

RESEARCH

Open Access



Comparative studies on seed protein characteristics in eight lines of two *Gossypium* species

SINGH Arvinder* and KAUR Amanjot

Abstract

Background: In order to achieve the targets aiming at the improvement of protein quality, knowledge regarding seed protein fractions and polypeptides constituting them in different crops is essential. Besides having high nutritional value as animal feed and human food, the protein isolates from cottonseed meal have also been proven promising as industrial raw materials for a number of applications. As far as Indian work on the characterization of cotton seed proteins is concerned, relatively meagre reports are available. Keeping in mind the importance of cotton seed proteins, lines belonging to *Gossypium arboreum* L. (Indian cotton) and *G. hirsutum* L. (American cotton) which are grown in all the major cotton growing states in India were selected for analysing their seed protein characteristics.

Results: Whereas *G. arboreum* (A-genome) lines revealed a lower range of seed protein content i.e. 19.5~24.3%, an upper range (21.8~29.5%) could be observed in lines of *G. hirsutum* (AD-genome). Globulins represented dominating fraction in both species followed by albumins, glutelins and prolamins. A significant positive correlation between albumins/globulins and seed protein content in *G. arboreum* /*G. hirsutum*, respectively, was observed. Intraspecific electrophoretic variation in seed protein extracts was observed in the region of molecular weight 22 kDa - 27 kDa in lines of both the species; however some lines with A-genome showed similarity in banding pattern with AD-genome. Four polypeptides with disulphide-linkages were also reported for the first time. Albumins were observed to reveal more variations in their electrophoretic pattern between the lines of two species followed by globulins.

Conclusion: On the basis of present and previous studies, screening the lines with low or high protein content will lead the selection of lines with superior polypeptide fraction important for nutritional and industrial purposes. On comparing the composition and behaviour of four 2-S linked polypeptides with other plant groups, these were suggested to be legumin-like in nature. The similarity in banding patterns between the lines of A-genome and AD-genome species marked towards the close evolutionary relationship between these two. Albumin fractions on the basis of our results could be taken for cultivar differentiation in cotton crop.

Keywords: *Gossypium*, SDS-PAGE, Seed proteins, Globulins

Introduction

Out of 50 species, there are four cultivated species of cotton, viz. *Gossypium arboreum* L., *G. herbaceum* L., *G. hirsutum* L. and *G. barbadense* L. (Wendel and Albert 1992). As former two species having A-genome ($2n = 26$) are mainly grown in Asia, these are termed as Asiatic cotton or Indian cotton. Remaining two are allotetraploid ($4n = 52$) with AD-genome, out of which *G. hirsutum* is

known as American cotton and *G. barbadense* is Pima or Egyptian cotton (Grover et al. 2015). *G. hirsutum* alone contributes 90% of the total global production of the cotton (Turley et al. 2007). Cotton is an important fibre crop cultivated in tropical and subtropical regions of the world. China, USA and India are the world's major cotton-producing countries, accounting for nearly 60% of world production (Yu et al. 2012; Fang 2015).

Cottonseeds in the form of whole cotton seed (WCS) and cottonseed meal (CSM) constitute the main source of oil, meal and protein for human consumption, livestock

* Correspondence: arvinder_bot@auts.ac.in

Department of Botany, Akal University, Talwandi Sabo, Bathinda, India



feed, and raw material for industrial applications, respectively (He et al. 2013, He et al. 2014a). According to Coppock et al. (1987), the nutritional protein degradability of CSM is similar to that of peanut meal, canola meal and soybean meal for lactating dairy cows, and to that of canola meal and soybean meal for young calves. Production performances of ruminant animals in terms of body weight gain, milk production, the fat content of the milk, wool production had been shown to improve when their diet was supplemented with WCS and CSM (Osti and Pandey 2006). Cottonseed protein food products have also been proven as a healthy addition to the diets of children, college-age women, and elder people. Baked goods, snack food, pet and livestock feed are just a few successful products developed utilizing cottonseed protein (Alford et al. 1996).

As far industrial applications are concerned, promising results of cottonseed protein isolates (CSPI) as well as protein extracted insoluble residue (CSIR) from the cottonseed meal have been revealed in the number of value-added products, viz. bio-based wood adhesive (Cheng et al. 2016a, b), bioplastics and films (Yue et al. 2014), and superabsorbent hydrogel (Zhang et al. 2010). The studies comparing the superiority of cottonseed protein-based wood adhesive over soy protein-based adhesive attributed this to the difference in the structural and functional properties of the proteins between these two crops (He et al. 2016a, b; Cheng et al. 2016a, b). Further, the effect of low and high seed protein content on the adhesive strength was also shown by Pradyawong et al. (2018) who suggested the use of the CSM with high protein content for good adhesive strength and spreadability.

Martinez (1964) had divided cotton seed proteins into two categories as: 1) water-soluble proteins having low molecular weight but high electrophoretic mobility, and 2) water-insoluble as true storage proteins with high molecular weight but low electrophoretic mobility. However, these were not classified as albumins, globulins or other classes of storage proteins at that time. Later on, the classification of these proteins as water-soluble albumins followed by alkali-soluble glutelins, salt-soluble globulins and alcohol-soluble prolamins was proposed by Sammour et al. (1995). Youle and Huang (1981) recognized three major types of proteins in cotton having sedimentation coefficient 2S, 5S and 9S in equal amount. Whereas 2S represented albumins, and 5S and 9S belonged to globulin proteins. In the same year, Dure and Chlan (1981), using SDS-PAGE and immuno-technique, purified and characterized the principal storage proteins in cotton having molecular weight 52 kDa and 48 kDa and designated these as α -globulins and β -globulins, respectively. In contrast to principal salt-soluble storage proteins (globulins) as reported by Dure and Chlan (1981), King

and Lefler (1979) concluded alkali-soluble proteins to be the major storage proteins in cotton seed. However, the molecular weights in both the cases for principal storage proteins were almost same, i.e. 52 kDa and 48 kDa, and 52.7 kDa and 46 kDa, respectively. The fate of other proteins again remained unexplored. In the sequence of this protein work, Marshall (1990) reported 98 kDa polypeptides with its two subunits 54 kDa and 48 kDa as 7S globulins of cotton; however, these two subunits were found to contain no covalent linkage between them. All the above-mentioned studies on seed proteins were mainly carried out in *G. hirsutum* L. cultivars, and concentrated on major storage protein – globulins. Still, very negligible work is there in other cultivated species of cotton in terms of investigating the components of seed proteins. By working on another cultivated species - *G. barbadense* L., Sammour et al. (1995) have reported a high percentage of alkali-soluble proteins than salt-soluble proteins. Further, using two-dimensional gel electrophoresis, they revealed the presence of a 2S-bonded polypeptide of molecular weight 45 kDa in total seed protein extract of cottonseeds but did not assign it to any protein fractions. Protein identification and genetic characterization of high abundance proteins in seeds of three *Gossypium* species belonging each to AD-genome, A-genome and D-genome have revealed two major families of globulin seed storage proteins, i.e. vicilin and legumin accounting for 60~70% of cotton seed proteins (Hu et al. 2011). A recent study on water- and alkali-soluble cottonseed proteins (He et al. 2018) in *G. hirsutum* exhibited the presence of 6 and 12 major protein bands on SDS-PAGE belonging to CSPa and CSPw, respectively. Among these proteins, the most abundant peptides were shown to be legumin and vicilin types. Cotton is a major crop of India after wheat and paddy but as far as Indian work on the characterization of seed proteins in cotton is concerned, relatively meagre reports are available; however, all the four cultivated species are grown here in India. A correlation study by Pandey and Thejappa (1975) on 97 varieties of cotton exhibited a negative correlation between protein and gossypol content. Goyal (1992) utilized cotton seed proteins as a marker to generate electrophoregram for identification of some cotton cultivars. Similarly, Kumar et al. (2007) also used SDS-PAGE technique for identification and genetic diversity estimation of six tetraploid and two diploid cotton cultivars using seed protein profile. Zymograms generated by using globulin protein fractions from three cotton hybrids and their parents were employed to test the genetic purity of the seed by Reddy et al. (2008).

