8-26-2017

Identification of hydrated and dehydrated lipids and protein secondary structures in seeds of cotton (Gossypium hirsutum) line

Diwas Kumar Silwal
Tennessee State University

Nsoki Phambu
Tennessee State University

Bharat Pokharel
Tennessee State University

Ahmad Naseer Aziz
Tennessee State University

Recommended Citation

Research Article

Identification of hydrated and dehydrated lipids and protein secondary structures in seeds of cotton (*Gossypium hirsutum*) lines

Diwas Kumar Silwal[1,2], Nsoki Phambu[3], Bharat Pokharel[1] and Ahmad Naseer Aziz[1*]

1. Department of Agricultural and Environmental Sciences, Tennessee State University, Nashville TN 37209-USA
2. Present Address: Systems Plant Physiology, Texas A & M Agri Life Research, Uvalde TX 78801-USA
3. Department of Chemistry, Tennessee State University, Nashville TN 37209-USA
Corresponding author's email: aaziz@tnstae.edu

Abstract
Cottonseeds from two parents (TM-1 and 3-79) and their 17 progeny (chromosomal substitution) lines were analyzed for various secondary structures of proteins and moisture content of lipids, separately in hulls and kernels. Fourier transform infrared spectroscopy (FTIR) was used on mature seeds from Upland cotton (*G. hirsutum*) progeny lines and parents. Based on secondary structures of proteins and hydration levels of lipids, differences were observed among the cottonseeds. The two progeny lines – CS-B12sh and CS-B22sh retained lipid moisture content and protein secondary structures similar to both parents, while CS-B06, CS-B15sh and CS-B16 remained distinct from either parent. On the other hand, CS-B05sh, CS-B07 and CS-B26lo were alike to TM-1 parent for lipid and protein profiles, whereas CS-B02 and CS-B04 were comparable with 3-79 parent. The capacity to detect hydrated and dehydrated lipids and different protein secondary structures using FTIR in these cottonseeds is the novel finding of this project for improving the seed nutritional traits in cotton towards cooking-oil and protein-feed usages.
Keywords: Upland cotton (*Gossypium hirsutum*); Pima cotton (*G. barbadense*); Fourier Transform Infrared Spectroscopy (FTIR); Lipid; Protein

Introduction
Upland cotton (*Gossypium hirsutum*) and Pima cotton (*G. barbadense*) are the two most commonly cultivated cotton species which attribute for 98% and 2% of the cotton acreage in the United States respectively. Current cotton yield in the United States is 1 1/3 bales of fiber and about 1,078 pounds of seeds per acres [1] and it generates more than $120 billion business revenue annually. Grown in less than 3% of world's total agricultural land, cotton farming supports 30% global need for textile fiber [2]. The largest uses of cotton are from its fiber that is used to produce clothes, towels, shoe strings, sheets, high

quality papers and cushions [3]. Besides fibers, cottonseeds are crushed to obtain oil and cottonseed meal for more than 100 years. Cottonseed kernels have been pressed for the extraction of oil which can be directly used as cooking oil for human consumption as it contains low level of saturated fatty acids and possesses a high level of natural antioxidants 'tocopherols' [4]. Kernels are the most nutritious part which contains 28.24 to 44.05% of oil and 27.83 to 45.6% of protein along with 17 different kinds of amino acids [5]. Whole cottonseed (WCS) feed increases milk production and fat test in high-producing dairy cow and does not interfere with forage digestion when fed at a reasonable level. When dried, WCS contains 23% protein, 25% crude fiber and 20% fat, so are termed as a cost-effective "Triple nutrient" [6]. Cottonseed meal contains around 41% proteins and thus can also be used as protein concentrate in the form of cake or pellet [7]. Previous gene introgression approaches in cotton were done only in whole genome level, which caused accumulation of unwanted DNA resulting in negative effects such as infertility, cytological abnormalities and distorted segregation [8]. This led to the development of backcrossed chromosome or chromosome arm substitution (CS) lines by replacing chromosomes or chromosomal arms of *G. hirsutum* with corresponding chromosomes or chromosomal arms from *G. barbadense* (B) lines. Upland parent (TM-1) had been derived as an inbred from Deltapine 14, a commercial cultivar, and maintained for 40 generations through self-pollination. The Pima parent (3-79) was developed as a doubled haploid from Pima germplasm. All their progeny aneuploid CS-B lines used in this study are nearly isogenic to TM-1 parent for 25 chromosomal pairs and with themselves for the 24 chromosomal pairs [9].

Infrared (IR) spectroscopy is a classically common and non-destructive experimental method for evaluation of organic compounds which can be used under wide variety of environments [10]. During IR procedure, samples absorb electromagnetic radiation based on molecular vibrational stretching and bending of the chemical bonds which can be used to determine membrane linked alignment and conformation of the functional groups. The absorbance of electromagnetic radiation in IR is directly proportional to the concentration of target molecules and the path length of the measuring cell which allows measurement with high precision and high throughput at low cost [11]. Sun *et al*. [12] studied chemical composition between transgenic and non-transgenic cottonseeds, while using of Fourier Transform (FT)-IR in four absorption regions, 1800 to 1720, 1720 to 1580, 1580 to 1480 and 1200 to 1130 cm^{-1}. They compared the protein secondary structures for cottonseeds, and their profiles showed large amounts of α-helices and β-sheets along with some random coils and β-turns.

