Cloning, expression and characterization of arcelin and its impact on digestive enzymes of the stored product insect pest, Callosobruchus maculatus (F.)

Karuppiah Hilda  
University of Madras

Sreeramulu Bhuvaragavan  
University of Madras

Ramakrishnan Kamatchi  
University of Madras

Mani Meenakumari  
University of Madras

Sundaram Janarthanan  
University of Madras  janas_09@yahoo.co.in

Research Article

Keywords: Phaseolus lunatus, arcelin, cloning, insect pest management, digestive enzymes, RT-PCR

DOI: https://doi.org/10.21203/rs.3.rs-664734/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Key message The defensive role of arcelin from *Phaseolus lunatus* for its protection against insect pest *Callosobruchus maculatus* was elucidated. The potent insecticidal property of arcelin and its impact on pest molecular physiology has identified the adaptive strategies.

Abstract The pulse beetle *Callosobruchus maculatus* causes potential damage to legume crops by infesting the seeds and causes a reduction in the content of total protein. Arcelin found in the wild accessions of the common bean, is an insecticidal protein that has the potency to hamper the metabolism of the bruchid beetle. The arcelin gene from the wild accession of *Phaseolus lunatus* was isolated and the ORF encoding 158 amino acids was cloned in pET-45b (+) vector. The recombinant was transformed into BL21 STAR (DE3) pLysS cells, and the expressed arcelin was purified using Ni-NTA column. The recombinant protein was used in preparing artificial diet and the insecticidal activity was elucidated against the bruchid pest *C. maculatus*. Adult emergence and seed damage were drastically reduced in the treated groups. The response towards ingested diet by digested enzymes was elucidated through quantitative gene expression. Maximum expression was observed in the aminopeptidase, followed by upregulation of alpha-amylase, glycoside hydrolase family 31 and cathepsin D-like aspartic protease, and downregulation of cathepsin L-like cysteine protease. These results showed the anti-metabolic nature of the recombinant arcelin. The changes in digestive enzymes to counteract the anti-nutritional nature of the protein were the strategies of the insect defense mechanism.

Introduction

Plants have evolved various defense strategies against insect herbivory based on physical barriers, constitutive chemical defenses, and direct and indirect inducible defenses (Berenbaum 1995). Legume plants adapt numerous defense mechanisms by developing distinct pod and seed characteristics that either act as a morphological barrier or possess an inherent mechanism of production of secondary metabolites and anti-nutritional compounds that cause anti-metabolic activity (antibiosis) against bruchid beetles which results in the death of the pest (Osborn et al. 1988; Shaheen et al. 2006). Anti-metabolic seed proteins include lectins, arcelins, vicilin, phaseolin, phytohemagglutinins (PHAs), trypsin inhibitors, and cyanogenic glycosides, accumulate in the cotyledons of seeds of wild varieties of pulses during the seed maturation (Shaheen et al. 2006; Lambrides and Godwin 2007). The exploitation of the applications of defense proteins has gained importance in the recent past and has substituted the use of synthetic insecticides. The utilization of such compounds in crop rotation by conventional plant breeding or genetic engineering has been practiced (Gatehouse et al. 1990). Several pieces of evidence are available for vegetal lectins’ defensive role in protecting plants against insect pests (Gatehouse et al. 1995; Jaber et al. 2010; Vandenborre et al. 2011). Apart from true lectins, lectin-related polypeptides (arcelin/phytohemagglutinin/α-amylase inhibitors) possess insecticidal activity against bruchid pests. Among these lectin-related genes, arcelin is an effective insecticidal protein resistant to bruchid beetles and is found in some wild accessions of the common bean, *Phaseolus vulgaris*. The presence of arcelin at high concentrations (30–50% w/w), seems to be correlated with low amounts of the storage protein in
those seeds. The absence or low expression of arcelin coding genes in some species might be due to chromosomal recombination (Osborn et al. 1986; Mirkov et al. 1994; Hartweck and Osborn 1997; Hartweck et al. 1997). To date, eight different allelic variants (designated Arc, Arc-1 to 8) of arcelin proteins have been described with subunit molecular weights in the range from 27 to 42 kDa in the seeds of *P. vulgaris* (Osborn et al. 1986; Lioi and Bollini 1989; Santino et al. 1991; Acosta-Gallegos et al. 1998; Zaugg et al. 2013). Apart from *P. vulgaris*, the arcelin-like sequence was reported in seeds of *P. acutifolius* (Mirkov et al. 1994) and *P. lunatus* (Sparvoli et al. 2001). These extensive studies on the arcelin molecule provided insight for designing a gene-specific primer for arcelin, which established the platform for cloning and expression of the arcelin gene from wild seeds of *P. lunatus* from Indian accession and looking at its impact on digestive enzymes of *Callosobruchus maculatus*.

**Materials And Methods**

**Collection of wild pulse variety**

The Indian pulse variety, *Phaseolus lunatus* (L.) (black) was collected from Vadakavunji village, Kodaikanal, Dindigul district, Tamil Nadu, India. The collected pulse variety was brought to the laboratory, cleaned, shade dried, and stored at room temperature (26 ± 2°C).

**RNA isolation and cDNA synthesis**

Total RNA was isolated from finely powered seed flour (50 mg) of *P. lunatus* seeds using TRIzol reagent. Total RNA was isolated following the manufacturer’s instructions in RNase-free microcentrifuge tubes, and the isolated RNA was stored at -80°C. Five micrograms of total RNA was used for cDNA synthesis using a RevertAid first-strand cDNA synthesis kit (Thermo Fisher Scientific Company, USA). Freshly synthesized cDNA was subjected to spectrophotometric analysis using a NanoDrop 2000c UV-Vis Spectrophotometer (Thermo Fisher Scientific Company, USA) for further PCR amplification.

**Specific amplification of the arcelin gene**

Bioinformatic analyses were carried out from the deposited arcelin variants (National Center for Biological Information) and their conserved domains were identified along the N-terminal and C-terminal regions to design gene-specific primers. Arcelin-specific oligonucleotide primers were: PlArcF: 5’-CTAGCCCTCTTTCCCTGT-3’ and PlArcR: 5’-ACCAAGAGAGCACGTC-3’. 100 ng of synthesized cDNA served as the template for the amplification reaction. PCR amplification was carried out using Taq polymerase master mix (Ampliqon) to a final concentration of 25 μl. The tubes were placed in a thermal cycler preheated at 94°C. The program set in the thermal cycler was as follows: 1 cycle at 94°C for 2 min (initial denaturation); 35 cycles of 94°C for 1 min (denaturation), 54.7°C for 45 sec (primer annealing) and 72°C for 1 min (extension); and a final extension cycle of 72°C for 7 min. It was then stored at 4°C. The amplified product was analyzed using a 1.2% agarose gel.
Sequencing of the arcelin gene and domain analysis

A FavorPrep™ Gel/PCR purification kit (Favorgen, Europe) was used to excise the amplified product from the agarose gel. The extraction method was followed according to the manufacturer's protocol. The purified product was sequenced using the Sanger sequencing method on an ABI 3730 XL (Applied Biosystems, USA). Sequenced data were compiled, analyzed, and submitted to GenBank, NCBI. The similarity analysis of nucleotide and protein sequence was carried out using BLAST at NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The nucleotide sequences were translated into their corresponding amino acid using the ExPASy translate tool, and the deduced amino acid was checked for its N-glycosylation sites. Subsequently, the nature of the protein, such as consensus sequences for legume lectins alpha and beta signatures, was identified using the ScanProsite tool (ExPASy).

Construction of plasmid pET-45b (+)/Arc

Bioinformatics analysis of the arcelin gene sequence and its corresponding amino acid sequence was performed before designing primers for cloning. NEB cutter (Version 2.0) was used to recognize the restriction sites in the arcelin gene and for adding restriction sites in the primer sequence. The oligonucleotide primers designed were: PlArcF_BamHI: 5’-AAAGCGGATCCAGGGACAGCACCGGC-3’ and PlArcR_HindIII: 5’-GAGCCCAAGCTTACCAAGAGACGACGTC-3’. The gene fragment encoding the arcelin gene was amplified with Phusion High-Fidelity DNA polymerase (New England Biolabs, UK). The tubes were then placed in a thermal cycler preheated at 94°C. The program set in the thermal cycler was: 1 cycle at 98°C for 30 sec (initial denaturation); 35 cycles of 98°C for 30 sec (denaturation), 72°C for 30 sec (primer annealing) and 72°C for 30 sec (extension); and a final extension cycle of 72°C for 10 min. It was then stored at 4°C. The amplified product was purified using the QIAquick PCR purification kit (Qiagen, USA) following the manufacturer's instructions and the purified product was analyzed using a 1.2% agarose gel before restriction enzyme digestion. The amplified arcelin gene fragment was digested with BamHI-HF and HindIII-HF (New England Biolabs, UK) and inserted using T₄ DNA ligase (Thermo Fischer Scientific Company, USA) into the predigested E. coli expression vector pET-45b(+) (Novagen) to construct pET-45b(+)/Arc.

