

Size-Selective Fractionation and Visual Mapping of Allergen Protein Chemistry in *Arachis hypogaea*

Christine M. Hebling, Mark M. Ross, John H. Callahan, and Melinda A. McFarland*

U.S. Food and Drug Administration, Center for Food Safety and Applied Nutrition,

5100 Paint Branch Parkway, College Park, MD 20740

Supporting Information

SUPPORTING INFORMATION TABLE AND FIGURE LEGENDS

Supplemental Table 1 – Compilation of identified protein allergens (A) and non-allergens (B) grouped into protein families that are obtained from the GELFrEE-LC-MS/MS and whole lysate LC-MS/MS digestion platforms. The parsimony type is indicated for each protein identified, with subset and subsumable proteins filtered from the parsimony report. For example, the Ara h 1 (7S Globulin) protein family is represented by the following SwissProt protein names: ALL11_ARAHY, ALL12_ARAHY, B3IXL2_ARAHY, Q6PSU3_ARAHY, Q6PSU4_ARAHY, Q6PSU5_ARAHY, and Q6PSU6_ARAHY. In this protein family, ALL11_ARAHY, B3IXL2_ARAHY and Q6PSU3_ARAHY are identified by equivalent peptides and ALL12_ARAHY, Q6PSU5_ARAHY, and Q6PSU6_ARAHY are differentiable by parsimony comparison for GELFrEE sample fractions analyzed. Protein sequences recognized as allergen standards in the allergen community are indicated (www.allergen.org).

Supplemental Figure 1 - Confirmation of disulfide-associated Ara h 3/4 protein subunits by GELFrEE-Western blot. (A) One-dimensional SDS-PAGE gel of reduced GELFrEE fractions from PBS-extracted raw peanut. (B) Western blot analyses using Chicken anti-Ara h 3 on reduced GELFrEE fractions indicate the presence of Ara h 3 in numerous gel bands. A SeeBlue Plus2 prestained protein ladder is provided for molecular weight reference.

Supplemental Figure 2 – Geneious Pro 5.5.6 created by Biomatters (available from <http://www.geneious.com/>) multiple sequence alignment for Ara h 3.0101 (O82580_ARAHY; Sequence 1), Ara h 3.0201 (Q9SQH7_ARAHY; Sequence 2), and eight additional protein isoforms (Sequences 3-10). Percent homology, or the percentage of columns in the alignment in which all residues are identical, is reported as 81% for all ten sequences. Global sequence alignment was accomplished using default cost matrix parameters. The consensus sequence is provided for comparison with the aligned protein sequences. Single letter amino acid notation is provided for the consensus sequence for identical residues. Ambiguous amino acids, asparagine/aspartic acid, glutamine/glutamic acid, and leucine/isoleucine are represented by single letter residues B, Z, or J, respectively. Conserved residues matching by identity or homology are shaded in yellow and variable regions in white. Gaps are noted with a dashed line or red bar.

Supplemental Figure 3 – Geneious Pro 5.5.6 created by Biomatters (available from <http://www.geneious.com/>) multiple sequence alignment for Ara h 3.0101 (O82580_ARAHY; Sequence 1), Ara h 3.0201 (Q9SQH7_ARAHY; Sequence 2), and two sub-classes of isoform alignments (A and B). Arachin isoforms (A), Q6IWG5_ARAHY, Q0GM57_ARAHY, and AHY3_ARAHY (Sequences 3, 4, and 5 respectively), share a 79% homology with one another and 57% homology as compared to Ara h 3 standard sequences. The Arachin sub-classification is shown to contain extended gap regions within the acidic chain leading to regions of variability within the combined sequence alignment. Panel B, includes the A1DZF1_ARAHY and Q8LL03_ARAHY protein sequences associated with the common names of Arachin 7 and Trypsin Inhibitor, respectively (Sequences 3 and 4). These protein fragments are closely related to the N-terminal domain of Ara h 3 (Sequences 1 and 2), with an 89% homology. Global sequence alignment is completed using default cost matrix parameters. The consensus sequence is provided for comparison with the aligned protein sequences. Single letter amino acid notation is provided for the consensus sequence for identical residues. Ambiguous amino acids, asparagine/aspartic acid, glutamine/glutamic acid, and leucine/isoleucine are represented by single letter residues B, Z, or J, respectively. Conserved residues matching by identity or homology are shaded in yellow and variable regions in white. Gaps are noted with a dashed line or red bar.

