Title

Investigating Genetic Components of the Symbiosis-associated Glycocalyx in Medicago truncatula

Author

Moore, William McKnight

Publication Date

2018

Investigating Genetic Components of the Symbiosis-associated Glycocalyx in *Medicago truncatula*

By

William McKnight Moore

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Plant Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Henrik V. Scheller, Chair
Professor Shauna Somerville
Professor Mathew Welch

Summer 2018

Abstract

Investigating Genetic Components of the Symbiosis-associated Glycocalyx in *Medicago truncatula*

by

William McKnight Moore

Doctor of Philosophy in Plant Biology

University of California, Berkeley

Professor Henrik V. Scheller, Chair

Plants form a variety of endosymbiotic relationships with bacteria and fungi that promote plant growth and fitness through reciprocal nutrient exchange. Central to the development and function of endosymbiotic relationships is the synthesis of a specialized host-derived membrane, which compartmentalizes the endosymbiont inside plant cells, and creates a dynamic interface for the exchange of nutrients and information. The development of these interfaces is dependent upon bidirectional signaling between the plant and microorganism, and is highly coordinated with the morphological differentiation of the endosymbiont within plant cells. Previous studies using glycan-directed monoclonal antibodies indicated that glycoproteins, glycolipids, and pectic polysaccharides localize to periarbuscular and symbiosome membranes. Among these epitopes were arabinogalactan proteins (AGPs) and glycosyl inositolphophorylceramides (GIPCs), which have potent signaling properties in plants. These epitopes appeared to be developmentally regulated, which led to the hypothesis that a membrane-associated glycocalyx of glycoproteins and glycolipids might be important for mediating interactions through these membrane interfaces. However, AGPs have not been well documented outside of *Arabidosis thaliana* and are intrinsically disordered, which make them difficult to identify and study in plants species capable of forming endosymbiosis. Here, we have developed a new bioinformatic search tool that identifies AGP-encoding genes based on the noncontiguous hydroxyproline motifs that direct AGP glycosylation. We used this tool to identify all putative AGP-encoding genes in the *Medicago truncatula* genome, which were cross-referenced to transcriptomic studies of roots engaged in symbiosis with *Sinohizobium meliloti* and the arbuscular mycorrhizal (AM) fungus *Rhizohagus irregularis*. Using this approach we identified a small three-member family of tandemly duplicated *SYMBIOSIS-ASSOCIATED ARABINOGALACTAN PEPTIDES* (*SAP*s) that were differentially expressed in root nodules and AM colonized roots. SAPs localized to symbiotic membranes and knockdown of *SAP* expression using RNAi-mediated gene silencing impaired the growth and differentiation of *Sinorhizobium meliloti* and *Rhizphagus irregularis* within these compartments. In parallel we also identified a glycosyltransferase gene highly expressed in root nodules and AM colonized roots as *GIPC GLUCOSAMINE TRANSFERASE 1* (*GINT1*), and showed that the corresponding GINT1 enzyme functions in the synthesis of HexN(Ac) decorated GIPCs *in planta.* Silencing of *GINT1* using RNAi impaired the development of symbiotic membranes, which resulted in the senescence of symbiosomes and arbuscules. Taken together these results provide genetic evidence to support that reprogramming of the membrane-associated glycocalyx with specific AGPs and GIPCs is necessary for endosymbiosis.

Dedication

This dissertation is dedicated to my mother Sandra Cornforth Moore, my father Stephen Moore, my stepmother Eva Moore, my grandparents, my girlfriend Devon Birdseye, my PhD advisor Dr. Henrik Scheller, my former undergraduate advisors Dr. Eugene Nothnagel and Dr. Darleen DeMason, Plant and Microbial Biology cohort 2012, and my colleagues at the Joint BioEnergy Institute – all of whom have inspired, supported, and driven me to be my best self. These have been the greatest years of my life. Thank you.

New Bioinformatic Tools Identify Novel Arabinogalactan Proteins in *Arabidopsis thaliana* and *Medicago truncatula*

William M. Moore[1,2,3], Paul Hussey[1,2,3], Oge Nnadi[1,2], Nathan Hillson[1,2], and Henrik V. Scheller[1,2,3]
[1]Joint BioEnergy Institute, Emeryville, California 94608
[2]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720
[3]Department of Plant and Microbial Biology, University of California, Berkeley, California 94720

Introduction

Arabinogalactan proteins (AGPs) belong to a large class of structurally complex cell surface proteoglycans characterized by extensive glycosylation with hydroxyproline *O*-linked arabinogalactan (AG) (1–3). The carbohydrate moieties of AGPs consist of a core β(1,3) galactan polysaccharide with branching β(1,6) galactan side chains that are highly decorated in with arabinose and contain other sugars such as rhamnose, glucuronic acid, and fucose (4, 5). AGPs are ubiquitously found throughout the green plant lineage and are believed to be secreted from all plant cells (6–8). Though individual AGPs often have cell-type specific expression patterns (9–11), and can be either GPI-anchored to the outer leaflet of the plasma membrane (12, 13), cross-linked into the extracellular matrix (14), or released in plant exudates (15, 16). Collectively, AGPs are involved in many aspects of plant growth and development, ranging from cell division, expansion (17), adhesion (18), proliferation (19), and differentiation (20–23), to reproduction (24–26), stress response (27–29), plant-microbe interactions (30), hormone signaling (28, 31), and influencing mechanical cell wall properties (14, 32, 33) – however, the mode in which they act still remains poorly understood.

While AGPs remain important regulators of plant growth and development, there are many research challenges that face our study of these enigmatic glycoproteins, which have been previously elaborated upon by Tan *et al.* (2012) (34). One of the key challenges is simply identifying AGP encoding proteins in plant genomes.

From a proteomics-based approach, AGPs are difficult to identify due to their high carbohydrate content, which can constitute up to of 90% of the total molecular weight of an individual AGP. Extensive glycosylation of the protein backbone makes these proteins resistant to trypsin digest and therefore they often go unseen in proteomic analyses. Indeed, shotgun proteomic studies of the plant cell wall have generally failed to identify any AGPs (35, 36). Extraction with the Yariv phenol β-glucoside reagent, which specifically binds to the core β(1,3) galactan backbone of AG glycomodules (37), followed by chemical deglycosylation has had some success (38). However, the number of unique proteins identified using this approach is much lower than the total number of predicted AGP-encoding genes in the genome. This is likely because individual AGPs can occur in low abundance, be specifically expressed within a small cell niche, or require developmental and environment cues to induce expression. Furthermore, it has been suggested that Yariv reagent may need at least 7 unbranched β(1,3) galactose residues to form an insoluble aggregate and therefore may only react with a subset of AGPs based branch substitution of the core β(1,3) galactan backbone (34, 37).

From a computational perspective, AGP-encoding genes are difficult to identify due to the low level of shared sequence similarity among individual proteins and inherent variability within the amino acid motifs that direct protein glycosylation. Even very closely related AGPs will typically have less than 30% shared sequence identity between them at the amino acid level. Therefore, unsurprisingly, alignment-based search methods generally fail to identify other AGPs. In total, AGPs comprise a large and diverse class of hydroxyproline-rich glycoproteins. From this perspective it is helpful to think of AGP glycosylation as a post-translational modification that occurs on a large diversity of proteins, as opposed to a characteristic representing a single group. For this reason, more tailored bionformatic tools are needed to identify proteins containing AG glycomodules.

The first major breakthrough came with the work of Schultz et al. (2002) who developed a computational program to identify AGPs based on biased amino acid composition by searching for proteins with high proline, alanine, serine, and threonine content (39). AGPs are rich in Pro, Ala, Ser, and Thr, which form the characteristic 'PAST' motifs found in AGPs. This approach identified a total of 52 AGP-encoding genes in the *Arabidopsis thaliana* genome, primarily from classical, AG peptide, lysine-rich, and chimeric fasciclin-like (FLA) subclasses. Improving upon this method, work done by Showalter et al. (2010) added a sliding window function to identify a total of 85 AGP-encoding genes in *A. thaliana*, including additional classical AGPs and AG peptides (2). However, this method still failed to recover the lipid transfer-like AGPs (LTP AGP), plastocyanin-like AGPs (PAGs) and other chimeric GPI-anchored proteins believed to contain AG glycomodules, such as COBRA, SKU5, (1-3)-glucanases, receptor-like peptides (RLPs) and aspartyl proteases described by Borner et al. (2003) (40, 41).

Here we have developed a new AGP prediction program (AGPredictor) guided by the non-contiguous hydroxyproline hypothesis (described below) that is able to rapidly identify AGPs in any plant genome. Using this program, we were able to identify all previously described AGPs in *A. thaliana* including the more cryptic chimeric AGPs. We also identified an additional 346 previously unreported chimeric AGPs, such as actin nucleating formin proteins, receptor-like kinases, GDSL lipases/acylhydrolases, carbohydrate binding proteins, pectin lyases, proteases, and many others that are believed to inhabit or interact with the plant extracellular matrix. We further show that the glycosylation motifs present in these lesser-known chimeric AGPs are also represented in *Medicago truncatula*. The AG glycosylation of some of the new identified AGPs was supported biochemically based on altered protein mobility when separated on reducing SDS-PAGE.

Materials and Methods

Developing the AGPredictor Program

AGPredictor was written as a Python script with two interfaces: a command line interface and an easy-to-use Web application (Fig. 1a) incorporating the Flask microframework (http://flask.pocoo.org/). In order to reduce false positives both interfaces automatically integrate SignalP (v4.1) as an initial filter to select for proteins entering the secretory pathway. This is necessary for protein glycosylation as the proline hydroxylases and glycosyltransferases involved in AGP biosynthesis are located in the endoplasmic reticulum and the Golgi apparatus (3, 42, 43). The AGPredictor source code is freely available for download online and published under the open source 2-clause BSD license. The core of the program is written as a regular expression:

[ASTV]?(?:P[ASTV])+(?:P(?!P))?

The underlying principle for how the AGPredictor program works is based on the non-contiguous hydroxyproline hypothesis as the primary search criterion. AGPS are rich in XP and PX proline dipeptide repeats (where X = Ala, Ser, Thr, or Val) that often occur in interspersed clusters known as the ‘PAST’ rich motifs that direct AGP-type hydroxyproline glycosylation. This hypothesis states that non-contiguous hydroxyproline residues (i.e. XPXPX) become glycosylated with arabinogalactan, while contiguous hydroxyproline residues (i.e. XPPP) become glycosylated with short arabinan oligosaccharides. This has been experimentally validated through carbohydrate profiling of both synthetic and native peptide sequences expressed in plant tissue that contained either contiguous or non-contiguous hydroxyproline motifs (44–49).