Cloning of the arcelin gene

The resultant construct was transformed into a non-expression host, NEB 10-beta cells (New England Biolabs, UK). A loopful of NEB 10-beta cells was selected by plating onto an agar plate supplemented with streptomycin (30 µg/ml) and incubated at 37°C, overnight. Later, the cells were made competent using 0.1 M ice-cold CaCl₂ and resuspended in 2 ml of ice-cold 0.1 M CaCl₂ containing 15% glycerol, and was kept overnight on the ice at 4°C. To these competent cells, 10 µl of ligation mixture was added and incubated on ice for 30 min, followed by heat shock for 30 sec at 42°C. The tubes were immediately transferred to ice for the next 5 min and 900 µl of sterile SOC medium was added to it and incubated at 37°C for 1 hr with constant shaking at 250 rpm. The transformed cells were plated onto an agar plate
supplemented with streptomycin (30 µg/ml) and ampicillin (50 µg/ml). The plates were incubated overnight at 37°C. Transformed colonies were processed for plasmid isolation, and the positive colonies were subcultured in agar plates for further use. A QIAprep® Spin Miniprep kit (Qiagen, USA) was used for isolation of the recombinant plasmid DNA following a standard protocol according to the manufacturer’s instructions. The isolated plasmid DNA was subjected to spectrophotometric analysis. Double digestion of plasmid DNA was performed using restriction enzymes (BamHI-HF and HindIII-HF) to screen the positive recombinant colonies. The digested product was then analyzed by 1% agarose gel electrophoresis along with insert DNA to confirm the correct insert fall out. Recombinant plasmid DNA harbouring the appropriate insert used was amplified using T7 promoter primers, the amplified product was sequenced, and a dendrogram based on the sequence obtained from the amplified fragment was constructed based on the neighbour-joining method.

**Expression of recombinant arcelin protein in BL21 STAR (DE3) pLysS cells**

A loopful of BL21 STAR (DE3) pLysS cells (Thermo Fisher Scientific Company, USA) was selected by plating onto an agar plate supplemented with chloramphenicol (34 µg/ml) and incubated at 37°C overnight, and the selected cells were made competent using 0.1 M ice-cold CaCl2. Finally, the cells were resuspended in 2 ml of ice-cold 0.1 M CaCl2 containing 15% glycerol and were kept overnight on the ice at 4°C.

Competent BL21 STAR (DE3) pLysS cells (100 µl) were thawed on ice for 20 min before the transformation. To this end, 100 ng of the recombinant plasmid (1 µl) was added and gently tapped. This mixture was incubated on ice for 30 min, and soon after incubation, heat shock was given for 30 sec at 42°C. The tubes were immediately transferred to the ice for the next 5 min, and 250 µl of sterile SOC medium was added and incubated at 37°C for 1 hr with constant shaking at 250 rpm. The transformed cells were plated onto an agar plate supplemented with chloramphenicol (34 µg/ml) and ampicillin (50 µg/ml). The plates were incubated overnight at 37°C. Transformed colonies were processed for expression of recombinant arcelin protein, and the positive colonies were subcultured in agar plates for further use, from which 15% glycerol stock was made and stored at -80°C.

BL21 STAR (DE3) pLysS cells containing the recombinant plasmids were grown overnight at 37°C with shaking at 150 rpm in 5 ml of LB broth supplemented with chloramphenicol (34 µg/ml) and ampicillin (50 µg/ml). The preinoculum (4 ml) was transferred to 400 ml of LB broth supplemented with the same antibiotics. The culture was allowed to grow at 37°C with shaking at 150 rpm till the absorbance at 600 nm reached 0.6 (~ 3 hrs). The uninduced cells (1 ml) was removed before isopropyl β-D-1-thiogalactopyranoside (IPTG) induction. Once the desired absorbance was reached, the expression of arcelin protein was induced with 1 mM IPTG for 3 hrs at 37°C with shaking at 150 rpm. Randomly, 1 ml of cells was harvested after 1 hr and 3 hrs of IPTG induction. After 3 hrs, the cells were harvested by centrifugation at 5000 rpm for 20 min at 4°C. The supernatant was discarded, and the cell pellet was further processed before the purification of recombinant arcelin protein.
Isolation and solubilization of expressed protein from inclusion bodies

The cell pellet was thawed on ice and weighed. To the pellet, 7 ml of lysis buffer (50 mM Tris HCl, pH 8.0; 1 mM EDTA, pH 8.0; 25 mg lysozyme and 1 mM PMSF) was added, and the suspension was sonicated for 10 min on ice (9 sec on and 9 sec off, 10 cycles). The lysed solution was centrifuged at 10,000 rpm for 20 min at 4°C. The pellet was resuspended using 5 ml of wash buffer (50 mM Tris HCl, pH 8.0; 1 mM EDTA, pH 8.0; 0.5% deoxycholic acid and 1 mM PMSF). The resuspended suspension was sonicated for 10 min on ice (9 sec on and 9 sec off, 10 cycles) and centrifuged at 10,000 rpm for 20 min at 4°C. The obtained pellet was dissolved in solubilization buffer (50 mM Tris HCl, pH 8.0; 1 mM EDTA, pH 8.0; 1 mM PMSF and 8 M urea) and incubated at room temperature for an hour. After incubation, the suspension was centrifuged at 12,000 rpm for 20 min at 4°C. The supernatant contained solubilized proteins and was stored at -20°C until use.

Purification of recombinant protein using a Ni-NTA column

Ni-NTA resin (1 ml) (Qiagen, USA) was added to the column and packed (6.5×1 cm diameter), furthermore, the resin was washed with five column volumes of sterile double distilled water at a 7.2 ml/hr flow rate. Following this, the matrix was equilibrated with five column volumes of binding buffer, pH 8 (50 mM Na$_2$HPO$_4$; 300 mM NaCl; 10 mM imidazole and 8 M urea) at a flow rate 7.2 ml/hr. Solubilized protein (0.4 ml) was mixed with an equal volume of binding buffer and applied to the matrix at a 3.6 ml/hr flow rate. The column was washed with 10 ml of wash buffer, pH 8 (50 mM Na$_2$HPO$_4$; 300 mM NaCl; 20 mM imidazole and 8 M urea) to completely remove unbound or free proteins at a flow rate of 7.2 ml/hr. Histidine tagged recombinant protein bound to the column matrix was eluted with 3 ml of elution buffer, pH 8 (50 mM Na$_2$HPO$_4$; 300 mM NaCl; 250 mM imidazole and 8 M urea) at a flow rate of 18 ml/hr. Fractions of 1 ml were collected, and absorption at 280 nm was measured for each fraction. The fractions were also analyzed using denatured gel electrophoresis (Laemmli 1970). The total protein concentration of the purified recombinant protein was estimated following the method of Lowry et al. (1951). The activity of eluted fractions was analyzed after dialyzing the samples against 1X TBS buffer, pH 7.5 (50 mM Tris-HCl and 110 mM NaCl) for 24 hrs to ensure complete removal of urea and other denaturants.

Elucidating the insecticidal activity of purified recombinant arcelin protein

Artificial insect bioassay

To examine the effects of the arcelin molecule from *P. lunatus* on the development of *C. maculatus*, an artificial seed system, was performed following Shade et al. (1986). The second elution of the recombinant protein fraction was dialyzed extensively against distilled water and lyophilized. Artificial seeds (250 mg each seed) containing the crude seed kernel extract (1%, 3%, and 5%) and isolated molecule (1%) were obtained by thoroughly mixing the lyophilized powder with seed flour of the most susceptible variety of cowpea seeds (*V. unguiculata*). The susceptible cowpea seeds were milled into a
powder. The resulting flour was mixed with TBS I buffer in a 2:1 ratio and was made until a smooth paste was formed. The paste was then transferred to an acrylic mould. These plates were frozen at -20°C for 12 hrs and lyophilized for 12 hrs. After lyophilization, the solid artificial seeds were removed from the wells of the acrylic mould by gentle pressure. The seeds were then placed in plastic Petri plates and maintained for hydration at a constant temperature (25°C) and relative humidity (60 ± 5%) for 48 hrs. During hydration, the plates were closed with fine mesh to prevent accidental infestation. The seeds were then coated with 8% gelatin to mimic the seed coat texture. The artificial seeds (10 seeds/treatment) were placed in glass jars for *C. maculatus* infestation. Each treatment had ten artificial seeds and was replicated three times for each of the above concentrations. Newly emerged adults (five pairs of both sexes) were introduced into the plastic container for oviposition. This setup was maintained in an insect growth chamber with constant temperature and humidity. The effect of various doses of crude seed protein and isolated arcelin molecule on oviposition, adult emergence, and percentage of infestation/seed damage was studied. Control artificial seeds were made with a susceptible variety of cowpea seed (*V. unguiculata*) flour with an appropriate buffer.