Keeping in view the importance of cottonseed meal (CSM) as animal feed, human food and in industries, it has been suggested from time to time to evaluate the chemical composition of cottonseeds especially for proteins and

dietary fibres. In addition, to utilize one plant species successfully in a breeding programme for improving some quality traits, genetic relationships among different species of the same genus as well as among different groups of the plants have also been evaluated employing protein characterization studies. Most of the studies on cottonseed proteins are concentrated on *G. hirsutum*, and explored mainly globulin fractions; very little reports are there on the characterization of other seed protein fractions which could be helpful in screening and better utilization of the cotton germplasm for nutraceutical and industrial end use. In this regard, the protein characterization studies on Indian cotton germplasm are very scanty. Varieties belonging to *G. arboreum* (Indian cotton) along with *G. hirsutum* (American cotton) are cultivated in all the major cotton growing states in India under varying environmental conditions. Therefore, it was planned to work out the seed protein analysis in *G. arboreum* in relation to the most widely studied species - *G. hirsutum* by studying variation in protein content, the proportion of four protein fractions and polypeptide patterns of eight lines of *G. arboreum* and *G. hirsutum*.

Materials and methods

Seeds of the following eight lines belonging to two *Gossypium* species, i.e. *G. hirsutum* (American cotton) and *G. arboreum* (Indian cotton), were used for the present study:

- a) *G. hirsutum* lines: 1) F-2228, 2) F-2383, 3) LH-2108, 4) LH-2076;
- b) *G. arboreum* lines: 1) LD-1019, 2) LD-949, 3) LD-327, 4) FDK-124.

The seeds were kindly supplied by Punjab Agricultural University Regional Station, Bathinda (Punjab), India.

Total seed protein extraction

Preparation of total seed protein extracts was based on the method employed by Singh and Matta (2011). Total seed protein extracts were prepared by mixing the defatted seed meal in Tris-HCl buffer solution (0.2 mol·L⁻¹, pH 6.8) containing 2% SDS. Forty mg of seed meal was suspended in 400 µL buffer solution heated at 80 °C in a water bath for 45 min. The contents were centrifuged at 2 000 g for 10 min and supernatant used for analysis. Glycerol was added to the sample meal containing the extracts so that it amounted to 10% of the final volume. To run the proteins under reducing conditions, 2-mercaptoethanol was added to limit its concentration to 2% in the total protein extracts.

Seed protein fractionation

Separation of four protein fractions was based on methods employed by Luthe (1983), and by Schaeffer and Sharpe (1990), with slight modifications. All aqueous extraction solvents were buffered with 10 mmol·L⁻¹ Tris-HCl (pH 7.5). After extraction of albumins in water, the residue was used for separation of globulins with 0.5 mol·L⁻¹ NaCl, followed by 55% n-propanol for prolamins and 0.5% SDS for the glutelin fraction. Each extraction was repeated twice and the supernatants were pooled.

Protein estimation

Protein content in the defatted seed meal was estimated by Semi-micro Kjeldahl method as suggested by Peach and Tracey (1956). Seed meal was digested with concentrated sulphuric acid in the presence of a catalytic mixture of copper sulphate, selenium dioxide and potassium dichromate. The digest was heated with 40% NaOH in Markham's distillation assembly and the ammonia so evolved was volumetrically titrated with N/40 HCl to determine the nitrogen present in the sample. The so determined nitrogen was multiplied by a conversion factor 6.25 to get the seed protein content value.

Protein concentration in the four separated fractions was determined using Bradford method (Bradford 1976). A volume of 100 µL of the given fraction representing an extract from 100 mg of seed meal was used and the proportion of each fraction calculated as 'g/100 g seed meal'.

SDS-polyacrylamide gel electrophoresis

SDS-polyacrylamide gel electrophoresis was carried out on 14% gels following the method of Laemmli (1970). For gel electrophoresis under reducing conditions, 2% 2-mercaptoethanol was added to the seed protein extracts and the samples were heated in an oven at 90 °C for 10 min before loading these onto the gels. The gel was run at 17 mA and after the tracking dye moved down into the separation gel, the current was increased to 25 mA. The gel was stained with Coomassie Brilliant Blue (0.05%) dissolved in a solvent containing methanol, acetic acid and distilled water in the ratio 50:7:43 (v/v), and destained in the same solvent mixture but lacking the dye.

Two-dimensional gel electrophoresis

Two-dimensional gel electrophoresis of total seed protein extract was carried out following the method as described by Singh and Matta (2008). The 1.5 mm thick gel strip with polypeptides separated under non-reducing conditions (1D, -2ME) was equilibrated for 2 h with gentle shaking in 0.2 mol·L⁻¹ Tris-HCl buffer (pH 6.8) containing 2% SDS and 2% 2-mercaptoethanol, and loaded onto another gel of 2 mm thickness for electrophoresis in the second dimension (2D, +2ME).

Molecular weight determination

To calculate the molecular weight of the bands appeared on SDS-gel, standard curve was drawn according to molecular weight protein markers and their pixel position on the gel using Total lab TL software.

Statistical methods

Mean and coefficient of correlation were calculated by using SPSS 18.0.

Results

Seed protein content

Semi-micro Kjeldahl method was employed to determine the total seed protein content of different lines. The four cotton lines belonging to *G. arboreum* species revealed the protein content in the range of 19.5% in line 'LD-1019' to 24.3% in line 'LD-327'. On the other hand, the cotton lines belonging to *G. hirsutum* showed the protein content in the range of 21.8% in line 'F-2228' to 23.4% in line 'LH-2076' (Table 1).

Proportion of four protein fractions

The relative distribution of four protein fractions, viz. albumins, globulins, prolamins and glutelins in the seeds of different lines is given in Tables 2 and 3. In all the cotton lines of the *G. arboreum* species, globulins represented the major fraction with their proportion varying from 36.2% in line 'FDK-124' to 63.7% in line 'LD-1019'. These were followed by albumins which were present in the range of 20.8% in line 'LD-1019' to 32.2% in line 'LD-327'. Remaining two fractions - glutelins and prolamins - were found to be present in proportion as varying from 9.1% in line 'LD-1019' to 24.5% in line 'FDK-124', and 5.4% in line 'LD-949' to 7.8% in line 'FDK-124', respectively.