Supplemental Figure 4 – One-dimensional SDS-PAGE gels (1, 3) and anti-Ara h 1 Western blots (2, 4) for unreduced (1, 2) and reduced (3, 4) raw peanut PBS extracts fractionated using the GELFrEE system. Protein bands from Western blots were visualized using enhanced chemiluminescence imaging. A SeeBlue Plus2 prestained protein ladder is provided for molecular weight reference.

Supplemental Figure 5 – Geneious Pro 5.5.6 created by Biomatters (available from <http://www.geneious.com/>) multiple sequence alignment for Ara h 2 (CONG7_ARAHY; 2.01-P2/P4 and 2.02-P3/P1), Ara h 6 (CONG_ARAHY), Ara h 7.0101 (Q9SQH1_ARAHY), and Ara h 7.0201 (B4XID4_ARAHY). Percent homology, or the percentage of columns in the alignment in which all residues are identical, is reported as 25%. Global sequence alignment is completed using default cost matrix parameters. The consensus sequence is provided for comparison with the aligned protein sequences. Single letter amino acid notation is provided for the consensus sequence for identical residues. Ambiguous amino acids, asparagine/aspartic acid, glutamine/glutamic acid, and leucine/isoleucine are represented by single letter residues B, Z, or J, respectively. Conserved residues matching by identity or homology are shaded in yellow and variable regions in white. Gaps are noted with a dashed line or red bar.

(A) *Arachis hypogaea* Protein Allergens Identified in GELFrEE and Whole Protein Lysate Digestions

