BIOCELL
2019 43(3): 155-166

Tech Science Press

# Cloning and characterization of 66 kDa streptavidin-binding peptides (SBP) of *Pisum sativum* L. embryo specific to var. Alaska

Mahmoud MOUSTAFA[1,2]; Saad ALAMRI[1]; Tarek TAHA[3]; Ali SHATI[1]; Sulaiman ALRUMMAN[1]; Mohamed AL-KAHTANI[1]

[1] Department of Biology, College of Science, King Khalid University, 9004, Abha, Kingdom of Saudi Arabia
[2] Department of Botany, Faculty of Science, South Valley University, Qena, Egypt
[3] Environmental Biotechnology Department, Genetic Engineering and Biotechnology Research Institute (GEBRI), City of Scientific Research and Technological Applications (SRTA-City), New Borg El-Arab City, 21934, Alexandria, Egypt

**Key words:** Cloning, Embryo, 66 kDa, *Pisum sativum*, Streptavidin-binding peptides (SBP)

**Abstract:** The aim of the current research was to clone and to characterize the partial 66 kDa streptavidin-binding peptide (SBP) found in the germinated embryos of *Pisum sativum* L. var. Alaska. The pea (*P. sativum* var. Alaska) embryos possess prominent 66 kDa SBPs that gradually disappeared after few hours of germination in germinated embryos, but not in the cotyledons. The total RNA was isolated from embryos of *P. sativum* but could not be isolated from the cotyledons. The partial nucleotides sequences of 66 kDa SBPs of embryonic stalk (*P. sativum* var. Alaska) were cloned and identified using *pMOSBlue* vector. 66 kDa (SBP) gene from the embryos of *P. sativum* var. Alaska possesses 327 bp having an open reading frame (ORF) region in a part of the gene that encoded for 108 amino acids. Alignment showed similarity among 66 kDa SBPs *P. sativum* var. Alaska, with *P. sativum* seed biotinylated protein (SBP65) and *P. sativum* sbp65a mRNA with DNA distance matrix between 0.0094 to 1.2676. MALDI-TOF mass spectrometry analysis of 66 kDa (SBP) proteins showed it had similar short peptides to 19 proteins found in different organisms, especially *Convicilin precursor*, and the seed biotinylated protein in *P. sativum*. The alignment results of both nucleotide sequences and amino acid residues either from cloning or MALDI-TOF-MS showed differences with related species, especially *P. sativum*. No mRNA was found in the cotyledons during seeds germination, which means no metabolic activities and this part may act only as food reservoirs for growing newly embryos.

## Introduction

Streptavidin (52.8 kDa), a nonglycosylated protein with a little acidic isoelectric point about 5, was isolated from *Streptomyces avidinii* (Green, 1975). Avidin (MW 66 kDa, diameter ~7 nm) is a globular and highly cationic glycoprotein with an isoelectric point of about 10.5 (Bajpayee *et al.*, 2014). An application of avidin (or streptavidin)-biotin interactions has been widely monitored to antibodies, to DNA and to immobilize enzymes activities. This highly specific identification is one of the fixed non-covalent interactions well known in biology science with an association constant (Ka) of $10^{15}$ $M^{-1}$ (Sassolas *et al.*, 2008). This may be due to its oligosaccharide component and to the highly positive-charged residues that can interact non-specifically with the nucleic acids and negative-charged cell surfaces, forming background issues (Bruch and White, 1982).

In general, the bioreceptors are adhered to the biotin molecules and then immobilized on streptavidin-modified surfaces. In the case of the DNA molecule, the biotin is perfectible introduced to the 5' or 3' end by cross-linking reactions, leaving the entire DNA ready for typical hybridization with its complementary target (Bonel *et al.*, 2011; Zhang *et al.*, 2011). The rigid stability nature of the biotin bonds might be due to the opposition of the complex against organic solvents, detergents, denaturants, extreme pH, proteolytic enzymes, and temperature conditions. This is widely applied in biological assays activities by coupling biotin to substrates either for proteins, DNA or RNA (Chevalier *et al.*, 1997; Chapman-Smith and Cronan, 1999).

Biotin was considered as an important cofactor of carboxylase enzyme in all aspects of life. For example, its cofactor is mainly involved in the oxidative metabolism reactions through the transfer of carbon dioxide. Biotin is bound through a biotin carboxyl carrier protein that could be a segment of a multi-domain carboxylase or by the replacement of individual subunit the enzyme of carboxylase complex (Wood and Barden, 1977; Fugate and Jarrett, 2012; Tong, 2013).

The cytoskeleton of plant cells has been found to have many streptavidin-binding carrier proteins and many

*Address correspondence to: Mahmoud Moustafa, mfmostfa@kku.edu.sa

Doi: 10.32604/biocell.2019.06814

other types, which showed various functions including the translation of mRNA, targeting, and transport (Davies *et al.,* 2001). Many of them were isolated and characterized, and various techniques were developed for such purpose (Abe *et al*., 1992, 1994; Abe and Davies, 1995). Nearly 15 specific proteins in the cytoskeleton fraction have been characterized at the level of its localization in pea (Davies *et al.*, 2001; Sami-Subbu *et al*., 2001). For example, 110 kDa (SBP) was identified by using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Abe and Davies, 1991; Abe *et al.*, 1992). Also, the rice endosperm cells having 120 kDa Tudor protein (SBP) with four SNc domains have been characterized (Sami-Subbu *et al*., 2001).

In databases (CAB87924, BAB10078), the genomic fragments of *Arabidopsis thaliana* plant encoding similar proteins, but its subcellular localization and the real function have not yet been identified. 66 kDa (SBP) cDNA sequences in cotyledons have not reported before; nevertheless, different scientists have shown that 66 kDa (SBP) in pea can be present in different subcellular locations, including the cell wall, nucleus, and in the cytoskeleton (Shibata *et al*., 1999, Moustafa *et al*., 2003; Moustafa, 2016).

MALDI-TOF-MS has received more attention from biologists and have proven the importance of the technique for rapid identification of clinical pathogens (Ferreira *et al*., 2010; Mellmann *et al*., 2009), plant-parasitic nematodes (Ahmad *et al*., 2012), and angiosperms (Kumar Vemuri *et al*., 2016).

In the present study, the DNA coding partial sequence for the 66 kDa (SBP) was cloned and sequenced. A protein band corresponding to 66 kDa (SBP) was excised from the gels and processed for MALDI-TOF mass spectrometry analysis. We showed there is no mRNA in cotyledons of pea seedlings, and we provided comparisons between nucleotides sequences and the deduced partial polypeptide sequences for 66 kDa (SBP) of *P. sativum* embryo specific to var. Alaska, resulting from cloning or MALDI-TOF-MS with the most similar sequenced gene available in the gene bank.

## Materials and Methods

*Plant materials and sample preparation*

Pea (*P. sativum* L. var. Alaska) seeds were germinated in the vermiculite for 7 days in the dark at 23°C. Three grams from cotyledons and embryos were dissected, and ground with a mortar and pestle at 4°C in cytoskeleton stabilizing buffer (CSB) made of 7 volumes of 5 mM HEPES-KOH (pH 7.5), 10 mM Mg $(OAc)_2$, 2 mM EGTA, and 1 mM PMSF with the addition of 0.5% plus polyoxyethylene-10-tridecyl ether (PTE). The homogenate from each sample was filtered by applying two layers of Miracloth (Calbiochem), and then centrifuged at 250 *g* for 15 min to furnish the nuclei pellets (Abe and Davies, 1991; 1995). The 250-*g* pellets were re-suspended using CSB in a volume equivalent to the supernatants for further use (Shibata *et al.,* 2002).

*SDS-PAGE and Western blotting*

Twenty microliters from nucleus fractions either from cotyledons or germinated embryos were analyzed by SDS-PAGE (Abe and Davies, 1991; 1995). Each fraction was mixed with 2X sample loading buffer to yield final concentrations of 2% LDS, 0.01 M Tris-HCl (pH 6.8), 20% glycerol, 1% β-mercaptoethanol, and 0.01% of bromophenol blue (BPB), heated at 95°C for 5 min, and separated by SDS-PAGE. After electrophoresis, each gel was plotted onto a PVDF membrane and subjected to the anti-apyrase antibody from rat as the primary antibody (Shibata *et al*., 1999; Moustafa *et al*., 2019) and a biotinylated anti-rat-Ig species-specific whole antibody obtained from sheep, as the secondary antibody.

Detection of binding the secondary antibody was done with streptavidin-alkaline phosphatase conjugate (Amersham Pharmacia Biotech) with 5-bromo-4-chloro-3-indolyl-phosphate (BCIP) and nitro blue tetrazolium (NBT) as substrates. In order to identify the SBPs, reactions were done without the anti-apyrase antibody (Moustafa *et al*., 2003).

*Isolation of total RNA from germinated embryos and cotyledons*

Specific weights (149 mg) from cotyledons and embryos of *P. sativum* L. var. Alaska at the intervals (24 h, 36 h, and 50 h; while, cotyledons at 1 day, 3 days, and 7 days) after germination were dissected and immediately placed in liquid nitrogen. Immediately, 450 µL Buffer RLT (10 µL B-ME was added per 1 mL Buffer RLT before use), was added and grinded thoroughly with a mortar and a pestle. Decant powder was placed into an RNase-free, liquid nitrogen-cooled, 2 mL microcentrifuge tube, shook vigorously, and incubated for 3 min at 56°C to disrupt the tissue.

The lysate was transferred onto a QIAshredder spin column (lilac), placed in 2 mL collection tube, and centrifuged for 2 min at maximum speed. The total RNA was isolated from the embryo and cotyledons by using an mRNA Purification Kit (Amersham Pharmacia Biotech) (Shibata *et al*., 2001). Finally, 50 µL RNase-free water was added directly onto the RNeasy silica-gel membrane and centrifuged for 1 min at 10000 rpm for elution. The extracted RNA was collected and tested for integrity by 1.0% (w/v) agarose gel. $A_{260/230}$ and $A_{260/280}$ values for RNA samples were determined and evaluated with an Analytic JENA Scandrop 200 to determine the purity and yield (Liu *et al.*, 2018).

