

Supplemental Figure 1. Protein sequences encoded by the predicted AGP genes. Colored sequences at the N and C terminus indicate predicted signal peptide (green) and GPI anchor (blue) addition sequences if present. AP, PA, SP, and TP repeats (yellow) and lysine-rich regions (olive) are also indicated.

>At1g24520-AGP50C 68414.m03087 anther-specific protein agp1 identical to agp1 GI:780177 from [Arabidopsis thaliana]
MGRQNIVVVVALVFMAIIGLAAAASSPSPSASPSKAPAASKTDHVEAPVTDDQIGTTDDDAAPTFGDGDVAVAGPLGSDSSYDNAPTGSADSAKSGAAALGVS AVVVGVHPLLVLSCYSKLGIVL

>At1g31250-AGP51C 68414.m03825 proline-rich family protein contains proline-rich extensin domains, INTERPRO:IPR002965; similar to prolin rich protein GB:S44189 GI:433706 from [Zea mays]
MANQYFRMAFLLCFLSLSYQSIRIEAREGKACIGNCYGSSAPVPPNASLSIPPNSSPSVKLTTPPYASPSVKLTTPPYASPSVRPAGTTPNASPSVKLTTPPYASPSMRPAGTPNASPSVKLTTPPYASPSVRPTGTTPNASPSLTTPPNPSPSEKFIPPNASPFIHT

>At1g35230-AGP5C 68414.m04369 arabinogalactan-protein (AGP5) identical to gi 3883128 gb AAC77827
MASKSVVVFLLFLLVASSVVAQAPGPAPTISPLPATPTPSQSPRATAPAPSPSANPPPSAPTTAPVVSQPPTESPPAPPTSTSPSGAPGTNVPSGEAGPAQSPLSGSPNAAAVSRVSLVGTTFAGVAVIAALLL

>At1g63530-AGP52C 68414.m07182 hypothetical protein
MRVDFSEPEYRNCYHCRDSGILNTQESEPNNVIGSSITSVPVSSGPFQALGHASAAIGSTSTSSPVQSSSSFSFGSPAATTSVGVSSSTSTPTPISFLFPFYVPPPPSSSLSGPAQNVWISAYSTAGRGVEWAWARPDVAVYPASSSSSTSTVVFGTTPVSVSSLFGPSQDFPKTTFACSSTPAPTRCECSFAKPFPGSSPASSSSNLFGPNPSTTTASVFGTSATSGSSLFGPAQHPLSYDSIGGSFAFAGVKTS SPQPLGFNGATLLPSSHEGQNISSFKPYVDKSHDELWYKKEAKGGSFPAAHASPIGSKPNAVFSPTVSTSLFASPSVDPDHHQRTNFSTTLGNFTGPAFGFFPAAAGGATTFSSSTGFGQNPAPFPFGFDSQFPFAFSSFSSTTYPAVSSSVQRPHETGAVSSPAFGCTACGATSSSSASDHFTFNGATTPPGLFFPTTSSGPMFMTLLAAQGTTPALQTYPVQGFILLPFAAMSLQ

>At1g63540-AGP53C 68414.m07183 hydroxyproline-rich glycoprotein family protein
MKIDFSEPECSNCYQFTNPDLLNTPSEQSNVIGSSITSVPVNDGPVPPLELDSAAVVSTSTSSPVQALGHDSGATVSTSTSSPVQIFSSPFSFGSAHAATPVSSGPAPSPPTFGEPRILLATSGSGASATSTSSSTSSPLHSSPFSFGSAPAAITSVSSGPAQSPASSPRLWIDRFATSSSASATSSSSTSSPFHSSLLGFAPA VTSVSSAPTACGPQTQAFQPTQAFGLSMFGSTPRFEITGFPFQASASRNSPSPSFGPAHNCGKPAFGSPFGNNVAFARPDVGISPVASSSTSTEIFGATPALSLSFPFGPMQAPVQASASSTSTFPFGCVPPSPSSGSLFNSAFGSLPAPSSSNFFGQSSSNLLGQNPSTTGVGYPG SPLNSSFPGFVGYPGSSSNLFRSNPPNFGGSGIGAGPQHFGFNGDASVLPSTPFSLSPAFSSNTNTGSPYFASH EWSRPTEQGSMNPGYAPTHEGDNSSGWSFPTSKGNIYISISASKPYLHKSHEELRWEDYKQGDGKGGPFPAAPASTIGSRPNAAFSPPTVSPPAHGCTACGATSSSSASRHFTFNGATSPPSAATTPPGLFFPSTGFAPMMFGTNLAVQGTSPALQAYPVQGYILLPFAAMTLQ

>At2g14890-AGP9C 68415.m01693 arabinogalactan-protein (AGP9) identical to gi|10880495|gb|AAG24277
MARSFAIAVICIVLIAGVTGQAPTSPTATPAPPTPTTPPPAATPPPVASPPVTTSPPPVTTAPPANPPPPVSSPPASPPPATPPPVASPPPPVASPPPATPPPVPATPPPAPLASPPAQVPAPAPTTKPDSPSPSPSSSPPLPSSDAPGPS TDSISPAPSPTDVNDQNGASKMVSSLVFGSVLVWFMI

>At2g22470-AGP2C 68415.m02664 arabinogalactan-protein (AGP2) identical to gi|3883122|gb|AAC77824; supported by cDNA gi|3883121|gb|AF082299
MNSKAMQALIFLGLFATSCLAQAPAPAPTTPVTPPTALPPVTAETPSPIASPVPVNEPTPAPTTSPPTSPVASPPQTDAPAPGPSAGLTPTSSPAPGPDGAADAPSAAWANKAFLVGTAVAGALYAVVLA

>At2g28440-AGP54C 68415.m03455 proline-rich family protein contains proline-rich extensin domains, INTERPRO:IPR002965; contains similarity to vegetative cell wall protein gp1 [Chlamydomonas reinhardtii] gi|12018147|gb|AAG45420; +
MAKKLCFIVMLSLICLLIFDFAGAQEESPSPAAVSPGREPSTDSPLSPSSSPEEDSPLSPSSSPEEDSPLPPSSSPEE
DSPLAPSSSPEVDSPLAPSSSPEVDSPQPPSSSPEADSPLPPSSSPEANSPQSPASSPKPESLADSPSPPPPPQPE
SPSSPSYPEPAPVPAPSDDDDSDDDPEPETEYFSPSPAPSPELGMAQDIKASDAAGEELNDERGEDYGMGSLKAGIAI
GTILGVGAIIVIGALVYKKRRDNMTRARYTYFTEGEFL

>At2g45000-AGP55C 68415.m05603 expressed protein contains Pfam profile:
PF05064 NSP1-like C-terminal region
MSGFPFGQNSVGGFSGSSSATNSSASSTTSPLSFSFNQSSNPSSSTGFGFGSSVSSTPASSTTPSFGFGASSTPS
FGFGSSASSSTPSFGFGSSASVTPASTTPSFGFGTAASSSAPAPSLFGSSTTNASSAAPGSSPFGFVTSSASSTATP
SSSLFGAPASSAATPSSSPFGAAPASGSTPLFGSSPSLFSAPSSASANSSSLFGASSAATSTSPFLFGAPSSATGAT
PSFSVASSAPGSSSIFGATGSSPSFSVASSASGSSPSIFGATGSSPFFGSSSSAGSTPSLFASSSSGATTSSPSPF
GVSTFNSSSTSNASASPFSASTGFSFLKSTASSTTSSTTPSAPPQTASSSSSFSFGTSANSGFNLSTGSSAAPA
SSTSGAVFSIATTTTTSSSTPAATSAPASSAPASTMAFSPFGVTSSATNTTPASSAATFSTTGFLASSTPATGSTN
SFTGFAVPKTSTPASSSQPQTTSPAFSFLPSSTSTTAPATSSATTTQTTLVVPSSTGTTAVAPVAGSPKLPSEIT
GKTVEEIIKEWNTLQERTGRFRKQANAIAEWDKRIQLNRDVLRLLEIEVAKVETQSSLERQLELIETHQOQEVDKA
LQSMEEEAERIYNDERKSLLDDEAASTRDAMYEQSELVERELEHMTEQIRSI IQSVNANQGGELEAIDGMSPLDVVV
RILNNQLSSLMWIDEKAEFSSRIQKIALQGGSGDRELMAPKHWMS