Similarly, the lines belonging to *G. hirsutum* revealed globulins to be the major protein fraction in the range of 33.0% in line 'F-2228' to 40.7% in line 'LH-2076'. It was followed by albumins, glutelins and prolamins as in *G. arboreum*. The water-soluble albumins exhibited

Table 1 Seed protein content in cotton lines of two *Gossypium* species

	S. No.	Cotton line	Protein content /%
<i>Gossypium arboreum</i> (Indian cotton)	1.	FDK-124	24.1
	2.	LD-1019	19.5
	3.	LD-949	20.0
	4.	LD-327	24.3
<i>Gossypium hirsutum</i> (American cotton)	1.	F-2228	21.8
	2.	F-2383	29.5
	3.	LH-2108	22.3
	4.	LH-2076	23.4

Table 2 Relative proportion of four protein fractions in seeds of cotton lines of *G. arboreum*

Relative proportion of four protein fractions (g/100 g)					
Sr. No.	Cotton line	Albumins	Globulins	Glutelins	Prolamins
1	FDK-124	31.3	36.2	24.5	7.8
2	LD-1019	20.8	63.7	9.1	6.6
3	LD-949	25.0	54.3	15.2	5.4
4	LD-327	32.2	44.4	16.6	6.6
	Mean	27.3 ± 3.5	49.6 ± 6.8	16.3 ± 2.5	6.6 ± 0.9

their proportion varying from 23.3 to 29.7% in lines 'F-2383' and 'F-2228', respectively. Prolamins, the lowest in proportion in all the lines, were observed to be present in the range of 5.5% in line 'F-2383' to 12.0% in line 'LH-2108'.

Electrophoretic variation studies

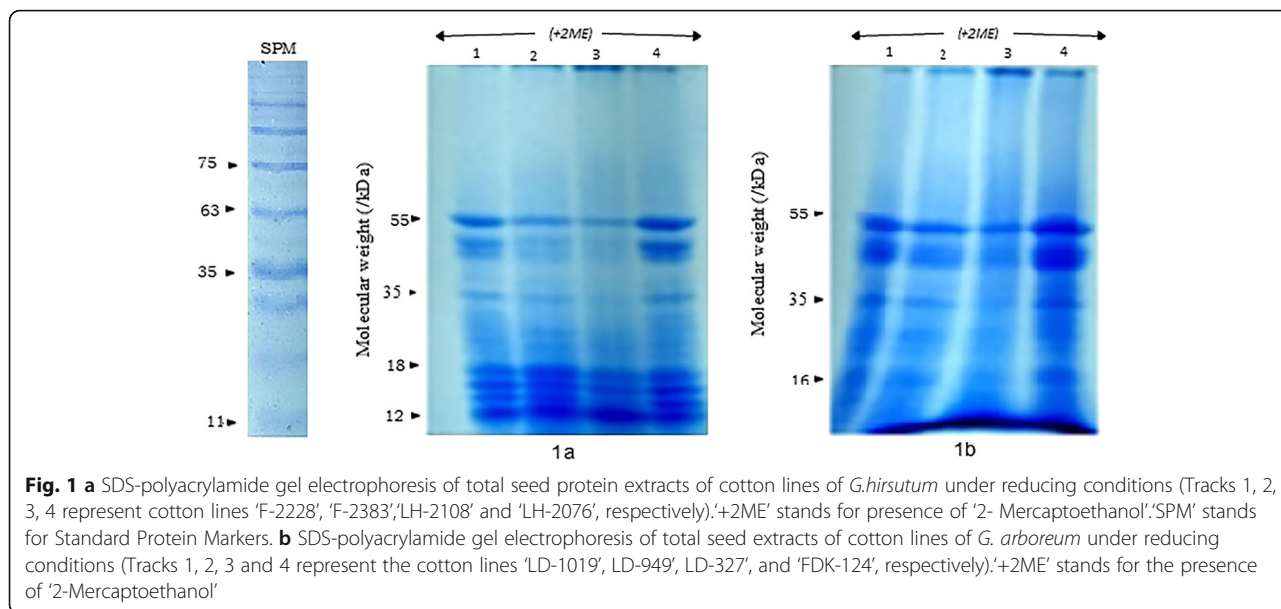
Total seed protein extracts

The polypeptide patterns of total seed protein extracts of each set of four cotton lines belonging to *G. hirsutum* and *G. arboreum* as analysed on SDS-gels under reducing conditions can be seen in Fig. 1a, b.

***G. hirsutum* lines:** A large number of polypeptides with molecular weight ranging between 10 kDa to 122 kDa were observed and can be seen in Table 4. The major polypeptides of 57 kDa, 55 kDa, 50 kDa, 47.5 kDa, 18 kDa, 17 kDa, 15 kDa, 14 kDa, 13 kDa and 12 kDa were intense and darkly stained; those of molecular weight 49 kDa, 46 kDa, 40.5 kDa, 38 kDa, 36 kDa, 32 kDa, 27 kDa, 26 kDa, 25 kDa, 24.5 kDa, 24 kDa, 23.5 kDa, 22 kDa, 14.5 kDa, 11.5 kDa and 10 kDa were prominent but of relatively lower intensity. Some other polypeptides of molecular weight 120 kDa, 115 kDa, 80 kDa, 75 kDa, 60 kDa were represented by lightly stained bands. On further comparing the variation in the polypeptide patterns of *G. hirsutum* lines on SDS-gels, the polypeptide region of molecular weight 22 kDa to 27 kDa was seen with three different patterns (designated as 'P' to 'R') as can be seen in Table 6. The pattern 'P' with molecular weight 27 kDa, 26 kDa, 25 kDa, 24 kDa and 22 kDa was

Table 3 Relative proportion of four protein fractions in seeds of cotton lines of *G. hirsutum*

Relative proportion of four protein fractions (g/100 g)					
S No.	Cotton line	Albumins	Globulins	Glutelins	Prolamins
1	F-2228	29.7	33.0	28.0	9.0
2	F-2383	23.3	51.5	19.6	5.5
3	LH-2108	25.0	38.0	25.0	12.0
4	LH-2076	26.2	40.7	23.3	9.7
	Mean	26.0 ± 2.7	40.8 ± 5.3	23.9 ± 1.7	9.0 ± 1.1



seen to occur in lines 'F-2228' and 'LH-2076' followed by the pattern 'Q' with molecular weight 25.5 kDa, 24.5 kDa and 23.5 kDa in line 'F-2383'. The third pattern 'R' was seen in line 'LH-2108' with molecular weight 24.5 kDa and 23.5 kDa.

G. arboreum lines: Almost similar kind of polypeptide pattern could be seen in these lines with polypeptides of molecular weight ranging between 10 kDa to 122 kDa as in American cotton. The molecular weights of different polypeptides present in seed extracts of these lines are given in Table 5. Like *G. hirsutum* lines, variation in electrophoretic pattern of these lines could also be observed in the region of molecular weight 22 kDa to 27 kDa. Unlike American cotton lines, only two polypeptide

patterns designated as 'P' and 'Q' could be seen in this region (Table 6). Three lines 'LD-1019', 'LD-327' and 'LD-949' were represented by a common polypeptide pattern 'Q' with molecular weight 25.5 kDa, 24.5 kDa and 23.5 kDa. Only one line 'FDK-124' revealed a single pattern 'P' with molecular weight 27 kDa, 26 kDa, 25 kDa, 24 kDa and 22 kDa.