*Amino acid sequences*

The same specified weight (149 mg) from embryos of P. sativum L. var. Alaska at intervals (24 h, 36 h, and 50 h) after germination were dissected, gathered on ice, and ground with a mortar and a pestle. Protein samples obtained were mixed with 5 mL of 2X sample loading buffer containing 4% LDS, 0.02 M Tris-HCl (pH 6.8), 40% glycerol, 2% β-mercaptoethanol, 0.02% BPB, and heated at 95°C for 5 min. For internal sequence determination of 66 kDa protein, 30 µL from prepared sample was subjected to SDS/PAGE (in each gel well 10 µL).

After staining the gel with Coomassie blue, the 66 kDa bands were cut and equilibrated for 10 min with 100 mm Tris/HCl (pH 6.8) containing 12% (v/v) glycerol, 50 mm 2-mercaptoethanol, and 2% (w/v) SDS (buffer A). The slices were then inserted into a gel well and covered with buffer A containing 20% glycerol (v/v). V8 protease solution (2 µL in 10 µL of buffer A) was layered onto the top (Cleveland *et al*., 1977). The sample was separated and electroblotted on PVDF membrane, then excised and subjected to N-terminal amino

acid sequencing using a protein sequencer.

*First-strand cDNA synthesis*

The first strand cDNA was synthesized by the tag-oligo-dT primers (5'-AAGAATTCTCGAGCTCCAGAATTTTTTT TTTTTTTTTTTTTTTTTTTT-3'). Starting from 10 µg of total RNA diluted with 10 µL of DEPC-water and incubated at 65°C for 5 min. After the incubation, the content was kept on ice a few minutes and then was mixed with a master mix composed of 10 µL of 5X RT buffer, 10 µL of 10 mM dNTPs, 0.5 µL of RNase Inhibitor (40 U/µL), 1 µL of Reverse Tra (100 U/µL), 8.5 µL of DEPC-water and 10 µL of a tag-oligo-dT primer (100 pmol/µL), and finally incubated at 42°C for 60 min. The synthesized first strand cDNA was purified by an equal volume of chloroform/isoamyl alcohol (24:1) and then re-purified by a Spin Column S-400 HR (Amersham Pharmacia Biotech, UK). Finally, the first strand cDNA was precipitated with 100% ice-cold ethanol and stored at 20°C until use.

*DNAs/RNAs precipitation*

Either DNAs or RNAs were precipitated by 100% ice-cold ethanol in the presence of 3 M $CH_3COONa$ (pH 5.2). The DNAs/RNAs were first added with 1/10 (v/v) of 3 M $CH_3COONa$, mixing well and were then precipitated with 2.5 volumes (v/v) of 100% ethanol, and stored at -85°C for about 30 min. The precipitated content was thawed well and warmed up at room temperature for 10 min, and then the DNAs/RNAs were collected by centrifugation at 15000 rpm for 15 min. The DNAs/RNAs were washed twice by equal volume of 70% ice-cold ethanol, centrifuged at 15000 rpm for 5 min, degassed by vacuum pump for 3 min and finally re-suspended in TE buffer.

*Gene amplification by RT-PCR*

To amplify the gene encoding the 66 kDa protein from *P. sativum*, RT-PCR was performed by using forward primer (5'-GCNCAYGCNGCNGCNAG-3) and reverse primer (5'-CTNGCNCCNCCYCTNGC-3'), which degenerated form amino acid sequence of 66 kDa protein. The PCR was performed in 25 µL reaction, using a XL-PCR kit (Amersham Pharmacia Biotech, UK). The rTth DNA polymerase of XL-PCR kit was added manually after DNA template in the reaction master mix and was preheated at 95°C for 30 s, as recommended from the kit protocol.

The polymerase starting reaction under the cycling temperature at 94°C for 15 s, annealing temperature at 47°C for 3 min, extension at 65°C for 2 min, the cycle was repeated 35 times and finally a extension phase at 67°C for 10 min was done to maximize its polymerization. 0.9% agarose gel was used to determine the DNA fragments using the mini apparatus (Bio Craft, model BE-560). The samples were mixed with 1/5 volume of loading buffer, running in 0.5X TBE electrophoresis buffer with the applied constant current of 50 V for 50 min. The gel was stained in 100 mL of ethidium bromide in electrophoresis buffer (0.75 mg/mL) for 15 min and visualized by UV light of Image Master VDS (Amersham Pharmacia Biotech).

*Subcloning of the DNA fragment of 66 kDa protein*

The PCR products were mixed with an equal volume of CIAA, vortexed for 1 min and centrifuged at 15000 rpm for 1 min. The aqueous top layer was used for subcloning. The aqueous phase containing DNA inserts was re-purified by a Spin Column S-400 HR (Amersham Pharmacia Biotech, UK) before subcloned into the *pMOSBlue vector*. After purification by CIAA that removes any residual DNA polymerase activity, the PCR products were directly ligated into *pMOSBlue* vector. The target band was isolated from agarose gel using a DNA and Gel Band Purification kit (Amersham Pharmacia Biotech, UK). The *pMOSBlue* blunt-ended cloning kit was used to clone the target DNA fragments.

*DNAs sequencing*

ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems) was used to generate the DNA strands from the purified plasmid products. The reaction contents were performed under cycling temperature of 25 cycles of 95°C for 30 s; 52°C for 15 s; 72°C for 1min. We used a vector primer U19 (5'-GTTTTCCCAGTCACGACGTTG-3') to read from downstream of the sequence. After finishing the PCR reaction, the content was purified by CIAA and then precipitated with 100% of ice-cold ethanol and was kept for 1 h at 20°C. The pellet was re-suspended in 25 µL of high formamide buffer (Short read sequence, POP4 ABI 310 Prism sequencer).

The DNA strands were sequenced by using ABI Prism 310 sequencer (Amersham Pharmacia Biotech). Sequence alignment was accomplished using ClustalW2 to estimate the match for the selected sequences, and the alignments were adjusted by using the Bioedit V7.2 and Mega 7 program for nucleotide sequence analysis.

*MALDI-TOF-MS analysis for 66 kDa proteins*

149 mg from intact embryonic tissue were dissected from the embryos, collected on ice, and ground with a mortar and a pestle. The obtained protein samples were mixed with 5 mL of 2X sample loading buffer containing 4% LDS, 0.02 M Tris-HCl (pH 6.8), 40% glycerol, 2% β-mercaptoethanol, 0.02% BPB, and heated at 95°C for 5 min. 10 µL were loaded on 10% polyacrylamide resolving gel and 5% stacking gel. The gel was fixed in 100 mL 46% MeOH, 7% HAc for 1 h followed by staining in 100 mL 46% MeOH, 7% HAc, 0.1% Coomassie Brilliant Blue for 1 h. The gel was distained in 100 mL 5% MeOH, 7.5% HAc for 24 h. 66 kDa protein was cut out from the gel by new razor blades. The band was stored in 1% acetic acid in a clean and sealed tube and analyzed by MALDI-TOF-MS (Kumar Vemuri and Veeravalli, 2014).

## Results

*66 kDa (SBP) in cotyledons and in embryos of* P. sativum *var. Alaska*

Pea seeds were used without any prior imbibition and sown directly in vermiculite and allowed to germinate for up to 7 days. Protein was extracted from embryos at 24 h, 36 h, and 50 h while cotyledons at 1 day, 3 days, and 7 days after

germination. The protein samples were subjected to SDS-PAGE, blotted and probed with an antibody to apyrase as the primary antibody and anti-rat as the secondary antibody. The binding of the anti-rat was detected with streptavidin-alkaline phosphatase conjugate; results are shown in Fig. 1.

1, 3, and 7-day germinated cotyledons revealed 66 kDa SBPs (Moustafa *et al.*, 2003) and many others SBPs, as noted previously (Abe *et al.*, 1992; Abe *et al.*, 1994; Moustafa, 2014). 1, 3, and 7-day germinated cotyledons (lanes 1, 2, and 3) revealed 66 kDa (SBP) and two prominent weakly antigen Ig to apyrase (WAA) at 47 and 51 kDa remaining constant up to 7 days after germinations. Also, cotyledons revealed 66 kDa streptavidin–binding proteins and also many others (Moustafa *et al.*, 2003; Abe *et al.*, 1992; Abe *et al.*, 1994; Moustafa, 2014). 24, 36, and 50 -h germinated embryos (lanes 4, 5, and 6) revealed 66 kDa (SBP) and two prominent weakly antigenic antiapyrase antibody (WAA) at 47 and 51 kDa that disappeared gradually at 50 h after germinations. SDS-PAGE revealed two undefined proteins (UDP) at 49 and 49.5 kDa that disappeared in embryos and in cotyledons through germinations.

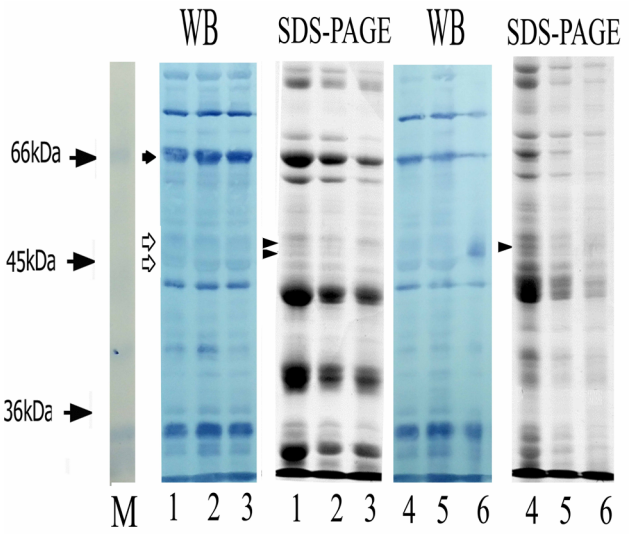


**FIGURE 1.** Western Blotting and SDS-PAGE of nucleus pellets. Germinated cotyledons of P. sativum L.var. Alaska at 1-d , Lane 1; 3-d, Lane 2; 7-d, Lane 3; and embryo at 24 -h, Lane 4; 36 -h, Lane 5; and 50 -h, Lane 6. Lane M, molecular weight markers. Open arrows are (WAA), 47 kDa and 50 kDa; closed arrow 66 (SBP); Triangle (UIP) 48 kDa and 49.5 kDa.