>At2g47930-AGP26C 68415.m05993 hydroxyproline-rich glycoprotein family
protein contains proline-rich extensin domains, INTERPRO:IPR002965
MSVSLFTAFTVLSLCLHTSTS EFQLSTISAAPSFLPEAPSSFSAS TPAMSPDTSPLFPTPGSSEMSPSPSESSIMPT
IPSSLSPNPDAVTPDPLLEVSPVGSPLPASSSVCLVSSQLSSLLLVLMLLLAFCSFF

>At3g01700-AGP11C 68416.m00103 arabinogalactan-protein (AGP11) identical to
gi|10880499|gb|AAG24279
MARLFVVVALLALAVGTVFAADAPSAAPTASPTKSPTKAPAAAPKSSAAAPKASSPVAEPTPEDDYSAASPSDSAE
APTSSPPAPTPEADGPSSDGPSSDGPAAESEPKSGATTNVKLSIAGTVAAAGFFIFSL

>At3g06360-AGP27C 68416.m00734 arabinogalactan-protein (AGP27)
MASSILLTLITFIFLSSLSLSSPTTNTIPSSQTI SPSEEKISPEIAPLLPSPAVSSTQTISSSTLPEPENDDVSA
PDPAFAPSASPPASSLASLS SQAPGVFIYFVFAAVYCFSLRLLAVSAI

>At3g22070-AGP56C 68416.m02785 proline-rich family protein contains proline-rich extensin domains, INTERPRO:IPR002965
MLCYVGKATKIFIFIVTVVFI GLVVA FGVLRRHSHHCSGDYCSSSTDPSSSSSSSSSSSTSPFITPFPNPNPNPNPN
PPVLGSSPPSPTDSSSSTSI SPNPPAPIVNPNP PPPPSTPNPPPEFSPPPPDLDTTTPPPPSTDIPIPPPP PAPVSA
SPPLTPSSSVTSPAPVHAKLVND

>At3g45230-AGP57C 68416.m04881 hydroxyproline-rich glycoprotein family
protein contains proline-rich extensin domains, INTERPRO:IPR002965; related
to vegetative cell wall protein gp1 [Chlamydomonas reinhardtii]
gi|12018147|gb|AAG45420
MKLEFIIVAMMLSLVLSGEILTKSSPAPSPDLADSPLIHASPPSKLGSNHNSPAESP I EYSSPPEPETEHSPPSPSPA
NSPSVSPPLPNDSQSPSSASPSPSPEASDVNHSDITGIEGKLPSPSGGGMGGKKGVAFAAIAAVCVVGVAGFV
YKKRQENIRRSRYGYAAREIL

>At4g09030-AGP10C 68417.m01490 arabinogalactan-protein (AGP10) identical to
gi|10880497|gb|AAG24278; supported by Ceres cDNA 265772
MASKSVVLLFLALIASSAIAQAPGPAPTRSPLPSAQPRTAAPTPTPTPTPSATPTAAPVSPAGSPLPSS
ASPPAPPTSLTPDGAPVAGPTGSTPVDNNAATLAAGSLAGFVVASLLL

>At4g16980-AGP58C 68417.m02560 arabinogalactan-protein family similar to
arabinogalactan protein [Arabidopsis thaliana] gi|10880495|gb|AAG24277;
contains proline-rich extensin domains, INTERPRO:IPR002965

MASSFSSQAFFLLTLSMVLIPFSLAQAPMMAPSGSMSMPMSSGGGSSVPPVMSMPMMTPPPMPMTPPMPMTTP
PMPMAPPPMPMASPPMPMTPTSTSPSPLTVPDMPSPMPMSGMESSPSPGMPMPAMAASPD SGAFNVRNNVTLSCVV
GVVAAHFLLV

>At4g40090-AGP3C 68417.m05675 arabinogalactan-protein (AGP3) InDels between the genome sequence and the cDNA prevent identical translations from being annotated; the basis for this is under investigation.
MALKTLQALIFLGLFAASCLAQAPAPAPITFLPPVESPSPVVTPTAEP PARCFTPYSSQORTHSGSNNSTNRLTPDHV
AHNFPCRFS

>At5g10430-AGP4C 68418.m01209 arabinogalactan-protein (AGP4) identical to gi 3883126 gb AAC77826
MGSKIVQVFLMLALFATSALAQAPAPTPTATPP PATPPPVATPPPVATPPPAATPAPATPP PAATPAPATTPPSVAP
SPADVPTASPPAPEGPTVSPSSAPGPSDASPAPSAAFSNKAFFAGTAFAAIMYAAVLA

>At5g14380-AGP6C 68418.m01680 hydroxyproline-rich glycoprotein family protein identical to gi|4775268|emb|CAB42531
MARQFVVLVLLTLTIATAFAADAPSASPKKSPSPTAAPTAKAPTATTKAPSAPTKAPAAAPKSSSASSPKASSPAAEG
PVPEDDYSASSPSDSAEAPT VSSPPAPT PDSTSAADGPSDGTAE SPKSGAVTTAKFSVVGTVATVGGFFFSF

>At5g18690-AGP25C 68418.m02218 hydroxyproline-rich glycoprotein family protein contains proline-rich extensin domains, INTERPRO:IPR002965; Common family member: At3g06360 [Arabidopsis thaliana]
MAFSFLNKLIIIFIFIFISLSSSPTISLVQQLSPEIAPLLPSPGDALPSDDGSGTIPSPSPDPDPTNDGSYPDPL
AFSPFASPPVSSPSPPSLPSAGVLLISLISSASFLAL

>At5g64310-AGP1C 68418.m08078 arabinogalactan-protein (AGP1) identical to gi:3883120 gb:AAC77823
MAFSKSLVFVLLAALLISSAVAQSPAPAPSNVGGRRISPAPSPKMMTAPAPAPEVSPSPSPAAALTPESSASPPSP
LADSP TADSPALSPSAISDSPTEAPGPAQGGAVSNKFASFGSVAVMLTAAVLVI

>At5g65390-AGP7 68418.m08224 arabinogalactan-protein (AGP7)
MNSKIIIEAFFIVALFTTSCLAQAPAPSPTTTVTPPPVATPPPAATPAPTTPPPAVSPAPTSSPPSSAPSPSSDAPT
ASPPAPEGPGVSPGELAPTPSDASAPPPNAALTNKAFVVGSLVAIIYAVVLA

>At1g68725-AGP19K 68414.m07853 arabinogalactan-protein, putative (AGP19) non-consensus splice site at the intron:exon boundary (AT:exon)
MESNSIIWSLLLASALISSFSVNAQGPAASPVSTTTTAPPPTTAAPPTTAAPPTTTTTPVSAQAQPPASPVTPPPAV
TPTSPAPKVPVISPATPPPQPPQSPASAPT VSPPPVSPPPAPTSPPTPASPPAPASPPPAPASPPPAPVSP
PVQAPSPISLPPAPAPAPT KHKRKHKHKRHHHAPAPAPIPPSPSPSPVLTDPQDTAPAPSNTGGNALNQLKGRAVM
WLNTGLVILFLLAMTA