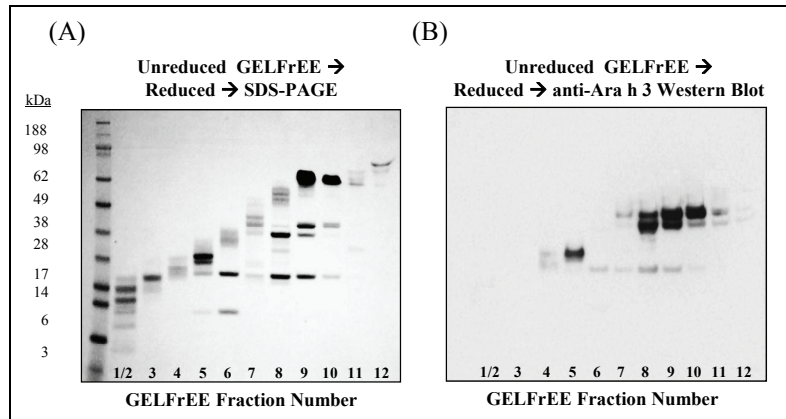
Protein Name	Mass	Protein Family	IUIS	GELFrEE FRACTIONATION			LYSATE DIGESTION		
				Parsimony Type	Peptides	% Coverage	Parsimony Type	Peptides	% Coverage
ALL11_ARAHY	70639.4	Ara h 1 (7S Globulin)	ARA H 1.0101	EQUIVALENT	52	59.3	EQUIVALENT	41	53.7
ALL12_ARAHY	71700.9	Ara h 1 (7S Globulin)		DIFFERENTIABLE	51	61.5	DIFFERENTIABLE	43	57.5
B3IXL2_ARAHY	70639.4	Ara h 1 (7S Globulin)		EQUIVALENT	52	59.3	EQUIVALENT	41	53.7
Q6PSU3_ARAHY	66933.6	Ara h 1 (7S Globulin)		EQUIVALENT	52	62.8	EQUIVALENT	41	56.9
Q6PSU4_ARAHY	48122.3	Ara h 1 (7S Globulin)	ARA H 2.0201	SUBSET	36	67.5	DIFFERENTIABLE	34	70.6
Q6PSU5_ARAHY	33641.0	Ara h 1 (7S Globulin)		DIFFERENTIABLE	24	55.2	DIFFERENTIABLE	20	51.8
Q6PSU6_ARAHY	34169.2	Ara h 1 (7S Globulin)		DIFFERENTIABLE	27	59.4	DIFFERENTIABLE	20	56.1
ASZ1Q8_9FABA	19202.1	Ara h 2.0101 and 2.0201 (2S Albumin)		SUPERSET	22	81.9	SUBSET	10	63.1
ASZ1Q9_9FABA	20557.7	Ara h 2.0101 and 2.0201 (2S Albumin)	ARA H 2.0101	EQUIVALENT	22	82.6	SUBSET	11	65.7
A8VT50_9FABA	19201.1	Ara h 2.0101 and 2.0201 (2S Albumin)		DIFFERENTIABLE	21	81.2	SUBSET	10	63.1
CONG7p1_ARAHY	20557.7	Ara h 2.0101 and 2.0201 (2S Albumin)		EQUIVALENT	22	82.6	SUBSET	11	65.7
CONG7p2_ARAHY	19202.1	Ara h 2.0101 and 2.0201 (2S Albumin)		SUPERSET	22	81.9	SUBSET	10	63.1
CONG7p3_ARAHY	20238.6	Ara h 2.0101 and 2.0201 (2S Albumin)	ARA H 2.0202	DIFFERENTIABLE	22	83.5	DIFFERENTIABLE	12	67.1
CONG7p4_ARAHY	18883.0	Ara h 2.0101 and 2.0201 (2S Albumin)	ARA H 2.0102	DIFFERENTIABLE	21	82.3	DIFFERENTIABLE	11	64.6
A1DZF0_ARAHY	60680.9	Ara h 3.0101 and 3.0201 (Ara h 3/4; 11S Globulin)	ARA H 3.0101	DIFFERENTIABLE	37	82.4	DIFFERENTIABLE	24	66.7
AHY3_ARAHY	54990.9	Ara h 3.0101 and 3.0201 (Ara h 3/4; 11S Globulin)		DIFFERENTIABLE	18	66.3	DIFFERENTIABLE	9	26.7
B5TYU1_ARAHY	60929.0	Ara h 3.0101 and 3.0201 (Ara h 3/4; 11S Globulin)		DIFFERENTIABLE	34	81.7	DIFFERENTIABLE	23	66.4
O82580_ARAHY	58542.8	Ara h 3.0101 and 3.0201 (Ara h 3/4; 11S Globulin)		DIFFERENTIABLE	25	61.5	DIFFERENTIABLE	15	45.6
Q0GM57_ARAHY	58569.8	Ara h 3.0101 and 3.0201 (Ara h 3/4; 11S Globulin)	ARA H 3.0201	EQUIVALENT	26	65.6	EQUIVALENT	22	67.8
Q516T2_ARAHY	61041.0	Ara h 3.0101 and 3.0201 (Ara h 3/4; 11S Globulin)		DIFFERENTIABLE	32	73.6	DIFFERENTIABLE	21	59.9
Q647H3_ARAHY	61836.5	Ara h 3.0101 and 3.0201 (Ara h 3/4; 11S Globulin)		DIFFERENTIABLE	33	75.2	DIFFERENTIABLE	21	60.0