The cottonseeds used in this study were from two parents - TM-1 and 3-79 as well as their 17 progeny lines with either substituted chromosomes or short (sh)/long (lo) chromosomal segments from *G. barbadense*. The progeny CS-B lines assayed were: CS-B01, CS-B02, CS-B04, CS-B05sh, CS-B06, CS-B07, CS-B11sh, CS-B12sh, CS-B14sh, CS-B15sh, CS-B16, CS-B17, CS-B18, CS-B22sh, CS-B22lo, CS-B25 and CS-B26lo. Total lipid contents of TM-1 and 3-79 are reported [5] to be 22.14% and 25.79%, respectively. Horn *et al*. [13] also estimated the total protein levels for TM-1 and 3-79 to be 15.14% and 21.69%, respectively. Thus, their 17 CS-B progeny lines used in this study were not expected to widely differ from either parent or from each other in terms of the total seed lipid and protein

contents. This research, therefore, compared the CS-B lines for their protein secondary structures (Table 1) as well as difference in

moisture content of lipids, reflecting the nutritional quality, separately for hulls and kernels.

Table 1. Common protein secondary structures and their respective reference frequencies for Infrared spectrometer analyses

Protein secondary structures	Wavenumber (cm^{-1})
β-Turns	1665 to 1680
α-Helices	1646 to 1660
Random Coils	1638 to 1645
β-Sheet	1615 to 1637 1685 to 1699

Adapted and modified from Tamm and Tatulian [11]. The above ranges are for guideline purposes only

Results and discussion

Cotton can be grown not only as one of the most important cash crop for its fiber but also for seed nutritional elements usage in vegetable oil and protein seed supplements. Both parents, TM-1 and 3-79 had detectable lipid and protein levels, however, based on their moisture contents or secondary structures 3-79 appears to have profiles of two nutritional elements that are more desirable (Tables 2 and 3). The comparison

of 17 progeny CS-B lines for nutritional seed traits (lipid and proteins) with their two parents and among themselves revealed that hulls of CS-B07 and CS-B25 were closely related to TM-1 while CS-B01, CS-B04, CS-B05sh and CS-B22sh hulls reflected 3-79 profiles (Figure 1). Kernels of CS-B01, CS-B11sh, CS-B12sh, CS-B22sh and CS-B26lo were closely related to TM-1 while CS-B02 and CS-B04 kernels were like that of 3-79 (Figure 2).

Table 2. Identification of areas under curve for hydrated and dehydrated lipids through Fourier Transform Infrared Spectrometry analyses of seeds from TM-1 (*G. hirsutum*) and 3-79 (*G. barbadense*) parents as well as their 17-progeny chromosomal substitution (CS) lines with *G. barbadense* (B) chromosome or segments in TM-1 background

Sample	Hull		Kernel	
	ṽ (Hy. /Dehy.) *	Area^{a}	ṽ (Hy. /Dehy.) *	Area^{a}
TM-1	1720.65 (Dehy.)	2.73	1743. 13 (Hy.)	10.28
3-79	1746.88 (Hy.)	3.18	1744.16 (Hy.)	8.07
CS-B01	1738.01 (Hy.)	5.45	1743.06 (Hy.)	12.52
CS-B02	1739.32 (Hy.)	2.10	1743.92 (Hy.)	6.89
CS-B04	1744.22 (Hy.)	1.20	1744.28 (Hy.)	8.60
CS-B05sh	1721.45 (Dehy.)	4.61	1742.45 (Hy.)	10.41
CS-B06	1742.84 (Hy.)	5.46	1742.62 (Hy.)	10.10
CS-B07	1716.79 (Dehy.)	4.27	1743.27 (Hy.)	9.86
CS-B11sh	1740. 93 (Hy.)	2.22	1742.78 (Hy.)	11.89
CS-B12sh	1744.32 (Hy.) 1719.13 (Dehy.)	7.34 2.51	1742.89 (Hy.)	10.87
CS-B14sh	1742.89 (Hy.)	5.44	1742.83 (Hy.)	10.51
CS-B15sh	1742.43 (Hy.)	5.78	1739.34 (Hy.)	7.82
CS-B16	1744.92 (Hy.)	2.65	1717.83 (Dehy.)	5.69

Silwal et al.