>At2g23130-AGP17K 68415.m02760 arabinogalactan-protein (AGP17) identical to gi 11935086 gb AAG41963
MTRNILLTVTLICIVFITVGGQSPATAPIHSPSTSPHKPKPTSPAISPAAPTPESTEAPAKTPVEAPVEAPPSPTPA
STPQISPPAPSPPEADTPSAPEIAPSADVAPALTKHKKKTKKKHTAPAPGPASELLSPPAPPGEAPGPGPSDAFSPA
ADDQSGAQRISVVIQMVGAAAIWSLLVLA

>At4g37450-AGP18K arabinogalactan-protein (AGP18) identical to gi 11935088 gb AAG41964
MDRNFLTTLICIVVAGVGGQSPISSPTKSPTPSAPTTSPTKSPA VTSPTTAPAKTPTASASSPVESPKSPAPVS
ESSPPPTPVPESSPPVPAPMVSSPVSPVPAPVADSPAPVAAPVADVPAPAPSKHKKTTKSKKHQAAPAPAPEL
LGPPAPPTESPGPNSDAFSPGPSADDQSGAASTRVLRNVAVGAVATAWAVLVMAF

>At1g51915-AGP42P 68414.m05852 cryptdin protein-related contains weak similarity to Swiss-Prot:P17533 cryptdin-related protein 1C precursor (CRS1C) [Mus musculus]

MATERFSTMLISVLVLAALVLSPIPCQA TRAHLDAETRMLRRVCPSCVCCAPAPRGACCPCRCPKNP

>At1g55330-AGP21P 68414.m06321 arabinogalactan-protein (AGP21)
MEAMKMKMMVFIMVVAFAAATAATVEAPAPSPTSDAAMFVPALFASVVALASGFIF

>At2g41905-AGP43P 68415.m05184 expressed protein
MEMKKIACGVVFAAASMTAVMAAEVGGAPAPGPAASGASVAVPALGSLVGLVSLFAYYLN

>At2g46330-AGP16P 68415.m05764 arabinogalactan-protein (AGP16) identical to
gi|10880509|gb|AAG24284
MASRNSVTGFALFSFVFAVILSLAGA QSLAPAPAPTS DGTSIDQGIAYLLMVVALVLTLYLIHPLDASSSYSFF

>At3g01730-AGP44P 68416.m00109 expressed protein
MRAFSLMICVVLVLA TLIMVAESKSSYSGNDETLTPPPMAPSPSPTGTDGGHFGEPKTLPPPNAAATFTTCPLLITS
TLISALAFVF

>At3g13520-AGP12P 68416.m01700 arabinogalactan-protein (AGP12) identical to
gi|10880501|gb|AAG24280
MESMKMKLIVVLMVAIVAFSAVGNVAAQTEAPAPSPTSDAAMFVPALFASVAALASGFLF

>At3g20865-AGP40P 68416.m02638 arabinogalactan-protein, putative (AGP)
similar to gi|10880503|gb|AAG24281
MEMKNIFVALFISAVLVSSVSAATMESPAPSPGASSASTVAFPVVGSIVAASLSAFLALLLQ

>At3g57690-AGP23P 68416.m06427 arabinogalactan-protein, putative (AGP23)
similar to arabinogalactan protein [Arabidopsis thaliana]
gi|10880503|gb|AAG24281
MEMKKIACGVLF A AASMTAVMAAEVGGAPAPGPASAASAALPALGSLVGLVSLFSSYYLH

>At3g61640-AGP20P 68416.m06908 arabinogalactan-protein (AGP20)
MASRNSVAVIALFAFVFAVISPFAGA QSLAPAPSPTSDGTSIDQGIAYLLMVVALVLTLYLIHPLDASSSYTFF

>At4g26320-AGP13P 68417.m03785 arabinogalactan-protein (AGP13) identical to
gi|10880503|gb|AAG24281
MEAMKMRFLFVAVLVAAMAFSAVQQA AAEVAPAPSPTS DASLAIPAFFASVATLAFGFLF

>At5g11740-AGP15P 68418.m01371 arabinogalactan-protein (AGP15) identical to
gi|10880507|gb|AAG24283
MAISKASIVVLMVVIISVVASAQSEAPAPSPTS GSSAISASFVSAGVA AVALVFGSALRI

>At5g12880-AGP45P 68418.m01478 proline-rich family protein similar to
hydroxyproline-rich glycoprotein gi|169345|gb|AAA33763
MRRSWPTRLILLMVLTVVITMIAAAYGYSSVSSSKHKFPHYKYKAPSPPTTYSPIRYFSPPPVTDSDSAAYVR

>At5g24105-AGP41P 68418.m02831 expressed protein
MSGSRLLFFGVSTIVSIIFAILLPMAHAQSAAPAPAPTS DGTTIDQGIAYVLMMLVALVLTLYLIH

>At5g40730-AGP24P 68418.m04943 arabinogalactan-protein (AGP24)
MMMMTKMFVQIAVVCLLATMAVVSAHEGHHHHAPAPAPGPASSSTVVSATNMFTVLAIAAVALVGSNH

>At5g53250-AGP22P 68418.m06619 arabinogalactan-protein, putative (AGP22)
similar to arabinogalactan protein [Arabidopsis thaliana]
gi|10880509|gb|AAG24284
MASLKFPLEILAVFVIISVILLPIAQS HSSSPAPAPTS DGTSIDQGIAYVLMMLVALTYFIH

>At5g56540-AGP14P 68418.m07056 arabinogalactan-protein (AGP14) identical to
gi|10880505|gb|AAG24282

MEAMKMKLYVVVVLVAVIAFSTVHQTVAVDAPAPSPTSDASSFIPTFFASVAVMAFGFFF

>At1g03870-FLA9 68414.m00371 fasciclin-like arabinogalactan-protein (FLA9)
identical to gi 13377784 gb AAK20861

MATTRLTLAPLLLIAAVLLATKATAQPAAPAPEPAGPINLTAILEKGGQFTTFIHLNITQVGSQVNIQVNSSSEGM
TVFAPTDNAFQNLKPGTLLNQLSPDDQVKLILYHVSPKYYSMDLLSVSNPVRTQASGRDNGVYGLNFTGQTINQINVS
TGYVETRISNSLRQORPLAVYVVDMLLPGEMFGEHKLSPAPAPKSKSGGVTTDDSGSTKKAASPDKSGSGEKKVG
LGFGLGLIVLCLKFLF

>At1g15190-FLA19 68414.m01816 hypothetical protein

MAKISSASCRAIFLGLALIILCLPHPSTGVPLEELERAIAILRVGRALFANAIITSDLLFDLLSDESLTLFAPTDS
MLFDLDMTHSLPFYVSTLRLHSLPLRLSLGLRSLPNSSSLPTLLPSHRLLLTKHSSSNDSIFLDGVQLLIPGLFDG
QHIAVHGLADLLPLTAPSSPNRLVEDSTALAKSPWFLGSRFSPAPEPYFAFMDLSPAESPVEEVSPSPSWGEGEED
FIVGDEGGPLDGRNNGF

>At2g04780-FLA7 68415.m00488 fasciclin-like arabinogalactan-protein (FLA7)
identical to gi 13377782 gb AAK20860

MAKMQLSIFIAVVALIVCSASAKTASPPAPVLPPTPAPAPAPENVNLTELLSVAGPFHTFLDYLLSTGVIETFQNOA
NNTTEGITIFVPKDDAFKAQKNPPLSNLTKDQLKQLVLFHALPHYYSLSEFKNLSQSGPVSTFAGGQYSLKFTDVS
TVRIDSLWTRTKVSSSVFSTDPVAVYQVNRVLLPEAIFGTDVPPMPAPAPAPIVSAPSDSPSVADSEGASSPKSSHK
NSGQKLLLAPISMVISGLVALFL