Further, on comparing the electrophoretic profile of all the cotton lines belonging to both the *Gossypium* species, the line 'FDK-124' of *G. arboreum* was observed similar in its polypeptide pattern with the lines 'F-2228' and 'LH-2076' of *G. hirsutum*. The rest of the lines 'LD-1019', 'LD-949' and 'LD-327' of *G. arboreum* were found similar to lines 'F-2383' and 'LH-2108' of *G. hirsutum* in their polypeptide patterns.

Table 4 Molecular weight of polypeptides in seed protein extracts of *G. hirsutum* under reducing conditions

Sr. No.	Cotton line	Molecular weight /kDa
1.	F-2228	122, 120, 115, 80, 75, 60, 55, 50, 49, 46, 40.5, 38, 36, 32, 27, 26, 25, 24, 22, 18, 17, 16, 15, 14.5, 14, 12, 11.5, 10
2.	F-2386	122, 120, 115, 80, 75, 60, 55, 50, 49, 46, 40.5, 38, 36, 32, 25.5, 24.5, 23.5, 18, 17, 16, 15, 14.5, 14, 12, 11.5, 10
3.	LH-2108	122, 120, 115, 80, 75, 60, 55, 50, 49, 46, 40.5, 38, 36, 32, 24.5, 23.5, 18, 17, 16, 15, 14.5, 14, 12, 11.5, 10
4.	LH-2076	122, 120, 115, 80, 75, 60, 55, 50, 49, 46, 40.5, 38, 36, 32, 27, 26, 25, 24, 22, 18, 17, 16, 15, 14.5, 14, 12, 11.5, 10

Table 5 Molecular weight of polypeptides in seed protein extracts of *G. arboreum* under reducing conditions

Sr. No.	Cotton line	Molecular weight /kDa
1.	FDK-124	122, 120, 115, 80, 75, 60, 55, 50, 49, 46, 40.5, 38, 36, 32, 27, 26, 25, 24, 22, 18, 16, 15, 14.5, 14, 12, 11.5, 10
2.	LD-1019	122, 120, 115, 80, 75, 60, 55, 50, 49, 46, 40.5, 38, 36, 32, 24.5, 23.5, 18, 16, 15, 14.5, 14, 12, 11.5, 10
3.	LD-949	122, 120, 115, 80, 75, 60, 55, 50, 49, 46, 40.5, 38, 36, 32, 24.5, 23.5, 18, 16, 15, 14.5, 14, 12, 11.5, 10
4.	LD-327	122, 120, 115, 80, 75, 60, 55, 50, 49, 46, 40.5, 38, 36, 32, 24.5, 23.5, 18, 16, 15, 14.5, 14, 12, 11.5, 10

Table 6 Polypeptide pattern types in seed protein extracts of different cotton lines under reducing conditions

Molecular weight region	Pattern type	Molecular weight /kDa	Cotton lines
<i>G. hirsutum</i> lines			
22–27 kDa	P	27, 26, 25, 24, 22	F-2228, LH-2076
	Q	25.5, 24.5, 23.5	F-2383
	R	24.5, 23.5	LH-2108
<i>G. arboreum</i> lines			
22–27kDa	P	27, 26, 25, 24, 22	FDK-124
	Q	25.5, 24.5, 23.5	LD-1019, LD-327, LD-949

Total seed proteins extracts were further analysed by using two-dimensional gel electrophoresis wherein proteins separated under non-reducing conditions in the first dimension were run under reducing conditions in the second dimension (Fig. 2). For the purpose, the line 'F-2228' of '*G. hirsutum*' representing common polypeptide pattern with *G. arboreum* line 'LD-327' was selected. The total seed protein extract of this line was first run under non-reducing conditions (-2ME) on SDS-gels in the first dimension, then the separated polypeptides were run in the second dimension after treating the gel strip with 2-ME as explained under the section - materials and methods. The polypeptides of molecular weight 122 kDa, 120

kDa, 115 kDa, 80 kDa, 75 kDa, 60 kDa, 50 kDa, 40.5 kDa, 38 kDa, 27 kDa, 26 kDa, 25 kDa, 18 kDa, 16 kDa, 14 kDa and 12 kDa, present under non-reducing conditions in first dimension, were found to resolve along the diagonal in second dimension. Some bands having disulphide linkages were reduced and moved down as spot off the diagonal. On the basis of bands seen under non-reducing conditions in first dimension and breaking off the diagonal under reducing conditions in the second dimension, following polypeptide pairs and their constituent subunits could be discerned:

Polypeptide pairs		Polypeptide subunits
52 kDa	=	32 kDa + 24 kDa
40 kDa	=	22 kDa + 17 kDa
36 kDa	=	20 kDa
32 kDa	=	22 kDa

Polypeptides of four protein fractions

For assigning various bands, as seen in the total seed protein extracts to different protein fractions, four protein fractions of two cotton lines, one each from *G. hirsutum* (AD-genome) and *G. arboreum* (A-genome) species, were analysed for their polypeptide composition using SDS-PAGE under reducing conditions (Fig. 3). The albumins and globulins which contributed 75% of the four protein fractions were represented by large number of dark and light intensity bands whereas glutelins were seen to possess fewer bands mainly light in intensity. Prolamins which represented just 5~10% of total protein fractions could not appear on the SDS-gel.

In both the lines, polypeptides belonging to albumin and globulin fractions could be seen in the range of molecular weight 12 kDa to 120 kDa and 11.5 kDa to 122 kDa, respectively (Table 7). The comparative analysis of polypeptide pattern of salt-soluble fraction (globulins) revealed the presence of some bands (27 kDa, 22 kDa, 20 kDa and 17 kDa) only in the line of A-genome species which could not be spotted in line of AD-genome species. Similarly, the bands of molecular weight 46 kDa, 27.5 kDa, 26 kDa and 25 kDa were seen only in the line of later but not in former species. Whereas water-soluble albumins of A-genome species exhibited the bands of molecular weight 34 kDa, 30 kDa, 26 kDa and 19 kDa in its polypeptide profile, the bands of molecular weight 45 kDa, 31 kDa, 25 kDa, 15 kDa, 14 kDa and 13 kDa could only be specified to AD-genome species. The polypeptide pattern of the alkali-soluble fraction (glutelins) with molecular weight ranging from 10 kDa to 54 kDa was found similar in the lines of both A- and AD-genome species.

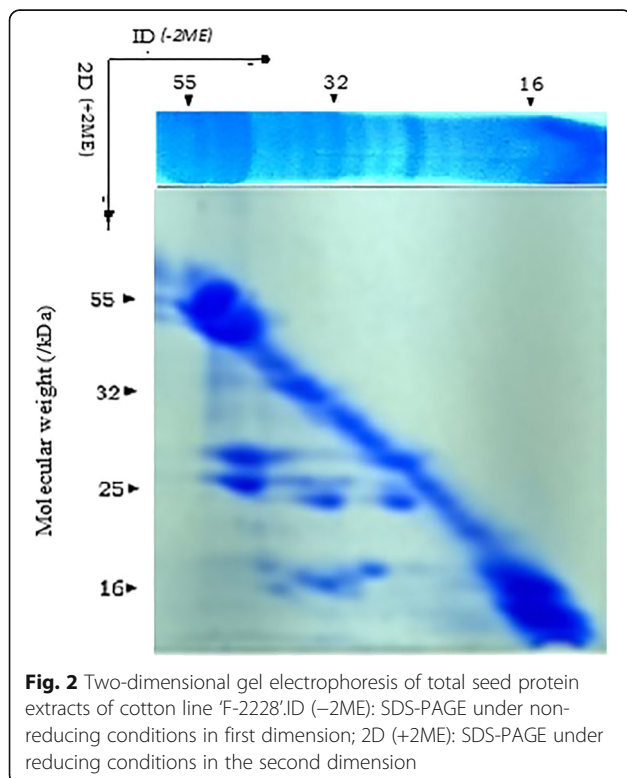


Fig. 2 Two-dimensional gel electrophoresis of total seed protein extracts of cotton line 'F-2228'. 1D (-2ME): SDS-PAGE under non-reducing conditions in first dimension; 2D (+2ME): SDS-PAGE under reducing conditions in the second dimension