Q647H4_ARAHY	61810.3	Ara h 3.0101 and 3.0201 (Ara h 3/4; 11S Globulin)		SUPERSET	33	76.7	SUPERSET	22	61.9
Q61WGS_ARAHY	58367.7	Ara h 3.0101 and 3.0201 (Ara h 3/4; 11S Globulin)	ARA H 3.0101	EQUIVALENT	26	65.9	EQUIVALENT	22	68.0
Q9FZ11_ARAHY	60753.9	Ara h 3.0101 and 3.0201 (Ara h 3/4; 11S Globulin)		DIFFERENTIABLE	34	76.4	DIFFERENTIABLE	22	60.5
Q9SQH7_ARAHY	61429.9	Ara h 3.0101 and 3.0201 (Ara h 3/4; 11S Globulin)		SUBSET	18	50.4	DIFFERENTIABLE	16	45.7
ASZ1Q5_9FABA	17480.2	Ara h 6 (2S Albumin)	ARA H 6.0101	DIFFERENTIABLE	15	78.6	DIFFERENTIABLE	7	57.9
ASZ1Q6_9FABA	17479.2	Ara h 6 (2S Albumin)		SUPERSET	15	78.6	SUPERSET	7	57.9
ASZ1R0_ARAHY	17479.2	Ara h 6 (2S Albumin)		SUPERSET	15	78.6	SUPERSET	7	57.9
CONG_ARAHY	17479.2	Ara h 6 (2S Albumin)		SUPERSET	15	78.6	SUPERSET	7	57.9
B4XID4_ARAHY	19782.5	Ara h 7.0101 and 7.0201 (2S Albumin)	ARA H 7.0101	DIFFERENTIABLE	6	39.6	DIFFERENTIABLE	4	26.8
Q647G8_ARAHY	18786.2	Ara h 7.0101 and 7.0201 (2S Albumin)		DIFFERENTIABLE	5	30.4	SUBSET	2	17.1
Q9SQH1_ARAHY	18748.1	Ara h 7.0101 and 7.0201 (2S Albumin)		DIFFERENTIABLE	2	13.1			
B1PYZ4_ARAHY	16899.7	Ara h 8.0101 (Pathogenesis-Related Protein)		SUPERSET	3	25.5	EQUIVALENT	3	24.2
B2ZGS2_ARAHY	16909.8	Ara h 8.0101 (Pathogenesis-Related Protein)	ARA H 8.0101	SUBSET	2	17.2	EQUIVALENT	3	24.2
Q0PKR4_ARAHY	16914.7	Ara h 8.0101 (Pathogenesis-Related Protein)		DIFFERENTIABLE	2	15.9			
Q2YH1R1_ARAHY	14376.6	Ara h 8.0101 (Pathogenesis-Related Protein)		SUBSET	2	20.0	EQUIVALENT	3	28.1
Q6VT83_ARAHY	16941.8	Ara h 8.0101 (Pathogenesis-Related Protein)		SUPERSET	3	25.5	EQUIVALENT	3	24.2
B0YIU5_ARAHY	16402.5	Ara h 8.0201 (Pathogenesis-Related Protein)	ARA H 8.0201	DIFFERENTIABLE	13	85.0	DIFFERENTIABLE	7	69.3
B6CEX8_ARAHY	12156.1	Ara h 9.0101 and 9.0201 (Non-Specific Lipid Transfer Protein)	ARA H 9.0101	DIFFERENTIABLE	5	45.7	DIFFERENTIABLE	3	31.9
B6CG41_ARAHY	9504.7	Ara h 9.0101 and 9.0201 (Non-Specific Lipid Transfer Protein)	ARA H 9.0201	DIFFERENTIABLE	5	71.7	DIFFERENTIABLE	2	33.7
Q647G4_ARAHY	15574.2	Ara h 10.0101 and 10.0102 (16 kDa Oleosin)	ARA H 10.0101	DIFFERENTIABLE	4	30.0	SUBSET	2	14.0
Q647G5_ARAHY	17798.4	Ara h 10.0101 and 10.0102 (16 kDa Oleosin)	ARA H 10.0102	DIFFERENTIABLE	7	41.4	DIFFERENTIABLE	4	26.0
Q45W86_ARAHY	14344.8	Ara h 11 (14 kDa Oleosin)	ARA H 11.0101	DIFFERENTIABLE	4	27.0	DIFFERENTIABLE	3	24.8
Q45W87_ARAHY	14298.8	Ara h 11 (14 kDa Oleosin)		DIFFERENTIABLE	4	27.0	DIFFERENTIABLE	3	24.8