>At2g20520-FLA6 68415.m02397 fasciclin-like arabinogalactan-protein (FLA6)
identical to gi|13377780 gb AAK20859

MSSSLFSYVLLIFLFTIPYIQSQPTAPAPTTEKSPINLTAILEAGHQFTTLIQLLNTTQVGFQVSVQLNSSDQGMT
IFAPTDNAFNLKPGTLLNSLTYQQIQLMLYHIIPKYYSLSDLLASNPVRTQATGQDGGVFGFLNFTGQAQSNQVNV
STGVVETRINNLRQQFPLAVYVVDVSVLLPEELFGTKTTPGTGAPAPKSSSTSSSDADSPAADDEHKSAGSSVKRTSLG
IVVSFALFCCSVIYIA

>At2g24450-FLA3 68415.m02922 fasciclin-like arabinogalactan family protein
similar to fasciclin-like arabinogalactan-protein 1 [Arabidopsis thaliana]
gi|13377776|gb|AAK20857

MGLKVSSSLCLTILLAVSSIVSAVNITRVLEKYPEFSTMTTELLAKTELTPPIINKRQTITVLALNNDIAIGSISGRPE
EEVKNILMNHVVLDYFDELKALKALKEKSTLLTTLTQSTGLGQQQNGFLNCTKSNKGIYFGSGVKGAPQTAEYITTVF
RNPYNLSVVQISMPIVAPGLGSPVKVPPPPMSSPPAPSPKKGAA TPAPAPADEGDYADAPPGLAPETAPASAPSES
DSPAPAPDKSGKKKMAAADEAEPPSSASNTGLSFGAVLVLGFVASFVGF

>At2g35860-FLA16 68415.m04403 beta-Ig-H3 domain-containing protein /
fasciclin domain-containing protein contains Pfam profile PF02469: Fasciclin
domain

MDSSYGATKFLLLLFLTTSIATA LPDNKPVPQGINSNSVLVALLD SHYTELAELVEKALLLQ TLEEAVGKHNITIFA
PRNDALERNLDPLFKSFLLEPRNLKSLQSLLMFHILPKRITSPQWPSLSHHHRTLSNDHLHLTVDVNTLKVDSAEII
RPDDVIRPDGIIHGIERLLIPRSVQEDFNRRRSLRSISAVIPEGAPEVDPRTHRLKKPSPAVPAGAPPVLPPIYDAMS
PGPSLAPAPAPGPGGPRGHFNQDAQVKDFIHTLLHYGGYNEMADILVNLTSLATMGRLVSEGYVLTVLAPNDEAMA
KLTTDQLSEPGAPEQIMYYHIIPEYQTEESMYNAVRRFGVKVYDSLRFPHKVLAQEADGSVKFGHGDGSAYLFDPI
YTDGRISVQIGDGVLFPKKE TPA TEIKPAAPVVKVSKSRRGKLMVACRMMGSRFIPCQ

>At2g45470-FLA8 68415.m05655 fasciclin-like arabinogalactan-protein (FLA8)

MAASQTFSLLAFTFSLLAFASTVSSHNITQILADSPDYSSFNSYLSQTKLADEINSRTTITVLVLNNGAMSALAGKH
PLSVIKSALSLLVLLDYYDPQKLHKISKGTTSTTLTSTLYQTTGNAPGNLGFVNITDLKGGKVGFGSAASGSKLDSSYTK
SVKQIPYNISILEIDAPIIAPGVLTAPAPSASLSNITGLLEKAGCKTFANLLVSSGVLKTYESAVEKGLTVFAPSDE
AFKAEGVPLTKLTQAEVVSLLLEYHALAEYKPKGSLKTNKNNISTLATNGAGKFDLTTSTSGDEVILHTGVAPSRLA
DTVLDA TPVVIIFTVDNVLPAELFGKSKSPSPAPAPEPVTAPTSPADAPSPTAASPPAPTDESPESAPSDSPTGS
ANSKSANA AVGVSTPSLFTALVTIAAIVSVSLCS

>At3g11700-FLA18 68416.m01434 beta-Ig-H3 domain-containing protein / fasciclin domain-containing protein contains Pfam profile PF02469: Fasciclin domain

MDRCIYGCSVITIFFSFFFLN~~ASA~~LES~~GH~~HNITGSGQINSNSVLVALLDSRYTELAELVEKALLLQ~~TLE~~DAVGRHN
ITIF~~AP~~RNEALERDLDPDFKRFL~~LQ~~PGNLKSLQ~~TLL~~LSHII PKRVGNSQWPEENSGRVKHVTLGHDQVLHLSK~~LK~~GT
NGKRLVNSAVITRPDDLTRPDGLIHGIERLLI PRSVQEDFNRRRNLRSISAVLPEG~~AP~~EIDPRTNRLKKSATAVSV~~P~~
~~AG~~SPVVLPIESAM~~AP~~GPSL~~AP~~~~AP~~~~AP~~GPGGAHKHFNGDAQVKDFIHTLLHYGGYNEMADILVNLTSLATEMGR~~L~~VSEG
YVLTVL~~AP~~NDEAMGKLT~~TD~~Q~~L~~SEPG~~AP~~EQIMYYHIIPEYQTEESMYNSVRRF~~G~~KVKYETLRFPHKVGAK~~E~~ADGSVKF
GSGDRSAYLFD~~PD~~IYTDGRISVQ~~G~~IDGVLFPEEKEEETVKKPTG~~P~~VKKV~~VQ~~PRRGK~~L~~LEVAC~~S~~MLGAIGK~~D~~SYL~~S~~R~~C~~

>At3g12660-FLA14 68416.m01578 fasciclin-like arabinogalactan family protein similar to fasciclin-like arabinogalactan-protein 1 [Arabidopsis thaliana] gi|13377776|gb|AAK20857;

MSSSLT~~IFF~~FFFF~~FAST~~FLYTSSNS~~FN~~ITNILNEHDDFSN~~F~~NQLLSETQLASTINKRQTITVLVVSNGALSSLSGQPTS
VIKKILSLHIVLDY~~YD~~QK~~KL~~KNLSKKT~~V~~LLTTLFQSSGLARGQ~~Q~~GF~~L~~NATVMKNGDVAF~~G~~SAVPGSS~~L~~DAQLQ~~D~~TVA
ALPFNISVLHISSAIMIDVKGD~~N~~APTAS~~PL~~SPVS~~SP~~PR~~PA~~ESP~~N~~DDGQDFDEPPSS~~AP~~GAAADEPSENAG~~S~~ANGV~~S~~R
NDSQPAFAFTLLMSFIW~~W~~F~~M~~AR~~L~~R

>At3g46550-FLA4 68416.m05053 fasciclin-like arabinogalactan family protein similar to fasciclin-like arabinogalactan protein FLA8 [Arabidopsis thaliana] gi|10880493|gb|AAG24276

MANVISISHFTLLALPYLLLLLSSTAAA~~IN~~VTA~~V~~LVSSFPNLS~~S~~FSNLLVSSGIAAELSGRNSLTLLAVPNSQFSSAS
LDL~~T~~RRLPPSALADLLRFHVLLQFLSD~~S~~DLRRIPPSGSAVTTLYEASGR~~T~~FFGSGSVNVTRD~~PA~~SGSVTIG~~S~~PATKN
VTVLK~~L~~LETKPPNITVLTVD~~S~~LIVPTGIDITASETL~~TP~~PPTSTSL~~S~~PP~~PA~~GINLTQILINGHNFNVALSLLV~~V~~ASGVI
TEFENDERGAGITV~~F~~VPTDSAFSD~~L~~PSNVNLQSL~~PA~~EQKAFVLK~~F~~HVLHSYYTLGSLESITNPVQPTLATEEMGAGS
YTLNISR~~V~~NGSIVTINS~~G~~VVLAVVTQTAFDQNPVSVF~~G~~VSKVLLPKELFPKSGQPVAT~~AP~~PQEISL~~S~~PESSESSEQPSR
LV~~SP~~PREIVSSGAVKRPLGFLVLWCWCIAFCYV~~L~~V