The AGPredictor program works by finding repetitive and interspersed non-contiguous XP and PX dipeptides while excluding contiguous hydroxyproline residues. Each protein is scored based on the length of the longest repetitive motif found (ie XPXPXPX). There is an extra option, *runwindow*, which, if set to a number, *N*, greater than its default of 0, will score the proteins based on the total length of motifs found that are within *N* amino acids of the longest motif. For instance, by default, the sequence “APAP□□□APAP” has a score of 4. However, when the run window is increased to 3, the score of the sequence becomes 8. This is because there are two AGP motifs separated by 3 (or fewer) non-motif amino acids. Proline residues need to be within 15 amino acid residues of each other in order to be eligible for glycosylation. Proteins are then parsed based on arbitrary user-defined cut-offs to generate output files enriched in AGP sequences. Output files contain FASTA formatted protein sequences with predicted non-contiguous hydroxyproline motifs highlighted in bold to assist manual sequence curation (Fig. 1b). The program also generates a csv table summarizing the data for all protein analyzed (Fig. 1c).

Source Files

All predicted protein sequences encoded by the *A. thaliana and M. truncatula* (v4.0) genomes were downloaded as single FASTA files from Phytozome (www.phytozome.org). Known *A. thaliana* AGP sequences Showalter et al. 2010 (2) and Borner et al. 2003 (40) were used as a test set to assess AGPredictor program function and are listed in Appendix I.

Testing the AGPredictor Program with Known AGPs

The AGPredictor program was assessed using a test set of 123 known or previously predicted AGPs (Appendix I). The list was comprised of 85 AGPs identified by including classical AGPs, AG peptides, and chimeric AGPs including fasciclin (FLA), plastocyanin (PLA), and lipid transfer protein (LTP) AGPs (2). An additional 38 chimeric proteins speculated to be AGP from Borner *et al*. (2003) were also added to the list, which include proteins like COBRA and pectin lyase (40). AGPredictor was able to recover all known AGPs from the test set and 65% of AGPs parsed together into a single file using the parameters: motif score ≥13, window =6. This demonstrates that the program successfully enriches for AGPs based on their non-contiguous hydroxyproline motifs (Fig. 2).

AGP Prediction of *A. thaliana* and *M. truncatula* genomes

The AGPredictor program was used for parsing of the *A. thaliana* and *M. truncatula* genomes using a motif score cutoff ranges (0-3, 4-5, 6-12, 13-∞) and window length 6. Output files were then manually curated to develop a comprehensive list of predicted AGPs. These lists were then cross-referenced to proteins recovered by other programs for comparison (2, 50). All predicted AGPs are listed in Tables I and II and sequences with highlighted glycosylation motifs are listed in Appendices II and III.

SDS-PAGE and Western blot

The extracellular domains of *AtALE2* (AT2g20300), *AtFH3* (AT4g15200), and *MtFH3* (Medtr4g081410) were cloned from cDNA into plant expression vector pGWB17 to generate a *C*-terminal fusion protein with a 4xMYC tag (51). *AtAGP9* was used as a control and expressed in pEARLEYGATE101 with a *C*-terminal YFP-HA tag (52).Proteins were transiently expressed in *Nicotiana benthamiana* by leaf infiltration using *Agrobacterium tumifaciens* GV3101 suspension cultures (OD 0.6) (53). Leaf tissue was collected at 4 dpi, ground in liquid nitrogen, and total protein was extracted by boiling in 2X reducing SDS-PAGE sample buffer. Proteins were run on 16-8% mini-PROTEAN TGX precast protein gels (Bio-Rad) and transferred to PVDF membranes. Membranes were blocked in 3% BSA TBST, probed with a mixture of mouse anti-HA (Invitrogen; cat# 26183) and anti-MYC (Invitrogen; cat# 13-2500) mouse primary antibodies (1:5,000), followed by (1:10,000) rabbit anti-mouse IgG HRP secondary antibody (Invitrogen; cat#61-6520), with subsequent washes in between. SuperSignal West Dura™ (Thermo) was used as chemical substrate and imaged on Amersham Imager 600 (General Electric). Primers used are listed in Table III.

Results

Re-evaluating AGPs in the *A. thaliana* Genome

We tested the AGPredictor program using all predicted protein sequences encoded by the *A. thaliana* genome as a larger dataset. All known AGPs were successfully recovered and enriched in output files separate from the majority of non-AGP proteins (Fig 3 *A*, *B*). Surprisingly, the program also recovered a large number of proteins that have not previously been reported to be AGP but contained convincing non-contiguous hydroxyproline motifs. All protein sequences were manually curated and we developed a comprehensive list of approximately 431 proteins that we predict to be glycosylated with AG (Table I). This is at least four times larger than previous estimates and suggests that AGP post-translational modification is more prevalent than previously thought. Proteins with a motif score ≥ 13s (window length =6) were considered high confidence canidates. Fifty percent of the proteins in this file were previously known

AGPs and an additional 40% contained convincing noncontiguous proline motifs (Fig. 3 *B*). Convincing motifs were judged based on both the spacing of the proline residues and the context in which they occur. Proteins with single XP or PX dipeptides that occurred interspersed at regular intervals near window length (max) or as part of a broader repetitive peptide sequence were manually excluded. This is included some extensins that have single noncontiguous prolines that interlink longer runs of contiguous prolines (i.e., SPPPSPSPPP), which are known not to contain AG (46). We also excluded proteins that contained transmembrane domains where the putative glycosylation motifs (SPSPSPSP) occurred on the cytosolic domain of the protein, as these sequences are not exposed to the prolyl-4 hydxoxylases and glycosyltransferases in the endomembrane lumen necessary for protein glycosylation. Approximately 10% of proteins in the high confidence file were not expected to be AGPs based on these criteria (Fig. 3 *B*).

Many of the proteins we identified with convincing non-contiguous proline motifs contain additional functional domains by which they can be classified. Among this group we identified 9 new plastocyanin-like AGPs (PAG) and 40 new lipid transfer protein (LTP) AGPs, which belong to previously established AGP subclasses. However, many of the proteins we identified contain functional domains that have not been previously reported to occur in conjunction with AGP. These proteins are highly diverse; however, most can be classified based on their putative annotation as being involved with cellular processes related to cell wall modification, lipid modification, and signal transduction (Fig. 4 *A*, *B*).

AGPs related to cell wall metabolism include pectin methyl esterase inhibitors (PMEI), pectin lyases, glycosyl hydrolase (GH) family 17 proteins, and carbohydrate-binding X8 domain proteins. Other proteins including COBRA-like proteins, a transglycosylase, and some peroxidases were also identified. Roughly 23% of AGPs identified have functional annotations related to lipid metabolism. Within this group we identified new LTP AGPs, indicating that this family was previously underestimated. We also identified several lipid binding and lipid recognition proteins, as well as 17 GDSL lipase/esterase/acylhydrolase proteins. GDSL enzymes have flexible catalytic site with broad substrate specificity known to act on lipids, carbohydrates, and proteins (54).

Diverse proteins related to defense response and signal transduction were also represented in this list. Defense-related proteins include thionins, defensins, and leucine-rich repeat (LRR) proteins. Signaling-related proteins included hormone regulated GASA/SNAKIN proteins, GPI-anchored receptor-like peptides (RLP), and a relatively large number of receptor like kinase (RLKs). Identifying RLKs within the AGPrediction set was surprising because transmembrane proteins have not previously been established as being glycosylated with AG. However, the RLKs we identified have convincing non-contiguous hydroxyproline motifs that preferentially occur in the extracellular domain of the receptor (Figure 8 *C*), which make them eligible for AG *O*-glycosylation. *N*-glycosylation was previously shown to be important for the function of immune receptors AtFLS2 and AtEFR (55), which opens up the possibility that other types of glycosylation of PTM may occur on RLKs. Receptors containing AG glycosylation sites include LRR, lysine motif (LysM), malectin/lectin domain, cysteine-rich and STRUBBELIG family receptors, as well as others. Most notably this list includes ABNORMAL LEAF SHAPE 2 (ALE2; AT2G20300.1), SOMATIC-EMBRYO

RECEPTOR-LIKE KINASE 5 (SERK5; AT2G13800.1), SUPPRESSOR OF BIR1 1 (SOBIR1; AT2G31880.1), and LysM RECEPTOR-LIKE KINASE 3 (LYK3; AT1G51940.1).

Lastly, we identified several plant-specific type-I formin proteins involved in cytoskeleton organization. Type-I formins are single pass integral plasma membrane proteins that have a cytosolic FH domain well described in nucleating actin and an extracellular proline-rich domain (56). Based on our analysis we found that half of *A. thaliana* type-I formins (FH3, FH4, FH5, FH6, FH8) contain AGP glycosylation sites with some occurring as chimeric AGP-extensin motifs. The remaining formins proteins in this family either have extensin or polyproline motifs. Formins have previously been shown to associate with the plant cell wall and plant-specific mechanisms for cytoskeletal and membrane anchoring have been proposed (57, 58).

Genome-wide Identification of AGPs in *M. truncatula*

AGPs have not been annotated in plant genomes outside of *A. thaliana,* therefore the fidelity, evolution, and phylogenic relationship of AGP-encoding genes is not well understood. As a test case, we used the AGPredictor program to identify AGP-encoding genes in the *M. truncatula* (barrel medic) genome. We identified 44 classical AGPs, 35 small AG peptides, 4 lysine-rich AGPs, 33 FLAs, 44 PLAs, and 62 LTPs, with an additional 141 chimeric AGPs (Table II). Chimeric AGP distribution in *M. truncatula* based on functional annotation and AGP subclass is similar to that of *A. thaliana* (Fig. 5). Most of the same chimeric AGPs identified in *A. thaliana* were also represented in *M. truncatula*, suggesting that AGP glycosylation may be more broadly conserved between these protein families.

Analysis of AGP Fusion Proteins by SDS-PAGE and Western blot

We chose one known AGP (AtAGP9) and three new putative AGPs for biochemical characterization. Transmembrane proteins were chosen for analysis because they have not previously been reported to contain AG glycosylation and may function in transmitting information from the cell wall to inside the plant cell. We transiently expressed the extracellular domains of *AtALE2* (AT2g20300), *AtFH3* (AT4g15200), and *MtFH3* (Medtr4g081410) as C-terminal fusion proteins with a 3xMYC tag in *Nicotiana benthamiana* and analyzed the proteins by SDS-PAGE and Westernblot (Fig. 6). *AtAGP9-YFP-HA* was expressed and analyze as a control for relative comparison. The AtAGP9-YFP-HA fusion protein is predicted to be approximately 45 kDa based on amino acid composition but produce two high molecular weight (MW) bands approximately near 250 kDa and 90 kDa. A lower MW band was also present at 35 kDa, which is the predicted size of the YFP-HA tag and therefore likely a degradation product. AtFH3ecd-4xMYC was predicted to be ~24 kDa but produced a high MW band at 250 kDa followed by two additional bands between 35-45 kDa. MtFH3ecd-4xMYC was predicted to be ~ 33 kDa and produced a high MW band around 120 kDa and a secondary band near 50 kDa. AtALE2ecd-4xMYC was predicted to be ~34 kDa and produced a smearing band between 90-70 kDa. All proteins tested produced bands much larger than their predicted weight. Protein extracts were boiled in SDS reducing buffer, therefore the changes in predicted weight are not likely due to the formation of protein dimers or complexes. The changes in protein mobility were not subtle and suggest that these fusion proteins contain post translation modifications.