SUPPLEMENTAL TABLE 1A

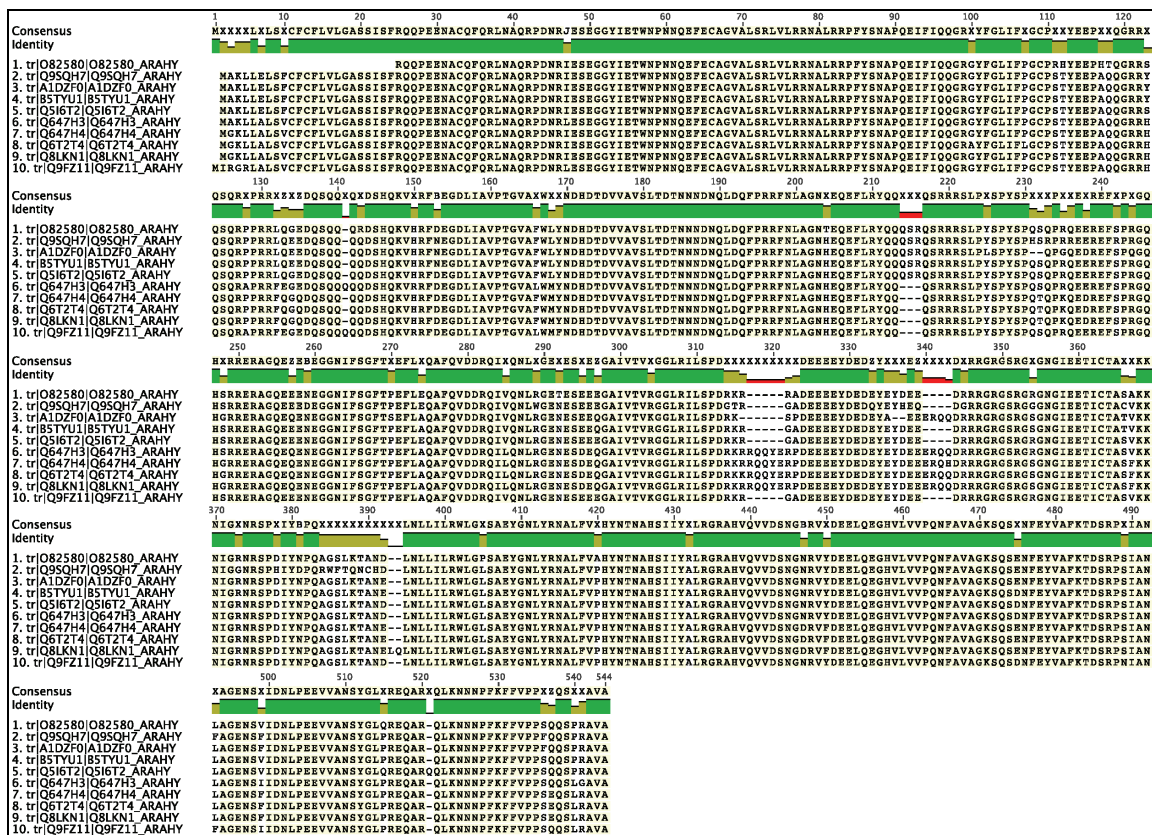
(B) Additional *Arachis hypogaea* Proteins Identified in GELFrEE and Whole Protein Lysate Digestions