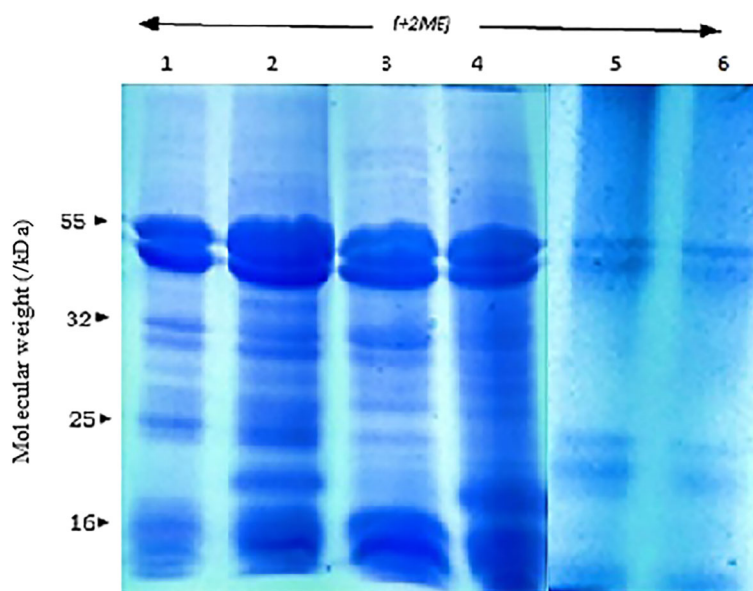


Fig. 3 SDS-polyacrylamide gel electrophoresis of seed protein fractions of cotton lines 'F-2228' and 'FDK-124' of *G. hirsutum* and *G. arboreum*, respectively under reducing conditions. '+2ME' stands for the presence of '2-Mercaptoethanol'

Track

1
2
3
4
5
6

Fraction

Globulins of 'F-2228'
Albumins of 'F-2228'
Globulins of 'FDK-124'
Albumins of 'FDK-124'
Glutelins of 'F-2228'
Glutelins of 'FDK-124'

Correlation studies

Coefficient of correlation, calculated between protein content and four protein fractions in lines of both the *Gossypium* species, could be seen in Tables 8 and 9. A positive significant correlation was found between protein content and albumin fraction in *G. arboreum* lines. On the other hand, globulins and glutelins were observed to be negatively correlated in this species.

In lines of *G. hirsutum*, a significant and positive correlation could be seen in protein content and salt soluble globulins. However, unlike *G. arboreum* lines, glutelins exhibited negative correlation with the albumins.

Discussion

Besides fibre, a major product obtained from the cotton plant, other by-products like cottonseed meal (CSM), cottonseed hull (CSH) and cottonseed oil also possess some good nutritional values. Whereas CSH is a conventional feed for cattle and rich in cellulose, CSM is an important protein source for the ruminants (Osti and Pandey 2006). The major limitation associated with cottonseed protein is the presence of polyphenolic toxic compound – gossypol – which can form a covalent

linkage with the epsilon group of lysine and arginine, thus, reducing the quality of protein (Price et al. 1993). In order to achieve the targets aiming at improving the seed protein quality, knowledge regarding its protein fractions and polypeptides constituting them in different crops is essential. Information on seed storage proteins and nutritional quality in cotton is available through the work mainly carried out by foreign scientists, though a few reports on cotton proteins from Indian labs are also available (Goyal 1992; Kumar et al. 2007; Reddy et al. 2008).

The present work involved the analysis of eight lines of two *Gossypium* species – *G. hirsutum* and *G. arboreum* – for variation in their seed protein content, four protein fractions, polypeptide patterns on SDS-gels. Seed protein content was estimated between 19.5 to 24.3% and 21.8 to 29.5% in *G. arboreum* and *G. hirsutum* lines, respectively, in contrast with the other workers who reported it in the range of 30 to 45% (Church 1991; Mujahid et al. 1999). The variation in protein content between the lines of two *Gossypium* species seems obvious due to the difference in their ploidy levels; the lines belonging to a particular species also displayed a fair variation in seed protein content within

Table 7 Polypeptide composition of seed protein fractions in selected lines of two *Gossypium* species under reducing conditions

Albumin polypeptide		Globulin polypeptide		Glutelin polypeptide	
		Line with Molecular weight /kDa			
'FDK-124' (A-genome)	'F-2228' (AD-genome)	'FDK-124' (A-genome)	'F-2228' (AD-genome)	'FDK-124' (A-genome)	'F-2228' (AD-genome)
120	120	122	122	54	54
80	80	115	115	49	49
75	75	80	80	25.5	25.5
60	60	75	75	21	21
57	57	60	60	19	19
55	55	54	54	12	12
47.5	47.5	52	52	10	10
35	45 ^b	35	46 ^b		
34 ^a	35	33	35		
30 ^a	31 ^b	32	33		
29	29	27 ^a	32		
26 ^a	27 ^b	24	27.5 ^b		
24	25 ^b	22 ^a	26 ^b		
19 ^a	24	20 ^a	25 ^b		
18	18	17 ^a	24		
12	15 ^b	14.5	14.5		
	14 ^b	11.5	11.5		
	13 ^b				

^a denotes the polypeptides present in albumin and globulin fractions of A-genome species only

^b denotes the polypeptides present in albumin and globulin fractions of AD-genome species only

these. The mature seeds of these lines, procured from Punjab Agricultural University in January 2019, were the harvest of previous crop (May to November, 2018). So, the variation in the seed protein content within the lines of the same species may attribute to genotypic differences in the varieties which may lead to varied gene expression for the seed proteins under same growing conditions.