Sample	ṽ	Area	ṽ	Area
CS-B17	1742.23 (Hy.)	2.99	1741.94 (Hy.)	9.12
	1706.18 (Dehy.)	3.18		
CS-B18	1746.44 (Hy.)	1.35	1742.18 (Hy.)	9.61
CS-B22Lo	1739.41 (Hy.)	4.06	1740.72 (Hy.)	8.86
CS-B22sh	1746.02 (Hy.)	3.18	1742.71 (Hy.)	10.45
	1725.86 (Dehy.)	1.06		
CS-B25	1741.58 (Hy.)	3.14	1742.78 (Hy.)	11.51
CS-B26Lo	1744.45 (Hy.)	4.28	1743.23 (Hy.)	10.31
	1718.99 (Dehy.)	1.07		

*Wave-numbers (ṽ) for detecting Hydrated (Hy) lipids (1750 to 1731 cm^{-1}) and Dehydrated (Dehy) lipids (1700 to 1730 cm^{-1}) in CS-B lines

[a]Area under curve for Fourier Transform Infrared spectra of lipids

Table 3. Determination of areas under curve for different protein secondary structures through Fourier Transform Infrared Spectrometry of seeds from 17 *G. hirsutum* cotton chromosomal substitution (CS) lines with *G. barbadense* (B) chromosome or segments as well as both that of TM-1 (*G. hirsutum*) and 3-79 (*G. barbadense*) parents

Sample	β-Turns		Random Coils		α-Helices		β-Sheets	
	ṽ (H/K)*	Area[a]	ṽ (H/K)*	Area[a]	ṽ (H/K)*	Area[a]	ṽ (H/K)*	Area[a]
TM-1	1673.14 (H)	36.79					1631.57(K)	32.12
3-79			1643.69(K)	64.2	1650.73(H)	44.2		
CS-B01					1647.17(H)	65.56	1632.25(K)	29.58
CS-B02			1642.95(K)	73.46	1663.81(H)	16.45	1635.07(H)	31.36
CS-B04			1643.75(K)	56.99	1649.86(H)	49.56	1685.11(H)	1.15
CS-B05sh	1683.90 (H)	0.125			1654.91(K)	13.50		
CS-B06	1673.48 (K)	21.94					1624.22(H)	61.75
CS-B07	1673.38 (H)	23.26					1631.25(K)	44.35
CS-B11sh	1680.87 (H)	10.05			1658.13(H)	11.33	1635.55(H)	3.24
							1626.45(K)	33.87
CS-B12sh	1678.29 (H)	11.05			1654.28(H)	18.51	1628.84(H)	23.74
							1630.54(K)	35.80
CS-B14sh					1646.58(H)	9.05	1618.31(K)	42.67
CS-B15sh	1670.44 (H)	30.26			1657.20(K)	35.37		
CS-B16			1644.44(H)	70.77	1652.94(K)	66.36		
CS-B17	1671.10 (H)	2.74			1647.85(K)	33.12		
CS-B18	1677.26 (H)	6.73			1655.81(H)	10.08	1696.07(H)	0.18
					1652.42(K)	13.50	1635.50(H)	12.26
							1615.22(H)	13.04
CS-B22lo					1661.18(H)	7.37	1633.57(K)	59.15
CS-B22sh	1670.31 (H)	4.27			1656.28(H)	43.05	1685.54(H)	4.24
							1627.10(H)	2.74
							1632.91(K)	32.84
CS-B25	1670.49 (H)	29.15					1627.71(K)	30.29
CS-B26lo	1677.55 (H)	12.65	1645.41(H)	15.32			1638.60(K)	24.92

*Wavenumber (ṽ) to identify β-Turns (1665 to 1680 cm^{-1}), α-helices (1646 to 1660 cm^{-1}), Random coils (1638 to 1645 cm^{-1}) and β-sheets (1610 to 1637 and 1685 to 1699 cm^{-1}) in Hulls (H) and Kernels (K) of CS-B lines

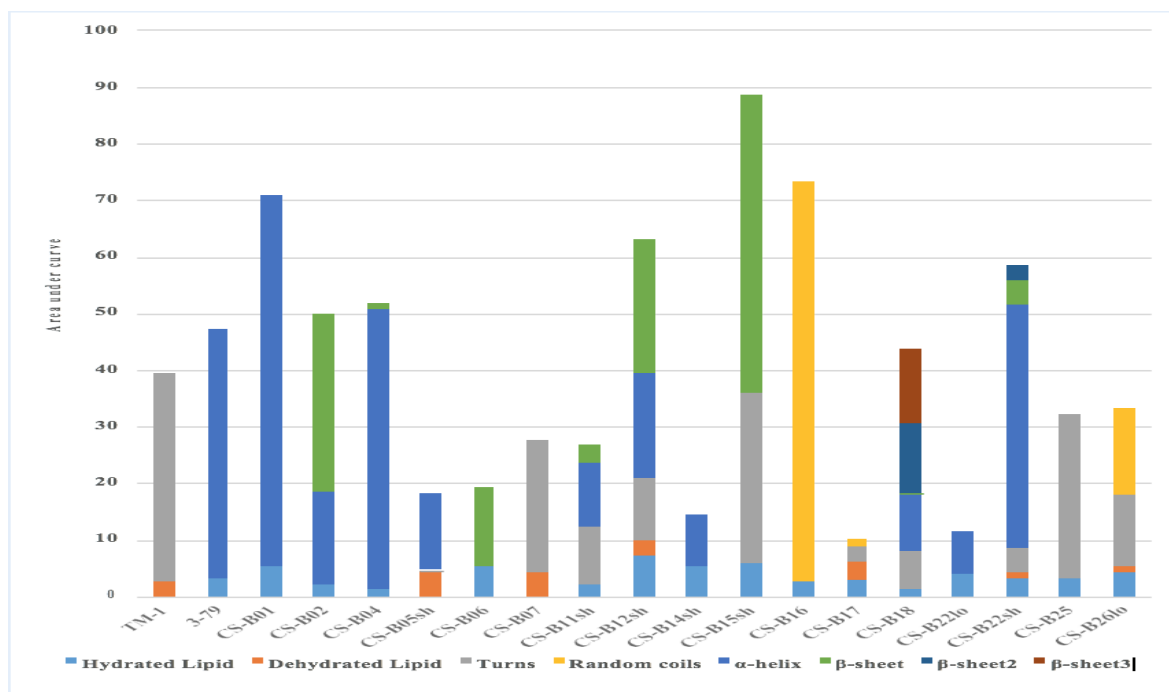[a]Area under curve for Fourier Transform Infrared spectra of protein secondary structures

