-1873-0> PMID: 31272394
39. Zhang L, Cai X, Wu J, Liu M, Grob S, Cheng F, et al. Improved *Brassica rapa* reference genome by single-molecule sequencing and chromosome conformation capture technologies. *Horticulture Research*. 2018; 5(1):50. <https://doi.org/10.1038/s41438-018-0071-9> PMID: 30131865
40. Baten A, Mauleon R, King G. Re-sequencing and comparative analysis of the *Brassica rapa* ssp. *trilocularis* (R-o-18) genome. 2020 (in preparation).
41. Rahman M, Liu L, Barkla BJ. A single seed protein extraction protocol for characterizing *Brassica* seed storage proteins. *Agronomy*. 2021; 11(1):107.
42. Barkla BJ, Vera-Estrella R, Raymond C. Single-cell-type quantitative proteomic and ionic analysis of epidermal bladder cells from the halophyte model plant *Mesembryanthemum crystallinum* to identify salt-responsive proteins. *BMC Plant Biology*. 2016; 16(110):16 pages. <https://doi.org/10.1186/s12870-016-0797-1> PMID: 27160145
43. Cárdenas C, Barkla BJ, Wachter C, Delgado-Olivares L, Rodríguez-Sanoja R. Protein extraction method for the proteomic study of a Mexican traditional fermented starchy food. *Journal of Proteomics*. 2014; 111:139–47. <https://doi.org/10.1016/j.jprot.2014.06.028> PMID: 25009144

44. Deutsch EW. File formats commonly used in mass spectrometry proteomics. *Mol Cell Proteomics*. 2012; 11(12):1612–21. <https://doi.org/10.1074/mcp.R112.019695> PMID: 22956731
45. Searle BC. Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies *PROTEOMICS*. 2010; 10(6):1265–9. <https://doi.org/10.1002/pmics.200900437> PMID: 20077414
46. Gulcicek EE, Colangelo CM, McMurray W, Stone K, Williams K, Wu T, et al. Proteomics and the analysis of proteomic data: an overview of current protein-profiling technologies. *Current protocols in bioinformatics*. 2005; 10(1):13.1. 1–1. 31. <https://doi.org/10.1002/0471250953.bi1301s10> PMID: 18428746
47. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*. 2009; 4(1):44–57. <https://doi.org/10.1038/nprot.2008.211> PMID: 19131956
48. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*. 2019; 47(W1):W636–W41. <https://doi.org/10.1093/nar/gkz268> PMID: 30976793
49. Brown NP, Leroy C, Sander C. MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics (Oxford, England)*. 1998; 14(4):380–1. <https://doi.org/10.1093/bioinformatics/14.4.380> PMID: 9632837
50. Burks AW, Shin D, Cockrell G, Stanley JS, Helm RM, Bannon GA. Mapping and mutational analysis of the IgE-binding epitopes on Ara h 1, a legume vicilin protein and a major allergen in peanut hypersensitivity. *European journal of biochemistry*. 1997; 245(2):334–9. <https://doi.org/10.1111/j.1432-1033.1997.t01-1-00334.x> PMID: 9151961
51. Jimenez-Lopez JC, Zafra A, Palanco L, Florido JF, Alché JdD. Identification and assessment of the potential allergenicity of 7S vicilins in olive (*Olea europaea* L.) seeds. *BioMed Research International*. 2016; 2016:4946872-. <https://doi.org/10.1155/2016/4946872> PMID: 27034939
52. Wang F, Robotham JM, Teuber SS, Tawde P, Sathe SK, Roux KH. Ana o 1, a cashew (*Anacardium occidentale*) allergen of the vicilin seed storage protein family. *Journal of Allergy and Clinical Immunology*. 2002; 110(1):160–6.
53. Baten A, Mauleon R, King G. Brassica rapa R-o-18 genome annotation version 2.2 (unpublished, the sequences are available on request). 2020 (unpublished).
54. Garibay-Hernández A, Barkla BJ, Vera-Estrella R, Martinez A, Pantoja O. Membrane proteomic insights into the physiology and taxonomy of an oleaginous green microalga. *Plant physiology*. 2017; 173(1):390–416. <https://doi.org/10.1104/pp.16.01240> PMID: 27837088
55. Murad AM, Rech EL. NanoUPLC-MS(E) proteomic data assessment of soybean seeds using the UniProt database. *BMC Biotechnology*. 2012; 12(82):17 pages.
56. Mohottalalage UPS. A comparative study of the structural and physicochemical properties of the major proteins from Camelina sativa (L.) Crantz and Brassica napus L: University of Saskatchewan; 2016.