Table 8 Bivariate correlation between protein content and four protein fractions in the cotton lines of *G. arboreum*

Protein content	Albumins	Globulins	Glutelins	Prolamins
1.000	0.970	-0.918	0.778	0.649
	0.030*	0.082	0.222	0.351
		-0.938	0.814	0.476
		0.061	0.186	0.524
			-0.962	-0.619
			0.038*	0.381
				0.599
				0.401

* indicates the coefficients are statistically significant at $P < 0.05$. Pair with positive correlation coefficients and $P < 0.05$ tend to increase together. For the pairs with negative correlation coefficients and $P < 0.05$, one variable tends to decrease while the other increases. For pairs with $P > 0.05$, there is no significant relationship between the two variables

Moreover, the physiological and morphological changes during seed development like fibre development and nutrient mobilization (mainly nitrogen) from leaves (source) to seed (sink) in cotton plant have also been shown to affect the final seed protein content in mature seeds (Bellaloui et al. 2015). Different cropping management practices like rate and time of fertilization treatments,

Table 9 Bivariate correlation between protein content and four protein fractions in the cotton lines of *G. hirsutum*

Protein content	Albumins	Globulins	Glutelins	Prolamins
1.000	-0.741	0.968	-0.914	-0.863
	0.259	0.032*	0.085	0.137
		-0.866	0.897	0.301
		0.134	0.103	0.699
			-0.986	-0.728
			0.014*	0.272
				0.633
				0.367

* indicates the coefficients are statistically significant at $P < 0.05$. Pair with positive correlation coefficients and $P < 0.05$ tend to increase together. For the pairs with negative correlation coefficients and $P < 0.05$, one variable tends to decrease while the other increases. For pairs with $P > 0.05$, there is no significant relationship between the two variables

use of cover crops, plant density and plant growth regulators could be the detrimental factors for the accumulation of the seed protein content (He et al. 2014; Yang et al. 2016). So, it may be drawn that the variation in protein content in the germplasm of a crop is the result of the cumulative effect of genotypes used, growing/environmental conditions, cultural practices, and methods used for its estimation.

Among the four protein fractions, separated on the basis of solubility criteria, globulins represented the major fraction with the proportion varying from 36.2 to 63.7% and 33.0 to 51.5% in *G. arboreum* and *G. hirsutum* lines, respectively. These were followed by albumins (20.8~32.2% and 23.3~29.7%) and glutelins (9.1~24.5% and 19.6~28.0%), respectively, in *G. arboreum* and *G. hirsutum* lines; prolamins being the lowest in both the lines. The pattern of distribution of albumin and globulin fractions in the seed protein was found in contradiction with the results of Gandhi et al. (2017) and Sammour et al. (1995) who reported albumins with 40 to 50% as the major dominating fraction followed by globulins with 21 to 42% in lines of *G. herbaceum* and *G. barbadense*. Like seed protein content as discussed earlier, the proportion of protein fractions and their individual polypeptides could also be altered under different conditions of genetic set up, plant nutrients and other growth conditions. The time period over which and the efficiency with which the genes for these protein fractions are expressed in the developing seed represent the important factors for variation in the ratio of different protein fractions in mature seeds. The existing difference in the proportion of seed protein fractions in the present study with other workers might be the result of different species involved, and also due to different extraction solvents as well as protocols followed.

The correlation studies carried out in any domain of biological research are always helpful in understanding the relationship between the genes governing two characters, whether they are linked or not. In this way, selection and screening for one character would indirectly work for the selection of another character. A positive significant correlation was revealed by us between protein content and albumin fractions in *G. arboreum* lines, and similarly between protein content and salt soluble globulins in *G. hirsutum* lines. On the other hand, the proportion of glutelins had a significant negative correlation with the proportion of albumins and globulins in Indian cotton and American cotton, respectively. Similarly, a negative correlation between seed protein content and gossypol content - a limiting factor in nutritional quality as mentioned earlier, has been shown by Pandey and Thejappa (1975). Recently, a study has demonstrated a high positive correlation of

seed protein content with the lint yield and fineness but at the same time negatively correlation with seed oil content in cotton (Cambell et al. 2016). In this way, considering all these correlation studies including the present study, it could be recommended to select the cotton lines with higher protein content which may prove better in term of improved agronomical (high fibre quality) and nutritional qualities (reduced gossypol content). Further, the lines with low protein content could also be selected as a possible source for producing the value-added product like bio-based wood adhesives as these lines will be having high percentage of water-insoluble/alkali soluble fraction (WIF) which has been suggested as potential protein fraction imparting this quality to the CSM (He et al. 2014).

In order to understand the genetic structure for ultimate applications in breeding programmes, studies on genetic diversity and relationships of the crop cultivars within the species, and with other species have been routinely emphasized. In continuation with this, comparison of polypeptide patterns for any variation and similarities in the lines of both the *Gossypium* species on SDS-gels was carried out. Occurrence of intraspecific variation in molecular weight region 22 kDa to 27 kDa with three banding patterns 'P', 'Q' and 'R' (as described in results) in *G. hirsutum* lines and two patterns 'P' and 'Q' in *G. arboreum* point towards the different rates of evolutionary mechanisms for the genes undergoing various changes; these changes being independent of the genes for polypeptides of one region to the genes for polypeptides of the other region. However, no other regions on SDS-gel were found showing variation in the polypeptide patterns. On further comparing the polypeptide pattern between lines of two species, the presence of similar banding pattern in some A-genome species with AD-genome species marked towards the close evolutionary relationship between these two species. The similar kind of study showing interrelationships among 18 *Oryza* species using seed proteins as markers has also been published by the author emphasizing the importance of these proteins in establishing the evolutionary relationships among other crop species also (Singh et al. 2018).

Further analysis of the polypeptide patterns of four seed protein fractions from the lines of these two species, specify the presence of some polypeptide bands of a fraction to the line of a particular genome group (A- genome) but absent in another (AD-genome). As *G. arboreum* (Diploid A-genome species) has been suggested as one of the progenitors of allotetraploid AD-genome species, i.e. *G. hirsutum* (Hu et al. 2011), the interspecific variation in expression patterns of some of the seed protein fractions between lines of these two species might be the result of interaction between the contributing genomes. The rapid adjustment to duplicated genome dosage is most probably through the control of

gene expression which may be due to gene silencing or gene activation. Transcriptome study has shown the absence of a storage protein subunit in the genome of wheat hexaploid species which otherwise was present in its tetraploid species, and it was explained as the result of inter-genomic suppression of transcript for that particular seed protein subunit after introgression of new diploid genome in allotetraploid genome (Kashkush et al. 2002). Similarly, the differences in the level of gene expression in diploid and its allotetraploid *Gossypium* species, manifested in term of developmental, biochemical changes, etc., could be due to unpredictable gene interaction at the time of genome merging, genome duplication and duplicate gene evolution. These phenomenon may further lead to favouring of one of the genome (A- or D-genome) more, suppressing other or equivalent expression of homeologous genes from both the parental species in allotetraploids as explained by number of workers (Yang et al. 2006; Flagel et al. 2008; Flagel and Wendel 2010). In this way, in the present study, the absence of some of the polypeptide bands of albumins and globulins fractions of diploid A-genome from allotetraploid AD-genome species attribute to the genome biasing during the process of allotetraploidization which favoured the selection of D-genome genes over A-genome or suppression of A-genome genes for these polypeptides by D-genome. The similar kind of polypeptide pattern for alkali-soluble protein fractions (glutelins) in lines of both diploid (A-genome) and tetraploid (AD-genome) species indicate towards the equal expression of the A- and D-genome or the dominance of A-genome over the D-genome for this protein fraction. So, it will be of interest to carry out further studies by examining protein profile and transcript levels of individual seed protein fractions from geographically varied accession of both the diploid (A- and D-genome) and allotetraploid (AD-genome) species for better understanding of the value of gene expression.