Figure 1. Fourier Transform Infrared Spectrometry based evidence of secondary structures of protein and lipid profiles in hulls of seeds from 17 *G. hirsutum* cotton chromosomal substitution (CS) lines with *G. barbadense* (B) chromosome or segments as well as both that of TM-1 (*G. hirsutum*) and 3-79 (*G. barbadense*) parents
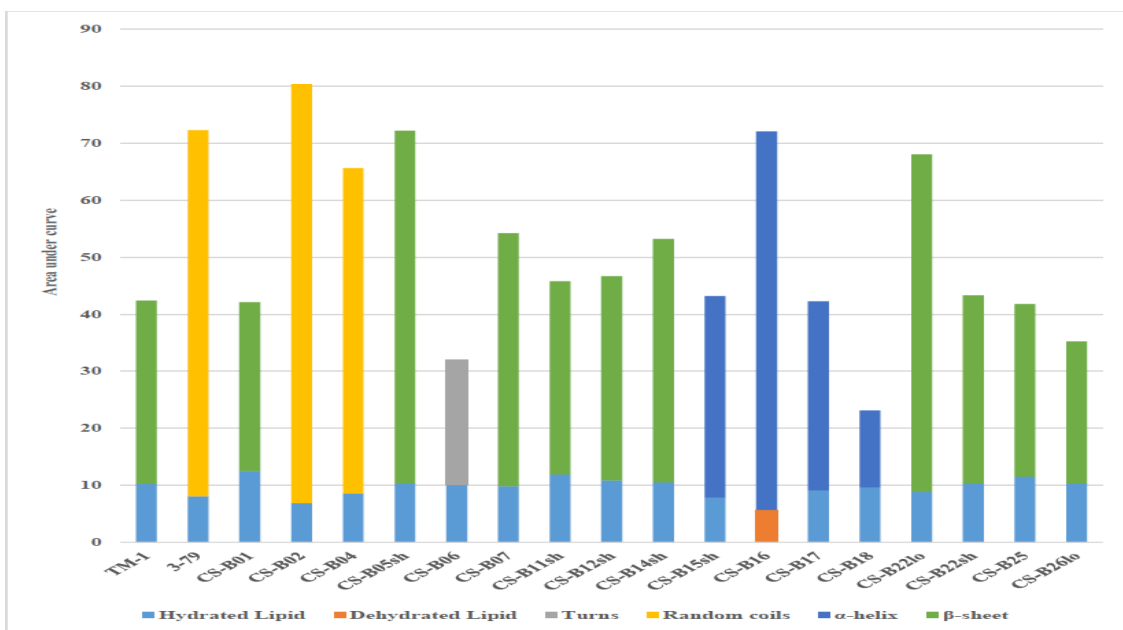

Figure 2. Detection of secondary structures of protein and lipid profiles through Fourier Transform Infrared Spectrometer in kernels of seeds from TM-1 (*G. hirsutum*) and 3-79 (*G. barbadense*) parents as well as their 17-progeny chromosomal substitution (CS) lines with *G. barbadense* (B) chromosome or segments in TM-1 background

Lipid profiles per moisture contents

The initial phases of seedling growth are supported by the nutritional reserves available in the cotyledons, which are primarily represented by lipid and protein contents in cottonseeds [14]. Both components are stored in cytosolic organelles of mature seeds and in cottonseeds the stored lipids are approximately 20% of the dry weight [13]. For the 19 cotton lines assayed, both the hydrated and dehydrated types of lipids were evident in cottonseed hulls as well as kernels. The wavenumbers ranging from 1731 to 1750 cm^{-1} were selected to detect hydrated type of lipids and 1700 to 1730 cm^{-1} were used for the dehydrated types. Pima parent (3-79) had relatively higher lipid area than Upland parent (TM-1) in hull but lower in the kernel (Table 2). Both parents had same kinds of lipids in their kernel but were found to be different in hull, since TM-1 hull showed evidence for dehydrated lipid while 3-79 hull had the hydrated type. The distinctive detection of hydrated and dehydrated lipids was possible non-destructively through FTIR analysis of cottonseed samples and this warrants further research for better understanding of mechanism and roles of lipid structures especially in the areas of seed physiology.