*Cloning of 66 kDa SBP protein*

RNA prepared from germinated embryos was examined by electrophoresis on 1% agarose gel and showed a 28s rRNA band equal to or more quantity than the 18s rRNA band, revealing that little or no RNA degradation happened during the extraction (Fig. 2(A)). While, RNA prepared from germinated cotyledons was also examined by electrophoresis on 1% agarose gel and showed neither 28s rRNA nor 18s rRNA bands (Fig. 2(D)).

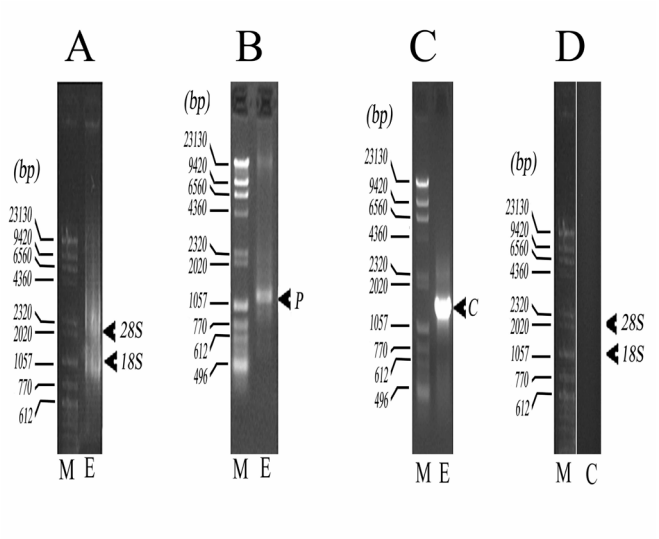


**FIGURE 2.** The cloning procedure of DNA fragment of 66 kDa protein. Panel A: Ethidium bromide stained gel showing total RNA from *P. sativum*. M = Molecular markers. Panel B: Determination of RT-PCR product resulted from degenerated primers as indicated in Material and Methods using 0.09% agarose gel electrophoresis (band p). M = Molecular markers. Panel C: Purified PCR fragment (band p) was cloned into the pMOSBlue vector according to the manufacturer's protocol and determined by 0.09% agarose gel electrophoresis (band c). M = Molecular markers. Panel D: Ethidium bromide stained gel showing no RNA from cotyledons of *P. sativum*. M = Molecular markers.

RT-PCR was performed by using synthetic primers based on the partial amino acid sequence degenerated from 66 kDa protein, cloned into the *pMOSBlue* vector, the product was determined by 0.9% agarose gel electrophoresis, and the results are shown in Figs. 2(B)-2(C). Based on the ABI Prism 310 sequencer program, raw data was assembled automatically, and edited to the message, which had been used to align in order to identify common structures (Fig. 3). The nucleotide sequences were compared with those from the Genbank database by using BLAST search program algorithm of NCBI (National Centre for Biotechnology Information) to align the homology and identify sequences. Sequence analysis demonstrated that these 327 bp contained open reading frame (ORF) and the predicted amino acid sequence is specific to 66 kDa protein of *P. sativum* embryo var. Alaska, and there was not a high degree of sequence homology with any other known proteins in the data base. These results prompt us to do MALDI-TOFF analysis for this protein.

1 TCACTCTTTCAGAAAGCTGATCATCTGGAGGATCTTGA
AGACGAGCAAGTCATAGCAAAA
S L F Q K A D H L E D L E D E Q V I A K
60 TTGGAACCCAATAGAAATCCATTTCACATTGCCATTAG
ATTAATGTCTGTTTTTATCTGT
L E P N R N P F H I A I R L M S V F I C
120 TGTGCGATTACCTCTTTCCAGGGATTGCTTCCTGATGT
AAATGCCTATTCAGAGGTCGTT
C A I T S F Q G L L P D V N A Y S E V V
240 CAATCAATAAAATCATGTTATTTAGTATATCTCTGTTTT
CATTTTCATTTTTCTGTTTTG

```
      Q S I K S C Y L V Y L C F H F H F S V L
300 TTTGCAAAAATGACGTCCGAATTTTTTGATAAACATTG
CATCATGACACATAAGAGCTTT
      F A K M T S E F F D K H C I M T H K S F
327 ACAGATACATGCTTAATAAACATTTAA
      T D T C L I N I *
```

**FIGURE 3.** Partial nucleotide sequence and predicted amino acid sequence (single-letter amino acid code) of 66 kDa protein. The translation stop codon is shown with an asterisk.

*Nucleotide sequence analysis*

The present phylogenetic investigations dependent on seed biotinylated protein either 66 or 65 kDa partial nucleotides sequences showed a minor genetic divergence between *P. sativum* seed biotinylated protein (SBP65) and *P. sativum* sbp 65a mRNA, however 66 SBP *P. satvium* var. Alaska showed distant relationship with them (Fig. 4).

A comparison of partial nucleotides sequence is shown in Fig. 5. The *P. sativum* seed biotinylated protein (SBP65) and *P. sativum* sbp 65a mRNA both differ from 66 (SBP) *P. satvium* var. Alaska sequence. DNA distance matrix between 66 (SBP) *P. satvium* var. Alaska *P. sativum* seed biotinylated protein (SBP65) and *P. sativum* sbp 65a mRNA (1.2473 to 1.2676, respectively); while, between *P. sativum* seed biotinylated protein (SBP65) and *P. sativum* sbp 65a mRNA is 0.0094 (Tab. 1).

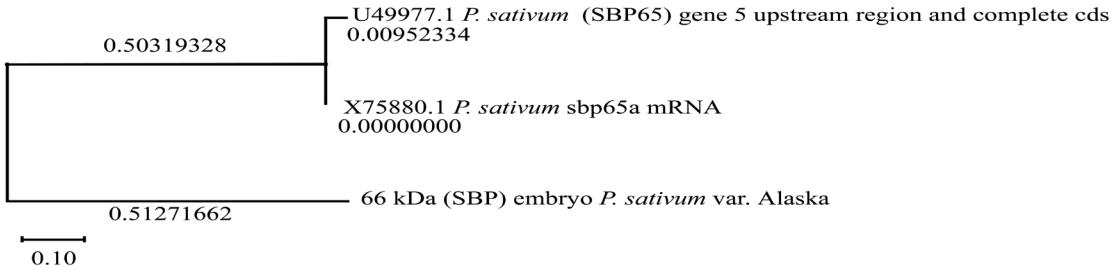


**FIGURE 4.** Phylogenetic analysis of *P. sativum* species based on aligned sequence of biotyinlated proteins (66 and 65 kDa SBP) using the ClustalW software. Phylogenetic analyses were conducted using the MEGA 5 software.

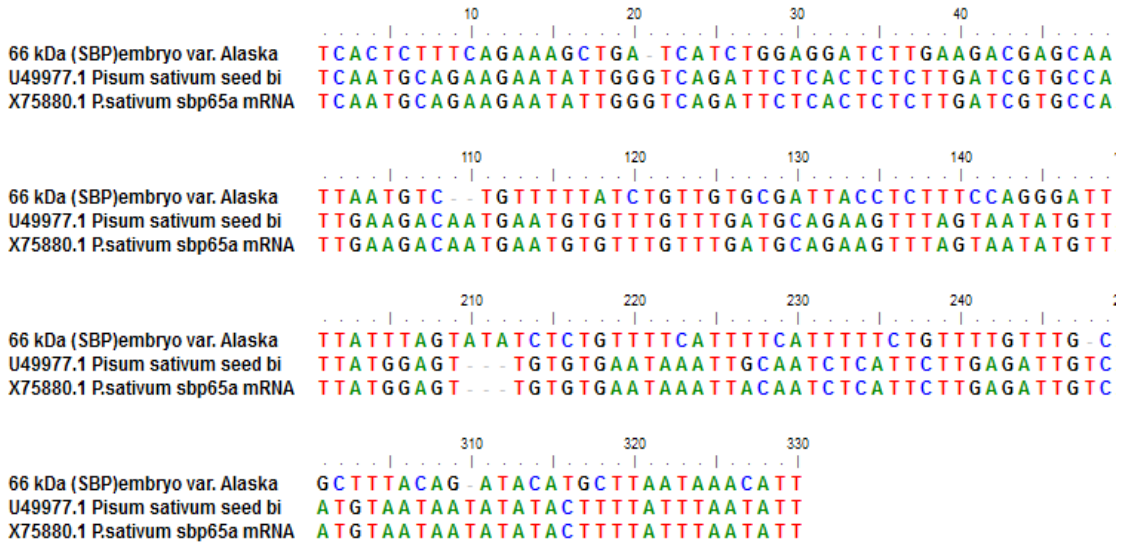


**FIGURE 5.** Partial nucleotides sequence of 66 (SBP) *P. satvium* var. Alaska, *P. sativum* seed biotinylated protein (SBP65) and *P. sativum* sbp 65a mRNA. Similar colours show homology among sequences.