Protein Name	Mass	Protein Family	IUIS	GELFrEE FRACTIONATION			LYSATE DIGESTION		
				Parsimony Type	Peptides	% Coverage	Parsimony Type	Peptides	% Coverage
Q4JME7_ARAHY	97819.8	Lipoxygenase		DIFFERENTIABLE	50	90.6	SUPERSET	31	48.6
B4UWB9_ARAHY	26777.7	Lipoxygenase - Fragment		DIFFERENTIABLE	6	51.0			
B4UWC0_ARAHY	20545.5	Lipoxygenase - Fragment		DIFFERENTIABLE	10	80.4	DIFFERENTIABLE	6	41.9
C6ZGE3_9FABA	22348.7	Lipoxygenase - Fragment		DIFFERENTIABLE	2	13.6			
Q2HWT8_ARAHY	92339.5	Phospholipase D		DIFFERENTIABLE	10	15.6			
Q647H1_ARAHY	76343.2	Conarachin (Q647H1_ARAHY)		DISCRETE	21	53.5	DISCRETE	12	25.4
Q20356_ARAHY	51949.2	Ribulose Biphosphate Carboxylase		DIFFERENTIABLE	10	28.6	DIFFERENTIABLE	13	35.7
A7LB59_ARAHY	39857.3	Sterokosin-B (A7LB59_ARAHY)		DISCRETE	4	14.4	DISCRETE	6	19.3
A7LB60_ARAHY	39009.2	Sterokosin-A (A7LB60_ARAHY)		DISCRETE	6	18.3	DISCRETE	6	23.8
A1E2B0_ARAHY	33840.9	11S Seed Storage Globulin (A1E2B0_ARAHY)		DISCRETE	15	44.3	DISCRETE	11	39.3
A1E2B1_ARAHY	31476.0	11S Seed Storage Globulin (A1E2B1_ARAHY)		DISCRETE	5	24.5	DISCRETE	5	26.7
Q43373_ARAHY	29605.0	Galactose-Binding Lectin (Q43373_ARAHY)		DISCRETE	2	6.5			
LECG_ARAHY	29363.8	Galactose-Binding Lectin (LECG_ARAHY)		DIFFERENTIABLE	10	64.1	DIFFERENTIABLE	10	64.1
B4UW79_ARAHY	24398.6	Glutathione Peroxidase		DISCRETE	4	19.4	DISCRETE	3	22.7
B4UWB4_ARAHY	23136.3	Lactoylglutathione Lyase		DISCRETE	3	17.5			
B4UWA1_ARAHY	22383.5	Putative Uncharacterized Protein (B4UWA1_ARAHY)		DISCRETE	5	45.9	DISCRETE	2	11.2
B4UWB2_ARAHY	22079.2	Kunitz Trypsin Inhibitor 4		DISCRETE	2	9.9			
B4UWA3_ARAHY	20364.4	Putative Uncharacterized Protein (B4UWA3_ARAHY)		DISCRETE	5	49.2			
B4UWE2_ARAHY	20302.5	Perchloric Acid Soluble Translation Inhibitor Protein		DISCRETE	4	29.6	DISCRETE	2	15.9
Q06H34_ARAHY	19248.2	Tonoplast Intrinsic Protein Alpha TIP		DISCRETE	5	61.2			
Q06H31_ARAHY	19172.6	Translationally Controlled Tumor-like Protein		DISCRETE	5	42.9	DISCRETE	2	17.9
B4UW70_ARAHY	18775.6	Fiber Annexin		DISCRETE	4	28.6	DISCRETE	9	62.7
Q6J1J8_ARAHY	18493.9	Oleosin 18 kDa		SUPERSET	3	12.5	SUBSUMABLE	2	11.4
Q9AXI0_ARAHY	18502.8	Oleosin 18 kDa		SUBSET	2	6.2	DIFFERENTIABLE	2	10.8
Q9AXI1_ARAHY	18480.9	Oleosin 18 kDa		DIFFERENTIABLE	3	12.5	DIFFERENTIABLE	2	11.4
Q06H21_ARAHY	17868.6	Ubiquitin-Ribosomal Protein S27a		SUPERSET	4	27.7	EQUIVALENT	2	11.6
Q06H40_ARAHY	21432.5	Ubiquitin-Ribosomal Protein S27a		SUPERSET	4	53.1	EQUIVALENT	2	18.8
Q45W78_ARAHY	14892.0	Ubiquitin-Ribosomal Protein S27a		SUPERSET	4	33.6	EQUIVALENT	2	14.1
B4UW81_ARAHY	20386.8	Glutathione S-Transferase 2		EQUIVALENT	3	25.7			
B4UW83_ARAHY	15191.2	Glutathione S-Transferase 2		EQUIVALENT	3	34.6			
Q06H32_ARAHY	17525.2	Thioredoxin Fold		DISCRETE	6	51.9	DIFFERENTIABLE	6	60.5