**Isolation of total RNA using TRizol reagent**

Total RNA was isolated from 100 mg of fourth instar larvae of *C. maculatus* (control), and the same treated with purified arcelin molecule (treated) using TRizol reagent. Total RNA was isolated following the manufacturer’s instructions in RNase free microcentrifuge tubes, and the isolated RNA was stored at -80°C. Total RNA (5 µg) was used for cDNA synthesis using a RevertAid first-strand cDNA synthesis kit. Freshly synthesized cDNA was subjected to spectrophotometric analysis using a NanoDrop 2000c UV-Vis Spectrophotometer for further PCR amplification.

**Sequence retrieval of digestive enzyme gene mRNA from NCBI and primer design for real-time RT-PCR**

Five different digestive enzyme mRNA sequences were retrieved from GenBank, NCBI, for designing primers for real-time RT-PCR. The primers were designed by IDT software (Coralville, IA, USA). Primer compatibility such as melting temperature, GC%, length of the primer, molecular weight, self, and 3’ complementarity was assessed using the same tool. To ensure optimal polymerization efficiency and reduce the impact of RNA integrity on gene expression in RT-qPCR, primers were selected with a Tm of 58–64°C, length of 20–24 bp, and GC content of 45–60%. Primers were designed to amplify products within the range of 153–229 bp. The designed primer was commercially synthesized from Eurofins Genomics India Pvt. Ltd.

For the normalization of the real-time RT-PCR, a housekeeping gene is required. Among various universal housekeeping genes, 18S rRNA was selected as a suitable internal control for the real-time RT-PCR gene expression study.
**Real-time PCR amplification**

Real-time PCR amplification of the digestive enzyme gene fragments was performed using Applied Biosystems StepOne™ Real-Time PCR Systems in 8-well optically active strips (Applied Biosystems) under aseptic conditions. The comparative $C_\text{T}$ ($\Delta \Delta C_\text{T}$) method was adopted for relative expression analysis. SYBR green master mix (Applied Biosystems) was used, and melt-curve analysis was performed at the end of each reaction. The following thermal cycling conditions were used for real-time RT-PCR: initial hold at 95°C for 10 min for the activation of Taq DNA polymerase present in the master mix, an amplification program repeated 40 times (95°C for 15 sec and 60°C for 1 min), and a melting curve program at 60 to 95°C with a warming of 0.2°C per sec. Negative controls (deionized water) were included in each run. The housekeeping gene, 18S rRNA was used for the normalization by real-time RT-PCR. The relative expression ratio was calculated for each gene of interest by using a mathematical model described by Pfaffl (2001).

**Results**

**Amplification and sequence analysis of cDNA encoding the arcelin gene**

Total RNA isolated from seeds of *P. lunatus* was used as the template for cDNA synthesis. The designed gene-specific primer from the arcelin variants amplified the arcelin gene in the seeds of *P. lunatus*. The PCR product of these primers resulted in a single band in a 1.2% agarose gel. This confirmed the suitability of primer sequences for PCR and the presence of a single PCR product in agarose gel electrophoresis (Fig. 1). The amplified PCR fragment (271.7 ng/µl of DNA) corresponding to 650 bp was excised and purified. The fragment was sequenced and was submitted to NCBI GenBank (accession number: MN219438). A total of 637 bp of nucleotides were obtained from the sequence, and the open reading frame (ORF) encoded 473 bp. Based on the deduced amino acid sequence, the open reading frame of arcelin from *P. lunatus* was found to be composed of 158 amino acids.

The results on the N-glycosylation site and functional domain in the amino acid sequence of the arcelin gene from *P. lunatus* using ScanProsite (ExPASy) are presented in Table 1. The consensus sequence for legume lectin signatures was present in the obtained amino acid sequence. A total of 63-69 amino acids were responsible for coding the legume lectin-beta domain, and 137-146 amino acids were responsible for coding the legume lectin-alpha domain.

BLAST analysis of the amino acid sequence of arcelin from the seed of *P. lunatus* showed close similarities of 98.10% with the arcelin-like protein of *P. lunatus* (CAB96393.1) and with the alpha-amylase inhibitor-like protein of *P. lunatus* (CAB96394.1 and CAB96395.1). Nearest similarities were also seen with *L. purpureus* arcelin (ABJ16470.1) and *P. lunatus* lectin sequences (CAA93828.1 and CAA93827.1). Additionally, various isoforms of arcelin from *P. vulgaris* showed high matching with the deduced amino acid sequence of *P. lunatus* in the present study.

**Cloning of the arcelin gene in the pET-45b (+) vector**
Specific amplification of the arcelin gene was carried out to clone the product in the expression vector system with a new set of primers designed, and the amplified PCR product was digested with BamHI-HF and HindIII-HF. The digested PCR product (500 bp) gave a single band in a 1.2% agarose gel. The above was successfully ligated with the predigested *E. coli* expression vector pET-45b(+) to construct pET-45b(+)/Arc. NEB 10-beta cells were transformed with the constructed expression vector and the plasmid DNA was isolated from each recombinant derivative from LB medium supplemented with appropriate antibiotics. Plasmid DNA digested with the restriction enzymes resulted in two fragments; one was similar to the size of pET-45b(+) vector DNA, and the other was in the size of the insert DNA. A clear fallout band at approximately 500 bp (size of insert DNA) indicates that the positive clones with the desired recombinant plasmid can subsequently express the desired protein. Similarly, sequencing of the T₇ promoter region in recombinant plasmid DNA confirmed the uptake of insert DNA.

BLAST analysis of the amino acid sequence of the arcelin clone showed close similarities of 99.33% with the arcelin-like protein of *P. lunatus* (CAB96393.1), alpha-amylase inhibitor-like protein of *P. lunatus* (CAB96394.1 and CAB96395.1) and 96.67% with *P. lunatus* lectin-2 (CAA93828.1. Additionally, a similarity of 82.76% was observed with *L. purpureus* arcelin (ABJ16470.1) and with various isoforms of arcelin from *P. vulgaris*. To examine the relationship between the *P. lunatus* arcelin clone gene sequence and other arcelin isoforms, a neighbour-joining (NJ) phylogenetic tree was drawn. The tree indicates the closeness of the related species with the isolated arcelin gene fragment (Fig. 2).

**Expression and purification of the recombinant protein**

Recombinant protein constructs of pET-45b(+)/Arc were tested for protein expression in BL21 STAR (DE3) pLysS cells. The cells were induced with 1 mM IPTG at 37°C for 3 hrs for overexpression of recombinant protein at the mid-exponential phase. Maximum expression was optimized at the third hour of IPTG induction. Fusion proteins usually include a partner or are "tag" linked to the passenger or target protein by a recognition site for a specific protease. Most fusion partners are exploited for specific affinity purification strategies (Sato and Hori 2009). The present study utilizes 6 X His-tag as the fusion partner. Since overexpression of arcelin protein was observed in pellet, the inclusion bodies found in the pellet was solubilized to isolate the intact protein fraction, and the solubilized inclusion bodies were purified using a Ni-NTA affinity column.

The histidine-tagged recombinant arcelin protein was eluted using 250 mM imidazole. The purified protein was dialyzed entirely against a suitable buffer to remove the urea that was used initially to solubilize the protein. 12% SDS-PAGE analysis of the purified recombinant protein revealed the presence of a single homogenous protein band just above 20 kDa. The total protein concentration of the purified recombinant protein was found to be 0.92±0.12 mg/ml (Fig. 3).

**Elucidating the insecticidal activity of purified recombinant arcelin protein**

**Artificial insect bioassay**
The results of the artificial bioassay justify the anti-metabolic nature of the purified recombinant arcelin protein. Although the number of eggs laid per seed was uniform in all the treatment and control groups, adult emergence was drastically reduced in the recombinant protein-treated groups when compared to control seeds. Adult emergence rates of 10.63% and 1.78% were observed in the 1% and 5%, respectively, of recombinant arcelin protein-treated groups. Additionally, seed damage was reduced from 100% to 15% and 1% in the treated groups. In addition to these observations, the average developmental period of the bruchid pest *C. maculatus* increased from 28 days to 40 days, which indicates the anti-metabolic potency of the purified recombinant arcelin protein fraction (Table 2). Interestingly, the larval mass from the treated seeds was significantly reduced compared to that of the control seeds (Fig. 4). These results conclude that the purified recombinant arcelin protein of *P. lunatus* demonstrates a potent insecticidal effect against the bruchid insect pest, *C. maculatus*.