Cottonseed hulls from 3-79 and its 11 progeny lines showed evidence for hydrated type of lipid (Table 2) while only seven lines including Upland parent TM-1, CS-B05sh, CS-B07, CS-B12sh, CS-B17, CS-B22sh and CS-B26lo showed presence for dehydrated type (Table 2). Four lines CS-B12sh, CS-B17, CS-B22sh and CS-B26lo showed evidence of both kinds of lipids in their hulls, of which CS-B12sh possessed the highest lipid profiles. These four progeny lines showed traits of both the parental hull types, while all other progenies were like either one of the parents (TM-1 or 3-79). Among the 12 cottonseed hulls showing evidence for hydrated lipids, 3-79, CS-B12sh had the highest relative amount (7.34) while CS-B04 had the lowest (1.20). CS-B22sh and CS-B25 were closely related to 3-79 in terms of lipid curve area (Table 2). Among the seven cottonseed hulls that showed evidence for dehydrated lipid, TM-1 and CS-B05sh had the highest area (4.61) whereas CS-B22sh had the lowest (1.06). CS-B12sh was like TM-1 while CS-B05sh, CS-B07 and CS-B17 had higher lipid areas than this parental line (Table 2). When cottonseeds were analyzed for kernels, all the samples except CS-B16 showed the evidence for hydrated lipids, of which CS-B01 had the highest (12.52) lipid area (Table 2). Hydrated lipids can be advantageous in seed germination as they exhibit stronger hydrogen bonds and act like "water bridges", thereby water molecules in the lipid bilayer tightly connect with the neighboring phosphate and carbonyl head-groups [15]. The most energy dense storage compound present in dormant cottonseed is lipids, which are utilized for the growth of the plant after its germination. Thus, presence of lipid in the seed kernels, as witnessed in these CS-B lines, boosts the energy required for the growing seedling [16]. CS-B04 and CS-B22lo showed evidence for similar structures to that of Pima parent (3-79) while CS-B05sh, CS-B06, CS-B12sh, CS-B14sh, CS-B22sh and CS-B26lo were closely related to Upland parent (TM-1) in terms of lipid areas in kernels (Table 2). The discrimination of lipid profiles in hulls and kernels by FTIR indicates the importance of this tool in selection of breeding lines for seed traits.

Protein secondary structures' comparison

Hinze *et al.* [17] have reported protein contents above 21% for both *G. hirsutum* and *G. barbadense*, thus both parental lines have high potentiality as protein source even for human consumption though the seeds are primarily fed to livestock. The lack of

detailed information about the structure of functional storage proteins of plant seeds' protein has hampered genetic transformation approaches for improving their nutritional traits [18]. The proteins which are stored in organelles of seeds [13] are degraded by endogenous protease enzymes upon germination for providing nutrition to the growing seedlings [19]. Storage proteins in seed tissues are highly expressed during later growth of the plants, and determine the seed nutritional value when used as food or as feed [20].

All analyzed samples for hulls and kernels showed the evidence for several protein secondary structures (Table 3). Most CS-B lines possessed multiple secondary structures in their hulls while kernels contained only a single kind of structure. α-helices and β-sheets were the major secondary structures found in these cottonseeds, and both structures are also highly responsible for the organization of three-dimensional proteins [21]. The two cotton parents showed completely different secondary structures profiles for both their hulls and kernel when compared with each other. In hulls, TM-1 showed evidence of turns while α-helix was evident in 3-79 (Table 3) suggesting that the latter had stable protein structure. For kernels, TM-1 was detected with β-sheets while random coils were evident in 3-79 (Table 3), suggesting that the former had more stable proteins. Hulls of CS-B18 and CS-B22sh showed evidence for all the four-different protein secondary structures, of which CS-B18 had three and CS-B22sh had two curves of β-sheets respectively (Table 3).

Turns and random coils
TM-1, CS-B05sh, CS-B07, CS-B11sh, CS-B12sh, CS-B15sh, CS-B17, CS-B18, CS-B22sh, CS-B25 and CS-B26lo hulls showed the evidence of turns (Table 3), of which TM-1 had the highest area (36.79) and CS-B05sh had the lowest (0.125). For kernels,

only CS-B06 had this protein secondary structure with an area of 21.94 (Table 3). Only two progeny lines, i.e., CS-B16, CS-B17 and CS-B26lo hulls showed evidence of random coils (Table 3), of which CS-B16 had the highest area (70.77) and CS-B17 had the lowest (1.27). Kernels of two progeny lines, CS-B02 and CS-B04 as well as 3-79 parent showed the evidence of random coils (Table 3), of which CS-B02 had highest area (73.46) and CS-B04 had the lowest area (56.99) of random coils. Due to their hydrophilic nature, these unordered protein structures serve as water binding proteins, interact with macromolecules as a water matrix, may act as hydration buffers that regulate the water in cells and help to resist protein denaturation when the tissues are dehydrated [22]. Thus, the above cotton lines with evidences for these unordered protein structures may have better seed germination rates.