ATPB_HUMAN	56524.6	ATP Synthase Beta Subunit		DIFFERENTIABLE	3	11.5	EQUIVALENT	2	4.5
D3Y4M7_BRAD	17501.0	ATP Synthase Beta Subunit		EQUIVALENT	2	14.7	EQUIVALENT	2	14.7
D3Y4M9_BRAEL	17501.0	ATP Synthase Beta Subunit		EQUIVALENT	2	14.7	EQUIVALENT	2	14.7
D3Y4N1_RHILE	16956.7	ATP Synthase Beta Subunit		SUBSET	2	15.1	EQUIVALENT	2	15.1
D3Y4N3_BRAD	17818.1	ATP Synthase Beta Subunit		EQUIVALENT	2	14.5	EQUIVALENT	2	14.5
D3Y4N4_BRAEL	17414.0	ATP Synthase Beta Subunit		EQUIVALENT	2	14.8	EQUIVALENT	2	14.8
D3Y4N5_BRAD	17521.1	ATP Synthase Beta Subunit		EQUIVALENT	2	14.7	EQUIVALENT	2	14.7
D3Y4N6_RHIZ	16768.7	ATP Synthase Beta Subunit		SUBSET	2	15.2	EQUIVALENT	2	15.2
D3Y4N7_RHIZ	16841.7	ATP Synthase Beta Subunit		SUBSET	2	15.1	EQUIVALENT	2	15.1
D3Y4N8_RHIZ	16768.7	ATP Synthase Beta Subunit		SUBSET	2	15.2	EQUIVALENT	2	15.2
D3Y4N9_BRAD	17415.9	ATP Synthase Beta Subunit		EQUIVALENT	2	14.8	EQUIVALENT	2	14.8
D3Y4P0_BRAD	17428.0	ATP Synthase Beta Subunit		EQUIVALENT	2	14.8	EQUIVALENT	2	14.8
D3Y4P2_RHIME	16824.7	ATP Synthase Beta Subunit		SUBSET	2	15.1	EQUIVALENT	2	15.1
D3Y4P8_BRAD	17507.0	ATP Synthase Beta Subunit		EQUIVALENT	2	14.7	EQUIVALENT	2	14.7
D5KZX9_RHIZ	17140.8	ATP Synthase Beta Subunit		SUBSET	2	14.7	EQUIVALENT	2	14.8
Q06H19_ARAHY	16908.1	UDP-Glucose Pyrophosphorylase		DISCRETE	8	65.3	DISCRETE	6	42.7
Q6PWX0_ARAHY	16714.8	Calmodulin		SUPERSET	5	42.6			
Q6R2U4_ARAHY	16676.8	Calmodulin		SUPERSET	5	42.6			
Q45W80_ARAHY	16428.6	Nucleoside Diphosphate Kinase		DISCRETE	7	55.7	DISCRETE	4	32.2
Q2PXN9_ARAHY	16159.3	Dessication Protectant Protein LEA 14		DISCRETE	2	19.0			
Q1PCR4_ARAHY	16062.1	Putative IN2-1 Protein		DISCRETE	2	22.3			
Q1HD87_ARAHY	15302.5	Superoxide Dismutase		DIFFERENTIABLE	3	38.8	EQUIVALENT	2	32.2
Q45W82_ARAHY	15203.4	Superoxide Dismutase		SUBSET	2	32.2	EQUIVALENT	2	32.2
B4UWD5_ARAHY	15118.0	Proteasome Subunit Alpha Type		DIFFERENTIABLE	2	26.3			
D3K177_ARAHY	14207.1	Profilin		DISCRETE	2	16.8			
B3GR01_9FABA	10760.5	Cystatin		EQUIVALENT	2	25.5			
Q647G6_ARAHY	10760.5	Cystatin		EQUIVALENT	2	25.5			
Q4U4M1_ARAHY	10408.1	LEA protein		SUPERSET	7	64.2	SUPERSET	4	52.6
Q850G8_ARAHY	10390.1	LEA protein		DIFFERENTIABLE	7	55.8	SUBSET	3	38.9
IBB1_ARAHY	8426.4	Bowman-Birk Type Proteinase Inhibitor		DIFFERENTIABLE	5	78.6	EQUIVALENT	2	48.6
Q0PKR5_ARAHY	7878.2	Bowman-Birk Type Proteinase Inhibitor		SUBSET	4	56.2	EQUIVALENT	2	53.1
Q2VMU0_ARAHY	12286.6	Bowman-Birk Type Proteinase Inhibitor		DIFFERENTIABLE	4	33.6			
Q7X973_ARAHY	9495.9	Bowman-Birk Type Proteinase Inhibitor		SUBSET	4	45.0	EQUIVALENT	2	42.5
Q0Q0Q8_ARAHY	9148.6	Type 4 Metallothionein (Q0Q0Q8_ARAHY)		DISCRETE	2	24.4			
Q0Q0Q9_ARAHY	9058.6	Type 4 Metallothionein (Q0Q0Q9_ARAHY)		DISCRETE	2	30.5			
Q2PXN4_ARAHY	5801.9	Seed Maturation Protein LEA 4		DISCRETE	2	29.1			