Discussion

Collectively, AGPs are a diverse group of cell surface glycoproteins that play critical roles in plant growth and development. However, research in this area has been stymied by the inability to identify AGP encoding proteins in plant genomes due to the broad diversity of these proteins and the intrinsically disordered nature of their amino acid sequences. Here we have developed AGPredictor as a new easy-to-use bioinformatic tool capable of rapidly identifying and enriching for AGP sequences in any plant genome. We used the *A. thaliana* genome as a benchmark to test program efficacy and show that known AGPs are successfully enriched in output files. We also identified many previously unreported proteins with convincing noncontiguous proline motifs, which suggest that AGP-type glycosylation may be more common than previously thought (Fig. 3).

The AGPredictor program functions better than previous search tools developed by Schultz (39) and Showalter (2) based on the different underlying methodologies on which these programs operate. The search methods used by both Schultz (2002) and Showalter (2010) rely on amino acid composition bias, where the total number of proline, alanine, serine, and threonine residues are calculated as a percentage relative to total protein length. While this approach has been successful in identifying some AGPs, it has two major problems. The first problem is that it doesn't take into account the sequence position and spacing of proline residues, which are necessary to direct AG glycosylation – as defined by the noncontiguous hydroxyproline hypothesis. Secondly, this approach is inherently constrained by protein size since the search criteria is a function of sequence length. This excludes many chimeric proteins that may contain AGP glycosylation subdomains. These previous methods had predicted 85 AGPs encoded in the *A. thaliana* genome, however it had been widely hypothesized that many more AGPs were likely to exist. Therefore, perhaps it is unsurprising that more refined search methods were able to uncover more AGP-encoding proteins. What was surprising was the diversity and total number of proteins recovered. Detailed characterization of the *A. thaliana* genome uncovered many novel chimeric AGP subclasses and suggested that the number of proteins potentially modified with AG is at least four times larger than previously estimated. Characterization of the *M. truncatula* genome revealed a very similar distribution of chimeric AGPs indicating that the glycosylation motifs on these proteins are not exclusive to *A. thaliana* and could be more broadly conserved throughout plants.

Recently, while in the midst of this dissertation a competing group developed a prediction program similar to ours and found many of the same chimeric AGPs as our program, including the formins and RLKs (50). Additionally, they were able to show that these proteins are conserved in 47 plant genomes. Identifying similar proteins in these analyses validates our approach and strengthens our hypothesis that AGP glycosylation motifs are more broadly distributed on a diversity of proteins. However, in detailed comparison we recovered an additional 280 and 126 proteins in the *A. thaliana* and *M. truncatula* genomes, respectively, that were not identified by Ma *et al.* 2017 (Fig. 7, 8, 9; Table I, II). Based on this comparison we believe the study by Ma *et al.* 2017 greatly underestimates the total size and diversity of AGP-encoding genes and have found a number of inconsistencies with their analysis. Examples of this can clearly be found in the *A. thaliana* genome with the X8 carbohydrate binding domain proteins and GDSL lipase proteins (Fig. 8). We identified at least five AtX8 and 17 AtGDSL proteins in the *A. thaliana* genome (Fig 8 *A, B;* Table I; Appendix I), while Ma *et al.* 2017 did not find any – yet inexplicably found at least one or two X8 and GDSL proteins in 32 of the 47 plant genomes analyzed. Another example is the RLK ALE2*,* which we identified in *A. thaliana* (AT2G20300) (Fig. 8 *C*) and *M. truncatula* (Medtr1g069340, Medtr2g039290, Medtr4g126270, Medtr8g016330), but was not reported by Ma *et al.* (2017). Furthermore we provide evidence that suggests the AtALE2 ectodomain contains post-translational modifications. These proteins clearly have noncontiguous AGP motifs and should have been identified in their analysis (Fig. 8, 9). Many other examples exist and are detailed in Table I and Appendix II. Due to these discrepancies we argue that our program operates with greater depth. Moreover, we have developed an easy-to-use graphical user interface similar to BLAST to facilitate adoption by the scientific plant biology community (Fig. 1). This enables users to search genomes or personal datasets and draw their own conclusions by accessing the primary data. This is important because Ma *et al.* (2017) has not released their program to the public and we have provided several examples where their program fails to find homologous AGPs leading to underestimation and inaccurate reporting AGP encoding gene families. This is critical when investigating AGP function where phenotypes can be masked by functional redundancy. We provide an illustrated example of in Chapter 2.

False Positives an Scope of Use

A key issue that we have not yet addressed in depth is the false discovery rate. AGPredictor is a simple customizable parsing-program that operates using arbitrary user-defined cut-offs. Therefore, the false discovery rate is a function of the parameters used. For instance, using the parameters: motif ≥ 1, window = 0, will return all proteins with at least one proline anywhere in the amino acid sequence. In this scenario AGPs will not parse from random protein sequences. Conversely, if the parameters: motif ≥ 20, window = 0 are used, very few proteins will be recovered, however all of them will be bonafied AGPs. While testing the AGPredictor program we found that the parameters: motif ≥ 13, window = 6, were sufficient for enriching for AGPs in *A. thaliana*. This takes an input list of 27,416 proteins and rapidly produces an output file of 138 proteins. Approximately 50% of these proteins are known AGPs, while an additional 40% have convincing motifs that we expect to be glycosylated. Each protein returned was manually curated and individually scrutinized. While some false positives can be identified based on the known properties of AGPs and the context in which motifs occur, ultimate determination between AGP and not-AGP is subjective and dependent on the user. We estimated a false discovery rate of at least 10% in this file by removing extensins, proteins with regularly interspersed prolines, and Ser/Thr RLKs that have (SPSPSPS) motifs near the cytoplasmic kinase domain that would not be exposed to glycosylation machinery in the endomembrane lumen. An example of a false positive with a high predictive score identified in both our analysis and that done by Ma *et al.* 2017 is the *M. truncatula* photosystem I protein (Medtr1g069235.1). PSI protein has an *N*-terminal chloroplast-targeting signal that is falsely flagged as a secretion signal by

SignalP, which is used as an initial filter. PSI protein has a strong non-contiguous proline motif (**AP**A**APAAPAAPA**ADA**APTP)** and without prior background knowledge regarding this protein, would make for a convincing AGP. This draws into question whether other false positives exist within the hypothetical and predicted proteins we identified that have no supporting information or publication history. False positives are expected in any *predictive* program, as we have just demonstrated for SignalP, and they are difficult to either prove or disprove without experimental evidence. We provided data to suggest that some of these proteins do contain post-translation modifications based on their altered mobility in SDS-PAGE reducing gels, which garner some support for our hypotheses. The majority of predicted AGPs that we identified have functions related cell wall modification, lipid transport and modification, and cell signaling, which does not seem to be coincidental as these functions are consistent with the roles and cellular location ascribed to AGPs. In our experience the best approach is to cast a wide net to examine all possibilities in order to make informed decisions. We demonstrate the utility of this in Chapter 2 by using he AGPredictor program to breakthrough functional redundancy in a previously unreported three-member AGP peptide family. Below, we discuss the group of type-I formin proteins as an example, where in identifying these proteins as putative AGPs, we provide new insights that explain previous observations made in the literature.

Formins as new chimeric AGPs

The majority of chimeric AGPs reported in our analysis can be functionally classified as related to cell wall related processes, lipid related processes, signal transduction, or other. Among these proteins we identified formins as one of the most interesting new classes of AGP. Formins are well known for their role in nucleating actin (58–61) and have a lesser-known role in microtubule organization (62–64). Type-I formins, which have a transmembrane domain and proline-rich extracellular domain, appear to be unique to plants and may be an evolutionary consequence of the plant cell wall (56, 65). Indeed, *A. thaliana* Formin 1 (AtFH1) interacts with the cell wall through its proline-rich domain which is necessary for AtFH1 dependent remodeling of the actin cytoskeleton (66). While the mechanism of this interaction is not known, we hypothesize that it is mediated through post-translation glycosylation of these proteins and there is evidence in the literature to support links between AGPs and the cytoskeleton.

Chemical disruption of AGPs with Yariv reagent, a synthetic dye that specifically binds to the β(1-3) galactan backbone of AGP glycans, disorganizes F-actin and cortical microtubule networks in tobacco BY2 cells and *A. thaliana* roots (67, 68). Similarly, monoclonal antibodies that specifically bind AGPs (JIM13, JIM14) disrupt microtubules while antibodies that recognize homogalacturonan (LM5) and xyloglucan (CCRM-1) do not, indicating that AGP glycans specifically influence cytoskeleton organization (67). Interestingly, chemical disruption of actin and microtubules with cytochalasin-D and amiprophosmethyl can also influence AGP localization at the cell surface, which had suggested that the cytoskeleton and AGPs might be intrinsically linked (68, 69). Identification of formins as transmembrane AGPs capable of cytoskeleton remodeling provides an explanation for these observations and has major implications for cell wall deposition, cellulose synthase guidance, polar secretion, and retrograde signaling between cell wall and cytoskeleton. Formins have been shown to have asymmetrical, polarized, or otherwise discrete distribution patterns in the plasma membrane specifically localizing to cell-cell junctions (70), pollen tube tips (71), and developing cell plates (72). Closer examination of the formin protein family in *A. thaliana* indicates that formins can also have extension-type glycosylation, hybrid AGP-extensin-like, and proline rich-repeats. Based on these observations, we hypothesize that the differential glycan patterning of these proteins may allow them to probe different architectural cell wall subdomains.