**Gene expression analysis of digestive enzymes**

To determine the response of ingested arcelin to host digestive enzymes, five different digestive enzyme sequences were retrieved from GenBank, NCBI. All the selected genes belong to the genus *Callosobruchus*, and these genes are responsible for sugar metabolism in the host insect. Primers designed to amplify various digestive enzymes in control and arcelin-treated fourth instar larvae of *C. maculatus* are tabulated in Table 3. Furthermore, the housekeeping gene 18S rRNA (endogenous control), whose gene product was estimated to be approximately 151 bp, was chosen. To perform this study, total RNA was isolated from control and purified recombinant arcelin-treated fourth instar larvae of *C. maculatus* and converted into cDNA. All PCR assays were conducted in duplicate to obtain reliable results and the mean was used for analysis.

1. **C<sub>T</sub>** value analysis of the alpha-amylase gene of *C. maculatus* (*Cm AMY*)

The mean C<sub>T</sub> values of the reference gene, 18S rRNA were 33.07 (control) and 33.90 (treated), while the mean C<sub>T</sub> values of the *Cm AMY* gene in control and arcelin protein treated larvae were 33.11 and 31.55, respectively. Threshold fluorescence for cDNA of arcelin-treated samples reached 31.55 cycles, indicating a high expression of the alpha-amylase gene compared with control larvae and the reference gene. A 5.21 fold increase in the alpha-amylase gene expression was observed in recombinant arcelin fed fourth instar larvae of *C. maculatus* compared to the control larvae.

2. **C<sub>T</sub>** value analysis of glycoside hydrolase family 31 gene of *C. maculatus* (*Cm GH-13*)

The mean C<sub>T</sub> values of the reference gene, 18S rRNA were 32.82 (control) and 32.32 (treated), while the mean C<sub>T</sub> values of the *Cm GH-13* gene in control and arcelin protein-treated larvae were 25.17 and 23.78, respectively. Threshold fluorescence for cDNA of arcelin treated samples reached 23.78 cycles, indicating a high level of glycoside hydrolase family 31 gene expression compared with control larvae and the reference gene. A 1.86 fold increase in the glycoside hydrolase family 31 gene expression was observed in recombinant arcelin fed fourth instar larvae of *C. maculatus* compared to the control larvae.
3. **C\textsubscript{T} value analysis of aminopeptidase N gene of *C. maculatus* (**Cm APN**)  

The mean C\textsubscript{T} values of the reference gene, 18S rRNA were 32.00 (control) and 32.36 (treated), while the mean C\textsubscript{T} values of the *Cm* APN gene in control and arcelin protein-treated larvae were 25.72 and 21.57, respectively. Threshold fluorescence for cDNA of arcelin-treated samples reached 21.57 cycles, indicating a high expression of the aminopeptidase gene compared with control larvae and the reference gene. A 22.73 fold increase in the expression of aminopeptidase gene was observed in recombinant arcelin fed fourth instar larvae of *C. maculatus* when compared to the control larvae.

4. **C\textsubscript{T} value analysis of cathepsin D-like aspartic protease gene of *C. maculatus* (**Cm CAT-D**)  

The mean C\textsubscript{T} values of the reference gene, 18S rRNA were 31.99 (control) and 33.32 (treated), while the mean C\textsubscript{T} values of the *Cm* CAT-D gene in control and arcelin protein-treated larvae were 17.32 and 15.48, respectively. Threshold fluorescence for cDNA of arcelin-treated samples reached 15.48 cycles, indicating a high level of expression of the cathepsin D-like aspartic protease gene upon arcelin treatment compared with control larvae and the reference gene. A fold increase of 8.94 times increase in the expression of the cathepsin D-like aspartic protease gene was observed in recombinant arcelin fed fourth instar larvae of *C. maculatus* when compared to the control larvae.

5. **C\textsubscript{T} value analysis of cathepsin L-like cysteine protease gene of *C. maculatus* (**Cm CAT-L**)  

The mean C\textsubscript{T} values of the reference gene, 18S rRNA were 35.02 (control) and 32.70 (treated), while the mean C\textsubscript{T} values of the *Cm* CAT-L gene in control and arcelin protein-treated larvae were 21.19 and 22.08, respectively. Threshold fluorescence for cDNA of arcelin-treated samples reached 22.08 cycles, indicating a significantly decreased level of expression of the cathepsin L-like cysteine protease gene compared with control larvae. A 0.11 fold decrease in the expression of the cathepsin L-like cysteine protease gene was observed in recombinant arcelin fed fourth instar larvae of *C. maculatus* when compared to the control larvae.

**Comparative relative quantization of gene expression of different classes of digestive enzymes**

The calculated relative quantity or fold changes of gene expression in five different classes of digestive enzymes using the method of Pfaffl (2001) are given in Fig. 5. Of five different digestive enzymes studied, maximum expression was observed in the aminopeptidase gene and other genes responsible for the alpha-amylase activity, glycoside hydrolase family 31 and cathepsin D-like aspartic protease were upregulated when fourth instar larvae were treated with 1% of recombinant arcelin protein. In contrast, cathepsin L-like cysteine protease was downregulated significantly.

**Discussion**

Members of the lectin and lectin-like protein family are considered to be toxic towards insects of stored product pests belonging to the Coleopteran family of Bruchidae that commonly infest leguminous seeds.
and form the basis for postharvest losses. Wild accessions of *P. vulgaris* contain the insecticidal protein; arcelin, which replaces the normal seed storage protein in the seed, and the presence of this protein is responsible for insect resistance against bruchid pests. Based on this background information, the present study aims to select one such bruchid-resistant pulse, screening the existence of potent insecticidal arcelin molecules and the expression of the arcelin gene in the bacterial host, thereby elucidating its insecticidal activity towards the stored product insect pest *C. maculatus*.

To isolate the arcelin gene from *P. lunatus*, a detailed perceptive of various isoforms of arcelin and their molecular size is necessary. To date, eight arcelin variants have been described from the wild accessions of the common bean *Phaseolus*. Each of these variants has genetically different alleles belonging to the same loci. Arc-1, Arc-2, and Arc-6 contribute to the same cluster, while Arc-3 and Arc-4 belong to the most ancient cluster and Arc-5 has two isoforms 5a and 5b which belong to the separate branch. Apart from these features, understanding the nucleotide variation among each sequence would assist in designing a suitable gene-specific primer to amplify and sequence the arcelin gene.

Additionally, the isoform of arcelin from Indian wild bean, *L. purpureus*, showed significant homology to Arc-3 and Arc-4 alleles from *Phaseolus* sp (Janarthanan et al. 2012). LALFL was chosen as the conserved region in the N-terminal region and DVLSW in the C-terminal region for designing gene-specific primers. The nucleotide sequences designed using these amino acid sequences amplified the arcelin gene in the resistant pulse variety, *P. lunatus*. These findings indicated the occurrence of chosen conserved sequences in all arcelin variants, but are lost in other lectin-related genes.

Plants use these broad varieties of lectin domains to counteract pathogen and predator attacks (Lannoo and Van Damme 2014). Some legume lectins are proteolytically processed to produce two chains, beta (which corresponds to the *N*-terminal) and alpha (*C*-terminal) (Chawla et al. 1993). These alpha and beta motifs are present in most of the variants of arcelin amino acids in *P. vulgaris* accessions. Likewise, the deduced amino acid sequence motifs of *P. lunatus* arcelin were similar to the previously reported domains of arcelin protein with similar functional domains and *N*-glycosylation sites. Two significant consensus sequences for legume lectin signatures, namely, the legume lectin-alpha domain and beta domain, are found in the deduced amino acid sequence. This indicates the functional potency of the isolated arcelin gene from the seeds of *P. lunatus*. Apart from this, arcelin from *P. lunatus* revealed three putative *N*-glycosylation sites, namely, Asn^7^-Ala^8^-Ser^9^-Phe^10^, Asn^13^-Phe^14^-Thr^15^-Phe^16^, and Asn^106^-Ser^107^-Ser^108^-Ser^109^.