57. Tzen J. Integral proteins in plant oil bodies. *ISRN Botany*. 2012; 16:16 pages.
58. Wanasundara JP, McIntosh TC, Perera SP, Withana-Gamage TS, Mitra P. Canola/rapeseed protein-functionality and nutrition. *OCL*. 2016; 23(4):16 pages.
59. Wanasundara JPD. Proteins of Brassicaceae oilseeds and their potential as a plant protein source. *Critical Reviews in Food Science and Nutrition*. 2011; 51(7):635–77. <https://doi.org/10.1080/10408391003749942> PMID: 21793726
60. Spencer D. The physiological role of storage proteins in seeds. *Philosophical Transactions of the Royal Society of London B, Biological Sciences*. 1984; 304(1120):275–85.
61. Astwood JD, Silvanovich A, Bannon GA. Vicilins: A case study in allergen pedigrees. *Journal of allergy and clinical immunology*. 2002; 110(1):26–7. <https://doi.org/10.1067/mai.2002.125690> PMID: 12110813
62. López-Torrejón G, Salcedo G, Martín-Esteban M, Díaz-Perales A, Pascual CY, Sánchez-Monge R. Len c 1, a major allergen and vicilin from lentil seeds: Protein isolation and cDNA cloning. *Journal of Allergy and Clinical Immunology*. 2003; 112(6):1208–15.
63. Mills EC, Jenkins JA, Alcocer MJ, Shewry PR. Structural, biological, and evolutionary relationships of plant food allergens sensitizing via the gastrointestinal tract. *Critical reviews in food science and nutrition*. 2004; 44(5):379–407. <https://doi.org/10.1080/10408690490489224> PMID: 15540651
64. Mills ENC, Johnson PE, Alexeev Y. Chapter 2—Food Antigens. In: James JM, Burks W, Eigenmann P, editors. *Food Allergy*. Edinburgh: W.B. Saunders; 2012. p. 15–32.
65. Taylor WR, Jones DT. Deriving an Amino Acid Distance Matrix. *Journal of Theoretical Biology*. 1993; 164(1):65–83. <https://doi.org/10.1006/jtbi.1993.1140> PMID: 8264244

66. Price MN, Dehal PS, Arkin AP. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology and Evolution*. 2009; 26(7):1641–50. <https://doi.org/10.1093/molbev/msp077> PMID: 19377059
67. Gueguen J, Popineau Y. *Plant proteins from European crops*: Springer; 1998.
68. Mattison CP, Grimm CC, Li Y, Chial HJ, McCaslin DR, Chung S-Y, et al. Identification and Characterization of Ana o 3 Modifications on Arginine-111 Residue in Heated Cashew Nuts. *Journal of Agricultural and Food Chemistry*. 2017; 65(2):411–20. <https://doi.org/10.1021/acs.jafc.6b04134> PMID: 27966914
69. Lawrence MC, Izard T, Beuchat M, Blagrove RJ, Colman PM. Structure of phaseolin at 2.2 Å resolution: implications for a common vicilin/legumin structure and the genetic engineering of seed storage proteins. *Journal of molecular biology*. 1994; 238(5):748–76. <https://doi.org/10.1006/jmbi.1994.1333> PMID: 8182747
70. Argos P, Narayana SV, Nielsen NC. Structural similarity between legumin and vicilin storage proteins from legumes. *EMBO J*. 1985; 4(5):1111–7. PMID: 4006908
71. Gane PJ, Dunwell JM, Warwick J. Modeling based on the structure of vicilins predicts a histidine cluster in the active site of oxalate oxidase. *Journal of molecular evolution*. 1998; 46(4):488–93. <https://doi.org/10.1007/pl00006329> PMID: 9541544
72. Shewry PR, Napier JA, Tatham AS. Seed storage proteins: structures and biosynthesis. *The plant cell*. 1995; 7(7):945. <https://doi.org/10.1105/tpc.7.7.945> PMID: 7640527
73. Che H, Zhang Y, Jiang S, Jin T, Lyu S-C, Nadeau KC, et al. Almond (*Prunus dulcis*) Allergen Pru du 8, the First Member of a New Family of Food Allergens. *Journal of Agricultural and Food Chemistry*. 2019a; 67(31):8626–31.
74. Che H, Zhang Y, Lyu S-C, Nadeau KC, McHugh T. Identification of Almond (*Prunus dulcis*) Vicilin As a Food Allergen. *Journal of Agricultural and Food Chemistry*. 2019b; 67(1):425–32.