**TABLE 1**

**DNA distance matrix**

| | 66 kDa (SPB) | U49977.1_P | X75880.1_P |
|---|---|---|---|
| 66 kDa (SBP) *P. sativum* embryo var. Alaska | 0.0000 | 1.2676 | 1.2473 |
| U49977.1 *P. sativum* seed biotinylated protein (SBP65) gene 5 upstream region and complete cds | 1.2676 | 0.0000 | 0.0094 |
| X75880.1 *P. sativum* sbp65a mRNA | 1.2473 | 0.0094 | 0.0000 |

**TABLE 2**

**Compute nucleotide composition**

| | T(U) | C | A | G | Total |
|---|---|---|---|---|---|
| 66 kDa (SBP) *P. sativum* embryo var. Alaska | 36.7 | 17.9 | 29.6 | 15.7 | 324.0 |
| U49977.1 *P. sativum* seed biotinylated protein (SBP65) gene 5 upstream region and complete cds | 38.9 | 10.8 | 31.5 | 18.8 | 324.0 |
| X75880.1 *P. sativum* sbp65a mRNA | 38.9 | 11.1 | 31.8 | 18.2 | 324.0 |
| Avg. | 38.2 | 13.3 | 31.0 | 17.6 | 324.0 |

The compute nucleotide composition in 66 (SBP) *P. satvium* var. Alaska, *P. sativum* seed biotinylated protein (SBP65) and *P. sativum* sbp65a mRNA are shown in (Tab. 2). The average of all the three *Pisum* species nucleotide sequences had total of 324 positions, in the final data set. It revealed that 66 (SBP) *P. satvium* var. Alaska had fewest Guanine (15.7%), Adenine (29.6%) and Thymine (36.7) and more Cytosine (17.9%) than *P. sativum* seed biotinylated protein (SBP65) and *P. sativum* sbp65a mRNA.

Also, the partial nucleotide sequence of 66 kDa (SBP) showed homology with *P. sativum* DNA sequence from clone JICPSV-297I9, complete sequence (Sequence ID: CU655882.7 Length: 108562) with identities 312/327 (95%) and Gaps 1/327 (0%). No significant similarity was found in the BLAST search engine for the deduced amino acid.

*Identification of 66 kDa (SBP) protein by MALDI-TOF-MS analysis*

Expected amino acid sequences of the 66 kDa (SBP) were investigated by MALDI-TOF mass spectrometry for the purpose of fingerprinting. Ionization spectrum for the expected peptides showed 19 prominent proteins and was applied to the National Centre for Biotechnology Information database for non-redundant protein sequence to match similar peptides utilizing the Mascot Peptide Search program (Perkins *et al.*, 1999).

Fig. 6 summarizes information of the expected amino acid sequences of 66 kDa (SBP) resulted from MALDI-TOF MS using Mascot as a search program with other organism partial amino acid sequences. This data demonstrated that nucleotides of 66 kDa gene is to somewhat similar by various degree to many other genes, especially near seed biotinylated proteins of 65 kDa of apparent molecular mass.

MISLKNSFLSGSCSSPVPGKQRAGLVVPTAPSSLKVVALN
PNRERNGSIMMESTLQEMRDGASVLDLDPKSTVAGGVR
DVYGEDTATEDQFVTPWSLSVASGYSLLRDPHHNKGLA
FNDKERDAHYLRGLLPPAIVSQELQVKKMMHIIRQYQL
PLQK**YMAMMDLQER**NERLFYKLLIQNVEEMLPIVYTPT
VGEACQKYGSIFGRPQGLYISLKEKGRILEVLRNWPEKNI
QVIVVTDGER**ILGLGDLGCQGMGIPVGK**LSLYTALGGV
RPSSCLPVTIDVGTNNEK**LLNDEFYIGLR**QRRATGQEYA
ELLHEFMTAVKQNYGERVLVQFEDFANHNAFDLLAKYG
TTHLVFNDDIQGTASVVLAGLVAALKLVGGSLADHRFL
FLGAGEAGTGIAELIALEMSKQTNMPVEETRKKIWLVDS
KGLIVSSRMDSLQHFKRPWAHEHEPIKTLLDAVNDIKPT
VLIGTSGVGRTFTKEVVEAMASFNEKPIILALSNPTSQSE
CTAEEAYTWSQGRAIFASGSPFAPVEYEGKVYVPGQAN
NAYIFPGFGLGLIMSGTIRVHDDMLLAASEALAAQVTQE
NFDKGLIYPPFTNIRK**ISANIAANVAAK**AYELGLASRLPQ
PKDLVKYAES CMYSPAYRSYR

> ***Sequence ID: CAB82855.1-Convicilin precursor.-Pisum sativum***

MATTIKSRFPLLLLLGIIFLASVVCVTYANYDEGSEPRVPA
QRERGRQEGEKEEKRHGEWRPSYEKEEDEEEGQRERGR
QEGEKEEKR**HGEWRPSYEKQEDEEEK**QKYRYQREKED
EEEKQKYQYQREKKEQKEVQPGRER**EREEDEEQVDEE**
**WRGSQR**REDPEERARLRHREERTKRDRRHQREGEEEER
SSESQER**RNPFLFK**SNK**FLTLFENENGHIR**LLQRFDK**RSD**
**LFENLQNYR**LVEYRAKPHTIFLPQHIDADLILVVLSGK**AI**
**LTVLSPNDR**NSYNLERGDTIK**LPAGTTSYLVNQDDEED**
**LRLVDLVIPVNGPGKFEAFDLAK**NKNQYLRGFSK**NILE**
**ASYNTR**YETIEK**VLLEEQEKDR**KRRQQGEETDAIVKVSR
EQIEELKKLAKSSSKK**SLPSEFEPINLR**SHKPEYSNKFGK**L**
**FEITPEKK**YPQLQDLDLFVSCVEINEGALMLPHYNSR**AI**
**VVLLVNEGKGNLELLGLKNEQQER**EDRKERNNEVQRY
EARLSPGDVVIIPAGHPVAITASSNLNLLGFGINAENNER
**NFLSGSDDNVISQIENPVKELTFPGSVQEINR**LIKNKQS
HFANAE PEQKEQGSQGKRSPLSSILGTFY

> ***Sequence ID: P13915.1-Convicilin precursor (clone pJC4)-Pisum sativum***

MATTVKSRFPLLLFLGIIFLASVCVTYANYDEGSETRVPG
QRERGRQEGEKEEKRHGEWRPSYEKEEHEEEKQKYRYQ
REKKEQKEVQPGRERWEREEDEEQVEEEWRGSQRREDP
EERARLRHREERTKRDRRHQREGEEEERSSESQEHRNPF
LFKSNK**FLTLFENENGHIR**RLQRFDK**RSDLFENLQNYR**L
VEYRAKPHTIFLPQHIDADLILVVLNGK**AILTVLSPNDR**
NSYNLERGDTIK**IPAGTTSYLVNQDDEEDLRVVDFVIPV**
**NRPGKFEAFGLSENK**NQYLRGFSK**NILEASLNTK**YETIE
K**VLLEEQEKKPQQLR**DRKRTQQGEERDAIIKVSREQIEE
LRKLAKSSSKK**SLPSEFEPFNLR**SHKPEYSNKFGK**LFEITP**
**EKK**YPQLQDLDILVSCVEINK**GALMLPHYNSRAIVVLLV**
**NEGKGNLELLGLKNEQQER**EDRKERNNEVQRYEARLSP
GDVVIIPAGHPVAISASSNLNLLGFGINAKNNQR**NFLSGS**
**DDNVISQIENPVKELTFPGSSQEVNR**LIKNQK**QSHFASA**
**EPEQKEEESQR**KRSPLSSVLDSFY

> ***Sequence ID: Q41060.1- Seed biotin-containing protein SPB65 [validated]-garden pea***

MASEQLSRRENITTERKIQNAEDSVPQRTTHFELR**ETHE**
**LGPNFQSLPR**NENQAYLDRGAPLSANVSESYLDRARVPL
NANIPEHRVREKEDFGGVRDMGKFQMESKGGNK**SLAE**
**DRETLDTR**SRMVTGTPHIKEASGKGQVVEERERARERA
MEEEEK**RLTMEEISK**YR**NQAQQSALEALSAAQEK**YERA
KQATNETLRNTTQAAQEKGEAAQAKDATFEKTQQGYE
MTGDTVSNSARTASEKAAQAKNTTLGKTQQGYEATRD
TVSNAARTAAEYATPAAEKARCVAVQAKDVTLETGKTA
AEKAKCAAEIAAKVAVDLKEK**ATVAGWTASHYATQLTV**
**DGTRAAANAVEGAVGYVAPK**ASELAAKSVETVKGLAA

SAGETAKEFTARKKEESWREYEAK**RASQLQEGEEILPST GIGK**VLPSGERTQAQGTNLQEKVQGKGSDILGAVTETV SDIGSSMIKPIDNANTKVKEHGGTTITPKGQDAGGVLD AIGETIAEIAHTTK **VIVVGEDDEV**

> ***Sequence ID: CAA72090.1-62K sucrose-binding protein homolog-Pisum sativum***

MAIKTKLSLTIFLFFLLALLCSNLAVGRKEKDPELTTCKD QCDMQRQYDEEDKRICMERCDYIKKKQERQKHKEHEE EEEQEQEEDENPYVFEDNDFETKIDTKDGRVLILNKFNE KSKLLK**NIENYGLVLEIKANAFLSPHHYDSEAILFNIK**G R**GIIGLVAEDR**TER**FNLEEGDIMRVPAGTPMYLNRDEN EK**LYIAAFHMPPSSGSAPVNLEPFFESAGR**KPESVLNTFS SK**VLQAALKSSK**GELETVLDEQKK**GRIFKIEKEDVRGLA PKKSLWPFGGPFK**SPFNIFSNNPAFSNKFGSLFEVGPSQ EK**SGLEGLNLMLTLANITKGSMSTIHNTNANKIALVIDG EGELEMACPHMPSSSSNSRQKKSSISYHNINAKLRPGVM FVVPAGH PFVNIASKKKNLIVVCFEVN AQRNKKLALA GKKNIVSALD KAAKEVAFDIAAEKVDEVFERK**EEFFFPYD NEER**KEEHGRAVV