Of these three sites, Asn-Phe-Thr-Phe is predominantly found in almost all the arcelin sequences reported earlier, and Asn-Ala-Ser-Phe was found in Arl-2 (arcelin-like) of *P. lunatus* (Sparvoli et al. 1998). From the present study, it was also clearly evident that apart from sequence homology with different variants of arcelins, the amino acid sequence of arcelin from the seed of *P. lunatus* also shares close similarities with the alpha-amylase inhibitor-like protein and lectin sequences. Such observations were also seen among the *N*-terminal sequence of arcelin from *P. vulgaris* which has a highly conserved amino acid sequence from position 16 to 22 with PHA and alpha amylase-inhibitor protein. These regions are 92% identical and are identified as functioning to target PHA protein with the vacuole in plant cells (Schoonhoven and Cardona 1982; Hartweck et al. 1991).
Among various variants of arcelin genes, Arc-1 was the first variant cloned using a suitable cDNA cloning vector. mRNA from the developing seeds of the resistant pulse variety of *P. vulgaris* (SARC1-7) was cloned into the pARC7 vector that encloses 265 amino acids in its open reading frame (Osborn et al. 1988). Later, Arc-2, Arc-5, Arc-3 and Arc-7 genes from *P. vulgaris* were also cloned using suitable vector systems (John and Long 1990; Goossens et al. 1994; Lioi et al. 2003). cDNA cloning with the pGEMT vector system was used to clone *L. purpureus* (Janarthanan et al. 2012). Sparvoli et al. (1998), for the first time, isolated and characterized two lectin-related proteins, and their cDNA clone showed 93.7% of sequence identity with arcelin-like protein (239 amino acids) and alpha-amylase inhibitor-like protein (233 amino acids) from Big lima (*P. lunatus*), where the ZAP II vector was used, and XLI blue cells were used as transformation hosts. Later, the APA locus of *P. lunatus* was cloned into the pBSKII+ vector system (Sparvoli et al. 2001). These observations formed the basis of further expression experiments using an appropriate expression vector and host system. Gerhardt et al. (2000) for the first time carried out the expression of the arcelin gene (Arc-5 III) from *P. vulgaris* in a prokaryotic expression system. The EcoRI-NotI fragment of Arc-5 was cloned into EcoRI-NotI sites of pGEX-4T-1, and the recombinant product was soluble GST fusion protein expressed in *E. coli* BL21 (DE3) with a molecular weight of 26.6 kDa (239 amino acids). The fusion protein was then purified using glutathione Sepharose 4B column matrix. The expressed protein shows 81% amino acid similarity with Arc-5a and Arc-5b polypeptides. These results provide insight into designing an appropriate expression protocol for the production of recombinant protein. The present study utilizes the pET-45b(+) vector to express arcelin protein isolated from the seeds of *P. lunatus*. The BamHI-HindIII fragment of arcelin was cloned into the BamHI-HindIII sites of the pET-45b(+) vector, and the arcelin insert was effectively transformed into a non-expression host to confirm the successful transformation. Additionally, amplification of the T7 region of the positive recombinant colonies was also carried out. Sequence analysis of arcelin clones from *P. lunatus* revealed very close similarity with Arl-2 of *P. lunatus* and with alpha-amylase inhibitor protein of *P. lunatus*. Apart from these, they also share similarities with various isoforms of arcelin isolated from *P. vulgaris*. Later, the recombinant plasmid comprising the arcelin insert was expressed in BL21 STAR (DE3) pLysS cells. The inclusion bodies were solubilized, the histidine-tagged recombinant protein was purified using a Ni-NTA column, and the purified recombinant arcelin protein was refolded to perform an artificial insect feeding bioassay experiment.

The purified recombinant arcelin protein was tested for its insecticidal activity by formulating an artificial diet out of lyophilized powder to perform an insect feeding bioassay against the bruchid pest, *C. maculatus*. The 1% treatment effectively reduced the larval mass, and adult emergence was significantly decreased compared to the control-treated group. Five per cent of treatment was much more effective in reducing the percentage of adult emergence. In addition, a delay in the developmental period was observed in the recombinant arcelin feed groups. These results correlate with the action of Arc-1, which reduces the percentage of adult emergence of *Z. subfasciatus* and causes a delay in its emergence up to 53 days (Osborn et al. 1988). An artificial diet containing 7.1% of the arcelin molecule would result in 48% survival of *Z. subfasciatus*, and its mean developmental period was slightly reduced compared to its control treatment. It was also reported that arcelin was resistant to digestion by gut proteases (Minney et

Page 13/31
al. 1990). Studies conducted on two resistant (G02771 and G12882) pulse varieties of *P. vulgaris* revealed that Arc-1 potently altered the gut structure and thereby penetrated the hemolymph of *Z. subfasciatus*, causing a growth inhibitory effect on the larvae (Paes et al. 2000). Among all the isoforms of arcelin, Arc-5 is a potent insecticidal protein that can reduce larval mass and adult emergence at low concentrations (Gerhardt et al. 2000). Apart from targeting Coleopteran pests, arcelin was proven effective against certain pests belonging to Hemiptera and Lepidoptera (Oriani and Lara 2000; Malaikozhundan et al. 2003). Harmsen (1989) postulated that differences in the days of emergence and percentage of the emergence of bruchids were due to differences in arcelin type and arcelin concentration in the seed.

Apart from the arcelin molecule, true lectins are also known to cause insecticidal activity against the bruchid pest, *C. maculatus*. Larval mass was significantly lower when *C. maculatus* was fed lectin isolated from *Crataeva tapia* (CrataBL) demonstrating the efficiency of the protein (Nunes et al. 2015). Additionally, BmoLL lectin, purified from *Bauhinia monandra* leaves reduced the larval mass of *C. maculatus* by 50% when incorporated at a concentration of 0.4% in artificial seeds (Andrade et al. 2005) whereas the TEL lectin purified from *Talassia esculenta* seeds reduced the larval mass by 50% when applied at a concentration of 1% (w/w) (Macedo et al. 2002). These effects could be due to the hypothesis made by Paiva et al. (2012) that the decrease in the average survival rate and larval mortality is due to the binding of these molecules to glycosylated proteins in the midgut of larvae which reduces the efficiency of nutrient uptake and diet utilization, thereby causing a drop in total larval mass. The binding of these molecules precisely explains the anti-nutrient activity of plant lectins and lectin-like molecules to the glycan receptors exposed on the epithelial cells lining the digestive tract of insects, thereby blocking the enzymes responsible for digestion. In addition to these responses, they also disrupt midgut cells lining the intestinal surfaces (Eisemann et al. 1994; Peumans and Van Damme 1995; Fitches et al. 1997; Matsushita et al. 2002). This specific carbohydrate-lectin interaction with glycoconjugates such as glycoproteins and glycolipids on the epithelial cells lining the digestive tract can be envisioned as a possible biotechnological strategy for insect pest management (Gatehouse and Gatehouse 1998; Fitches et al. 2001; Macedo et al. 2004; Sauvion et al. 2004). If arcelins enact as these lectins or lectin-like proteins, then two types of interactions could be possible: binding of arcelins to glycosylated digestive enzymes and binding of arcelins to glycoconjugates exposed on the epithelial cells along the digestive tract (Peumans and Van Damme 1995). Minney et al. (1990) suggested that arcelin-containing beans may be resistant to *Z. subfasciatus* because larvae lack the necessary gut proteases to digest this abundant protein, resulting in starvation of larvae.

Gaining insight into detailed molecular mechanisms and possible interactions of defense proteins will significantly help us tailor and design defense proteins aimed at higher anti-insect efficacy and durability (Amirhusin et al. 2004). The present study focuses on regulating the gene expression level of major and minor digestive enzymes present in the fourth instar larvae of *C. maculatus*. In general, it was reported by many that the digestive enzyme pool of Coleopteran insects primarily consists of α-amylases along with proteases to break down starch, proteins and other carbohydrate components that are present in food grains. In addition to α-amylases, the larvae of *C. maculatus* rely on aspartic and
cysteine peptidases to digest proteins in their diet (Nogueira et al. 2012). This study observed varied expression of various digestive enzymes of *C. maculatus* larvae upon potent anti-metabolic protein treatment, suggesting that it may have a vital role in imparting insecticidal activity to arcelin by modulating digestive physiology.

Whole larvae of *C. maculatus* were used in the study to better understand the total enzymatic activity of the host insect upon treatment with arcelin protein. This is because most of the larval digestive enzyme activities were found in the luminal contents, where the majority of glycosidase and proteolytic activities are localized, and final digestion of proteins probably occurs at the surface of midgut cells under the action of a putative microvillar aminopeptidase (Kitch and Murdock 1986; Silva et al. 1999). Additionally, it is reported that the spatial organization of digestion in larvae of *C. maculatus* and *Z. subfasciatus* is so peculiar that they appear to lack digestive salivary carbohdrases, except for α-galactosidases and a soluble aminopeptidase, and there seems to be no contribution of enzymes from dietary seeds to luminal digestion. Thus, both species can produce their own digestive enzymes from midgut tissues, as is the rule among other insects (Terra and Ferreira 1994).