>At3g52370-FLA15 68416.m05756 beta-Ig-H3 domain-containing protein / fasciclin domain-containing protein weak similarity to osteoblast S~~P~~ecific factor 2 [Homo s~~AP~~iens] GI:393319; contains Pfam profile PF02469: Fasciclin domain; supporting cDNA gi|26450295|dbj|AK117608.1|

MDDL~~SK~~LLFFLLLTISITTA~~LP~~DKPGSGQINSNSVLVALLDSHYTELAELVEKALLLQ~~TLE~~EAVGQHNITIF~~AP~~RND
ALEKNLDPEFKS~~F~~LLQPKNLKSLQ~~S~~LLMFHILPKRIT~~SP~~QFSSAVVSHRTLSNDHLHFTNGK~~V~~NSAEITK~~P~~DDLTRP
DGI~~I~~HGIERLLI PRSVQEDFNRRRSLRSIAAVLPEG~~AP~~EVDPRTHRLKKK~~PAP~~I~~PA~~G~~AP~~PVLPVYDAM~~SP~~GPSL~~APA~~
~~PAP~~GPGGPRHHFN~~G~~E~~A~~QVKDFIHTLLHYGGYNEMADILVNLTSLATEMGR~~L~~VSEGYVLTVL~~AP~~NDEAMAKLT~~TD~~Q~~L~~S
EPG~~AP~~EQIMYYHIIPEYQTEESMYNSVRRF~~G~~KIRYDSLRFPHKVEAQEADGSVK~~F~~GHDGSAYLFD~~PD~~IYTDGRISV
Q~~G~~IDGVLFPEEK~~TP~~V~~E~~KKTGVPV~~V~~KK~~AP~~K~~PR~~R~~G~~KLM~~E~~VACTMLGSQFPTCQ

>At3g60900-FLA10 68416.m06813 fasciclin-like arabinogalactan-protein (FLA10)

MATSRAFTLFAFTLSLLT~~V~~ASTVSG~~HN~~ITQILSD~~TP~~EYSSFN~~Y~~LSQTKLADEINSR~~T~~ITVLV~~L~~NNGAMSSLAGKH
PLSVVKNALSLLVLLDYDPLKLHQLSKGTTLT~~T~~T~~T~~LYQTTGHALGNLGFVNVTDLKG~~G~~KVGF~~G~~SA~~AP~~GSKLDSSYTK
SVKQIPYNISVLEIN~~APII~~~~AP~~GIL~~T~~APAPSSAGVSNITGLLEKAGCKTFANLLVSSGVIKTFESTVEKGLT~~V~~F~~AP~~SD
EAFKARGVPDLT~~N~~LQAEVVS~~L~~LLEYHALAEYKPKGSLKTNKDAISTLATNGAGKYDLT~~T~~STSGDEVILHTGVG~~P~~SRL
ADTVVDE~~TP~~VVIFTVDNVLL~~PA~~ELFGKSS~~SP~~AP~~AP~~EPV~~S~~APT~~TP~~PAK~~SP~~SPVE~~AP~~SPTAA~~SPP~~APVDES~~S~~PEG~~AP~~S
DSPTSSSENSNAKNAAFHV~~N~~APALFTALVTIAAT~~S~~LLL

>At4g12730-FLA2 68417.m01999 fasciclin-like arabinogalactan-protein (FLA2) identical to gi 13377778 gb_AAK20858

MAYL~~RRA~~ATA~~LV~~LIFQLHLFLSLSNA~~HN~~ITRILAKDPDFSTFNHYLSATHLADEINRRQTITVLAVDNSAMSSILSN
GYSLYQIRN~~I~~LSLHVLDYDFG~~T~~KKLHQITDGSTSTASMFQSTGSATGTSGYINITDIKGGK~~V~~AFGVQDD~~D~~SKLTAHY
VKS~~V~~FEKPYNISVLHISQVLT~~S~~PEAE~~APT~~AS~~SP~~SDLILTILEKQ~~G~~CKAFSDILKSTGADKTFQ~~D~~TV~~D~~GGLTVFCPSD
SAVGK~~F~~MPK~~F~~KS~~L~~SPANKTALVLYHGMPVYQSLQMLRSGN~~G~~AVNTLATEGNNK~~F~~DFTVQNDGEDVTLETDV~~V~~TAKVM
GTLK~~D~~QEPLIVYKIDK~~V~~LLPREIYKAVKTS~~AP~~APKSSKKKPKNAEADADG~~P~~SAD~~AP~~SDDDVEVAD~~D~~KNGAVSAMITR
TSNVVTAIVGLCFGVWLM

>At4g31370-FLA5 68417.m04448 fasciclin-like arabinogalactan family protein similar to fasciclin-like arabinogalactan-protein 1 [Arabidopsis thaliana] gi|13377776|gb|AAK20857

MGLKASLSLLSLTILLVFSKVVTA
NNITLAFQKYSKFSTMRDLFIKTKLIAAIDKYQTITVLAVSNDAISSITNRSE
VELRNILMTHVILDYYDELKLGQGMREKSI MLTTLTYQT TGLGEMNGFLNVSKSKGRVYFGSEVKNSPLNAEYVSTVY
HNPYNLSIIQITMPIVAPGLSLAIFPPPPYVHVAPYPTPMDASVVPAPGPA
ADDNSPDSAVPKTTPAPATDTPEAD
SPAPAP
SADNEKIEAADKAKPSS
SASKAGWSFDVILLFLASFAF

>At5g03170-FLA11 68418.m00265 fasciclin-like arabinogalactan-protein (FLA11)
MATSRTFIFSNLFIFFLVIATTYQ
APAPGPGSGPTNITAILEKAGQFTLFI RLLKSTQASDQINTQLNSSSSNGLTV
FAPTDNAFNSLKSGLNLSLSDQKQVQLVQFHVLP
TLITMPQFQTVSNPLRTQAGDQNGKFP
LNITSSGNQVNIITG
VVSATVANSVYSDKQLAVYQVDQVLLPLAMFGSSV
APAPAPEKGGSVSKGSASGGDDGGDSTDS
SDAERTGFGFGIR
ITTVAAIAASSLWI

>At5g06390-FLA17 68418.m00715 beta-Ig-H3 domain-containing protein / fasciclin domain-containing protein contains Pfam profile PF02469: Fasciclin domain

MDRRIYGGSAVIHLFLFFSVLIFSAASALS
KNQSPSSGSGQINSNSVLVALLDSRYTELAELVEKALLLQTL
EDAVG
RHNITIFAPRNEALERDLDFEFKRFLL
EPGNLKSQTLLMFHII
PNRVGSNQPSEESGRVKHHTLGN
DQVRLSNGQ
GKKMVDLAEIIRPDDLTRPDGLIHGIERLLI
PRSVQEDFNRRRSLSISAVLPEGA
PEVDPRTNRLKKPAAPVPAGS
PPALPIQSAMAPGPSLAPAPAP
GPGGKQHFDGEAQVKDFIHTLLHYGGY
NEMADILVNLTSLATEMGRVSEGYVL
TVLAPNDEAMAKLTTDQLSEPGA
PEQIVYYHIIPEYQTEESMYNSVRR
FGVKVFDTLRFPHKVAAKEADGSVK
FGDG
EKSAYLFDPIYTDGRISVQIGDGVLP
QEEVSVKPKPKIVQPRRGLLEVAC
SMLGAFGKDTYLSKCR