> ***Sequence ID: CAA96513.1-Convicilin precursor.-Vicia narbonensis (Narbonne vetch)***

MATAMKSRFPVLLLLGIIILASLCVTYANYDEGTEPRVPG QRERGRQEGEKEEKRHGEWRPSHEKEAQPGRRERWETS EEEERVDEEWRGSQRHEDPEERARERYRAEERERRRQW EGEEKEGSSKSQER**RNPFLFK**SNK**FLTLFENENGHIR**RLQ RFDK**RSDLFENLQNYR**LVEYRAKPHTIFLPQHIDADLILT VLSGR**AILTVLSPNDR**NSYNLERGDTIKLPAGTTSYLLNQ DDEEDLRVVDLSISVNRPGKVESFGLSGSKNQYLRGFSK **NILEASLNTK**YETIEKVLLEEPQQSIGQKRRSQRQETNAL VKVSREQVEELKRLAKSSSKKGVSSEFEPFNLRSQNPKYS NKFGK**LFEITPEKK**YPQLQDLDIFVSSVEINEGGLMLPH YNSRAIVILLVNEGKGNLELVGLKNEQQEQRERERDEQQV QRYEARLSPGDVVIIPAGHPVAVSASSNLNLLGFGINAEN NQRNFLTGSDDNVISQIENPVKELTFPGSAQEVNRLLKN QEHSHFAN AEPEQKGEES QRKRSPISSILGTFN

> ***Sequence ID: CAB89812.1-Convicilin (Fragment).-Lens culinaris (Lentil)***

MATTIKSRFPLLLLLGIIFLAFVCVAYANDDEGSEPRVTGQR ERGRQEGEKAEQSREQSPGQWRPSHGKEEDEEEKEQKEA QSGREKWERKEDEEKVVEEEEGEWRGSQRHGDPEERSRQ RHREEKTKRQVEEQTEEKDRRYQHEGKEEETSSESQER**RN PFLFK**SNK**FLTLFENENGHIR**RLQRFDK**RSDLFENLQNYR** LVEYRAKPHSIFLPQHIDAEFIVVVLSGK**AILTVLSPNDR**NS YNLERGDAIKSPAGATYYLVNPDDEEDLRVVDFVISLNRPG KFEAFDLSANRRQYLRGFSKSVLEASLNTKYDTIEKVLLEEQ ENEPHQRRDRKGRPQGQEKHAIVKVSREQIEELRRLAKSSS KK**SLPSEFEPFNLR**SQNPKYSNKFGKFFEVTPEKKYPQLQD LDLLVSSVEINEGGLLLPHYNSR**AIVVLLVNEGK**GNLELVG FKNEQQEREDNKERNNEVQRYEARLSPGDVVIIPAGHPVSI SASSNLNLLGFGINAENNERNFLTGSDDN

> ***Sequence ID: P25698.2-GMTEFS1 NID: -Glycine max***

MGKEKVHISIVVIGHVDSGKSTTTGHLIYKLGGIDKRVI ERFEKEAAEMNKRSFKYAWVLDKLKAERERGITIDIALW KFETTK**YYCTVIDAPGHR**DFIKNMITGTSQADCAVLIID STTGGFEAGISKDGQTREHALLSFTLGVKQMICCCNKM DATTPKYSKARYDEIVKEVSSYLKKVGYNPDKIPFVPISG FEGDNMIER**STNLDWYK**GPTLLDALDQISEPKRPSDKPL R**LPLQDVYKIGGIGTVPVGRVETGVLKPGMVVTFAPT GLTTEVK**SVEMHHESLTEAHPGDNVGFNVKNVAVKDL KRGYVASNSKDDPAKEAANFTAQVIIINHPGQIGNGYAP VLDCHTSHIAVKFAELMTKIDRRSGKELEKEPKFLKNGD AGFVKMIPTKPMVVETFSEYPPLGRFAVRDMR**QTVAVG VIK**NVE KKDPTGAKVT KAAQKKK

> ***Sequence ID: BAA34348.1-Elongation factor-1 alpha.-Nicotiana paniculata***

MGKEKVHINIVVIGHVDSGKSTTTGHLIYKLGGIDKRVI ERFEKEAAEMNKRSFKYAWVLDKLKAERERGITIDIALW KFETTK**YYCTVIDAPGHR**DFIKNMITGTSQADCAVLIID STTGGFEAGISKDGQTREHALLAFTLGVKQMICCCNKM DATTPKYSKARYDEIKEVSSYLKKVGYNPDKIPFVPISGFE GDNMIER**STNLDWYK**GPTLLEALDQINEPKRPTDKPLR **LPLQDYKIGGIGTVPVGRVETGVLKPGMLVTFGPTGLT TEVK**SVEMHHEALQEALPGDNVGFNVKNVAVKDLKRG FVASNSKDDPAKGASSFTSQVIIMNHPGQIGNGYAPVLD CHTSHIAVKFAEILTKIDRRSGKELEKEPKFLKNGDAGM VKMIPTKPMVVETFSEYPPLGRFAVRDMR**QTVAVGVIK** NVD KKDPTGAKVT KAAQKKK

> ***Sequence ID: AAF60327.1-Keratin-(67 kDa cytokeratin) (Hair alpha protein)***

SRQFSSRSGYRSGGGFSSGSAGIINYQRRTTSSSTRRSGGG GGRFSSCGGGGGSFGAGGGFGSRSLVNLGGSKSISISVARG GGRGSGFGGGYGGGGFGGGGFGGGGFGGGGIGGGGFG GFGSGGGGFGGGGFGGGGYGGGYGPVCPPGGIQEVTIN QSLLQPLNVEIDPEIQKVKSREREQIKSLNNQFASFIDKVR **FLEQQNQVLQTK**WELLQQVDTSTRTHNLEPYFESFINNL RRRVDQLKSDQSRLDSELKNMQDMVEDYRNKYEDEINK R**TNAENEFVTIKK**DVDGAYMTKVDLQAKLDNLQQEID FLTALYQAELSQMQTQISETNVILSMDNNRSLDLDSIIAEV KAQNEDIAQKSKAEAESLYQSKYEELQITAGRHGDSVRNS KIEISELNRVIQRLRSEIDNVKKQISNLQQSISDAEQRGENA LKDAKNKLNDLEDALQQAKEDLARLLRDYQELMNTKLA LDLEIATYR**TLLEGEESR**MSGECAPNVSVSVSTSHTTISGG GSR**GGGGGGYGSGGSSYGSGGGSYGSGGGGGGGR**GSY GSGGSSYGSGGGSYGSGGGGGGHGSYGSGSSSGGYRGGS GGGGGGSSGGRGSGGGSSGGSIGGRGSSSGGVKSSGGSSS VRFVSTTYSG VTR

> ***Sequence ID: O49169.1-AF041463 NID: -Manihot esculenta***

MGKEKVHINIVVIGHVDSGKSTTTGHLIYKLGGIDKRV IERFEKEAAEMNKRSFKYAWVLDKLKAERERGITIDIAL WKFETTK**YYCTVIDAPGHR**DFIKNMITGTSQADCAVL IIDSTTGGFEAGISKDGQTREHALLAFTLGVKQMICCCN KMDATTPKYSKARYDEIVKEVSSYLKKVGYNPDKIPFVP ISGFEGDNMIER**STNLDWYK**GPTLLEALDQIQEPKRPSD KPLR**LPLQDVYKIGGIGTVPVGRVETGILKPGMVVTF GPTGLTTEVK**SVEMHHEALQEALPGDNVGFNVKNVAV KDLKRGIVASNSKDDPAKEAANFTSQVIIMNHPGQIGN GYAPVLDCHTSHIAVKFAEILTKIDRRSGKELEKEPKFLK NGDAGFV KMIPTKPMVVETFSAYPPLG RFAVRDMR**QT VAVGVIK**SVE KKDPSGAKVT KSAAKKGGK

> ***Sequence ID: BAC22126.1-Eukaryotic elongation factor 1A.-Suaeda japonica***

MGKEKIHISLVVIGHVDSGKSTTTGHLIYKLGGIDKRVIE RFEKEAAEMNKRSFKYAWVLDKLKAERERGITIDIALWK FETNK**YYCTVIDAPGHR**DFIKNMITGTSQADCAILIIDS TTGGFEAGISKDGQTREHALLSFTLGVRQMICCCNKMD ATTPKYSKARYDEIVKEVSSYLKKVGYNPEKVPFVPISGF EGDNMIER**STNLDWYK**GPTLLEALDMINEPKRPSDKPL R**LPIQDVYKIGGIGTVPVGR**IETGVLKPNMVVTFGPTG

LTTEVKSVEMHHESLPEALPGDNVGFNVKNVSIKDLKR
GYVASDSKNDPAKGASSFTAQVIIMNHPGQIGNGYAPVL
DCHTSHIAVKFAELLTKIDRRSGKELEKEPKFLKNGDAG
MVKMIPTKPMVVETFAEYSPLGRFAVRDMR**QTVAVGVI**
**K**SVD KKEPTSAKVT KAAMKKK

*> Sequence ID: O24534.1-VFEF1A NID: -Vicia faba*

MGKEKVHINIVVIGHVDSGKSTTTGHLIYKLGGID
KRVIERFEKEAAEMNKRSFKYAWVLDKLKAERERGI
TIDIALWKFETSK**YYCTVIDAPGHR**DFIKNMITGTS
QADCAVLIIDSTTGGFEAGISKDGQTREHALLAFTL
GVKQMICCCNKMDATTPKYSKGRYEEIVKEVSSYL
KKVGYNPDKIPFVPISGFEGDNMIER**STNLDWYK**G
PTLLDALDNINEPKRPSDKPLR**LPLQDVYK**IGGIGIV
PVGR**VETGVVKPGMLVTFAPTGLTTEVK**SVEMHH
EALTEALPGDNVGFNVKNVAVKDLKRGFVASNSKD
DPAKEAANFTSQVIIMNHPGQIGNGYAPVLDCHTS
HIAVKF AELITKIDRRSGKEIEKEPKFLKNGDAGMV
KMIPTKPMVVETFAEYPPLGRFAVRDMR**QT**
**VAVGVIK**SVE KKDPTGAKVT KAAAKKK