α-Helices
Hulls of 3-79, CS-B01, CS-B02, CS-B04, CS-B05sh, CS-B11sh, CS-B12sh, CS-B14sh, CS-B18, CS-B22sh and CS-B22lo showed the evidence of α-helices (Table 3). Among these lines, CS-B01 had the highest area (65.56) for α-helices while CS-B22lo had the lowest (7.37). The area under α-helix curves for CS-B01 and CS-B04 hulls were higher than 3-79 parent while that for other eight progeny lines had lower. Kernels of CS-B15sh, CS-B16, CS-B17 and CS-B18, showed evidence of α-helices (Table 3), among which CS-B16 had the highest area (66.36) while CS-B18 had the lowest (13.50). From the above CS-B lines, only CS-B18 had α-helical structure both in its hull as well as kernel (Table 3). α-helix rich proteins may provide fundamental mechanical support in cells, outline cell's stretchiness, enable binding with other signaling proteins, and help in cell motility as well as biochemical signaling [23].

β-Sheets

CS-B02, CS-B04, CS-B06, CS-B11sh, CS-B12sh, CS-B15sh, CS-B18 and CS-B22sh hulls showed evidence of β-sheets (Table 3). Among these eight lines CS-B18 had three curves of β-sheets at 1696.07, 1635.5 cm^{-1} and 1615.22 cm^{-1} with areas of 0.18, 12.26 and 13.04 respectively. Also, CS-B22sh showed evidence for two curves of β-sheets at 1685.54 and 1627.10 cm^{-1} with areas of 4.24 and 2.74 respectively (Table 3). Among these eight hulls tested, CS-B15sh had the highest area (52.51) while CS-B18 had the least (0.18) β-sheets. Kernels of 11 cottonseeds showed the evidence of β-sheets (Table 3), of which CS-B05sh had the highest (61.75) while CS-B17 had the lowest area (8.82). Kernels of CS-B05sh, CS-B07, CS-B11sh, CS-B12sh, CS-B14sh, CS-B22sh and CS-B22lo had higher area of β-sheets than TM-1 parent (32.12) while CS-B01, CS-B25 and CS-B26lo had lower (Table 3). β-sheets are formed from the simultaneous uncoiling of α-helices, the phenomenon termed as α-β transition [23]. Activities such as heating and roasting have been reported [14] to increase the percentage of β-sheets in cellular level. This protein structure is un-degradable and un-digestible, which lowers the feed value and its access to gastrointestinal digestive enzymes in ruminants [24]. Thus, CS-B progeny lines higher in β-sheets may not be desirable as feedstock development choice.

In this paper FTIR spectroscopy was found reliable and convenient for analyzing the secondary structures of lipids and proteins, non-destructively from all cottonseeds of these CS-B lines, which may also be used as source of food and feed [25, 26]. The CS-B lines used in these studies had background of TM-1 parent, therefore, as expected most of the progeny lines (CS-B01, CS-B05sh, CS-B07, CS-B11sh, CS-B12sh, CS-B14sh, CS-B22lo, CS-B22sh, CS-B25 and CS-B26lo) clustered around *G. hirsutum* parent based on their lipid and protein secondary structures (Figure 3). The remaining seven progeny lines (CS-B02, CS-B04, CS-B06, CS-B15sh, CS-B16, CS-B17 and CS-B18) had the lipid and protein FTIR profiles like 3-79 parent probably reflecting their *G. barbadense* foreground (Figure 3). Further understanding the mechanism of variation pertaining to these nutritional components between the lines could be utilized for selecting effective candidates in crop improvement programs.
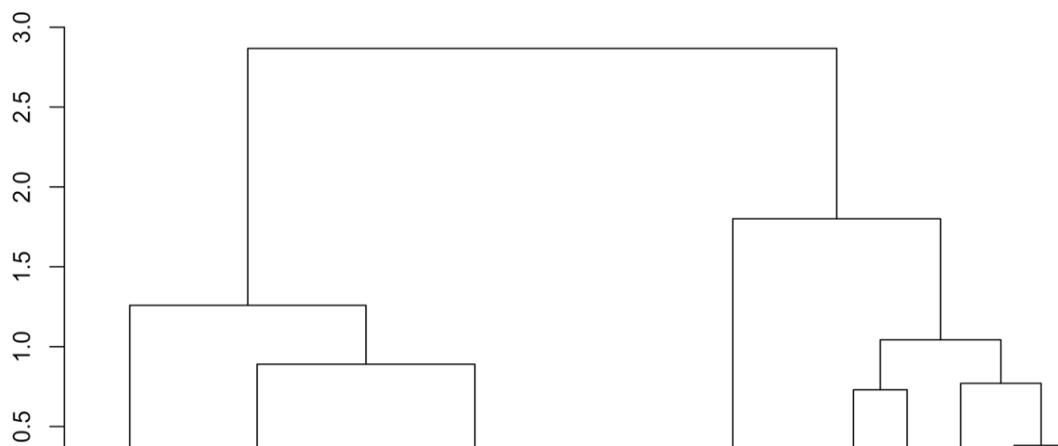