75. Capriotti AL, Caruso G, Cavaliere C, Samperi R, Stampachiachiere S, Zenezini Chiozzi R, et al. Protein profile of mature soybean seeds and prepared soybean milk. *Journal of agricultural and food chemistry*. 2014; 62(40):9893–9. <https://doi.org/10.1021/jf5034152> PMID: 25229310
76. Capriotti AL, Cavaliere C, Piovesana S, Stampachiachiere S, Ventura S, Zenezini Chiozzi R, et al. Characterization of quinoa seed proteome combining different protein precipitation techniques: Improvement of knowledge of nonmodel plant proteomics. *Journal of Separation Science*. 2015; 38(6):1017–25. <https://doi.org/10.1002/jssc.201401319> PMID: 25580831
77. Ye Z, Wu Y, Ul Haq Muhammad Z, Yan W, Yu J, Zhang J, et al. Complementary transcriptome and proteome profiling in the mature seeds of *Camellia oleifera* from Hainan Island. *PLoS One*. 2020; 15(2):e0226888. <https://doi.org/10.1371/journal.pone.0226888> PMID: 32027663
78. Koller A, Washburn MP, Lange BM, Andon NL, Deciu C, Haynes PA, et al. Proteomic survey of metabolic pathways in rice. *Proceedings of the National Academy of Sciences*. 2002; 99(18):11969–74. <https://doi.org/10.1073/pnas.172183199> PMID: 12163647
79. Mahalingam R. Shotgun proteomics of the barley seed proteome. *BMC genomics*. 2017; 18(1):44. <https://doi.org/10.1186/s12864-016-3408-5> PMID: 28061743
80. Hajduch M, Casteel JE, Hurrelmeyer KE, Song Z, Agrawal GK, Thelen JJ. Proteomic Analysis of Seed Filling in *Brassica napus*. Developmental Characterization of Metabolic Isozymes Using High-Resolution Two-Dimensional Gel Electrophoresis. *Plant Physiology*. 2006; 141(1):32–46. <https://doi.org/10.1104/pp.105.075390> PMID: 16543413
81. Garg H, Li H, Sivasithamparam K, Barbetti MJ. Differentially expressed proteins and associated histological and disease progression changes in cotyledon tissue of a resistant and susceptible genotype of *Brassica napus* infected with *Sclerotinia sclerotiorum*. *PLoS One*. 2013; 8(6):e65205. <https://doi.org/10.1371/journal.pone.0065205> PMID: 23776450
82. Kotecka-Majchrzak K, Sumara A, Fornal E, Montowska M. Proteomic analysis of oilseed cake: a comparative study of species-specific proteins and peptides extracted from 10 seed species. *Journal of the Science of Food and Agriculture*. 2020.
83. Kubala S, Garnczarska M, Wojtyła Ł, Clippe A, Kosmala A, Żmieńko A, et al. Deciphering priming-induced improvement of rapeseed (*Brassica napus* L.) germination through an integrated transcriptomic and proteomic approach. *Plant Science*. 2015; 231:94–113. <https://doi.org/10.1016/j.plantsci.2014.11.008> PMID: 25575995
84. Lorenz C, Rolletschek H, Sunderhaus S, Braun H-P. Brassica napus seed endosperm—metabolism and signaling in a dead end tissue. *Journal of Proteomics*. 2014; 108:382–426. <https://doi.org/10.1016/j.jprot.2014.05.024> PMID: 24906024
85. Yin X, He D, Gupta R, Yang P. Physiological and proteomic analyses on artificially aged *Brassica napus* seed. *Frontiers in plant science*. 2015; 6:112. <https://doi.org/10.3389/fpls.2015.00112> PMID: 25763006

86. Li W, Gao Y, Xu H, Zhang Y, Wang J. A proteomic analysis of seed development in *Brassica campestris* L. *PLoS One*. 2012; 7(11):e50290. <https://doi.org/10.1371/journal.pone.0050290> PMID: 23189193
87. Campbell L, Rempel CB, Wanasundara JP. Canola/rapeseed protein: future opportunities and directions—Workshop proceedings of IRC 2015. Multidisciplinary Digital Publishing Institute; 2016. <https://doi.org/10.3390/plants5020017> PMID: 27135237
88. DeLisle AJ, Crouch ML. Seed storage protein transcription and mRNA levels in *Brassica napus* during development and in response to exogenous abscisic acid. *Plant physiology*. 1989; 91(2):617–23. <https://doi.org/10.1104/pp.91.2.617> PMID: 16667077
89. Wallowitz M, Peterson WR, Uratsu S, Comstock SS, Dandekar AM, Teuber SS. Jug r 4, a legumin group food allergen from walnut (*Juglans regia* Cv. Chandler). *Journal of Agricultural and Food Chemistry*. 2006; 54(21):8369–75. <https://doi.org/10.1021/jf061329s> PMID: 17032053