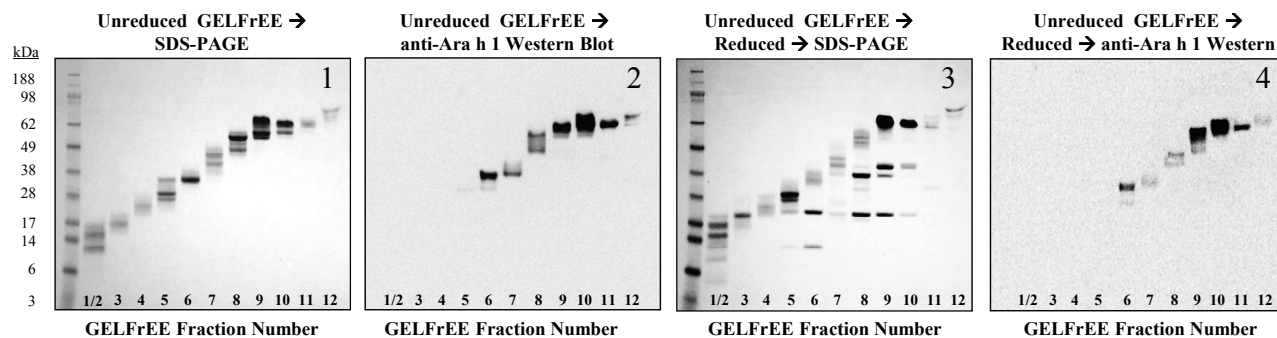
SUPPLEMENTAL TABLE 1B



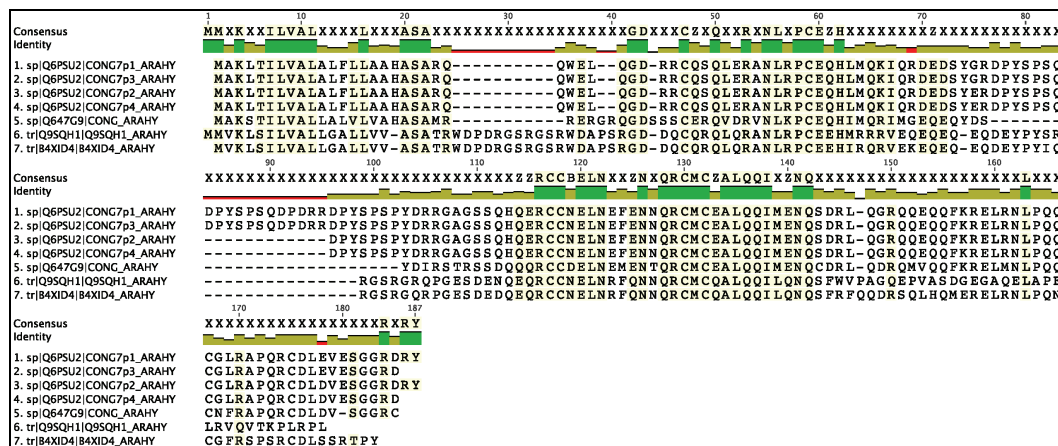
SUPPLEMENTAL FIGURE 1



SUPPLEMENTAL FIGURE 2



SUPPLEMENTAL FIGURE 4



SUPPLEMENTAL FIGURE 5