Two-dimensional gel electrophoresis in which polypeptides are separated according to their apparent molecular weight under non-reducing conditions in the first dimension, followed by reducing conditions in the second, has proven a valuable tool for analysing the occurrence of any disulphide-linked polypeptides in seed protein extracts. In the present study, a large number of bands with molecular weight 122 kDa, 120 kDa, 115 kDa, 80 kDa, 75 kDa, 60 kDa, 48 kDa, 40.5 kDa, 38 kDa, 27 kDa, 26 kDa, 25 kDa, 18 kDa, 16 kDa, 15 kDa, 14.5 kDa, 14 kDa, 12 kDa, 11.5 kDa and 10 kDa resolved as spots at the same molecular weight positions, as under non-reducing conditions on SDS-gel, along the diagonal on 2-D gels, thus, lacking any disulphide linkages. Meanwhile, the spots occupying the positions below the diagonal could also be seen arising due to the reduction of different kinds of disulphide-linkages in some other

polypeptides. The bands with molecular weight 52 kDa, 40 kDa, 36 kDa and 32 kDa were observed to be having inter-polypeptide disulphide-linkages. Like major 11S globulin sub-fraction in pea family as well as in *Cucumis* (Matta et al. 1981; Singh and Matta 2008), two polypeptides with molecular weight 52 kDa and 40 kDa were shown to consist of heterodimeric subunit pairs with heterogeneity in their respective subunits. Previously, Sammour et al. (1995) have reported only one such type of polypeptide band of molecular weight 45 kDa with disulphide-linkages in total seed protein extracts without assigning it to any protein fraction. Similarly, recent protein characterization studies on *Gossypium* species have reported 60~70% of the total seed proteins belonging to vicilin and legumin families; the later contributed more. These workers have also shown more heterogeneity in the molecular weights of the legumin A subunits (30 kDa, 17~20 kDa and 11~12 kDa) and less in legumin B subunits (11~13 kDa) (Hu et al. 2011; He et al. 2018). Further analysis indicated the polypeptide of molecular weight 58 kDa as a precursor of legumin A subunits. Our results using 2-D gel electrophoresis also exhibited the generation of peptide subunits in the range of molecular weight 17-32 kDa (equivalent to range of legumin A type subunits), from the precursors 52 kDa, 40 kDa, 36 kDa and 32 kDa, as spot below the diagonal. So, these legumin-like polypeptides in the previous studies and the present study may be suggested as the component of globulin fractions of seed storage proteins in *Gossypium*.

Conclusion

The work was carried out to explore the seed protein characteristics in diploid and allotetraploid *Gossypium* species. The end use of cottonseed, for nutritional and industrial purposes, depends upon its seed protein quality. On the basis of our and other workers' studies, it may be stated that the selection of cotton lines with low protein and high protein content will ultimately help in selecting the better lines with improved quality protein seed fraction important in industrial and nutritional uses, respectively. The combined approach of proteomics as well as transcriptomics, involving the accessions of both diploid (A- and D-genome) and allotetraploid (AD-genome) species representing wide geographical areas, could be applied to fully understand the mechanism of differential gene expression for each seed protein fractions in diploids and their allotetraploids species. The percent homology of four legumin-like subunits reported in our study with the legumin subunits further, could be confirmed by purification, sequencing and comparing the peptides for these in databases. The albumin fractions exhibiting maximum variation between lines of two species could be used for diversity analysis in cotton cultivars.

Acknowledgements

Kind supply of cotton seeds from the Punjab Agricultural University, Regional Station, Bathinda (Punjab) is gratefully acknowledged. Authors are grateful to Akal University, Talwandi Sabo for providing necessary infrastructure and research facilities. The corresponding author is very thankful to Dr. Barjinder Singh, Assistant Professor in English, Akal University for his painstaking efforts in revising the MS for its language improvement.

Authors' contributions

The co-author Kaur A is a MS student who carried out the experimental work and generated the data. Singh A who is the supervisor analysed and interpreted the data by reviewing it critically. All authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

The dataset supporting the conclusions of this article is included in the article.

Ethics approval and consent to participate

Not applicable.

Consent for publication

The co-author has consent for submission of manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 8 March 2019 Accepted: 30 May 2019