90. Jin T, Albillos SM, Chen YW, Kothary MH, Fu TJ, Zhang YZ. Purification and characterization of the 7S vicilin from Korean pine (*Pinus koraiensis*). *Journal of Agricultural and Food Chemistry*. 2008; 56(17):8159–65. <https://doi.org/10.1021/jf801138q> PMID: 18690685
91. Jin T, Fu TJ, Kothary MH, Howard A, Zhang YZ. Crystallization and initial crystallographic characterization of a vicilin-type seed storage protein from *Pinus koraiensis*. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2007; 63(Pt 12):1041–3. <https://doi.org/10.1107/S1744309107054310> PMID: 18084088
92. Sanchez-Monge R, Lopez-Torrejón G, Pascual C, Varela J, Martin-Esteban M, Salcedo G. Vicilin and convicilin are potential major allergens from pea. *Clinical & Experimental Allergy*. 2004; 34(11):1747–53. <https://doi.org/10.1111/j.1365-2222.2004.02085.x> PMID: 15544600
93. Bar-El Dadon S, Pascual CY, Eshel D, Teper-Bamnlker P, Paloma Ibáñez MD, Reifern R. Vicilin and the basic subunit of legumin are putative chickpea allergens. *Food Chemistry*. 2013; 138(1):13–8. <https://doi.org/10.1016/j.foodchem.2012.10.031> PMID: 23265449
94. Chen K, Yin Y, Liu S, Guo Z, Zhang K, Liang Y, et al. Genome-wide identification and functional analysis of oleosin genes in *Brassica napus* L. *BMC Plant Biology*. 2019; 19(1):294. <https://doi.org/10.1186/s12870-019-1891-y> PMID: 31272381
95. Huang M-D, Huang AHC. Bioinformatics Reveal Five Lineages of Oleosins and the Mechanism of Lineage Evolution Related to Structure/Function from Green Algae to Seed Plants. *Plant physiology*. 2015; 169(1):453–70. <https://doi.org/10.1104/pp.15.00634> PMID: 26232488
96. Siloto RMP, Findlay K, Lopez-Villalobos A, Yeung EC, Nykiforuk CL, Moloney MM. The accumulation of oleosins determines the size of seed oilbodies in *Arabidopsis*. *The Plant Cell*. 2006; 18(8):1961–74. <https://doi.org/10.1105/tpc.106.041269> PMID: 16877495
97. Czerniawski P, Bednarek P. Glutathione S-Transferases in the Biosynthesis of Sulfur-Containing Secondary Metabolites in Brassicaceae Plants. *Frontiers in Plant Science*. 2018; 9(1639). <https://doi.org/10.3389/fpls.2018.01639> PMID: 30483292
98. Noctor G, Mhamdi A, Chaouch S, Han Y, Neukermans J, Marquez-Garcia B, et al. Glutathione in plants: an integrated overview. *Plant, Cell & Environment*. 2012; 35(2):454–84.
99. Galland M, Huguet R, Arc E, Cueff G, Job D, Rajjou L. Dynamic proteomics emphasizes the importance of selective mRNA translation and protein turnover during *Arabidopsis* seed germination. *Molecular & Cellular Proteomics*. 2014; 13(1):252–68. <https://doi.org/10.1074/mcp.M113.032227> PMID: 24198433
100. Ogé L, Bourdais G, Bove J, Collet B, Godin B, Granier F, et al. Protein repair L-isoaspartyl methyltransferase 1 is involved in both seed longevity and germination vigor in *Arabidopsis*. *The Plant Cell*. 2008; 20(11):3022–37. <https://doi.org/10.1105/tpc.108.058479> PMID: 19011119
101. Wei Y, Xu H, Diao L, Zhu Y, Xie H, Cai Q, et al. Protein repair L-isoaspartyl methyltransferase 1 (PIMT1) in rice improves seed longevity by preserving embryo vigor and viability. *Plant molecular biology*. 2015; 89(4):475–92. <https://doi.org/10.1007/s11103-015-0383-1> PMID: 26438231
102. Pantoja-Uceda D, Bruix M, Giménez-Gallego G, Rico M, Santoro J. Solution Structure of RicC3, a 2S Albumin Storage Protein from *Ricinus communis*. *Biochemistry*. 2003; 42(47):13839–47. <https://doi.org/10.1021/bi0352217> PMID: 14636051
103. Kim C-K, Seol Y-J, Perumal S, Lee J, Waminal NE, Jayakodi M, et al. Re-exploration of U's Triangle Brassica Species Based on Chloroplast Genomes and 45S nrDNA Sequences. *Scientific Reports*. 2018; 8(1):7353. <https://doi.org/10.1038/s41598-018-25585-4> PMID: 29743507
104. Ostergaard L, King GJ. Standardized gene nomenclature for the Brassica genus. *Plant Methods*. 2008; 4:10. <https://doi.org/10.1186/1746-4811-4-10> PMID: 18492252