

Table 3 Protein family distributions in the proteins encoded by all genes and by genes with pre-mRNA alternative splicing (AS genes) in cotton

PFam ID	Total	AS	%	Pfam	Pfam description
pfam00078	3419	470	13.7	RVT_1	Reverse transcriptase
pfam00069	1538	590	38.4	Pkinase	Protein kinase domain
pfam07727	1201	142	11.8	RVT_2	Reverse transcriptase
pfam08284	1191	43	3.6	RVP_2	Retroviral aspartyl protease
pfam03732	1049	220	21.0	Retrotrans_gag	Retrotransposon gag protein
pfam07714	1014	326	32.1	Pkinase_Tyr	Protein tyrosine kinase
pfam13041	667	191	28.6	PPR_2	PPR repeat family
pfam00067	655	108	16.5	p450	Cytochrome P450
pfam00931	531	108	20.3	NB-ARC	NB-ARC domain
pfam13639	520	174	33.5	zf-RING_2	Ring finger domain
pfam14223	485	42	8.7	UBN2	gag-polypeptide of LTR copia-type
pfam13456	445	62	13.9	RVT_3	Reverse transcriptase-like
pfam00249	428	127	29.7	Myb_DNA-binding	Myb-like DNA-binding domain
pfam00076	410	233	56.8	RRM_1	RNA recognition motif
pfam13976	398	13	3.3	gag_pre-integr	GAG-pre-integrase domain
pfam00847	354	40	11.3	AP2	AP2 domain
pfam12776	292	143	49.0	Myb_DNA-bind_3	Myb/SANT-like DNA-binding domain
pfam13359	254	33	13.0	DDE_Tnp_4	DDE superfamily endonuclease
pfam03171	251	70	27.9	2OG-FeII_Oxy	2OG-Fe (II) oxygenase superfamily
pfam00201	240	27	11.3	UDPGT	UDP-glucuronosyl and UDP-glucosyl transferase
pfam00481	226	108	47.8	PP2C	Protein phosphatase 2C
pfam02365	221	86	38.9	NAM	No apical meristem (NAM) protein
pfam00071	216	75	34.7	Ras	Ras family
pfam03106	200	74	37.0	WRKY	WRKY DNA-binding domain
pfam00010	193	77	39.9	HLH	Helix-loop-helix DNA-binding domain
pfam10536	190	17	8.9	PMD	Plant mobile domain
pfam00153	187	66	35.3	Mito_carr	Mitochondrial carrier protein
pfam13966	186	52	28.0	zf-RVT	zinc-binding in reverse transcriptase
pfam13499	186	48	25.8	EF-hand_7	EF-hand domain pair
pfam02519	185	24	13.0	Auxin_inducible	Auxin responsive protein
pfam00141	178	47	26.4	peroxidase	Peroxidase
pfam12796	175	54	30.9	Ank_2	Ankyrin repeats (3 copies)
pfam14432	175	32	18.3	DYW_deaminase	DYW family of nucleic acid deaminases
pfam00106	171	49	28.7	adh_short	short chain dehydrogenase
pfam00004	168	56	33.3	AAA	ATPase family associated with various cellular
pfam00083	164	66	40.2	Sugar_tr	Sugar (and other) transporter
pfam02458	162	15	9.3	Transferase	Transferase family
pfam00665	160	25	15.6	rve	Integrase core domain
pfam00225	159	75	47.2	Kinesin	Kinesin motor domain
pfam03004	158	80	50.6	Transposase_24	Plant transposase (Ptta/En/Spm family)
pfam14244	155	72	46.5	UBN2_3	gag-polypeptide of LTR copia-type
pfam01095	155	20	12.9	Pectinesterase	Pectinesterase
pfam00657	150	30	20.0	Lipase_GDSL	GDSL-like Lipase/Acylhydrolase

Continued Table 3

PFam ID	Total	AS	%	Pfam	Pfam description
pfam00270	149	70	47.0	DEAD	DEAD/DEAH box helicase
pfam03514	147	46	31.3	GRAS	GRAS domain family
pfam00561	147	45	30.6	Abhydrolase_1	alpha/beta hydrolase fold
pfam00226	146	58	39.7	DnaJ	DnaJ domain
pfam00179	143	85	59.4	UQ_con	Ubiquitin-conjugating enzyme
pfam00854	139	44	31.7	PTR2	POT family
pfam13855	137	29	21.2	LRR_8	Leucine rich repeat
pfam04564	136	35	25.7	U-box	U-box domain
pfam01582	135	44	32.6	TIR	TIR domain
pfam01554	129	39	30.2	MatE	MatE
pfam00168	127	42	33.1	C2	C2 domain
pfam00295	127	30	23.6	Glyco_hydro_28	Glycosyl hydrolases family 28
pfam13839	124	29	23.4	PC-Esterase	GDSL/SGNH-like Acyl-Esterase family found
pfam00046	121	37	30.6	Homeobox	Homeobox domain
pfam01490	121	37	30.6	Aa_trans	Transmembrane amino acid transporter protein
pfam00005	120	37	30.8	ABC_tran	ABC transporter
pfam05699	119	53	44.5	Dimer_Tnp_hAT	hAT family C-terminal dimerization
pfam04043	119	7	5.9	PMEI	Plant invertase/pectin methylesterase inhibitor
pfam00501	117	31	26.5	AMP-binding	AMP-binding enzyme
pfam00332	117	27	23.1	Glyco_hydro_17	Glycosyl hydrolases family 17
pfam10250	112	49	43.8	O-FucT	GDP-fucose protein O-fucosyltransferase
pfam00400	111	45	40.5	WD40	WD domain
pfam02362	110	47	42.7	B3	B3 DNA binding domain
pfam03195	110	19	17.3	DUF260	Protein of unknown function DUF260
pfam00085	107	33	30.8	Thioredoxin	Thioredoxin
pfam00230	106	35	33.0	MIP	Major intrinsic protein
pfam01501	106	28	26.4	Glyco_transf_8	Glycosyl transferase family 8
pfam00319	105	28	26.7	SRF-TF	SRF-type transcription factor (DNA-binding and
pfam03018	105	9	8.6	Dirigent	Dirigent-like protein
pfam00248	104	44	42.3	Aldo_ket_red	Aldo/keto reductase family
pfam13414	104	32	30.8	TPR_11	TPR repeat
pfam03151	103	28	27.2	TPT	Triose-phosphate Transporter family
pfam14226	103	25	24.3	DIOX_N	non-haem dioxygenase in morphine synthesis
pfam08263	103	20	19.4	LRRNT_2	Leucine rich repeat N-terminal domain
pfam01370	102	28	27.5	Epimerase	NAD dependent epimerase/dehydratase family
pfam07732	102	23	22.5	Cu-oxidase_3	Multicopper oxidase
pfam00170	101	32	31.7	bZIP_1	bZIP transcription factor
pfam03552	100	39	39.0	Cellulose_synt	Cellulose synthase
Total	57900	18218	31.5		
Unique Pfam	3505	2454			

Note: This is only a partial list; The information for accessing the complete list can be found in the main text