Bibliography

1. Seifert GJ, Roberts K (2007) The biology of arabinogalactan proteins. *Annu Rev Plant Biol* 58:137–161.
2. Showalter AM, Keppler B, Lichtenberg J, Gu D, Welch LR (2010) A bioinformatics approach to the identification, classification, and analysis of hydroxyproline-rich glycoproteins. *Plant Physiol* 153(2):485–513.
3. Basu D, et al. (2013) Functional identification of a hydroxyproline-o-galactosyltransferase specific for arabinogalactan protein biosynthesis in Arabidopsis. *J Biol Chem* 288(14):10132–10143.
4. Tryfona T, et al. (2010) Carbohydrate structural analysis of wheat flour arabinogalactan protein. *Carbohydr Res* 345(18):2648–2656.
5. Tryfona T, et al. (2012) Structural characterization of Arabidopsis leaf arabinogalactan polysaccharides. *Plant Physiol* 160(2):653–666.
6. Popper ZA, Tuohy MG (2010) Beyond the green: understanding the evolutionary puzzle of plant and algal cell walls. *Plant Physiol* 153(2):373–383.
7. Popper ZA, et al. (2011) Evolution and diversity of plant cell walls: from algae to flowering plants. *Annu Rev Plant Biol* 62:567–590.
8. Majewska-Sawka A, Nothnagel E The Multiple Roles of Arabinogalactan Proteins in Plant Development1.
9. Nguema-Ona E, Coimbra S, Vicré-Gibouin M, Mollet J-C, Driouich A (2012) Arabinogalactan proteins in root and pollen-tube cells: distribution and functional aspects. *Ann Bot* 110(2):383–404.
10. Pielach A, Leroux O, Domozych DS, Knox JP, Popper ZA (2014) Arabinogalactan protein-rich cell walls, paramural deposits and ergastic globules define the hyaline bodies of rhinanthoid Orobanchaceae haustoria. *Ann Bot* 114(6):1359–1373.
11. Ito S, Suzuki Y, Miyamoto K, Ueda J, Yamaguchi I (2005) AtFLA11, a fasciclin-like arabinogalactan-protein, specifically localized in sclerenchyma cells. *Biosci Biotechnol Biochem* 69(10):1963–1969.
12. Oxley D, Bacic A (1999) Structure of the glycosylphosphatidylinositol anchor of an arabinogalactan protein from Pyrus communis suspension-cultured cells. *Proc Natl Acad Sci USA* 96(25):14246–14251.
13. Svetek J, Yadav MP, Nothnagel EA (1999) Presence of a glycosylphosphatidylinositol lipid anchor on rose arabinogalactan proteins. *J Biol Chem* 274(21):14724–14733.
14. Tan L, et al. (2013) An Arabidopsis cell wall proteoglycan consists of pectin and arabinoxylan covalently linked to an arabinogalactan protein. *Plant Cell* 25(1):270–287.
15. Akiyama Y, Eda S, Katō K (1984) Gum arabic is a kind of arabinogalactan–protein.

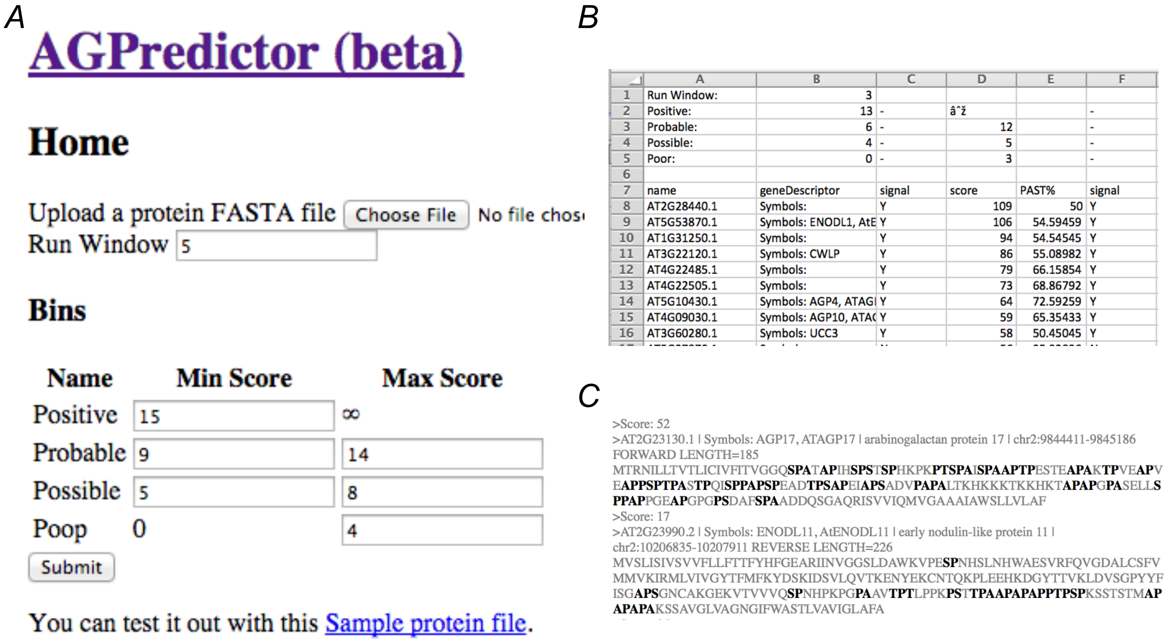
Agric Biol Chem 48(1):235–237.

16. Smallwood M, Yates EA, Willats WGT, Martin H, Knox JP (1996) Immunochemical comparison of membrane-associated and secreted arabinogalactan-proteins in rice and carrot. *Planta* 198(3):452–459.
17. Willats WG, Knox JP (1996) A role for arabinogalactan-proteins in plant cell expansion: evidence from studies on the interaction of beta-glucosyl Yariv reagent with seedlings of Arabidopsis thaliana. *Plant J* 9(6):919–925.
18. Huang Y, et al. (2016) Nanospherical arabinogalactan proteins are a key component of the high-strength adhesive secreted by English ivy. *Proc Natl Acad Sci USA* 113(23):E3193-202.
19. Serpe MD, Nothnagel EA (1994) Effects of Yariv phenylglycosides onRosa cell suspensions: Evidence for the involvement of arabinogalactan-proteins in cell proliferation. *Planta* 193(4):542–550.
20. Motose H, Sugiyama M, Fukuda H (2001) An arabinogalactan protein(s) is a key component of a fraction that mediates local intercellular communication involved in tracheary element differentiation of zinnia mesophyll cells. *Plant Cell Physiol* 42(2):129–137.
21. Motose H, Sugiyama M, Fukuda H (2004) A proteoglycan mediates inductive interaction during plant vascular development. *Nature* 429(6994):873–878.
22. van Hengel A, et al. N-Acetylglucosamine and Glucosamine-Containing Arabinogalactan Proteins Control Somatic Embryogenesis1.
23. Kreuger M, van Holst G-J (1993) Arabinogalactan proteins are essential in somatic embryogenesis of Daucus carota L. *Planta* 189(2):243–248.
24. Coimbra S, Costa M, Jones B, Mendes MA, Pereira LG (2009) Pollen grain development is compromised in Arabidopsis agp6 agp11 null mutants. *J Exp Bot* 60(11):3133–3142.
25. Demesa-Arévalo E, Vielle-Calzada J-P (2013) The classical arabinogalactan protein AGP18 mediates megaspore selection in Arabidopsis. *Plant Cell* 25(4):1274–1287.
26. Costa M, et al. (2013) Expression-based and co-localization detection of arabinogalactan protein 6 and arabinogalactan protein 11 interactors in Arabidopsis pollen and pollen tubes. *BMC Plant Biol* 13:7.
27. Lamport DTA, Kieliszewski MJ, Showalter AM (2006) Salt stress upregulates periplasmic arabinogalactan proteins: using salt stress to analyse AGP function. *New Phytol* 169(3):479–492.
28. Seifert GJ, Xue H, Acet T (2014) The Arabidopsis thaliana FASCICLIN LIKE ARABINOGALACTAN PROTEIN 4 gene acts synergistically with abscisic acid signalling to control root growth. *Ann Bot* 114(6):1125–1133.
29. Shi H (2003) The Arabidopsis SOS5 Locus Encodes a Putative Cell Surface Adhesion Protein and Is Required for Normal Cell Expansion. *THE PLANT CELL ONLINE* 15(1):19–32.
30. Nguema-Ona E, Vicré-Gibouin M, Cannesan M-A, Driouich A (2013) Arabinogalactan proteins in root-microbe interactions. *Trends Plant Sci* 18(8):440–449.
31. Park MH, Suzuki Y, Chono M, Knox JP, Yamaguchi I (2003) CsAGP1, a gibberellin-responsive gene from cucumber hypocotyls, encodes a classical arabinogalactan protein and is involved in stem elongation. *Plant Physiol* 131(3):1450–1459.
32. Wang H, et al. (2015) Antisense expression of the fasciclin-like arabinogalactan

protein FLA6 gene in Populus inhibits expression of its homologous genes and alters stem biomechanics and cell wall composition in transgenic trees. *J Exp Bot* 66(5):1291–1302.
33. Zhou Y, et al. (2009) BC10, a DUF266-containing and Golgi-located type II membrane protein, is required for cell-wall biosynthesis in rice (Oryza sativa L.). *Plant J* 57(3):446–462.
34. Tan L, et al. (2012) Arabinogalactan-proteins and the research challenges for these enigmatic plant cell surface proteoglycans. *Front Plant Sci* 3:140.
35. Gokulakannan GG, Niehaus K (2010) Characterization of the Medicago truncatula cell wall proteome in cell suspension culture upon elicitation and suppression of plant defense. *J Plant Physiol* 167(18):1533–1541.
36. Albenne C, Canut H, Jamet E (2013) Plant cell wall proteomics: the leadership of Arabidopsis thaliana. *Front Plant Sci* 4:111.
37. Kitazawa K, et al. (2013) β-galactosyl Yariv reagent binds to the β-1,3-galactan of arabinogalactan proteins. *Plant Physiol* 161(3):1117–1126.
38. Schultz CJ, Johnson KL, Currie G, Bacic A (2000) The classical arabinogalactan protein gene family of arabidopsis. *Plant Cell* 12(9):1751–1768.
39. Schultz CJ, et al. (2002) Using genomic resources to guide research directions. The arabinogalactan protein gene family as a test case. *Plant Physiol* 129(4):1448–1463.
40. Borner GHH, Lilley KS, Stevens TJ, Dupree P (2003) Identification of glycosylphosphatidylinositol-anchored proteins in Arabidopsis. A proteomic and genomic analysis. *Plant Physiol* 132(2):568–577.
41. Borner GHH, Sherrier DJ, Stevens TJ, Arkin IT, Dupree P (2002) Prediction of glycosylphosphatidylinositol-anchored proteins in Arabidopsis. A genomic analysis. *Plant Physiol* 129(2):486–499.
42. Geshi N, et al. (2013) A galactosyltransferase acting on arabinogalactan protein glycans is essential for embryo development in Arabidopsis. *Plant J* 76(1):128–137.
43. Showalter AM, Basu D (2016) Glycosylation of arabinogalactan-proteins essential for development in Arabidopsis. *Commun Integr Biol* 9(3):e1177687.
44. Kieliszewski MJ, Shpak E (2001) Synthetic genes for the elucidation of glycosylation codes for arabinogalactan-proteins and other hydroxyproline-rich glycoproteins. *Cell Mol Life Sci* 58(10):1386–1398.
45. Zhao ZD, Tan L, Showalter AM, Lamport DTA, Kieliszewski MJ (2002) Tomato LeAGP-1 arabinogalactan-protein purified from transgenic tobacco corroborates the Hyp contiguity hypothesis. *Plant J* 31(4):431–444.
46. Shpak E, Barbar E, Leykam JF, Kieliszewski MJ (2001) Contiguous hydroxyproline residues direct hydroxyproline arabinosylation in Nicotiana tabacum. *J Biol Chem* 276(14):11272–11278.
47. Estévez JM, Kieliszewski MJ, Khitrov N, Somerville C (2006) Characterization of synthetic hydroxyproline-rich proteoglycans with arabinogalactan protein and extensin motifs in Arabidopsis. *Plant Physiol* 142(2):458–470.
48. Tan L, Leykam JF, Kieliszewski MJ (2003) Glycosylation motifs that direct arabinogalactan addition to arabinogalactan-proteins. *Plant Physiol* 132(3):1362–1369.
49. Shpak E, Leykam JF, Kieliszewski MJ (1999) Synthetic genes for glycoprotein design and the elucidation of hydroxyproline-O-glycosylation codes. *Proc Natl Acad Sci USA*