*> Sequence ID: Q41011.1-PSEF1ALPH NID: -Pisum sativum*

MGKEKVHINIVVIGHVDSGKSTTTVHVIYKLGGIDKRV
IERFEKEADEMNKRSFKYAWLLDKLKAERERGITIDIAL
LKFETTKYYSTVMDAPGHRDFIKNMITGTSQADCAVLI
IDSTTGGFEAGISKDGQTREHALLAFTLGVKQMICCCN
KMDATTPKYSKGRYEEIVKEVSSYLKEVGYNPDKIPFVPI
SGFEGDNMIER**STNLDWYK**GPTLLDALDNINEPKRPSD
KPLR**LPLQDVYKIGGIGTVPVGRVETGVVKPGMLVTF**
**APTGLTTEVK**SVEMHHEALTEALPGDNVRFNVKNVAV
KDLKHGLVASNSKDDPAKDAANFTSQVIIMNHPGQIGN
GYAPVLDCHTSHIAVKFAELITKIDRRSGKEIEKEPKFLK
NGDAGMVKMIPTKPMVVETFAEYPPLGRFAVRDMR**QT**
**VAVGVIK**SVE KKDPTGAKVT KAAAKKK

*> Sequence ID: AAD56019.1-Elongation factor-1 alpha 2.-Lilium longiflorum*

MGKEKVHINIVVIGHVDSGKSTTTGHLIYKLGGIDKRV
IERFEKEAAEMNKRSFKYAWVLDKLKAERERGITIDIAL
WKFETTK**YYCTVIDAPGHR**DFIKNMITGTSQADCAILI
IDSTTGGFEAGISKDGQTREHALLAFTLGVKQMICCCN
KMDATTPKYSKARYDEIVKEVSSYLKKVGYNPDKIPFVP
ISGFEGDNMIERSINLDWYKGPTLLEALDMINEPKRPSD
KPLR**LPLQDVYKIGGIGTVPVGRVETGVLKPAMVVTF**
**GPTGLTTEVK**SVEMHHEALVEALPGDNVGFNVKNVAV
KDLKRGFVASNSKDDPAKEAANFTSQVIIMNHPGQIGN
GYAPVLDCHTSHIAVKFNEILTKIDRRSGKELEKEPKFLK
NGDAGMIKMIPTKPMVVETFSEYPPLGRFAVRDMR**QT**
**VAVGVIK**NVE KKEPTGAKVT KSAVKKK

*> Sequence ID: CAA68708.1-Vicilin precursor (clone pDUB9)--Pisum sativum*

SSRSDPQNPFIFKSNKFQTLFENENGHIRLLQKFDQRSK**IF**
**ENLQNYR**LLEYKSKPHTIFLPQHTDADYILVVLSGKAILT
VLKPDDRNSFNLERGDTIK**LPAGTIAYLVNR**DDNEELRV
LDLAIPVNRPGQLQSFLLSGNQNQQNYLSGFSK**NILEASF**
**NTDYEEIEK**VLLEEHEKETQHRRSLKDKRQQSQEENVIV
KLSRGQIEELSKNAKSTSKKSVSSESEPFNLRSRGPIYSNEF
GKFFEITPEKNPQLQDLDIFVNSVEIK**EGSLLLPHYNSR**A
IVIVTVNEGKGDFELVGQRNENQQEQRKEDDEEEEQGE
EEINKQVQNYKAKLSSGDVFVIPAGHPVALKASSNLDLL
GFGINAENNQRNFLAGDEDNVISQVQRPVKELAFPGSA
QEVDRILENQKQSHFADAQPQQRE RGSRETRDRL SSV

*> Sequence ID: AAD05191.1-Keratin, type II cytoskeletal 1 (Cytokeratin 1) (67 kDa cytokeratin).-Mus musculus*

SLQCSSRSLCRGGGGSRNFSSGSAGLVSFQRRSTSSSMR
RSGGGGGGRFSGGGFCGSSGSGFGSKSLMNLGGGRSI
SKSVAGGGGSFCGGFGGGSYGRGGFGGGSYGGGGFG
GGSFGGGGFGGSGFGGGSGGGGGFGSGGGFGGGRFG
GYGPVCSPSGIQEVTLNQSLLQPLNVKVDPQIQKVKSQ
EREQIKSLNDKFASFIDKVR**FLEQQNKVLQTK**WELLQ
QVDTTTRTQNLDPFFENYISILRRKVDSLKSDQSRMES
ELKNMQDLVEEYRTKYEDEINKR**TNAENEFVTIKK**DV
DSAYMTKVELQAKRDALQQDINFFSTLYQMEMSQMQ
TQISETNVVLSMDNNRQFDLDGIISEVKAQYDSICQRS
KAEAETFYQSKYEELQITAGKHGDSVRNTKMEISELNR
MIQRLRSEIDGCKKQISQIQQNINDAEQRGEKALKDAQ
NKLNEIEDALSQCKEDLARLLRDFQELMNTKLALDME
IATYKKLLEGEEIRMSGECTPNVSVSVSTSHTSMSGSSSR
GGGSGGGRYGGGGSYGGGSGGGSYGGSSGGGGSGGS
YGGGSGGGSSGSHRGGSGGGGGSSGGSYGGSSGGGRG
GSSSGGGGVKSSGSSTVKFVSTSYSRGTK

*> Sequence ID: CAB65347.1-Translation elongation factor 1 alpha (Fragment).-Phytophthora infestans (Potato late blight fungu)-Phytophthora infestans*

VIGHVDAGKSTTTGHLIYKCGGIDKRTIEKFEKEAAELG
KTSFKYAWVLDNLKAERERGITIDIALWKFESPKYFFTVI
DAPGHRDFIKNMITGTSQADCAILVVASGVGEFEAGISK
EGQTREHALLAFTLGVKQMVVAINKMDDSSVMYGQAR
YEEIKSEVTTYLKKVGYKPAKIPFVPISGWEGDNMIDR**S**
**TNMPWYK**GPFLLEALDNLNAPKRPSDKPLR**LPLQDVY**
**KIGGIGTVPVGR**VETGVIKPGMVATFGPVGLSTEVKSV
EMHHESLPEAVPGDNVGFNVKNVSVKELRRGFVASDSK
NDPAKATQDFTAQVIVLNHPGQIGNGYSPVLDCHTAH
VACKFKEITEKMDRRSGKVLETAPKFVKSGDACMVILE
PSKPMTVESFQEYPPLGRFAVRDMR**QTVAVGVIK**SVNK
KEASGKGGAK KK

*> Sequence ID: P14856.1-lipoxygenase (EC 1.13.11.12) 2 [similarity]-Pisum sativum*

MFPNVTGLLNKGHKIRGTVVLMRKNVLDFNTIVSIGGG
NVHGVIDSGINIIGSTLDGLTAFLGRSVSLQLISATKSDAN
GKGKVGKDTFLEGVLASLPTLGAGESAFNIHFEWDHEM
GIPGAFYIKNYMQVEFFLKSLTLEDVPNHGTIRFVCNSW
VYNSKLYKSPRIFFANKSYLPSETPSPLVKYREEELQTLRG
DGTGERKLHERIYDYDVYNDLGNPDHGEHLARPILGGS
STHPYPRRGRTGRYPTRKDPNSEKPATETYVPRDENFGH
LK**SSDFLAYGIK**SVSQCVVPAFESAFDLNFTPNEFDSFQD
VRNLFEGGIKLPLDVISTLSPLPVVKEIFRTDGEQVLKFTP
PHVIRVSKSAWMTDEEFAREMLAGVNPCMIRGLQEFPP
K**SNLDPAEYGDHTSK**ISVDVLNLDGCTIDEALASGRLFIL
DYHDTFIPFLRRINETSAKAYATRTILFLKENGTLKPVAIE
LSLPHPDGDKSGFVSKVILPADEGVESTIWLLAKAYVVV
NDSCYHQLMSHWLNTHAVIEPFVIATNR**QLSVVHPINK**
LLAPHYRDTMMNINALAR**DSLINANGLIER**SFLPSKYAV
EMSSAVYKYWVFTDQALPNDLIKRNMAVKDSSSPYGLR
LLIEDYPYAVDGLEIWTAIKTWVQDYVSLYYATDNDIKN
DSELQHWWKEVVEKGHGDLKDKPWWPKLQTFDELVE
VCTIIIWTASALHAAVNFGQYPYGGLILNRPTLSRRLLPEE
GTAEYDEMVKSSQKAYLRTITPKFQTLIDLSVIEILSRHAS
DEVYLGQRENPHWTSDSKALQAFQKFGNKLAEIEAKLT
NKNNDPSLYHRVGPVQLPYTLLHPSSKEGLTFRGIPN SISI

**FIGURE 6.** MALDI TOFF for predicted amino acid sequences in similar proteins (in red colours similar peptides).

## Discussion

*Status of 66 (SBP) kDa*

In *P. sativum* L. var. Alaska seeds, the amount of 66 kDa and the two 47 and 51 kDa in embryos at 24, 36, and 50 h and for the cotyledons at 1, 3, and 7 days after germination were investigated using SDS-PAGE and Western blotting. In cotyledons, in both SDS-PAGE and Western blotting, there were no dramatic changes observed at 7 days after germination for 66 kDa (SBP), 47 kDa (WAA) and 51 kDa (WAA). Also, 49 kDa apyrase has been shown not to be induced at all in the cotyledons of *P. sativum* var. Alaska. In embryos during germinations up to 50 hours after germinations, the level of 66 kDa (SBP) and 47 kDa (WAA) and 51 kDa (WAA) were changed significantly. 47 kDa (WAA) and 51 kDa (WAA) found to be as lower and upper regulating genes that might play a crucial role in germination of 49 kDa apyrase, as those two proteins were disappeared totally once the 49 kDa apyrase had been induced.