Figure 3. Dendrogram based on Fourier Transform Infrared Spectrometry detection of protein secondary structures and lipid moisture contents for seeds of *G. hirsutum* and *G. barbadense* (B) parent along with their 17 chromosomal substitution (CS) progeny lines with TM-1 background and *G. barbadense* chromosome or segments as foreground

General experimental procedures
Sample preparation
Mature delinted cottonseeds were obtained from Dr. Sukumar Saha (USDA-ARS, Genetics and Sustainable Agriculture Research Unit, MS, 39762). The seeds from each cotton line were randomly selected for the analysis and cut into thin pieces with help of a razor blade, while separating the hulls (outer coverings) and kernels (inner meats). Each sliced sample was placed on the sample holder of a Nicolet^{TM}is^{TM}10 spectrometer (Thermo Fisher Scientific, Waltham, MA, USA), firmly clamped and analyzed through OMNIC (v8.0) software as per the method of Silwal *et al.* [27].

Lipid and protein profiling by FTIR spectroscopy
Sixteen scans were collected for each sample, corrected against the background and then smoothened in SPECTRUM (v3.02, Sunnyvale, CA, USA) software three times to improve the quality. For determination of lipid moisture content and protein secondary structures, peak fitting was done in Gaussian, after setting full width at half height (FWHH) of 15 and the noise target at 10 for each spectrum. Area under the curve was measured for lipid in the region of 1700 to 1750 cm^{-1} and for the four secondary structures of protein (turns, α-helices, random coils and β-sheets) the regions of 1665 to 1680 cm^{-1}, 1646 to 1660 cm^{-1}, 1638 to 1645 cm^{-1} and 1615 to 1637 and 1685 to 1699 cm^{-1} were used (Table 1).

Statistical analysis
Seeds from some lines were analyzed twice to confirm the reproducibility of FTIR spectra. Data was analyzed in Microsoft (Redmond, WA) Excel (v2016) and chart diagrams were created for comparison with both parents and among other progeny lines for the two seed nutritional components. Dendrogram was created for both hulls and kernels in R-statistical computing environment (v3.1.3) [28] using "ClustOfVar" package [29].

Conclusion
Fourier transform infrared spectroscopy (FTIR) was found an appropriate and precise tool for non-destructive lipid moisture content and protein secondary structures from mature seeds to profile cotton lines. FTIR effectively and conveniently analyzed cotton lines to select candidate germplasm as source of food and feed in breeding programs. Therefore, such tools are also relevant to discover mechanisms of variation pertaining to seed nutritional components in cotton.

Authors' contributions
Conceived and designed the experiments: AN Aziz & N Phambu, Performed the Experiments: DK Silwal, AN Aziz & N Phambu, Analyzed the Data: DK Silwal, B Pokharel, AN Aziz & N Phambu, Contributed reagents/ materials/ analysis tools: AN Aziz & N Phambu, Wrote the paper: DK Silwal & AN Aziz.

References
1. National Cotton Council of America (2015). The story of cotton: Where cotton grows.
   https://www.cotton.org/pubs/cottoncounts/

story/where.cfm/, Accessed September 13, 2015.

2. Cotton Incorporated (2016). *http://cottontoday.cottoninc.com/wp-content/uploads/2016/08/Land-Fact-Sheet1.pdf*, Accessed July 17, 2017.

3. National Cotton Council of America (2015). The story of cotton: The importance of cotton. *https://www.cotton.org/pubs/cottoncounts/story/importance.cfm,* Accessed September 13, 2015.

4. Cotton Incorporated (2017). Cottonseed. *http://cottontoday.cottoninc.com/cotton-byproducts/cottonseed/,* Accessed July 17, 2017.

5. Liu H, Quampah A, Chen J, Li J, Huang Z, He Q, Zhu S & Shi C (2013). QTL mapping based on different genetic systems for essential amino acid contents in cottonseeds in different environments. *PLoS ONE.* 8: e57531. https://doi.org/10.1371/journal.pone.0057531

6. Cotton Incorporated (2017). Benefits of feeding cottonseed *http://www.wholecottonseed.com/cottonseed-101/benefits-of-feeding-wcs,* Accessed July17, 2017.

7. National Cottonseed Products Association, *http://www.cottonseed.com/publications/cottonseedanditsproducts.asp/,* Accessed September 13, 2015.