96(26):14736–14741.
50. Ma Y, et al. (2017) Bioinformatics prediction and evolution analysis of arabinogalactan proteins in the plant kingdom. *Front Plant Sci* 8:66.
51. Nakagawa T, et al. (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J Biosci Bioeng* 104(1):34–41.
52. Earley KW, et al. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J* 45(4):616–629.
53. Van Haute E, et al. (1983) Intergeneric transfer and exchange recombination of restriction fragments cloned in pBR322: a novel strategy for the reversed genetics of the Ti plasmids of Agrobacterium tumefaciens. *EMBO J*.
54. Akoh CC, Lee G-C, Liaw Y-C, Huang T-H, Shaw J-F (2004) GDSL family of serine esterases/lipases. *Prog Lipid Res* 43(6):534–552.
55. Häweker H, et al. (2010) Pattern recognition receptors require N-glycosylation to mediate plant immunity. *J Biol Chem* 285(7):4629–4636.
56. van Gisbergen PAC, Bezanilla M (2013) Plant formins: membrane anchors for actin polymerization. *Trends Cell Biol* 23(5):227–233.
57. Martinière A, Gayral P, Hawes C, Runions J (2011) Building bridges: formin1 of Arabidopsis forms a connection between the cell wall and the actin cytoskeleton. *Plant J* 66(2):354–365.
58. Cheung AY, Wu H (2004) Overexpression of an Arabidopsis formin stimulates supernumerary actin cable formation from pollen tube cell membrane. *Plant Cell* 16(1):257–269.
59. Pruyne D, et al. (2002) Role of formins in actin assembly: nucleation and barbed-end association. *Science* 297(5581):612–615.
60. Vidali L, et al. (2009) Rapid formin-mediated actin-filament elongation is essential for polarized plant cell growth. *Proc Natl Acad Sci USA* 106(32):13341–13346.
61. Kovar DR (2006) Molecular details of formin-mediated actin assembly. *Curr Opin Cell Biol* 18(1):11–17.
62. Deeks MJ, et al. (2010) The plant formin AtFH4 interacts with both actin and microtubules, and contains a newly identified microtubule-binding domain. *J Cell Sci* 123(Pt 8):1209–1215.
63. Wang J, Xue X, Ren H (2012) New insights into the role of plant formins: regulating the organization of the actin and microtubule cytoskeleton. *Protoplasma*.
64. Li Y, et al. (2010) The type II Arabidopsis formin14 interacts with microtubules and microfilaments to regulate cell division. *Plant Cell* 22(8):2710–2726.
65. Cvrcková F, Novotný M, Pícková D, Zárský V (2004) Formin homology 2 domains occur in multiple contexts in angiosperms. *BMC Genomics* 5(1):44.
66. Deeks MJ, Hussey PJ, Davies B (2002) Formins: intermediates in signal-transduction cascades that affect cytoskeletal reorganization. *Trends in plant science*.
67. Nguema-Ona E, Bannigan A, Chevalier L, Baskin TI, Driouich A (2007) Disruption of arabinogalactan proteins disorganizes cortical microtubules in the root of Arabidopsis thaliana. *Plant J* 52(2):240–251.
68. Sardar HS, Yang J, Showalter AM (2006) Molecular interactions of arabinogalactan proteins with cortical microtubules and F-actin in Bright Yellow-2 tobacco cultured cells. *Plant Physiol* 142(4):1469–1479.

69. Driouich A, Baskin TI (2008) Intercourse between cell wall and cytoplasm exemplified by arabinogalactan proteins and cortical microtubules. *Am J Bot* 95(12):1491–1497.
70. Deeks MJ, et al. (2005) Arabidopsis group Ie formins localize to specific cell membrane domains, interact with actin-binding proteins and cause defects in cell expansion upon aberrant expression. *New Phytol* 168(3):529–540.
71. Cheung AY, Niroomand S, Zou Y, Wu H-M (2010) A transmembrane formin nucleates subapical actin assembly and controls tip-focused growth in pollen tubes. *Proc Natl Acad Sci USA* 107(37):16390–16395.
72. Ingouff M, et al. (2005) Plant formin AtFH5 is an evolutionarily conserved actin nucleator involved in cytokinesis. *Nat Cell Biol* 7(4):374–380.



AGPredictor (beta)

Home

Upload a protein FASTA file
Run Window 5

Bins

Name	Min Score	Max Score
Positive	15	∞
Probable	9	14
Possible	5	8
Poop	0	4

You can test it out with this Sample protein file.

	A	B	C	D	E	F
1	Run Window:	3				
2	Positive:	13	-	âˆž		-
3	Probable:	6	-	12		-
4	Possible:	4	-	5		-
5	Poor:	0	-	3		-
6						
7	name	geneDescriptor	signal	score	PAST%	signal
8	AT2G28440.1	Symbols:	Y	109	50	Y
9	AT5G53870.1	Symbols: ENODL1, AtE	Y	106	54.59459	Y
10	AT1G31250.1	Symbols:	Y	94	54.54545	Y
11	AT3G22120.1	Symbols: CWLP	Y	86	55.08982	Y
12	AT4G22485.1	Symbols:	Y	79	66.15854	Y
13	AT4G22505.1	Symbols:	Y	73	68.86792	Y
14	AT5G10430.1	Symbols: AGP4, ATAGI	Y	64	72.59259	Y
15	AT4G09030.1	Symbols: AGP10, ATA(	Y	59	65.35433	Y
16	AT3G60280.1	Symbols: UCC3	Y	58	50.45045	Y

```
>Score: 52
>AT2G23130.1 | Symbols: AGP17, ATAGP17 | arabinogalactan protein 17 | chr2:9844411-9845186
FORWARD LENGTH=185
MTRNILLTVTLICIVFITVGGQSPATAPIHSPSTSPHKPKPTSPAISPAAPTPESTEAPAKTPVEAPV
EAPPSPTPASTPQISPPAPSPEADTPSAPEIAPSADVPAPALTKHKKKTKKHKTAPAPGPASELLS
PPAPPGEAPGPGPSDAFSPAADDQSGAQRISVVIQMVGAAAIAWSLLVLAF
>Score: 17
>AT2G23990.2 | Symbols: ENODL11, AtENODL11 | early nodulin-like protein 11 |
chr2:10206835-10207911 REVERSE LENGTH=226
MVSLISIVSVVFLLFTTFYHFGEARIINVGGSLDAWKVPESPNHSLNHWAESVRFQVGDALCSFV
MMVKIRMLVIVGYTFMFKYDSKIDSVLQVTKENYEKCNTQKPLEEHKDGYTTVKLDVSGPYYF
ISGAPSGNCAKGEKVTVVVQSPNHPKPGPAAVTPTLPPKPSTTPAAPAPAPPTPSPKSSTSTMAP
APAPAKSSAVGLVAGNGIFWASTLVAVIGLAFA
```

Figure 1. AGPredictor web application graphical user interface and output function. (*A*) AGPredictor GUI; Up load FASTA protein input file upload. Run window creates an additive score by joining interspersed motifs separated ≤ *n* amino acids. Bins sets cut off values for parsing output sequence files based on AGP motif score. (*B*) Data output summary csv file. (*C*) Parsed output files with highlighted AGP motifs.

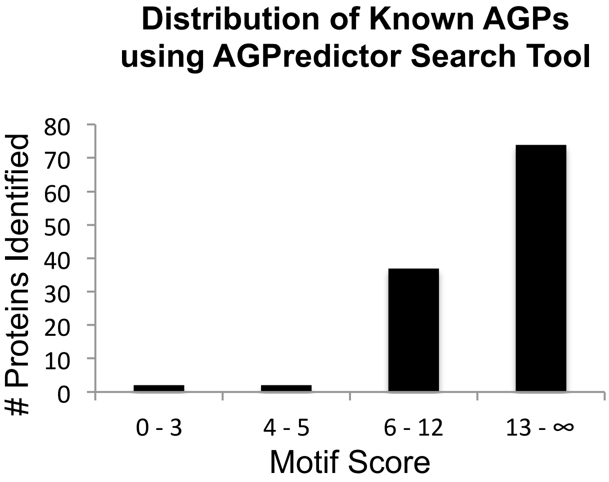


Figure 2. Parsed file distribution of known *A. thaliana* AGPs using AGPredictor. A list containing 123 AGPs from Showalter *et al.* 2010, Borner et al 2003, and LTP et al was used as a training set to develop the AGPredictor program. Using a run window length *n*=5 allows significant enrichment of AGPs with motif score ≥ 6. Two Fasciclin proteins from the training set were not enriched. We determined that these proteins are not AGP based on their absence of proline motifs and conclude that they were erroneously missclassified in previous studies based on their homology to true FLA-AGPs through their shared fasciclin domain.