This finding may be supported as there was no expression of 49 kDa apyrase in the *P. sativum* L. var. Alaska cotyledons. 47 kDa (WAA) and 51 kDa (WAA) were noted previously (Moustafa *et al*., 2003; Moustafa, 2014) and considered to play a role in the expression of 49 kDa apyrase. It appears that, seed cotyledons biotinylated proteins have no contribution to 49 kDa induction. A correlation between the disappearance of 66 kDa SBP and accumulation of the 49 kDa apyrase was observed during germination.

Several studies have examined the situation of many proteins involved through the course of plant germinations, as for example, Gallardo *et al*. (2002) demonstrated that methionine synthase protein was present at low levels in dry mature seeds and the level increased greatly at 1-day imbibition, prior to radical development. A large portion of the proteins found by Gallardo *et al*. (2001) were available in dry seeds, and their plenitude stayed consistent through the germination process. We believe, therefore, that these data overestimate the difference in cotyledons and embryos during the course of germinations in regard to the change in the amount of 66 kDa SBP, both 47 kDa and 51 kDa (WAA), and the two undefined proteins (UDP) (48 kDa and 49.5 kDa). To date, this is the most primary study on the probable function of those proteins and their expression profiling in *P. sativum* L. var. Alaska in cotyledons and embryos. The diversified features and expression patterns of those two proteins are inferred to be associated with the induced expression of 49 kDa apyrase and may be other key genes through germinations. Our findings provide a base for functional research on the specific weakly antigenic genes (WAA) or streptavidin binding proteins (SBP) or undefined proteins (UDP) for other specific genes, a better understanding that these types of proteins act as a clues to other important genes during the course of initial plant germinations.

It is possible that the relative contribution of nutrition flow from cotyledons as storage proteins to the embryos would be significant to other protein expression that will be important during plant germinations. Recognition of this may indicate that 66 kDa SBP may contributes also in the emerging new protein or might be included in other advantageous changes for plant germinations or may be involved in breaking the seeds dormancy. The importance of the biotinylated protein has been shown as the prosthetic group of the housekeeping biotin-dependent carboxylases (Dakshinamurti *et al*., 1985; Dakshinamurti and Chauhan, 1989), the biotin bound to this protein is used by the embryo during germination and the beginning of seedling growth (Dehaye *et al*., 1997). By using 2D-PAGE analysis of many proteins, they were found to possess a range of isotypes that different not only in their p*I* values but also in their molecular mass, as for example 49 kDa apyrase (Abe *et al*., 2002) that may have various function during pea seeds germinations. Thus, these results support other mechanisms in which this protein somehow furnishes the developing embryonic proteins.

The biochemical mechanism that leads to the disappearance of 66 kDa (SBP) and 47 kDa (WAA) and 51 kDa (WAA) during germination and undefined proteins (UDP) associated with the appearance of 49 kDa in embryos only requires additional research. This may shed light into the basic processes of this very important phenomenon of down/up regulated genes at the level of appearance/disappearance of proteins and may be a novel finding.

*Cloning of 66 kDa (SBP) gene from germinated embryos of P. sativum var. Alaska*

We successfully cloned 66 kDa (SBP) gene from germinated embryos of *P. sativum* var. Alaska and the sequence analysis indicated that it contains 327 bp with an open reading frame (ORF) encoded for 108 amino acids. The deduced amino acid sequence was compared with other amino acids in BLAST and found no conserved sequences with other BLAST proteins. The same is in agreement with the findings of Duval *et al*. (1999) who also cloned and characterized SBP65 previously, and reported it acts as a putative sink for the free vitamin, representing a high amount of total protein-bound biotin in seeds. Also, similar to the Scan of the entire amino acid sequences of SBP 65 through several data (Swiss-Prot, GenPept banks, NBRF) which showed there was no homology to other known proteins.

Phylogenetic analysis of 66 kDa genes of *P. sativum* var. Alaska with previously cloned (SBP) genes was done by constructing distance tree in the Clustal W2 software. The cluster showed that 66 SBP *P. satvium* var. Alaska had a separate group than other seed biotinylated proteins in *P. satvium*. The differences also confirmed by applying the DNA distance matrix and from computing the nucleotide composition of 66 kDa SBP *P. satvium* var. Alaska.

The genome of *P. sativum* var. Alaska appears to be at least 5% versatile than the homology region in the complete sequences (Sequence ID: CU655882.7), hence, these regional can create chance for novel analyses for each plant variety. For instance, sequence comparisons can be conducted between the same regions within varieties, indicating the rates of change that can found in a specific gene within nucleus. Comparisons the nucleotides sequence between any of the grasses and Arabidopsis plants have shown that few colinear segments can be present at adjacent genes (Tikhonov *et al.,* 1999; van Dodeweerd *et al.,* 1999). Bennetzen (2000) found that comparisons of very closely related species or

species populations, can only show differences that have been resolved over the past several thousand to millions of years. Research in genome size and chromosome number at the level of species suggested that there was a drastic change in genome composition within a short evolutionary time of plant life span. Comparative analysis of sequences of genes and of repetitive DNAs demonstrated that these developed at distinct rates, with genes that are more in conserved region (Chen, 1998; SanMiguel *et al.,* 1998). In addition, the genome re-arrangement in closely related plants may be due to translocations, large deletions and heterozygous inversions. From this finding, a strong evidence that specific gene at the level of plant variety have its own nucleotide sequences.

*66 kDa (SBP) partial amino acid sequence*

The proteomic analysis of 66 kDa proteins using molecular cloning and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI–TOF-MS) generates different data. Our results found that the MALDI-TOFF analysis of amino acids sequences of 66 kDa SBP could not be determined exactly. However, MALDI-TOFF analysis had been used in describing the discrimination between *vanB* and *non-vanB* carrying enterococci in Australia (Griffin *et al.,* 2012), separation of *Bacteroides fragilis* strains into two divisions (Nagy *et al.,* 2011), and in detection the acetylation of norfloxacin (Pardo *et al.,* 2016).

Different information of data probably due to applying the proteome with trypsin (cleaves C-terminal specific to lysine or arginine, except before proline), Endo Asp N (separate N-terminal specific to aspartic acid), or cyanogen bromide (separate C-terminal specific methionine), chymotrypsin (cleaves C-terminal to hydrophobic residues), V8 (separates the peptide to bonds exclusively on the carbonyl side of aspartate and glutamate residues) result in different product properties (Cagney *et al.,* 2003). In addition, certain protein or peptide enrichment protocols may disentangle the peptide distributions adequately that proteins can be unambiguously determined from or more daughter peptides.

The efficiency of protein identifications in proteomics is dependent on many factors, for instance the exactness, affectability and goals of the estimating instrument, and furthermore the size and appropriation of peptide and protein properties in proteome, yet the chemical or physical basis of this phenomenon is poorly understood (Kinter and Sherman, 2000). It is also demonstrated that a solitary bacterial cell may deliver 4000 proteins whose plenitudes and activities may change through an experiment, while the quantity of proteins expressed in higher eukaryotes is probably going to be at least 10-fold greater (Cagney *et al.,* 2003).

Identification of 66 kDa proteins by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) revealed that predicted amino acids might be same as to 65 SBP, since it has almost the same molecular weight, the same reaction with streptavidin-alkaline phosphatase conjugate and exhibits the same germination behavior (Duval *et al.*, 1994). Also, the predicted amino acids is somewhat similar to *convicilin* in *P. sativum*. This family contains 7S plant seed stockpiling proteins and revealed that plant seed stockpiling proteins give the significant nitrogen source to the developing plant. In addition, the similar amino acid in lipoxygenase and 62K sucrose-binding protein homolog found in *P. sativum* support our findings that 66 SBP works as nutrient supply for the growing embryos. During the pea seeds germinations, there is not any amount of the mRNA were detected in the cotyledons; while, mRNA found to be in the embryo stalks, what this means? Not any metabolic process in the cotyledons and this part is used only as reservoirs for nutrient supply to the new plant i.e., embryos. There was a detailed understanding about mRNA level, on how mRNAs are made, individual steps of transcription and regulation of splicing and maturation and less information about the decay regulation of mRNA (Chan *et al*., 2018).

The gene expression is the main process that cause all other cellular metabolism required for life. The modification states and amounts of the mRNA and protein gene products are what mainly determine the function, identity, fate of a given cell. Previous studies showed that mRNA is required for cell compartment formation and further infers that mRNA can be restricted for preparing bodies (PB) formation when translation is quickly down-regulated while the cell is subjected to the stress.

*Vice versa*, legume seed formation various part is attributed to metabolism and nutrient transport. Many processes during cotyledon differentiation depend on the mRNA gene expression, as for example, the increased sucrose levels are correlated with increased levels of ADP-Glc pyrophosphorylase mRNAs and sucrose synthase (Borisju *et al.*, 2003).

**Acknowledgement**

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through research group (R.G.P.1/26 /40).

**References**

Abe S, Davies E (1991). Isolation of F-actin from pea stems: evidence from fluorescence microscopy. *Protoplasma* **163**: 51-61.

Abe S, Moustafa MFM, Shibata K, Yoneda M, Davies E (2002). Purification and characterization of the major isotypes of apyrase from the cytoskeleton fraction in *Pisum sativum*. *Plant Physiology and Biochemistry* **40**: 1019-1023.

Abe S, Ito Y, Davies E (1992). Co-sedimentation of actin, tubulin and membranes in the cytoskeleton fractions from peas and mouse 3T3 cells. *Journal of Experimental Botany* **43**: 941-949.