8. Stelly DM, Saha S, Raska D, Jenkins J, McCarty JC & Gutierrez O (2005). Registration of 17 Upland (*Gossypium hirsutum*) germplasm lines disomic for different *G. barbadense* chromosome or arm substitutions *Crop Sc* 45: 2663-2665.

9. Saha S, Stelly DM, Raska DA, Wu J, Jenkins JN, McCarty JC, Makamov A, Gotmare V, Abdurakhmonov JY & Campbell BT (2012). Chromosome substitution lines: Concept, development and utilization in the genetic improvement

of upland cotton. In: Abdurakhmonov JY, Editor. *Plant Breeding* Chapter 6, ISBN: 978-953-307-932-5.

10. Kong J & Yu S (2007). Fourier transform infrared spectroscopic analysis of protein secondary structures. *Acta Biochimica. et Biophysica. Sin.* 39: 549-559.

11. Tamm LK & Tatulian SA (1997). Infrared spectroscopy of proteins and peptides in lipid bilayers. *Quart Rev of Biophy* 30: 365-429.

12. Sun C, Wu X, Wang L, Wang Y, Zhang Y, Chen L & Wu Z (2012). Comparison of chemical composition of different transgenic insect-resistant cotton seeds using Fourier transform infrared spectroscopy (FTIR) *Afri J of Agri Res* 7: 2918-2925.

13. Horn PJ, Neogi P, Tombokan X, Ghosh, S, Campbell BT & Chapman KD (2011). Simultaneous quantification of oil and protein in cottonseed by low-field time-domain nuclear magnetic resonance. *J Am Oil Chem Soc* 88: 1521-1529.

14. Turley RB & Chapman KD (2010). Ontogeny of cotton seeds: gametogenesis, embryogenesis, germination, and seedling growth. In: Stewart JM, Oosterhuis DM, Heitholt JJ, & Mauney JR, Editors. *Physiology of Cotton* Springer, New York, Chapter 29, pp 332-341.

15. Binder H (2007). Water near lipid membranes as seen by infrared spectroscopy. *Eur Biophys J* 36: 265-279.

16. Bartee SN & Krieg DR (1974). Cottonseed density: associated physical and chemical properties of 10 cultivars. *Agron J* 66: 433-435.

17. Hinze LL, Horn PJ, Dever JK, Frelichowski J, Chapman KD & Percy RG (2015). Non-destructive measurements of cottonseed nutritional trait diversity in the U.S. National

Cotton Germplasm Collection. *Crop Sc* 55: 770-782.

18. Derbyshire E, Wright DJ & Boulter D (1976). Legumin and vicilin, storage proteins of legume seeds. *Phytochemistry* 15: 3-24.

19. Millerd A (1975). Biochemistry of legume seed proteins. *Plant Physio* 26: 53-72. 10.1146/annurev.pp.26.060175.000413. http://www.annualreviews.org/doi/abs/10.1146/annurev.pp.26.060175.000413.

20. Lawrence MC, Suzuki E, Varghese JN, Davis PC, Van Donkelaar A, Tulloch PA & Colman PM (1990). The three-dimensional structure of the seed storage protein phaseolin at 3 Å resolution. *EMBO J* 9: 9–15.

21. Branden C & Tooze J (1999). Introduction to Protein Structure. Garland Publication, New York, pp. 13-33.

22. Qin Z, Krepak L & Buehler MJ (2009). Hierarchical structure controls nanomechanical properties of vimentin intermediate filaments. *PLoS ONE*. 4: e7294.

23. Soulages JL, Kim K, Walters C & Cushman JC (2002). Temperature-induced extended helix/random coil transitions in a group 1 late embryogenesis-abundant protein from soybean. *Plant Physio* 128: 822-832.

24. Yu P (2005). Protein secondary structures (a-helix and b-sheet) at a cellular level and protein fractions in relation to rumen degradation behaviours of protein: a new approach. *British J of Nutri* 9: 655-665.

25. Wu J, Jenkins JN, McCarty JC & Thaxton P (2009). Seed trait evaluation of *Gossypium barbadense* L. chromosomes/arms in a *G. hirsutum* L. background. *Euphytica* 167: 371-380.

26. Wu J, McCarty JC & Jenkins JN (2010). Cotton chromosome substitution lines crossed with cultivars: genetic model evaluation and seed trait analyses. *Theor Appl Genet* 120: 1473-1483.

27. Silwal D. K, Pokharel B, Phambu N & Aziz AN (2015). Fourier Transform Infrared (FTIR) Spectroscopy based Seed Trait Profiles of Upland Cotton Chromosomal Substitutions Lines. *J of App. Global Res* 8: 80-95.

28. R Core Team (2015). R: A Language and Environment for Statistical Computing, *http//: www.R-project.org/*, Accessed September 13, 2015.

29. Chavent M, Liquet B, Kuentz-Simonet V & Saracco J (2012). ClustOfVar: An R package for the clustering of variables. *J. of Stat. Software*. 50: 1-16.