Alpha-amylases belong to the GH13 glycoside hydrolases family of enzymes that hydrolyze the α-1,4-glicosidic bond in starch to produce glucose, maltose and other products, and these serve as energy sources for insect pests (Svensson 1994; Janecek et al. 1997; MacGregor et al. 2001; Cantarel et al. 2009). Expression of the α-amylase gene was significantly increased upon recombinant arcelin treatment; these results substantiate the previous finding where α-amylase activity was increased in the five consecutive generations of *C. maculatus* when treated with purified arcelin protein isolated from the wild seeds of *L. purpureus* when compared to the control treatment (Janarthanan and Suresh 2003). In both cases, a significant reduction in seed infestation was observed. These observations conclude that there is a significant impact of arcelin protein on the α-amylase of *C. maculatus*, which could be exploited in plant protection measures. Also, it was evident that in the presence of starch granules of *P. vulgaris*, larvae of *Z. subfasciatus* secrete more α-amylase. These observations demonstrate that α–amylase induction in bruchid larvae represents another case of induction of an initial phase digestive enzyme in response to different diets (Silva et al. 2001). These observations are also true in the case of some lectins when *Bauhinia monandra* leaf lectin (BmoLL) incubated with midgut homogenates of *C. maculatus* showed an increase in α-amylase activity. However, it has been proposed that lectin may affect the enzyme activities binding to these at places other than their substrate-binding site (Kim et al. 1976). As mentioned earlier, the anti-nutritional activity of lectins and lectin-related proteins is mainly mediated by the binding of these molecules to glycan receptors, increasing the number of enzyme active sites, increasing enzyme activity (Erickson et al. 1985; You and Chang 1992; Matsushita et al. 2002). Such changes were also observed when α-amylase inhibitors were given in larval diets of three species of leaf rollers (Lepidoptera: Tortricidae), resulting in overproduction of α-amylase (Markwick et al. 1996).

In addition, to increase α-amylase expression, moderately increased expression of glycoside hydrolase was found upon recombinant arcelin treatment, because glycoside hydrolase (GH) families aid in the digestion of cellulose and other polysaccharides. Moreover, in contrast, the highest expression of the N-
aminopeptidase gene was observed in the present study. These results correspond with the observation that short-term exposure of insects to dietary lectins [GNA (Galanthus nivalis agglutinin) and Con A (C. ensiformis)] drastically increases the activity of aminopeptidase and trypsin activities associated with the gut, indicating some kind of compensatory response of the gut epithelial cells that produce these enzymes (Powell et al. 1998).

Cathepsin D, a class of aspartic proteases, is secreted into the midgut lumen of many Coleopterans and is regarded as a minor digestive enzyme, whereas cathepsin L-like cysteine proteases are often the major digestive enzymes (Silva and Xavier-Filho 1991; Blanco-Labra et al. 1996; Brunelle et al. 1999). The aspartic protease cathepsin D functions in the intracellular and extracellular degradation of proteins in insects, and it is also essential in metamorphosis and food degradation (Ahn and Zhu-Salzman 2009). The present study elucidates that gene expression of cathepsin D-like aspartic protease was increased approximately nine times compared to control treatment, whereas the major digestive enzyme was downregulated. These results suggest that recombinant arcelin molecule were influential in the host and was not degraded by major digestive proteases. These results correlate with the findings of Minney et al. (1990), who postulated that the arcelin-containing beans might be resistant to Z. subfasciatus because the larvae lack the necessary gut proteases to digest this abundant protein and may result in starvation. Studies on Arc-1 revealed that this protein is highly resistant to gut proteolysis (Fabre et al. 1998). Similar results were observed when cowpea bruchid larvae were reared on a diet containing a sustained and sublethal dose of soybean cystatin (scN) where the midgut protease activity profile was dramatically remodelled (Zhu Salzman et al. 2003). The impact of this minor digestive enzyme could become significant when the major digestive enzymes are inhibited (Ahn and Zhu-Salzman 2009). In addition to proteases and peptidases, scN-adapted larvae induced amylases, galactosidases, glucosidases, and polygalacturonase, possibly to help meet their nitrogen and carbon requirements in the presence of a protease inhibitor (Moon et al. 2004).

From the above findings, it could be concluded that in the common insect digestive system of herbivorous insects acts as the first line of defense mechanism and exploits a suite of endo- and exopeptidases not only for the breakdown of a variety of dietary proteins but also as a counter-defense reservoir to assist them in coping with anti-nutritional compounds and dietary toxins that they may encounter. Far more than providing a mere physical barrier, this protection takes the form of changes in gene expression and protein accumulation in the cells lining the digestive tract. Changes in the composition of digestive enzymes can be considered an adaptive strategy to address unusual diets, as observed for Z. subfasciatus larvae, which evolved in the presence of P. vulgaris, but not in the presence of the introduced host V. unguiculata (Broadway 2000; Silva et al. 2001). These observations hold true in the present study; however, there are considerable changes in the gene expression of digestive enzymes because of uptake of anti-nutritional protein, larval mortality and decreased body mass were observed in recombinant arcelin of P. lunatus treated fourth instar larvae.

These results justified the anti-metabolic nature of the recombinant arcelin protein. In the future, studies could be carried out on extensive biophysical characterization in terms of
crystallization which would elucidate the structural conformations, and on the other hand, this antibiosis nature of recombinant arcelin protein of *P. lunatus* could be transferred by transgenesis into cultivated pulse varieties to produce a variety resistant to bruchid infestation that can enrich nutritional intake in humans.

**Declarations**

**Funding**

The work was supported by the Department of Science & Technology, New Delhi, with a DST-INSPIRE fellowship (IF140977) provided to Karupppiah Hilda.

**Conflicts of interest:** The authors declare that they have no conflict of interest.

**Availability of data and material:** GenBank accession number is given in the manuscript.

**Code availability:** Not applicable.

**Authors' contributions:** KH and SB have their contribution in conceptualization, investigation and drafting the manuscript. RK provided her immense support in insights into cloning. MM was involved in the investigation of the study. SJ has super visioned and resourced the entire work.

**Ethics approval:** Not applicable.

**Consent to participate:** There were no humans participants in the study.

**Consent for publication:** Authors can provide the signed approval upon publisher’s request.

**References**

Acosta-Gallegos JA, Quintero C, Vargas J, Toro O, Tohme J, Cardona C (1998) A new variant of arcelin in wild common bean, *Phaseolus vulgaris* L, from southern Mexico. Genet Resour Crop Ev 45:235-242. https://doi.org/10.1023/A:1008636132108

Ahn JE, Zhu-Salzman K (2009) CmCatD, a cathepsin D-like protease has a potential role in insect defense against a phytocystatin. J Insect Physiol 55:678-685. https://doi.org/10.1016/j.jinsphys.2009.04.016

Amirhusin B, Shade RE, Koiwa H, Hasegawa PM, Bressan RA, Murdock LL, Zhu-Salzman K (2004) Soyacystatin N inhibits proteolysis of wheat α-amylase inhibitor and potentiates toxicity against cowpea weevil. J Econ Entomol 97:2095-2100. https://doi.org/10.1093/jee/97.6.2095

Andrade CA, Baszkin A, Santos-Magalhaes NS, Coelho LC, de Melo CP (2005) Dielectric properties of *Bauhinia monandra* and concanavalin A lectin monolayers, part I. J Colloid Interf Sci 289:371-378. https://doi.org/10.1016/j.jcis.2005.01.076
moth (Lacanobia oleracea) larvae; mechanisms of insecticidal action. J Insect Physiol 47:777-787. https://doi.org/10.1016/S0022-1910(01)00068-3

Gatehouse AMR, Barbieri L, Stirpe F, Croy RRD (1990) Effects of ribosome inactivating proteins on insect development-differences between Lepidoptera and Coleopteran. Entomol Exp Appl 54:43-51. https://doi.org/10.1111/j.1570-7458.1990.tb01310.x

Gatehouse AMR, Gatehouse JA (1998) Identifying proteins with insecticidal activity: use of encoding genes to produce insect resistant transgenic crops. Pestic Sci 52:165-175. https://doi.org/10.1002/(SICI)1096-9063(199802)52:2<165::AID-PS679>3.0.CO;2-7

Gatehouse AMR, Powell KS, Peumans WJ, Van Damme JM, Gatehouse JA (1995) Insecticidal properties of plant lectins: their potential in plant protection. In: Pusztai A, Bardocz S (ed) Lectins: biomedical perspectives, CRC press, Taylor & Francis, pp 35-58.

Gerhardt IR, Paes NS, Bloch C, Mendes PA, Leite A, Chrispeels MJ, Grossi-de-Sa MF (2000) Molecular characterization of a new arcelin-5 gene. Biochim Biophys Acta Gene Struct Expression 1490:87-98. https://doi.org/10.1016/S0167-4781(99)00219-5

Goossens A, Geremia R, Bauw G, Van Montagu M, Angenon G (1994) Isolation and characterization of arcelin-5 proteins and cDNAs. Eur J Biochem 225:787-795. https://doi.org/10.1111/j.1432-1033.1994.0787b.x

Harmsen RH (1989) Agronomic traits and resistance to Mexican bean weevil (Zabrotes subfasciatus) and common bean weevil (Acanthoscelides obtectus) (Bruchidae) in cultivated bean lines containing arcelin seed-protein alleles from wild beans (Phaseolus vulgaris L.) Ph.D Dissertation, University of Wisconsin, Madison.