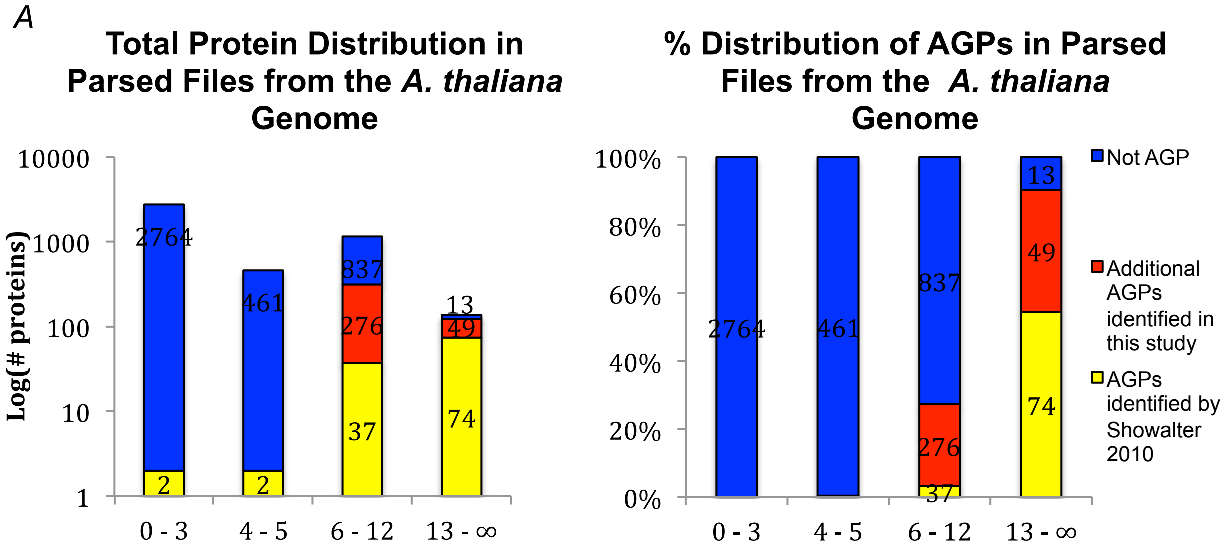


Figure 3. Parsed distribution of AGPs from the *A. thaliana* genome. (*A*) Total proteins in parsed files. (*B*) Data represented as % distribution. Files containing proteins parsed at AGP motif score ≥ 13 are almost exclusively AGP.

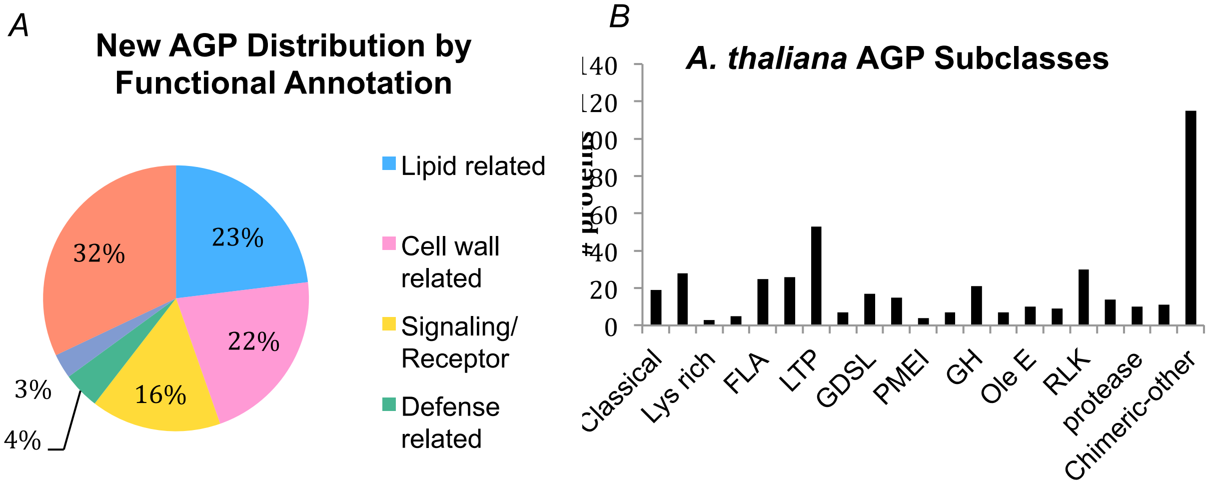


Figure 4. Distribution of *A. thaliana* AGPs based on chimeric subclass and cellular process. (*A*) Distribution of new chimeric AGPs identified in this study based on functional annotation and putative cellular process. (*B*) All AGP subclasses identified in this study.

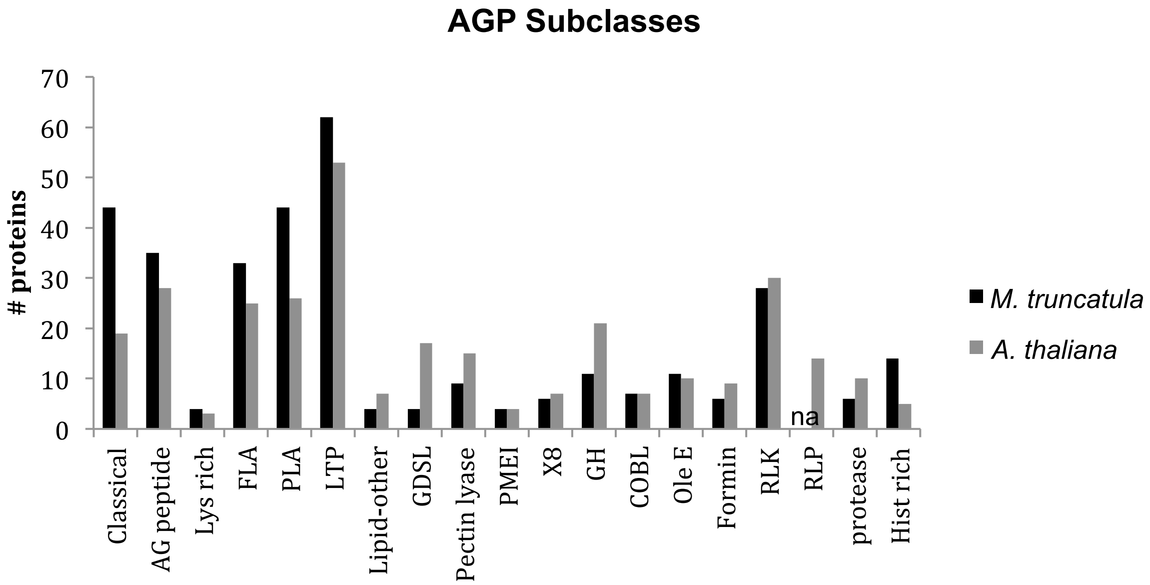


Figure 5. AGP subfamily distribution in *M. truncatula* and *A. thaliana.*

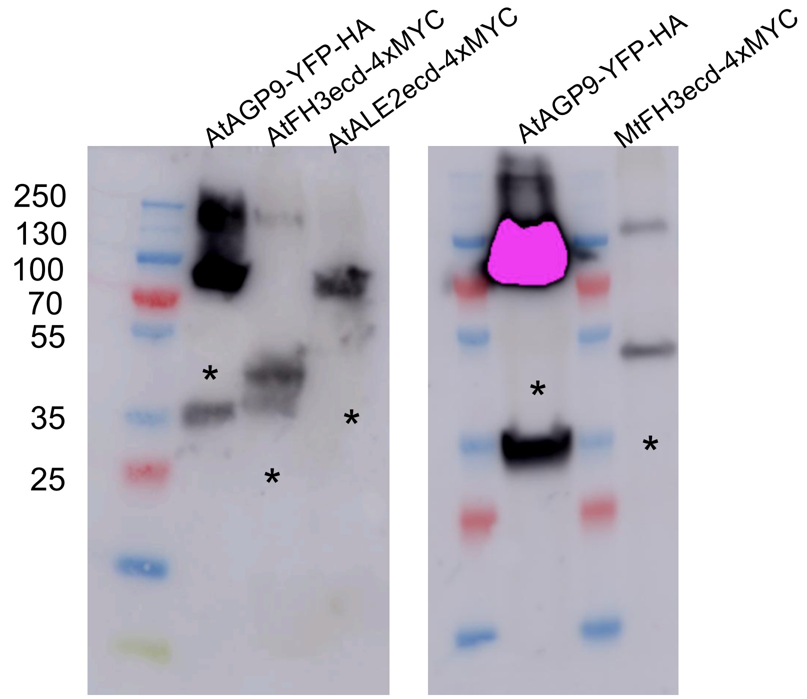


Fig 6. SDS-PAGE and Western blot analysis of putative AGPs. Asterisks indicated the predicted hypothetical size of the fusion protein based on primary amino acid sequence. AtAGP9-YFP-HA, ~45 kDa; AtFH3-4xMYC, ~24 kDa;

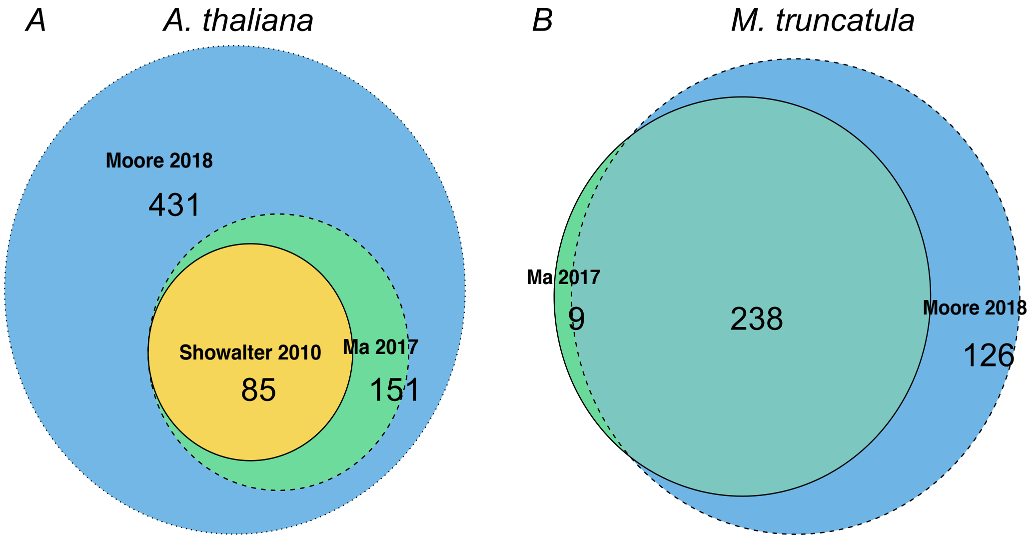


Figure 7. Comparison of AGPs reported by different prediction methods in *A. thaliana* (*A*) and *M. truncatula* (*B*). Proteins identified by Ma *et al.* in *M. truncatula* that we do not predict to be AGP are listed in appendix IV. Figures are drawn to scale.