Published online: 19 June 2019

References

- Alford B, Liepa G, Vanbeber A. Cottonseed protein: what does the future hold? *Plant Foods Hum Nutr.* 1996;49:1–11.
- Bellaloui N, Turley RB, Stetina SR. Water stress and foliar boron application altered cell wall boron and seed nutrition in near-isogenic cotton lines expressing fuzzy and fuzzless seed phenotypes. *PLoS One.* 2015;10:e0130759. <https://doi.org/10.1371/journal.pone.0130759>.
- Bradford MA. Rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding. *Anal Biochem.* 1976;72:248.
- Cambell T, Chapman K, Sturtevant D, et al. Genetic analysis of cottonseed protein and oil in a diverse cotton germplasm. *Crop Sci.* 2016;6:2457–64. <https://doi.org/10.2135/cropsci2015.12.0742>.
- Cheng HN, Ford CV, Dowd MK, He Z. Soy and cottonseed protein blends as wood adhesives. *Ind Crop Prod.* 2016a;85:324–30. <https://doi.org/10.1016/j.indcrop.2015.12.024>.
- Cheng HN, Ford CV, Dowd MK, He Z. Use of additives to enhance the properties of cottonseed protein as wood adhesives. *Int J Adhes Adhes.* 2016b;68:156–60. <https://doi.org/10.1016/j.jadhadh.2016.02.012>.
- Church DC. Livestock feeds and feeding. 3rd ed. New Jersey: Prentice Hall Incorporation; 1991. p. 546.
- Coppock CE, Lanham JK, Horner JL. A review of the nutritive value and utilization of whole cottonseed, cottonseed meal and associated by-products by dairy cattle. *Animal Feed Sci Tech.* 1987;18:89–129.
- Dure LS, Chlan CA. Development biochemistry of cottonseed embryogenesis and germination. XII. Purification and properties of the principle storage proteins. *Plant Physiol.* 1981;68:180–6.
- Fang DD. Molecular breeding. In: Fang DD, Percy RG, editors. *Cotton*, 2nd ed. Madison: ASA, CSSA, and SSSA; 2015. p. 57. <https://doi.org/10.2134/agronmonogr57.2013.0027>.
- Flagel L, Udall J, Nettleton D, Wendel J. Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. *BMC Biol.* 2008;6:16. <https://doi.org/10.1186/1741-7007-6-16>.
- Flagel L, Wendel J. Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. *New Phytol.* 2010;186:184–93. <https://doi.org/10.1111/j.1469-8137.2009.03107.x>.
- Gandhi K, Litoriya NS, Shah A, Talati JG. Extraction, fractionation and characterization of cotton (*Gossypium herbaceum* L.) seed proteins. *Ind J Agri Sci.* 2017;30(1):21. <https://doi.org/10.5958/0974-4479.2017.00003.X>.
- Goyal KC. Identification of some cotton cultivars using PAGE. *Ind J Plant Physiol.* 1992;35(4):405–7.
- Grover CE, Gallagher JP, Jareczek JJ, et al. Re-evaluating the phylogeny of allopolyploid *Gossypium*. *Mol Phylo Evol.* 2015;92:45–52. <https://doi.org/10.1016/j.jympev.2015.05.023>.
- He Z, Cao H, Cheng HN, et al. Effects of vigorous blending on yield and quality of protein isolates extracted from cottonseed and soy flours. *Modern Appl Sci.* 2013;7(10):79–88. <https://doi.org/10.5539/ma.v7n10p79>.
- He Z, Chapital DC, Cheng HN. Comparison of the adhesive performances of soy meal, water washed meal fractions, and protein isolates. *Modern Appl Sci.* 2016a;10(5):112–20. <https://doi.org/10.5539/ma.v10n5p112>.
- He Z, Chapital DC, Cheng HN, Dowd MK. Sequential fractionation of cottonseed meal to improve its wood adhesive properties. *J Am Oil Chem Soc.* 2014a;91:151–8. <https://doi.org/10.1007/s11746-013-2349-2>.
- He Z, Klasson KT, Wang D, et al. Pilot-scale production of washed cottonseed meal and co-products. *Modern Appl Sci.* 2016b;10(2):25–33. <https://doi.org/10.5539/ma.v10n2p25>.
- He Z, Zhang D, Cao H. Protein profiling of water and alkali soluble cottonseed protein isolates. *Sci Report.* 2018;8(1):2045–322. <https://doi.org/10.1038/s41598-018-27671-z>.
- He Z, Zhang H, Olk DC, et al. Protein and fiber profiles of cottonseed from upland cotton with different fertilizations. *Modern Appl Sci.* 2014;8(4):97–105. <https://doi.org/10.5539/ma.v8n4p97>.
- Hu G, Houston NL, Pathak D, et al. Genomically biased accumulation of seed storage proteins in allopolyploid cotton. *Genetics.* 2011;189:1103–15. <https://doi.org/10.1534/genetics.111.132407>.
- Kashkush K, Feldman M, Levy AA. Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics.* 2002;160:1651–9.
- King FE, Lefler HR. Cotton seed protein - its status and potential for improvement in cereals and grain legumes. Vol II. Vienna: International Atomic Energy Agency; 1979. p. 385–91.
- Kumar V, Singh G, Sharma R, Sharma SN. RAPD and protein profiles of cotton varieties. *Ind J Plant Physiol.* 2007;12(2):115–9.
- Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature.* 1970;227:680–8.
- Luthe DS. Storage protein accumulation in developing rice (*Oryza sativa* L.) seeds. *Plant Sci Lett.* 1983;32:147–58.
- Marshall HF. Isolation and purification of cottonseed 7S storage protein and its subunits. *J Agric Food Chem.* 1990;38(7):1454–7.
- Martinez WH. Cotton seed proteins. In: *Proc. Conf. Cotton seed protein concentrates*, USDA-ARS; 1964. p. 51–4.
- Matta NK, Gatehouse JA, Boulter D. Molecular and subunit heterogeneity of legumin of *Pisum sativum* L. – a multidimensional gel electrophoretic study. *J Exp Bot.* 1981;32:1295–307.
- Mujahid A, Abdullah M, Barque AR, Gilani AH. Nutritional value of cottonseeds and its derived products: Physical fractionations and proximate composition. Faisalabad: Department of Animal Nutrition, University of Agriculture; 1999.
- Osti NP, Pandey SB. Use of whole cotton seed and cotton seed meal as a protein source in the diet of ruminant animals: Prevailing situation and opportunity. In: *proceedings of the 6th National Workshop on Livestock and Fisheries Eeasrch*. Lalitpur: National Animal Science Research Institute (NASRI), Nepal; 2006. p. 111–9.
- Pandey SN, Thejappa N. Study on relationship between oil, protein, and gossypol in cottonseed kernels. *J Am Oil Chem Soc.* 1975;52:312–5.
- Peach K, Tracey MV. *Modern methods of plant analysis*, vol. 1. Berlin, Gottingen, Heidelberg: Springer Verlag; 1956.
- Pradyawong S, Li J, He Z, et al. Blending cottonseed meal products with different protein contents for cost-effective wood adhesive performances. *Ind Crop Prod.* 2018;126:31–7. <https://doi.org/10.1016/j.indcrop.2018.09.052>.
- Price WD, Lovell RA, McChesney DG. Naturally occurring toxins in feedstuffs: Centre for Veterinary Medicine Perspective. *J Animal Sci.* 1993;71(9):2556–62.
- Reddy KL, Zullo JM, Bertolino E, Singh H. Identification of cotton cultivars using electrophoresis technique. *Agric Sci Digest.* 2008;28(2):105–8.
- Sammour RH, Elshourbagy MN, Aboshady AM, Abasary AM. Proteins of cottonseed (*Gossypium barbadense*) extraction and characterization by electrophoresis. *Qatar Univ Sci Journal.* 1995;15(1):77–82.

- Schaeffer GW, Sharpe FT. Modification of amino acid composition of endosperm proteins from *in vitro* selected high lysine mutants in rice. *Theor App Genet.* 1990;80:841–6.
- Singh A, Kumar Y, Matta NK. Electrophoretic variation in seed proteins and interrelationships of species in the genus *Oryza*. *Genet Res Crop Evol.* 2018; 65(7):1915–36. <https://doi.org/10.1007/s10722-018-0665-y>.
- Singh A, Matta NK. Disulphide linkages occur in many polypeptides of protein fractions: a two-dimensional gel electrophoretic study. *Rice Sci.* 2011;18:86–94. [https://doi.org/10.1016/S1672-6308\(11\)60013-9](https://doi.org/10.1016/S1672-6308(11)60013-9).
- Singh NP, Matta NK. Variation studied on seed storage protein and phylogenetics of the genus *Cucumis*. *Plant Syst Evol.* 2008;275(3):209–18. <https://doi.org/10.1007/s00606-008-0063-6>.
- Turley RB, Vaughn KC, Scheffler JA. Lint development and properties of fifteen fuzzless seed lines of upland cotton (*Gossypium hirsutum*). *Euphytica.* 2007; 156:57–65. <https://doi.org/10.1007/s1068100693519>.
- Wendel JF, Albert VA. Phylogenetics of the cotton genus (*Gossypium*) character state weighted parsimony analysis of chloroplast DNA restriction site data and its systematic and biogeographic implications. *Syst Bot.* 1992;17:115–43.
- Yang HK, Meng YL, Chen BL, et al. How integrated management strategies promote protein quality of cotton embryos: high levels of soil available N, N assimilation and protein accumulation rate. *Front Plant Sci.* 2016;7:1118. <https://doi.org/10.3389/fpls.2016.01118>.
- Yang S, Cheung F, Lee JJ, et al. Accumulation of genome-specific transcripts, transcription factors and phytohormonal regulators during early stages of fiber cell development in allotetraploid cotton. *Plant J.* 2006;47:761–75.
- Youle RJ, Huang AHC. Occurrence of low molecular weight and high cysteine containing albumin storage proteins in oil seeds of diverse species. *Am J Bot.* 1981;68:44–8.
- Yu J, Yu S, Fan S, et al. Mapping quantitative trait loci for cottonseed oil, protein and gossypol content in *Gossypium hirsutum*/*Gossypium barbadense* backcross inbred line population. *Euphytica.* 2012;187:191–201. <https://doi.org/10.1007/s1068101206303>.
- Yue HB, Fernandez-Blazquez J, Shuttleworth P, et al. Thermomechanical relaxation and different water states in cottonseed protein derived bioplastics. *RSC Adv.* 2014;4:32320–6. <https://doi.org/10.1039/C4RA01794C>.
- Zhang B, Cui Y, Yin G, et al. Synthesis and swelling properties of hydrolyzed cottonseed protein composite superabsorbent hydrogel. *Int J Polym Mat Polym Biomat.* 2010;59:1018–32. <https://doi.org/10.1080/00914031003760709>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