Hartweck LM, Cardona C, Osborn TC (1997) Bruchid resistance of common bean lines having an altered seed protein composition. Theor Appl Genet 95:1018-1123.

Hartweck LM, Osborn TC (1997) Altering protein composition by genetically removing phaseolin from common bean seeds containing arcelin or phytohemagglutinin. Theor Appl Genet 95:1012-1017. https://doi.org/10.1007/s001220050655

Hartweck LM, Vogelzang RD, Osborn TC (1991) Characterization and comparison of arcelin seed protein variants from common bean. Plant Physiol 97:204-211.

Jaber K, Haubruge E, Francis F (2010) Development of entomotoxic molecules as control agents: illustration of some protein potential uses and limits of lectins. Biotechnol Agron Soc Environ 14:225-241.

Janarthanan S, Sakthivelkumar S, Veeramani V, Radhika D, Muthukrishanan S, (2012) A new variant of antimetabolic protein, arcelin from an Indian bean, Lablab purpureus (Linn.) and
its effect on the stored product pest, *Callosobruchus maculatus*. Food Chem 135:2839-2844. https://doi.org/10.1016/j.foodchem.2012.06.129

Janarthanan S, Suresh P (2003) Insecticidal potential of wild bean seed protein, arcelin. Nat Prod Rad 2:243-245. http://nopr.niscair.res.in/handle/123456789/12285

Janecek S, Svensson B, Henrissat B (1997) Domain evolution in the α-amylase family. J Mol Evol 45:322-331.

John ME, Long CM (1990) Sequence analysis of arcelin 2, a lectin-like plant protein. Gene 86:171-176. https://doi.org/10.1016/0378-1119(90)90276-W

Kim YS, Brophy EJ, Nicholson JA (1976) Rat intestinal brush border membrane peptidases. II. Enzymatic properties, immunochemistry, and interactions with lectins of two different forms of the enzyme. J Biol Chem 251:3206-3212. https://doi.org/10.1016/S0021-9258(17)33424-5

Kitch LW, Murdock LL (1986) Partial characterization of a major gut thiol proteinase from larvae of *Callosobruchus maculatus* F. Arch Insect Biochem Physiol 3:561-575. https://doi.org/10.1002/arch.940030607

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680-685. https://doi.org/10.1038/227680a0

Lambrides CJ, Godwin ID (2007) Mungbean. In: Pulses, sugar and tuber crops, Springer, Berlin, Heidelberg, pp 69-90.

Lannoo N, Van Damme EJ (2014) Lectin domains at the frontiers of plant defense. Front Plant Sci 5:387. https://doi.org/10.3389/fpls.2014.00397

Lioi L, Bollini R (1989) Identification of a new arcelin variant in wild bean seeds. Annu Rep Bean Improv Coop 32:28-29.

Lioi L, Sparvoli F, Galasso I, Lanave C, Bollini R (2003) Lectin-related resistance factors against bruchids evolved through a number of duplication events. Theor Appl Genet 107:814-822. https://doi.org/10.1007/s00122-003-1343-8

Lowry OH, Rosenbrough NJ, Farr AL, Randall RJ (1951) Protein measurement with the folin phenol reagent. J Biol Chem 193:265-275.

Macedo MLR, Freire MDGM, Novello JC, Marangoni S (2002) *Talisia esculenta* lectin and larval development of *Callosobruchus maculatus* and *Zabrotes subfasciatus* (Coleoptera: Bruchidae). Biochim Biophys Acta Gen Subj 1571:83-88. https://doi.org/10.1016/S0304-4165(02)00155-1
Macedo MLR, Sá CMD, Freire MDGM, Parra JRP A (2004) Kunitz-type inhibitor of coleopteran proteases, isolated from *Adenanthera pavonina* L. seeds and its effect on *Callosobruchus maculatus*. J Agric Food Chem 52:2533-2540. https://doi.org/10.1021/jf035389z

MacGregor EA, Janecek S, Svensson B (2001) Relationship of sequence and structure to specificity in the α-amylase family of enzymes. Biochim Biophys Acta Protein Struct Mol Enzymol 1546:1-20. https://doi.org/10.1016/S0167-4838(00)00302-2

Malaikozhundan B, Suresh P, Seshadri S, Janarthanan S (2003) Toxicity assessment of wild bean seed protein-arcelin on Asian armyworm, *Spodoptera litura* (Fabricius). Indian J Exp Biol 41:1463-1465. http://nopr.niscair.res.in/handle/123456789/23420

Markwick NP, Laing WA, Christeller JT, Reid SJ, Ruth Newton M (1996) α-Amylase activities in larval midgut extracts from four species of Lepidoptera (Tortricidae and Gelechiidae): Response to pH and to inhibitors from wheat, barley, kidney bean, and *Streptomyces*. J Econ Entomol 89:39-45. https://doi.org/10.1093/jee/89.1.39

Matsushita H, Takenaka M, Ogawa H (2002) Porcine pancreatic α-amylase shows binding activity toward N-linked oligosaccharides of glycoproteins. J Biol Chem 277:4680-4686. https://doi.org/10.1074/jbc.M105877200

Minney BHP, Gatehouse AM, Dobie P, Dendy J, Cardona C, Gatehouse JA (1990) Biochemical bases of seed resistance to *Zabrotes subfasciatus* (bean weevil) in *Phaseolus vulgaris* common bean; a mechanism for arcelin toxicity. J Insect Physiol 36:757-767. https://doi.org/10.1016/0022-1910(90)90049-L

Mirkov TE, Wahlstrom JM, Hagiwara K, Finardi-Filho F, Kjemtrup S, Chrispeels MJ (1994) Evolutionary relationships among proteins in the phytohemagglutinin-arcelin-α-amylase inhibitor family of the common bean and its relatives. Plant Mol Biol 26:1103-1113. https://doi.org/10.1007/BF00040692

Moon J, Salzman RA, Ahn JE, Koiwa H, Zhu-Salzman K (2004) Transcriptional regulation in cowpea bruchid guts during adaptation to a plant defense protease inhibitor. Insect Mol Biol 13:283-291. https://doi.org/10.1111/j.0962-1075.2004.00485.x

Nogueira FC, Silva CP, Alexandre D, Samuels RI, Soares EL, Aragão FJ, Palmisano G, Domont GB, Roepstorff P, Campos FA (2012) Global proteome changes in larvae of *Callosobruchus maculatus* Coleoptera: Chrysomelidae: Bruchinae) following ingestion of a cysteine proteinase inhibitor. Proteomics 12:2704-15. https://doi.org/10.1002/pmic.201200039

Nunes NN, Ferreira RS, Silva-Lucca RA, de Sá LF, de Oliveira AEA, Correia MTDS, Paiva PMG, Wlodawer A, Oliva MLV (2015) Potential of the lectin/inhibitor isolated from *Crataeva tapia* bark (CrataBL) for controlling *Callosobruchus maculatus* larval development. J Agric Food Chem 63:10431-10436. https://doi.org/10.1021/acs.jafc.5b03634
Oriani MADG, Lara FM (2000) Antibiosis effects of wild bean lines containing arcelin on *Bemisia tabaci* (Genn.) biotype B (Homoptera: Aleyrodidae). An Soc Entomol Bras 29:573-582. https://doi.org/10.1590/S0301-80592000000300020

Osborn TC, Alexander DC, Sun SS, Cardona C, Bliss FA (1988) Insecticidal activity and lectin homology of arcelin seed protein. Science 240:207-210. https://doi.org/10.1126/science.240.4849.207

Osborn TC, Blake T, Gepts P, Bliss FA (1986) Bean arcelin. Theor Appl Genet 71:847-855. https://doi.org/10.1007/BF00276428

Paes NS, Gerhardt IR, Coutinho MV, Yokoyama M, Santana E, Harris N, Chrispeels MJ, Grossi de Sa MF (2000) The effect of arcelin-1 on the structure of the midgut of bruchid larvae and immunolocalization of the arcelin protein. J Insect Physiol 46:393-402. https://doi.org/10.1016/S0022-1910(99)00122-5

Paiva PM, Pontual EV, Napoelão TH, Coelho LCBB (2012) Effects of plant lectins and trypsin inhibitors on development, morphology and biochemistry of insect larva. In: Pourali K, Raad VN (ed) Larvae: morphology, biology and life cycle, Nova Science Publishers, USA, pp 37-55.