A

\>AT1G20130.1 | Symbols: | GDSL-like Lipase/Acylhydrolase superfamily protein

MKRSSLVDSCSYSRIFRSIFCLLSFCIFFLTTTNAQVMHRRLWPWPLWPRPYPQPWPMNPP**TP**D**PSP**K
PVAPPG**PS**SK**PVA**PPG**PSP**C**PS**PPPKPQPKPPP**APSPSP**C**PS**PPPKPQPK**PV**PPPACPPTPPKPQPK**P**
APPPEPK**PA**PPP**AP**K**PVP**C**PS**PPKPP**APTP**K**PV**PPHGPPPK**PAPAPTPAPSP**K**PAPS**PPKPENKTI**PA**
VFFFGDSVFDTGNNNNLETKIKSNYRPYGMDFKFRVATGRFSNGMVASDYLAKYMGVKEIV**PA**YLDPK
IQPNDLLTGVSFASGGAGYN**PT**TSEAANAIPMLDQLTYFQDYIEKVNRLVRQEKSQYKLAGLEKTNQL
ISKGVAIVVGGSNDLIITYFGSGAQRLKNDIDSYTTIIADSAASFVLQLYGYGARRIGVIGTPPLGC**V**
PSQRLKKKKICNEELNYASQLFNSKLLLILGQLSKTLPNSTFVYMDIYTIISQMLE**TPA**AYGFEETKK
PCCKTGLLSAGALCKKSTSKICPNTSSYLFWDGVH**PT**QRAYKTINKVLIKEYLHVLSK

B

\>AT1G09460.1 | Symbols: | Carbohydrate-binding X8 domain superfamily protein

MAKQSYFFVFTFLFLSLLSYCSSTTTHHDVLNPPTVF**PT**N**PT**T**TPT**ATFPPVTI**TPT**N**PA**TT**VP**IVPP
VTTIPPPTLTPPPVITIPPPTLTPPVTN**PV**TN**PV**TQYPPTQ**PS**GT**VPVPVPV**VAPPVVSN**SPS**VSGQS
WCVAKPGASQVSLQQALDYACGIADCSQLQQGGNCY**SP**ISLQSHASFAFNSYYQKN**PSP**QSCDFGGAA
SLVNTN**PS**TGSCIYQTGSSTS**TP**MTAGTT**TPTPS**TQTVNQPPVTS**TP**II**PT**GGGIIGVGTPPAIFN**PA**
N**PT**SNTLNN**PS**SGGLAGGYGFDG**SP**NENN**PT**SSDSTHLKIHFGHAMVATLILHAVLFH

C

\>AT2G20300.1 | Symbols: ALE2 | Protein kinase superfamily protein

MRNFAMLLLLILLLHSLASFPICFARLFPMSLPFTRSKAHQMHFFHPYLN**PS**V**APTPSPA**F**SP**N**PS**RI
PPLRHKGHHRHRRWHLRRNATAV**SPS**SHDCQQTCVEPLTS**TP**FG**SP**CGCVFPMKVQLLLSV**AP**FSIF**P**
VTNELEIEVAAGTYLEQSQVKIMGASADSENQGKTVVDINLVPLGEKFDNTTATLIYQRFRHKKVPLN
ETVFGDYEVTHISYPGI**PS**S**SP**NGDVTGD**AP**GGLPIPINATTFANKSQGIGFRT***IAIIALSGFVLILV***
LVGAISIIVKWKKIGKSSNAVGPALAPSINKRPGAGSMFSSSARSSGSDSLMSSMATCALSVKTFTLS
ELEKATDRFSAKRVLGEGGFGRVYQGSMEDGTEVAVKLLTRDNQNRDREFIAEVEMLSRLHHRNLVKL
IGICIEGRTRCLIYELVHNGSVESHLHEGTLDWDARLKIALGAARGLAYLHEDSNPRVIHRDFKASNV
LLEDDFTPKVSDFGLAREATEGSQHISTRVMGTFGYVAPEYAMTGHLLVKSDVYSYGVVLLELLTGRR
PVDMSQPSGEENLVTWARPLLANREGLEQLVDPALAGTYNFDDMAKVAAIASMCVHQEVSHRPFMGEV
VQALKLIYNDADETCGDYCSQKDSSVPDSADFKGDLAPSDSSWWNLTPRLRYGQASSFITMDYSSGPL
EDMENRPHSASSIPRVGGLILPNRSGPLRPMRSRRNFFRLRGSMSEHGGPSSSRHLWSGNGDWL

Fig 8. Examples of proteins identified in this study but not reported by Ma *et al.* 2017. Bold red characters indicate noncontiguous proline motifs eligible for AGP glycosylation. Underlined sequences indicate contiguous proline motifs eligible for extensin-type glycosylation. Bold blue characters indicate transmembrane domains.

A

>AT2G42840.1 | Symbols: PDF1 | protodermal factor 1 | chr2:17826327-17827426 REVERSE LENGTH=306
MRGMVSFAVWALFAALLSQQLFASVASVRFEDAKTYYLSPPSGSHGTPPSHTPPSSNCGSPPYD**PSPST**
PSH**PS**PPSH**TPTPSTPS**H**TPTP**H**TPS**H**TPTP**HTPPCNCGSPPSH**PSTPS**H**PSTPS**H**PTPS**HPPSGGYYS
SPPPR**TPV**VVTPP**SP**IVDPG**TP**IIGGSPP**TP**IIDPG**TP**G**TP**FI**PAP**FPPITGTCDYWRNH**PT**LIWGLLG
WWGTVGGAFGTVSIPSSIPGFDPHMNLLQALSNTRSDPIGALYREGTASWLNSMVNHKFPFTTPQVRDH
FVAGLSSNKAATKQAHTFKLANEGRLKPRV

B

>AT5G14920.1 | Symbols: | Gibberellin-regulated family protein |
chr5:4826598-4827761 FORWARD LENGTH=275
MALSLLSVFIFFHVFTNVVFAASNEESNALVSL**PTPT**L**PSPSPA**TKPP**SPA**LKPP**TPS**YKPPT**LPTTP**IK
PPTTKPPVKPPTI**PVTPV**KPPVSTPPIKLPPVQPPTYKPPT**PT**VKPPSVQPPTYKPP**TPT**VKPPTT**SPV**K
PPTTPPVQSPPVQPPTYKPPT**SPV**KPPTTTPPVKPPTTTPPVQPPTYNPPT**TPV**KPPTAPPVKPPTPPPV
RTRIDC**VP**LCGTRCGQHSRKNVCMRACVTCCYRCKCVPPGTYGNKEKCGSCYANMKTRGGKSKCP

C

>AT4G20420.1 | Symbols: | Tapetum specific protein TAP35/TAP44 |
chr4:11017041-11017469 FORWARD LENGTH=142
MSISRVYLCLIFLTFVS**SP**LVLCSR**SP**KLAAASAAIGKKHGKEHVH**SPA**MLFSEFPKVDSSSSMTHIDE**P**
ATKSAIAGFFRYRLPFQGWPFHKY**AP**FPMG**TPT**N**PSVPV**TS**TPS**SGAAAAEEEETEK**VPSAPS**KGNRDGG
NA

Fig 9. Additional examples of proteins identified in this study but not reported by Ma *et al.* 2017. Bold red characters indicate noncontiguous proline motifs eligible for AGP glycosylation. Underlined sequences indicate contiguous proline motifs eligible for extensin-type glycosylation.

Table I. AGPs identified in the *A. thaliana* genome.

AGI	Showalter 2010	Ma 2017	Moore 2018	AA length	PAST %	AGP class	Gene name	Annotation
AT1G24520.1	Y	Y	Y	119	47	Classical	AGP50C	Arabinogalactan protein
AT3G27416.1	-	Y	Y	171	68	Classical		
AT1G31250.1	Y	Y	Y	165	55	Classical	AGP51C	Arabinogalactan protein
AT1G35230.1	Y	Y	Y	133	64	Classical	AGP5C	Arabinogalactan protein
AT2G14890.1	Y	Y	Y	191	72	Classical	AGP9C	Arabinogalactan protein
AT2G22470.1	Y	Y	Y	131	64	Classical	AGP2C	Arabinogalactan protein
AT2G28440.1	Y	Y	Y	268	50	Classical	AGP54C	Arabinogalactan protein
AT2G47930.1	Y	Y	Y	136	54	Classical	AGP26C	Arabinogalactan protein
AT3G01700.1	Y	Y	Y	136	62	Classical	AGP11C	Arabinogalactan protein
AT3G06360.1	Y	Y	Y	125	54	Classical	AGP27C	Arabinogalactan protein
AT3G45230.1	Y	Y	Y	175	43	Classical	AGP57C	Arabinogalactan protein
AT4G09030.1	Y	Y	Y	127	66	Classical	AGP10C	Arabinogalactan protein
AT4G16980.1	Y	Y	Y	164	55	Classical	AGP58C	Arabinogalactan protein
AT4G40090.1	Y	Y	Y	139	62	Classical	AGP3C	Arabinogalactan protein
AT5G10430.1	Y	Y	Y	135	73	Classical	AGP4C	Arabinogalactan protein
AT5G18690.1	Y	Y	Y	116	50	Classical	AGP25C	Arabinogalactan protein
AT5G64310.1	Y	Y	Y	131	60	Classical	AGP1C	Arabinogalactan protein
AT5G65390.1	Y	Y	Y	130	67	Classical	AGP7C	Arabinogalactan protein
AT5G14380.1	Y	Y	Y	150	65	Classical	AGP6C	Arabinogalactan protein
AT1G76955.1	-	-	Y	78	35	AG peptide		NA
AT1G12665.1	-	-	Y	89	43	AG peptide		Encodes a Plant thionin family protein
AT2G28405.1	-	-	Y	83	19	AG peptide	LCR32	Cysteine-rich 32
AT1G51913.1	-	-	Y	77	32	AG peptide		NA
AT5G10946.1	-	-	Y	55	30	AG peptide		NA
AT3G47295.1	-	-	Y	71	41	AG peptide		NA
AT5G04030.1	-	-	Y	76	29	AG peptide		NA
AT1G51915.1	Y	Y	Y	67	36	AG peptide	AGP24P	cryptdin protein-related
AT5G10946.1	-	-	Y	55	30	AG peptide		NA
AT3G13275.1	-	-	Y	69	12	AG peptide		NA
AT1G77093.1	-	-	Y	78	27	AG peptide		Defensin-like (DEFL) family protein