>At5g06920-FLA21 68418.m00782 hypothetical protein

MGCCSSDCFVYFILSIALAFMAISTTLRSP
PDSEPTIPIAFSSSSP
SLSLNASNTLRQSNFKAIATLLHI
SPEIFLS
SSPNTTLFAIEDASFFNTSSLHPLFLKQL
LLHYHTLPLMSMDLLKKPQGTCLPTLL
HHKSVQISTVNQESRTAEVN
HVRITHPDMFLGDSLVIHGVI
GPFSP
LQPHSDHLIHTPLCQSDTTNKT
SNNEVPVSI
DWTRIVQLLSSNGFV
PFAI
GLHSVLNRIVNDHNNHKNLTGV
TILATPNLVSLSAS
SPFLYEVRHHILVQRLTYKDFASMS
DKATVKTLDPYQDLT
ITRRNVNSSGGDFMISGVEIVDPDM
FSSSNFVIHGISHLTLEIPHV

>At5g40940-FLA20 68418.m04971 hypothetical protein

MNKKPSKYFPSLNRTDYKNCNFCNGFSLK
PNEKSSKNASMASKLLTTF
FLIFFVLDIDLVATSMTSVSSAVEVLS
DSGYLSMGLTLKLANQDLNLEDWQELTLF
APSDQSFSKFGQPSLLDMKYQLSP
TRLPGETLRNLPNGAKIPTLR
SNY
SLTVTNSSRFGGKTSINNVVVQD
SPVFDGYYVVIYGSDEFFT
SPTKISDDSSSSSSIPSTTS
TGSIPSSATQTP
PSPNIASDSTRNLPNRSKPVNRFNIF
ESASRLMSRGFVIIATFLALQ
LEDNTSGNDTKITVFAPI
IDEAIPNPTTKF
SDYVTIFRGHVVSQLLLWKDLQKFA
KEGSILQTVLKGYEIEISLSG
DILLNGVPLIYPDLVNDWIAVH
GFNQMIV
TKEKQVDVGDSTITV
LNNGEQEEGVHGEYSSELG
DYGLH

>At5g44130-FLA13 68418.m05401 fasciclin-like arabinogalactan-protein, putative similar to gi 13377784 gb_AAK20861

MATPLLLLLLTAVFLSTEITAQRAAPAPG
PAGPINITAILEKGGQFVTLIRLLNTTQ
IGNQINIQINSSSEGMTVL
APTDFNAFQNLKPGTLNKLSP
DDQVKLILYHVS
PKFYTLEDLLSVSNPVRTQASGRD
VGGVYGLNFTGQGNQVNVSTG
VVETRLSTSLRQERPLAVYV
VDMVLLPEEMFGERKI
SPMAPPKSKSP
DVSDSESSKAAAPSESEKS
GSGEMNTG
LGLGLGLVVLCKFL

>At5g55730-FLA1 68418.m06947 fasciclin-like arabinogalactan-protein (FLA1) identical to gi|13377776||AAK20857|13377775|gb|AF333970

MAKKMSSLIIFNILLTQT
THANVTRLLANHPSFSSFS
HFLTQTHLADEINRRRTITV
CAVDNAAMSALTSKGY
TLSTLKNILSLHVLLDYFGTK
KLHQIRDGSALAATLFQATGA
APGTSGFVNITDLRGGKVG
FGPDGGDLSFFVKSIE
EVPYNISIIQISRVL
PSETAAAPTAPAE
MNLTGIMSAHGCKVFA
ETLLTNP
GASKTYQESLEGGMTV
FCPGDDAM
KGFLPKYKNLTAPKKEA
FLDFLAVPTYY
SMAMLSNNGPMNT
LATDGANKFELTVQ
NDGEKVT
LKRINTVKIVDTL
IDEQPLAIYATDKVLL
PKELFKASAVE
APAPAPAPEDGDVAD
SPKAAK
GKAKGKKKAA
APSPDNDPFGDSD
SPA
E
G
D
G
E
A
D
D
A
T
A
D
D
A
G
A
V
R
I
I
G
G
A
K
A
G
L
V
V
S
L
L
C
L
F
A
S
S
W
L
L

>At5g60490-FLA12 68418.m07586 fasciclin-like arabinogalactan-protein (FLA12)

MEHSLIILLFTVLLLLLTTTPGILS QPSPAVAPAPPGPTNVTKILEKAGQFTVFIRLLKSTGTVANQLYGQLNNSDNGI
TIFAPSDSSFTGLKAGTLNSLTDEQQVELIQFHVIPSYSVSSSNFQTI SNPLRTQAGDSADGHFPLNVTTSGNTVNI
SGVTNTTVSGNVYSDGQLAVYQVDKVLPPQVFDPRP PAPA PAPS VSKSKKKKDDSDSSSDS SPADASFALRNVGSV
CDAVSFCVMSVMLAWFYL

>At2g23990-PAG1 68415.m02865 plastocyanin-like domain-containing protein
MVSLISIVSVVFLFTTFYHFGEAR IINVGGLDAWKVPE SPNHLNHWAESVRFQVGDALLFKYDSKIDSVLQVTK
ENYEKNTQKPLEEHKDGYYTTVKLDVSGPYYFISGAPSGNCAKGEKVTVVVQ SPNHPKPG PAAVTPTLPPKPS TPA
APAPAPPTPSPKSSTSTM APAPAPA KSSAVGLVAGNGIFWASTLVAVIGLAF

>At2g25060-PAG2 68415.m02997 plastocyanin-like domain-containing protein
MFLSASMASSSLHVAIFSLIFLFLSLAAANEVTVGGKSGDWKIPSSSYSTFEWAQKARFKVGFIVFRYESGKDSVL
EVTKEAYNSCNTTNPLANYTDGETKVKLD RSGPFYFISGANGHCEKGQKLSLVVI SPRHSVI SPAPSPVEFEDGPAL
APAPI SGSVRLGGCYVVLGLVGLCAWF

>At2g26720-PAG3 68415.m03205 plastocyanin-like domain-containing protein /
mavicyanin, putative similar to mavicyanin SP:P80728 from [Cucurbita pepo]
MALIKNNIFFTSLILFVTLFGVAVG GTVHKVGNTKGWTMIGGDYEAWASSRVFQVGDTLVFAYNKDYHDVTEVTHND
FEMCESSKPLRRYKTGSDSISLTKPGLQHFICGVPVGHCKKQKQLQIHVL PASLGHVAVPVPGPVRSSSSSS SPSQSP
LVDPPVNN APQYQMGPTPASHSAASADFIFTFSFDLTLIDLCTFFILFFILV

>At2g31050-PAG4 68415.m03788 plastocyanin-like domain-containing protein
contains plastocyanin-like domain Pfam:PF02298
MALIKSNAFFTSLILVALFGISVGTVHKVGDSDGWTIMSVNYETWASTITFQVGDLSLVFKYNKDFHDVTEVTHND
YEMCEPSKPLARYETGSDIVILTKPGLQHFICGFPVGHCDMGKQLQIHVL PASLGPVA APVPGPVRPPSSSFS SPSQSP
LAESPVNH APVQYQMGPTPAPH SAASNSNVWIGLCFLPLLSLILV

>At2g32300-PAG5 68415.m03949 uclacyanin I identical to uclacyanin I
GI:3399767 from [Arabidopsis thaliana]; contains Pfam profile PF02298:
Plastocyanin-like domain; identical to cDNA uclacyanin I GI:3399766
MASREMLIIISVLATTILGLTVA TDHTIGGSPGWTVGASLRTWAAGQTFVAVGDNLVFSY PAAFDVVEVTKPEFDSC
QAVKPLITFANGNSLVPLTTPGKRYFICGMPGHCSQGMKLEVNVPVPTATVAPTAPLPNTVPSLNAPSPSSVLPIQPL
LPLNPVPVLS PSSSTPLPSSSLPLIPPLSPAL SPATAAGTSLPLFPSPGSSSSTTSTKTVTGTFPSSTTGTTADLAG
ADSPPADSSSAKTLVLGFGFMVAMMLHLF