Abe S, Ito Y, Davies E (1994). Association of cytoskeletal proteins in the membrane-bound polysome fraction from peas using conventional polysome isolation buffers. *Plant Physiology and Biochemistry* **32**: 547-554.

Abe S, Ito Y, Davies E (1995). Isolation of a heparin sensitive, ribosome sedimenting factor from the cytoskeleton fractions of peas and corn. *Plant Physiology and Biochemistry* **33**: 463-470.

Ahmad F, Gopal J, Wu HF (2012). Rapid and highly sensitive detection of single nematode via direct MALDI mass spectrometry. *Talanta* **93**: 182-185.

Bajpayee AG, Wong CR, Bawendi MG, Frank EH, Grodzinsky AJ (2014). Avidin as a model for charge driven transport into cartilage and drug delivery for treating early stage post-

traumatic osteoarthritis. *Biomaterials* **35**: 538-549.

Bennetzen JL (2000). Comparative sequence analysis of plant nuclear genomes microcolinearity and its many exceptions. *The Plant Cell* **12:** 1021-1030.

Bonel L, Vidal JC, Duato P, Castillo JR (2011). An electrochemical competitive biosensor for ochratoxin A based on a DNA biotinylated aptamer. *Biosensors and Bioelectronics* **26**: 3254-3259.

Borisjuk L, Rolletschek H, Wobus U, Weber H (2003). Differentiation of legume cotyledons as related to metabolic gradients and assimilate transport into seeds. *Journal of Experimental Botany* **54**: 503-512.

Bruch RC, White HB (1982). Compositional and structural heterogeneity of avidin glycopeptides. *Biochemistry* **21:** 5334-5341.

Cagney G, Amiri S, Premawaradena T, Lindo M, Emili A (2003). In silico proteome analysis to facilitate proteomics experiments using mass spectrometry. *Proteome Science* **1**: 5.

Chan LY, Mugler CF, Heinrich S, Vallotton P, Weis K (2018). Non-invasive measurement of mRNA decay reveals translation initiation as the major determinant of mRNA stability. *eLife* **7**: e32536.

Chapman-Smith A, Cronan Jr JE (1999). Molecular biology of biotin attachment to proteins. *Journal of Nutrition* **129**: 477-484.

Chen M (1998). The Structure of *sh2/a1*-Homologous Regions in Maize, Rice and Sorghum. Ph.D. Dissertation. Purdue University. West Lafayette.Chevalier J, Yi J, Michel O, Tang XM (1997). Biotin and digoxigenin as labels for light and electron microscopy in situ hybridization probes: where do we stand? *Journal of Histochemistry and Cytochemistry* **45**: 481-491.

Cleveland DW, Fischer SG, Kirschner MW, Laemmli UK (1977). Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. *Journal of Biological Chemistry* **252**: 1102-1106.

Dakshinamurti K, Chalifour L, Bhullar RP (1985). Requirement for biotin and the function of biotin in cells in culture. *Annals of the New York Academy of Sciences* **447**: 38-55.

Dakshinamurti K, Chauhan J (1989). Biotin. *Vitamins & Hormones* **45**: 337-384.

Davies E, Stankovic B, Azama K, Shibata K, Abe S (2001). Novel components of the plant cytoskeleton: a beginning to plant 'cytomics'. *Plant Science* **160**: 185-196.

Duval M, De Rose RT, Job C, Faucher D, Douce R, Job D (1994). The major biotinyl protein from *Pisum sativum* seeds covalently binds biotin at a novel site. *Plant Molecular Biology* **26**: 265-273.

Ferreira L, Sánchez-Juanes F, González-Ávila M, Cembrero-Fuciños D, Herrero-Hernández A, González-Buitrago JM, Muñoz-Bellido JL (2010). Direct identification of urinary tract pathogens from urine samples by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Journal of Clinical Microbiology* **48**: 2110-2115.

Fugate CJ, Jarrett JT (2012). Biotin synthase: Insights into radical-mediated carbon-sulfur bond formation. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1824**: 1213-1222.

Gallardo K, Job C, Groot SPC, Puype M, Demol H, Vandekerckhove J, Job D (2001). Proteomic analysis of Arabidopsis seed germination and priming. *Plant Physiology* **126**: 835-848.

Gallardo K, Job C, Groot SP, Puype M, Demol H, Vandekerckhove J, Job D (2002). Importance of methionine biosynthesis for Arabidopsis seed germination and seedling growth. *Physiologia Plantarum* **116**: 238-247.

Green NM (1975). Avidin. *Advances in Protein Chemistry* **29**: 85-133.

Griffin PM, Price GR, Schooneveldt JM, Schlebusch S, Tilse MH, Urbanski T, Hamilton B, Venter D (2012). Use of matrix-assisted laser desorption ionization-time of flight mass spectrometry to identify vancomycin-resistant enterococci and investigate the epidemiology of an outbreak. *Journal of Clinical Microbiology* **50**: 2918-2931.

Kinter M, Sherman NE (2000). Protein sequencing and identification using tandem mass spectrometry. 1st Edition. Wiley-Interscience, New York.

Kumar Vemuri P, Talluri B, Panangipalli G, Kadiyala SK, Veeravalli S, Bodiga VL (2016). Purification and identification of 20 kDa protein from *Parthenium hysterophorus*. *International Journal of Pharmacognosy and Phytochemical Research* **8**: 827-830.

Kumar Vemuri P, Veeravalli S (2014). Expression, purification and characterization of human recombinant galectin 3 in *Pichia pastoris*. *Iranian Journal of Biotechnology* **12**: 3-8.

Liu L, Han R, Yu N, Zhang W, Xing L, Xie D, Peng D (2018). A method for extracting high-quality total RNA from plant rich in polysaccharides and polyphenols using *Dendrobium huoshanense*. *PLoS One* **13**: e0196592.

Mellmann A, Bimet F, Bizet C, Borovskaya AD, Drake RR, Eigner U, Fahr AM, He Y, Ilina EN, Kostrzewa M, Maier T, Mancinelli L, Moussaoui W, Prévost G, Putignani L, Seachord CL, Tang YW, Harmsen D (2009). High interlaboratory reproducibility of matrix-assisted laser desorption ionization-time of flight mass spectrometry-based species identification of nonfermenting bacteria. *Journal of Clinical Microbiology* **47**: 3732-3734.

Moustafa MFM (2014). Apyrase, streptavidin-binding protein, and antimicrobial activity in *Pisum sativum*. *Russian Journal of Plant Physiology* **61**: 496-502.

Moustafa M, Alamri S, Zaki H, Loutfy N, Taha T, Shati A, Al-kahtani M, Siddiqi S (2019). Gene expression of 49 kDa apyrase, cytoskeletal proteins, ATPase, ADPase and amino acid contents of *Pisum sativum* (L.) cells germinated in *Euryops arabicus* (Steud. ex Jaub. & Spach) water extract. *Biocell* **43**: 81-87.

Moustafa MFM, Yoneda M, Abe S, Davies E (2003). Changes in isotypes and enzyme activity of apyrase during germination of dark-grown pea (*Pisum sativum*) seedlings. *Physiologia Plantarum* **119**: 146-154.

Nagy E, Becker S, Sóki J, Urbán E, Kostrzewa M (2011). Differentiation of division I (*cfiA*-negative) and division II (*cfiA*-positive) *Bacteroides fragilis* strains by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Journal of Medical Microbiology* **60**: 1584-1590.

Pardo CA, Tan RN, Hennequin C, Beyrouthy R, Bonnet R, Robin F (2016). Rapid detection of AAC(6')-Ib-cr production using a MALDI-TOF MS strategy. *European Journal of Clinical Microbiology & Infectious Diseases* **35**: 2047-2051.

Perkins DN, Pappin DJC, Creasy DM, Cottrell JS (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**: 3551-3567.

Sami-Subbu R, Muench DG, Okita TW (2000). A cytoskeleton-

associated RNA-binding protein binds to the untranslated regions of prolamine mRNA and to poly(A). *Plant Science* **152**: 115-122.

SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998). The paleontology of intergene retrotransposons of maize. *Nature Genetics* **20**: 43-45.

Sassolas A, Leca-Bouvier BD, Blum LJ (2008). DNA biosensors and microarrays. *Chemical Reviews* **108**: 109-139.

Shibata K, Abe S, Davies E (2001). Structure of the coding region and mRNA variants of the apyrase gene from pea (*Pisum sativum*). *Acta Physiologiae Plantarum* **23**: 3-13.

Shibata K, Abe S, Yoneda M, Davies E (2002). Sub-cellular distribution and isotypes of a 49-kDa apyrase from *Pisum sativum*. *Plant Physiology and Biochemistry* **40**: 407-415.

Shibata K, Morita Y, Abe S, Stanković B, Davies E (1999). Apyrase from pea stems: isolation, purification, characterization and identification of an NTPase from the cytoskeleton fraction of pea stem tissue. *Plant Physiology and Biochemistry* **37**: 881-888.

Tikhonov AP, SanMiguel PJ, Nakajima Y, Gorenstein NM, Bennetzen JL, Avramova Z (1999). Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 7409-7414.

Tong L (2013). Structure and function of biotin-dependent carboxylases. *Cellular and Molecular Life Sciences* **70**: 863-891.

van Dodeweerd AM, Hall CR, Bent EG, Johnson SJ, Bevan MW, Bancroft I (1999). Identification and analysis of homoeologous segments of the genomes of rice and *Arabidopsis thaliana*. *Genome* **42**: 887-892.

Wood HG, Barden RE (1977). Biotin enzymes. *Annual Review of Biochemistry* **46**: 385-413.

Zhang B, Petersen R, Hussain S, Ye JY, Dallo S, Weitao T (2011). Detection of anthrax *lef* with DNA-based photonic crystal sensors. *Journal of Biomedical Optics* **16**: 127006.