Peumans WJ, Van Damme EJM (1995) Lectin as plant defense proteins. Plant Physiol 109:347-352. https://doi.org/10.1104/pp.109.2.347

Pfaiff MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29:e45. https://doi.org/10.1093/nar/29.9.e45

Powell KS, Spence J, Bharathi M, Gatehouse JA, Gatehouse AM (1998) Immunohistochemical and development studies to elucidate the mechanism of action of the snowdrop lectin on the rice brown planthopper *Nilaparvata lugens* (Stal.). J Insect Physiol 44:529-539. https://doi.org/10.1016/S0022-1910(98)00054-7

Santino A, Valsasina B, Lioi L, Vitale A, Bollini R (1991) Bean (*Phaseolus vulgaris* L.) seed lectins a novel electrophoretic variant of arcelin. Life Sci Adv Plant Physiol 10:7-11.

Sato T, Hori K (2009) Cloning, expression, and characterization of a novel anti-HIV lectin from the cultured cyanobacterium, *Oscillatoria agardhii*. Fish Res 75:743-753. https://doi.org/10.1007/s12562-009-0074-4

Sauvion N, Nardon C, Febvay G, Gatehouse AMR, Rahbe Y (2004) Binding of the insecticidal lectin Concanavalin A in pea aphid, *Acyrthosiphon pisum* (Harris) and induced effects on the structure of midgut epithelial cells. J Insect Physiol 50:1137-1150. https://doi.org/10.1016/j.jinsphys.2004.10.006

Schoonhoven AV, Cardona C (1982) Low levels of resistance to the Mexican bean weevil in dry beans. J Econ Entomol 75:567-569. https://doi.org/10.1093/jee/75.4.567

Shade RE, Murdock LL, Foard DE, Pomeroy MA (1986) Artificial seed system for bioassay of cowpea weevil (Coleoptera: Bruchidae) growth and development. Environ Entomol 15:1286-1291.
Shaheen FA, Khaliq A, Aslam M (2006) Resistance of chickpea (*Cicer arietinum* L.) cultivars against pulse beetles. Pak J Bot 38:1237.

Silva CP, Terra WR, Xavier-Filho J, Grossi-de-Sá MF, Isejima EM, Da Matta RA, Miguens FC, Bifano TD (2001) Digestion of legume starch granules by larvae of *Zabrotes subfasciatus* (Coleoptera: Bruchidae) and the induction of α-amylases in response to different diets. Insect Biochem Mol Biol 31:41-50. https://doi.org/10.1016/S0965-1748(00)00103-X

Silva CP, Terra WR, Xavier-Filho J, Grossi-de-Sá MF, Lopes AR, Pontes EG (1999) Digestion in larvae of *Callosobruchus maculatus* and *Zabrotes subfasciatus* (Coleoptera: Bruchidae) with emphasis on α-amylases and oligosaccharidases. Insect Biochem Mol Biol 29:355-366. https://doi.org/10.1016/S0965-1748(99)00010-7

Silva CP, Xavier-Filho J (1991) Comparison between the levels of aspartic and cysteine proteinases of the larval midguts of *Callosobruchus maculatus* (F.) and *Zabrotes subfasciatus* (Boh.) (Coleoptera: Bruchidae). Comp Biochem Physiol B Comp Biochem 99:529-533.

Sparvoli F, Gallo A, Marinelli D, Santucci A, Bollini R (1998) Novel lectin-related proteins are major components in Lima bean (*Phaseolus lunatus* L.) seeds. Biochim Biophys Acta Protein Struct Mol Enzymol 1382:311-323. https://doi.org/10.1016/S0167-4838(97)00168-4

Sparvoli F, Lanave C, Santucci A, Bollini R, Lioi L (2001) Lectin and lectin-related proteins in Lima bean (*Phaseolus lunatus* L.) seeds: biochemical and evolutionary studies. Plant Mol Biol 45:587-597. https://doi.org/10.1023/A:1010647310311

Svensson B (1994) Protein engineering in the α-amylase family: catalytic mechanism, substrate specificity, and stability. Plant Mol Biol 25:141-157. https://doi.org/10.1007/BF00023233

Terra WR, Ferreira C (1994) Insect digestive enzymes: properties, compartmentalization and function. Comp Biochem Physiol B Biochem Mol Biol 109:1-62. https://doi.org/10.1016/0305-0491(94)90141-4

Vandenborre G, Smagghe G, Van Damme EJ (2011) Plant lectins as defense proteins against phytophagous insects. Phytochemistry 72:1538-1550. https://doi.org/10.1016/j.phytochem.2011.02.024

You XM, Chang SKC (1992) Effect of purified lectins on pancreatic alpha-amylase activities. J Agric Food Chem 40:638-641.

Zaugg I, Magni C, Panzeri D, Daminati MG, Bollini R, Benrey B, Bacher S, Sparvoli F (2013) QUES, a new *Phaseolus vulgaris* genotype resistant to common bean weevils, contains the arcelin-8 allele coding for new lectin-related variants. Theor Appl Genet 126:647-661. https://doi.org/10.1007/s00122-012-2008-2
Zhu Salzman K, Koiwa H, Salzman RA, Shade RE, Ahn JE (2003) Cowpea bruchid Callosobruchus maculatus uses a three-component strategy to overcome a plant defensive cysteine protease inhibitor. Insect Mol Biol 12:135-145. https://doi.org/10.1046/j.1365-2583.2003.00395.x

Tables

Table 1 N-glycosylation sites and functional domains in the amino acid sequence of the arcelin gene from P. lunatus using ScanProsite (ExPASy)

| mRNA         | Legume lectin signature | N-glycosylation site | Protein kinase C phosphorylation site |
|--------------|-------------------------|----------------------|---------------------------------------|
| P. lunatus   | Legume lectin-beta      | 7 - 10: NASF         | 21 - 23: SnK                         |
|              | 63 - 69: VAVVFDT        | 13 - 16: NFTF        | 71 - 73: SnR                         |
|              | Legume lectin-alpha     | 106 - 109: NSSS      | 108- 110: SsK                        |
|              | 137- 146: LDDWVSFGS     |                      | 128 - 130: SaK                       |

Table 2 Effects of purified recombinant arcelin protein from P. lunatus in artificial seeds on the development of C. maculatus
| Concentration of protein (w/w) in artificial seeds | No. of eggs laid per seed* | Developmental period (days) | Adult emergence (%) | Seed damage (%) |
|--------------------------------------------------|---------------------------|----------------------------|---------------------|-----------------|
| Control (0.00%)                                  | 16.1±4.3                  | 28                         | 39.94%              | 100             |
| Purified recombinant Arc (1%)                    | 14.1±1.73                 | 40                         | 10.63%              | 15              |
| Purified recombinant Arc (5%)                    | 16.8±4.57                 | 40                         | 1.78%               | 7               |

*Values represent the mean ± SD of duplicates of 10 seeds each
Table 3 Primers designed to amplify various classes of digestive enzymes in fourth instar larvae of *C. maculatus*

| Gene name                        | Primer | 5’-3’ Primer Sequence     | Accession number |
|----------------------------------|--------|---------------------------|-----------------|
| Cathepsin L-like cysteine protease | F’     | TTGGTGAGGAGAAACGTCG       | FK668948.1      |
|                                  | R’     | GCTCGGTAGACAGGTACAC       |                 |
| Aminopeptidase N (APN2)          | F’     | AGTCTCGCTAGGCTGAAGGA      | HM347433.1      |
|                                  | R’     | TCCCAAGGTGGATGACTCC       |                 |
| Glycoside Hydrolase family 31    | F’     | GTCACTGGAGGTCCGGAAGAG     | KC776577.1      |
|                                  | R’     | TCAGTATGCCCTTGGGCCA       |                 |
| Alpha amylase                    | F’     | GGCTTCGGAAACGTACTCGA      | HO204643.1      |
|                                  | R’     | TGAAAGCGACCGGCTATTTTTC    |                 |
| Cathepsin D-like aspartic protease| F’     | CACTGATCATGTGCTTTTGGTG    | FJ360435.1      |
|                                  | R’     | ACTGGGGGTACTCCATCTACAGC   |                 |

Figures
Figure 1

Specific amplification of the arcelin gene. Lane 1: 3 kb marker (5 µl), Lane 2: PCR product (5 µl), Lane 3: Negative control (5 µl)
Figure 2

Neighbor-joining tree obtained by illustrating relationships between the deduced amino acid sequences of *P. lunatus* arcelin (clone) and other closely related lectin-like polypeptides
Figure 3

SDS-PAGE (12%) analysis of Ni-NTA affinity purified recombinant arcelin protein. Lane 1: Protein standard molecular weight (kDa) marker (mid-range), Lane 2: Purified recombinant arcelin protein (70 µg)
Figure 4

Effect of purified recombinant arcelin protein from P. lunatus on the development of C. maculatus larvae at 19 days in artificial seeds
Figure 5

Gene expression plot (RQ vs different classes of digestive enzymes)