>At2g44790-PAG6 68415.m05574 uclacyanin II strong similarity to uclacyanin
II GI:3399769 from [Arabidopsis thaliana]; contains Pfam profile PF02298:
Plastocyanin-like domain; identical to cDNA uclacyanin II GI:3399768
MAMNGLSKMAVAATAALLLVLTIVGAVA VTYTIEWTTGVDYSGWATGKTFRVGDILEFKYGSSTHTVDVVDKAGYDG
CDASSSTENHSDGDTKIDLTKVGINYFICS TPGHCRTNGGMKLA VNVVAGSAGP PATPTPPSSSTPGTPTTPESPSPSG
GSPTPTTPTPGAGSTSPPPPKASGASKGVMSYVLVGVSMVLGYGLWM

>At3g20570-PAG7 68416.m02604 plastocyanin-like domain-containing protein
MARNLKSMMMLCGFLLCFLMIVDRAYA REFTVGGATGWTVPVSGSQVYSQWAEQSRFQIGDSLFLVYQSNQDSVLQVT
RDAYDSCNTDSPTAKFADGKTSVTLNHS GPYYFISGNKDNCKKNEKLVVIVMADRSGNKNTASSP SPAPAPSGESA
PSPPVSGTFEMTPAPTPTTSED TPNSAASSLSFVAALLGAALASTLFLH

>At3g60270-PAG8 68416.m06737 uclacyanin, putative similar to uclacyanin 3
GI:3395770 from [Arabidopsis thaliana]; contains Pfam profile PF02298:
Plastocyanin-like domain
MGSTAATAALLLLLLLVAVPAVFA VTFQVGDNDGWTIGVEYTSWVSEKTFRVGDTLEFKYGPSHSVAVVNKADYDGC
TSRPTQSFSDGDTKIDLTKVGAIFLCLTPGHCSLGMKLA VQVLA AVSLEPPP SPSAPSPSPSAPSPSPSAPSPSPS
PGNAENLKN AASKGIMS YGKIVVTMVL MYGVLN

>At3g60280-PAG9 68416.m06738 uclacyanin 3 (UCC3) identical to uclacyanin 3
GI:3395770 from [Arabidopsis thaliana]; contains Pfam profile PF02298:
Plastocyanin-like domain; identical to cDNA uclacyanin 3 (UCC3)GI:3395769

MGSTVAAALLLFLAAVPAVFAATFKVGDISGWTSNLDYTVWLTGKTFRVGDTLEFVYGLSHSVSVVDKAGYDNCDDSS
GATQNFADGDTKIDLTTVGTMHFLCPTFGHCCKNGMKLAVPVLAAAPSPSTPSSPPSPTSPSPSPSPTSPSPSPSP
PSPSPSPSLPPSSLPPSASPPTNGTPDSETLTPPAPLPPSLSPNAASKGVMSYGIIGVTMILMYAVMT

>At4g27520-PAG10 68417.m03952 plastocyanin-like domain-containing protein
similar to PIR|JC7196 phytocyanin-related protein Pn14 {Ipomoea nil};
contains Pfam profile PF02298: Plastocyanin-like domain

MTFLKMKSLSEFFFTILLSLSTLFTISNARKFNVGGSGAWVTNPPENYESWSGKNRFLVHDTLYFSYAKGADSVLEVN
KADYDACNTKNPIKRVDDGDSEISLDRYGPFFYFISGNEDNCKKGQKLNVVVISARIPSTAQSPHAAAPGSSTPGSMT
PPGGAHSPKSSSPVSPPTTSPPGSTTPPGGAHSPKSSSAVSPATSPPGSMAPKSGSPVSPPTTSPAPPKSTSPVSPSS
APMTSPAPMAPKSSSTIPPSSAPMTSPPGSMAPKSSSPVSNPTVSPSLAPGGSTSSSPSDSPSGSAMGPSGDGPS
AAGDISTPAGAPGQKKS SANGMTVMSITTVLSLVLTIFLSA

>At4g28365-PAG11 68417.m04060 plastocyanin-like domain-containing protein
MGLVMRFDLYLMFVMLMGLGFTISNGYKFYVGGKDGWVPTPSEDYSHWSHRNRFQVNDTLHFKYAKGKDSVLEVTEQ
EYNTCNTTHTPLTSLSDGDSLFLLSHSGSYFFISGNSQNCLKGQKLAVKVLSTVHSHSPRHTSPSPSPVHQELSSPG
PSPGVEPSSDSNSRVPAPGPATAPNSAGLVGPGMVVLVIMISSLF

>At4g30590-PAG12 68417.m04340 plastocyanin-like domain-containing protein
MGIIVPVLTFLVLLFAKVSHGASNPRVILVGGSVGSWKVPDSPNNTLNHWAENNRFKVGDIVWKYDMKVDSVLQVT
KEDYESCNTANPLKQYNDGNTKVALDKSGPYFFISGAPGNCAKGEKITLVVLAERKSGGGSSSGDAPKVPSPVSPTAQ
TPAPAPGPAAAHNAAVGLKVASGWFLTAVVVGLAMA

>At4g31840-PAG13 68417.m04524 plastocyanin-like domain-containing protein
MASSLLVTIFLCISVFFFSSVNA NEVTVGGKSGDWKIPSSSFSFNEWAQKARFKVGDIVFKYEAGKDSVLQVTR
EAYEKCNTTSPKASYTDGNTKVKLDQAGPVYFVSGTEGHCQKQKLRLLVITPRNSAFSPGSPSEFDGPAVAPTSG
AAKLAGGFSVVFGLVLGLWAFFF

>At4g32490-PAG14 68417.m04625 plastocyanin-like domain-containing protein
MVFVKMTDVYLMIVMLMGLGFSIELSNHGKFYVGGRDGWVLTPTSEDYSHWSHRNRFQVNDTLFYKYVKGKDSVLEVS
EKEYNTCNTTHTPLTSLSDGDSLFLLSRSDPFFVSGNSGSLKGQKLAVTVMSTGHSHTPRHSPSPSPSASPVRK
ALLSPAPIPVHKALSSPAPTGVDPHSEVLAPAPGPAAAVRNLAGSVAPGVISLGLVLVIMISSMV

>At5g25090-PAG15 68418.m02973 plastocyanin-like domain-containing protein
MAQRTLVAFFLIFLLTNLVCSKEIIVGGKTSSWKIPSSPSESLNKWAE SLRFRVGD TLVWKYDEEKDSVLQVTKD
AYINCNTTNPAAANYSGDTKVKLERSGPYFFISGSKSNCVEGEKLIHIVMSSRGGHTGGFFTSSPSPAPSPALLGA
PTVAPASGGSASSLTRQVGLGVGLLAIVLL

>At5g26330-PAG16 68418.m03147 plastocyanin-like domain-containing protein /
mavicyanin, putative similar to mavicyanin SP:P80728 from [Cucurbita pepo]
MAAIIVAALACIVVMLRLSEAAVYKVGDSAGWTTIANVDYKLWASTKTFHIGDVLFEYNPQFHNVMRVTHPMYRSC
NTSKPISTFTTGNDISITLNTNHGHHFFFCGVPVGHCLAGQKLDLHVLLPASSTPLSDPPTSSSSSPSTTIIPAGVPGP
SPSLAASLPSMVTAQIVAVVTLVSLAFTNFAS

>At5g53870-PAG17 68418.m06701 plastocyanin-like domain-containing protein
contains similarity to SP|Q02917 Early nodulin 55-2 precursor {Glycine max};
PF02298: Plastocyanin-like domain
MSAIMKSLCFSFLILASFATFFSVADAWRFNVGGNGAWVTNPQENYNTWAERNRFQVNDLSLYFKYAKGSDSVQQVMK
ADFDGCNVRNPIKFNENGESVVTLDKRGAFYFISGNQDHCQKQKLI VVVLA VRNQPSAPAHSPVPSVSPPTQPPKSH
SPVSPVAPASAPSKSQPPRSSVSPAQPPKSSSPISHTPALSPSHATSHSPATPSPSPKSPSPVSHSPHSPHTPSH
SPAHTPSHSPAHAPSHSPAHAHSPAHAPSHSPAHSHPATPKSPSPSSSPAQSPATPSPMTPQSPSPVSSPSP
DQSAAPSDQSTPLASPSETTPTADNITAPAPSPRTNSASGLAVTSVMSTLFSATFTFLMFA

>At1g03820-AGP28I 68414.m00363 expressed protein similar to arabinogalactan-
protein; AGP (GI:1087015) {Pyrus communis}; similar to Protein E6.
(SP:Q01197) {Gossypium hirsutum}

MASFALKPIFCFIAVFCFIVHNVEA REGKLFSSKFTTHIDRPNNKDVALSPAPAPGLAQANGRLGNGSFGPGSGMIPO
TKESWPPSSSTTTDEEFELKMATFDEEKNTKLPEAFEEEESESDLNPKDKYNNNNNNNGYTYTTNNYNDNGRGY
GNEEEKQGMSDTRVMENGYFYDTRGRNSENTPSRGYENARGNDHTNEFETMEEYYKSLEGSQEEYEP

>At1g28290-AGP31I 68414.m03472 pollen Ole e 1 allergen and extensin family protein similar to arabinogalactan protein [Daucus carota] GI:11322245; contains Pfam profile PF01190: Pollen proteins Ole e I family

MGFIGKSVLVSVALWCFTSSVFT EEVNHTQTPSLAPAPAPYHHGHHHPHPHHHPHPHPHPHPHPAKSPVKPPVK
APVSPPAKPPVKPPVYPPTKAPVKPPTKPPVKPPVSPPAKPPVKPPVYPPTKAPVKPPTKPPVKPPVYPPTKAPVKP
PTKPPVKPPVYPPTKAPVKPPTKPPVKPPVSPPAKPPVKPPVYPPTKAPVKPPVSPPPTKPPVTPVYPPKFNRLVA
VRGTVYCKSCKYAAFNTLLGAKPIEGATVKLVCKSKKNITAETTTDKNGYFLLLAPKTVTNTFGFRGRCRVYLVKSKDY
KCSKVSCLFGGDVGAELKPEKLGKSTVVVNVKLVYGLFVNGPFAFNPSCP

>At1g36150-AGP29I 68414.m04494 protease inhibitor/seed storage/lipid transfer protein (LTP) family protein low similarity to glucoamylase S1/S2 [Precursor] from Saccharomyces cerevisiae [SP|P08640], proteophosphoglycan from Leishmania major [GI:5420387]; contains Pfam protease inhibitor/seed storage/LTP family domain PF00234

MKPSFVLLSIVLLSSLSDA ADFGSPSQPPSMAPTQPSNSTDCSSVIYSMDCLSFLTVMGSTDPSPTKTCCVGVK
TVLNYSKCLCSALESSREMGFVLDLTKALAMPKICNPIDPNCVSTPAASTPVSPPVESPTTSPSSAKSPAITPS
SPAVSHSPPVVRHSSPPVSHSSPPVSHSSPPTSRS SPAVSHSSPVVAASSPVKAVSSSTASSPRAASPSPSPSIS
SSGILLVSKLFIAVVMVSSFLYILA

>At2g33790-AGP30I 68415.m04144 pollen Ole e 1 allergen and extensin family protein similar to arabinogalactan protein [Daucus carota] GI:11322245, SP|Q03211 Pistil-specific extensin-like protein precursor (PELP) {Nicotiana tabacum}; contains Pfam profile PF01190: Pollen proteins Ole e I family

MGIIGKSVSLTLFALLCFTSSVFT LGVNQPGSSDPFHSLPQHLPLPPIKLPPLPAKAPIKLPAYPKAPIKLPTL
PPAKAPIKLPTLPPPIKPPVLPVYPPKYNKTLVAVRGGVYCKACKYAGVNNVQGAQKPVKDAVVRLVCKNKNKNSISET
KTDKNGYFMLLAPKTVTNYDIKGCRAFLVKS PDKTKCSKVSLLHDGGKGSVLKPVLPKPGFSSTIMRWFKYSVYNVGGP
AFEPTCPK

>At5g21160-AGP32I 68418.m02528 La domain-containing protein / proline-rich family protein contains proline-rich extensin domains, INTERPRO:IPR002965, PF05383: La domain

MMAETEGSVADDRELITREGGIGTKSPWKTTTSPVETIDAPVMGAHSWPALADAAQQPRPKNPAPAPAPPSKNIPT
SIPITPAVGTQAKSKGGKANPGHKNPSGRHSGKPGPRSNQNGPPPPYLVHAVPYHPPPPPMVPLPHAAGPDFPY
APYPPYPVPPVPTESGNEKQVQASPLPPVL PAPQGDGPKPWP HQRFDPNMPQAGPRNFRPFPMPGAPGFLVG
PGPGFPGPVYLLPGPPGAIRGPYPPRFAPYPVNQGPPI LSP EKL DLRDRVLKQVEYYFSDENLENDHYLISLMDEE
GWVPTKIIAGFKRVKAMTMDVDFIVYALGFSNSVEVQGDQIRKRDKWSDWI PA SKKSTSAETIGDGDKDSPKSITSG
DNFGNPSKGSKPTVSDFSSEGAQSSRTNNYKSGNLKSSADEKRNVEDLSNDFSNTFLLDEELDLEHRS PRKSGLSM
SKSIEYEDDDMAVDDQDIQKLVIVTQNSGKSDGAGIGGTEAKNIPKELASTINDGLYYFEQELKKKRSGRKNNSHL
DTKDGKIKSGEGLNTKLGENSAANDGEEHGIITSRRKQNKGTGTHKHTAHARRFFSSNIRNNGNISESPSSSIGFF
FGSTPDSHGPRLSKLSSSPQCTLSGSSPPVGSPLKSFPPFQHPSHQLLEENGFKQEKYLKRYKRLNERKKGSGC
SEEMNHLYRFWSYFLRDTFVLSMYDDFQKFALEDAAGNYDYGLECLFRFYSGLEKHFDELDYKDFEKLSDLDFYHKG
NLYGLEKYWAFHHYRGKEEPI TKHPELEKLLKEEFRSIDD FRAKETITNQKENKSH

>At5g56330-AGP33I 68418.m07031 carbonic anhydrase family protein contains proline-rich extensin domains, INTERPRO:IPR002965; contains Pfam profile PF00194: Eukaryotic-type carbonic anhydrase

MKISSLGWVVLVLFISITIVSSAPAPKPPKPKPAPAPT PPKPKP TPAPT PPKPKP PAPTPPKPKPAPAPT PPKPKP
APAPT PPKPKPAPTPPNPKP TPAPT PPKPKPAPAPAPT PPKPKPAPKPKPAPGGEVEDETEFSYETKGNKGP AKWG
TLDAEWMCMGIGKMQSPIDLRDKNVVVSNKFGLLRSQYLP SNTTIKNRGHDIMLKFKGGNKGIGVTIRGTRYQLQQL
HWHSPSEHTINGKRFAL EHLVHESKDKRYAVVAFLYNLGASDPFLFSLEKQLKKITDTHASEEHIRTVSSKQVKLL
RVAVHDASDSNARPLQAVNKRKVYLYKPKVKLMKKYCNISY