AT1G51920.1	-	-	Y	78	38	AG peptide		NA
AT1G51915.1	Y	Y	Y	67	36	AG peptide	AGP42P	AGP peptide
AT2G41905.1	Y	Y	Y	61	45	AG peptide	AGP43P	AGP peptide
AT2G46330.1	Y	Y	Y	73	42	AG peptide	AGP16P	AGP peptide
AT3G01730.1	Y	Y	Y	87	46	AG peptide	AGP44P	AGP peptide
AT3G13520.1	Y	Y	Y	60	44	AG peptide	AGP12P	AGP peptide
AT3G20865.1	Y	Y	Y	62	49	AG peptide	AGP40P	AGP peptide
AT3G57690.1	Y	Y	Y	61	46	AG peptide	AGP23P	AGP peptide
AT3G61640.1	Y	Y	Y	74	44	AG peptide	AGP20P	AGP peptide
AT4G26320.1	Y	Y	Y	59	48	AG peptide	AGP13P	AGP peptide
AT5G11740.1	Y	Y	Y	61	51	AG peptide	AGP15P	AGP peptide
AT5G12880.1	Y	Y	Y	73	44	AG peptide	AGP45P	AGP peptide
AT5G24105.1	Y	Y	Y	63	39	AG peptide	AGP41P	AGP peptide
AT5G53250.1	Y	Y	Y	63	39	AG peptide	AGP22P	AGP peptide
AT5G56540.1	Y	Y	Y	60	42	AG peptide	AGP14P	AGP peptide
AT1G55330.1	Y	Y	Y	58	47	AG peptide	AGP21P	AGP peptide
AT5G40730.1	Y	Y	Y	69	41	AG peptide/His-rich	AGP24P	AGP peptide
AT1G68725.1	Y	Y	Y	248	69	Lys rich	AGP19K	Lysine-rich AGP
AT2G23130.1	Y	Y	Y	185	57	Lys rich	AGP17K	Lysine-rich AGP
AT4G37450.1	Y	Y	Y	209	62	Lys rich	AGP18K	Lysine-rich AGP
AT1G03870.1	Y	Y	Y	247	32	FLA	FLA9	Fasciclin-like AGP
AT1G15190.1	Y	Y	Y	248	34	FLA	FLA19	Fasciclin-like AGP
AT2G04780.1	Y	Y	Y	254	40	FLA	FLA7	Fasciclin-like AGP
AT2G20520.1	Y	Y	Y	247	35	FLA	FLA6	Fasciclin-like AGP
AT2G24450.1	Y	Y	Y	280	39	FLA	FLA3	Fasciclin-like AGP
AT2G35860.1	Y	Y	Y	445	29	FLA	FLA16	Fasciclin-like AGP
AT2G45470.1	Y	Y	Y	420	44	FLA	FLA8	Fasciclin-like AGP
AT3G11700.1	Y	Y	Y	462	26	FLA	FLA18	Fasciclin-like AGP
AT3G12660.1	Y	Y	Y	255	36	FLA	FLA14	Fasciclin-like AGP
AT3G46550.1	Y	Y	Y	420	38	FLA	FLA4	Fasciclin-like AGP
AT3G52370.1	Y	Y	Y	436	29	FLA	FLA15	Fasciclin-like AGP
AT3G60900.1	Y	Y	Y	422	42	FLA	FLA10	Fasciclin-like AGP
AT4G12730.1	Y	Y	Y	403	32	FLA	FLA2	Fasciclin-like AGP

AT4G31370.1	Y	Y	Y	278	38	FLA	FLA5	Fasciclin-like AGP
AT5G06390.1	Y	Y	Y	458	27	FLA	FLA17	Fasciclin-like AGP
AT5G06920.1	Y	-	Y	353	33	FLA	FLA21	Fasciclin-like AGP
AT5G44130.1	Y	Y	Y	247	31	FLA	FLA13	Fasciclin-like AGP
AT5G55730.1	Y	Y	Y	424	34	FLA	FLA1	Fasciclin-like AGP
AT5G60490.1	Y	Y	Y	249	36	FLA	FLA12	Fasciclin-like AGP
AT5G03170.1	Y	Y	Y	246	37	FLA	FLA11	Fasciclin-like AGP
AT5G16920.1	-	Y	Y	256	33	FLA		Fasclin-like protein
AT3G55820.1	-	-	Y	204	23	FLA		Fasclin-like protein
AT4G12950.1	-	-	Y	176	30	FLA		Fasclin-like protein
AT4G29980.1	-	-	Y	169	34	FLA		Fasclin-like protein
AT1G30800.1	-	-	Y	239	40	FLA		Fasclin-like protein
AT1G45063.1	-	Y	Y	369	24	PAG		Copper ion binding;electron carriers
AT1G08500.1	-	Y	Y	228	33	PAG	ENODL18	Early nodulin-like protein 18
AT5G57920.1	-	-	Y	182	23	PAG	ENODL10	Early nodulin-like protein 10
AT1G22480.1	-	Y	Y	174	42	PAG		Cupredoxin superfamily protein
AT5G07475.1	-	Y	Y	192	29	PAG		Cupredoxin superfamily protein
AT5G20230.1	-	Y	Y	196	40	PAG	SAG14	Blue-copper-binding protein
AT1G79800.1	-	Y	Y	192	32	PAG	ENODL7	Early nodulin-like protein 7
AT1G64640.1	-	Y	Y	191	36	PAG	ENODL8	Early nodulin-like protein 8
AT2G23990.2	Y	Y	Y	226	37	PAG	PAG1	Early nodulin-like protein 11
AT2G25060.1	Y	Y	Y	182	32	PAG	PAG2	Early nodulin-like protein 14
AT2G26720.1	Y	Y	Y	206	31	PAG	PAG3	Cupredoxin superfamily protein
AT2G31050.1	Y	Y	Y	200	32	PAG	PAG4	Cupredoxin superfamily protein
AT2G32300.1	Y	Y	Y	261	47	PAG	PAG5	UCC1 uclacyanin 1
AT2G44790.1	Y	Y	Y	202	43	PAG	PAG6	UCC2 uclacyanin 2
AT3G20570.1	Y	Y	Y	203	39	PAG	PAG7	Early nodulin-like protein 9
AT3G60270.1	Y	Y	Y	187	39	PAG	PAG8	Cupredoxin superfamily protein
AT3G60280.1	Y	Y	Y	222	51	PAG	PAG9	UCC3 uclacyanin 3
AT4G27520	Y	Y	Y	349	53	PAG	PAG10/ENODL	Early nodulin-like protein

.1							2	2
AT4G28365 .1	Y	Y	Y	199	32	PAG	PAG11/ENODL 3	Early nodulin-like protein 3
AT4G30590 .1	Y	Y	Y	190	32	PAG	PAG12/ENODL 12	Early nodulin-like protein 12
AT4G32490 .1	Y	Y	Y	221	34	PAG	PAG14/ENODL 4	Early nodulin-like protein 4
AT5G26330 .1	Y	Y	Y	187	41	PAG	PAG16	Cupredoxin superfamily protein
AT5G53870 .1	Y	Y	Y	370	55	PAG	PAG17/ENODL 1	Early nodulin-like protein 1
AT1G21090 .1	-	Y	Y	242	39	PAG		Cupredoxin superfamily protein
AT4G31840 .1	Y	Y	Y	177	32	PAG	PAG13/ENODL 15	Early nodulin-like protein 15
AT5G25090 .1	Y	Y	Y	186	33	PAG	PAG15/ENODL 13	Early nodulin-like protein 13
AT4G12500 .1	-	-	Y	177	43	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT1G18280 .1	-	Y	Y	180	47	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT2G48140 .1	-	Y	Y	200	43	LTP	EDA4	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT4G12490 .1	-	-	Y	182	42	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT1G03103 .1	-	Y	Y	171	43	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT1G05450 .2	-	Y	Y	205	48	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT2G45180 .1	-	-	Y	134	36	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein

AT1G27950.1	-	Y	Y	193	37	LTP	LTPG1	Glycosylphosphatidylinositol-anchored lipid protein transfer 1
AT4G12480.1	-	-	Y	168	40	LTP	pEARLI 1	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT4G22460.1	-	-	Y	133	34	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT3G57310.1	-	-	Y	103	29	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT3G18280.1	-	-	Y	96	41	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT1G32280.1	-	-	Y	112	24	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT4G14805.1	-	Y	Y	219	44	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT4G12360.1	-	Y	Y	161	43	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT1G62510.1	-	-	Y	149	36	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT3G22620.1	-	Y	Y	203	47	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT5G56480.1	-	-	Y	113	22	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein

AT3G52130.1	-	-	Y	125	35	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT1G43665.1	-	-	Y	123	31	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT1G73560.1	-	Y	Y	147	36	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT2G37870.1	-	-	Y	115	32	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT5G46890.1	-	-	Y	127	32	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT5G46900.1	-	-	Y	127	33	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT1G73780.1	-	-	Y	98	28	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT4G12470.1	-	-	Y	161	38	LTP	AZI1	Azelaic acid induced 1
AT1G73890.1	-	Y	Y	193	45	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT5G48490.1	-	-	Y	101	32	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT1G62790.1	-	Y	Y	150	38	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein

AT1G48750 .1	-	-	Y	94	41	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT5G48485 .1	-	-	Y	102	35	LTP	DIR1	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT3G59455 .1	-	-	Y	96	23	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT5G38160 .1	-	-	Y	103	35	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT4G22490 .1	-	-	Y	120	31	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT1G12100 .1	-	-	Y	132	31	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT5G38197 .1	-	-	Y	96	28	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT5G38195 .1	-	-	Y	95	29	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT5G55450 .1	-	-	Y	104	34	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT4G22666 .1	-	-	Y	160	36	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein

AT3G51590.1	-	-	Y	119	37	LTP	LTP12	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT5G38170.1	-	-	Y	103	34	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT1G36150.1	Y	Y	Y	256	55	LTP	AGP29I	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily
AT2G13820.1	-	Y	Y	169	45	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily
AT2G27130.1	-	Y	Y	176	36	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily
AT2G44290.1	-	-	Y	205	34	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily
AT2G44300.1	-	Y	Y	204	31	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily
AT2G48130.1	-	Y	Y	183	45	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily
AT3G22600.1	-	Y	Y	170	36	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily
AT3G43720.1	-	Y	Y	193	37	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily

AT4G08670 .1	-	Y	Y	208	49	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily
AT4G14815 .1	-	-	Y	156	33	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily
AT5G09370 .1	-	Y	Y	158	43	LTP		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily
AT5G64080 .1	-	Y	Y	182	48	LTP	XYP1	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily
AT5G23820 .1	-	-	Y	164	28	Lipid-related		MD-2-related lipid recognition domain-containing protein
AT5G23840 .1	-	-	Y	167	30	Lipid-related		MD-2-related lipid recognition domain-containing protein
AT5G48605 .1	-	-	Y	91	25	Lipid-related		Putative membrane lipoprotein
AT3G44100 .1	-	-	Y	152	34	Lipid-related		MD-2-related lipid recognition domain-containing protein
AT5G17340 .1	-	-	Y	160	35	Lipid-related		Putative membrane lipoprotein
AT2G26370 .1	-	-	Y	173	30	Lipid-related		MD-2-related lipid recognition domain-containing protein
AT2G39560 .1	-	-	Y	233	35	Lipid-related		Putative membrane lipoprotein
AT1G20130 .1	-	-	Y	534	40	GDSL		GDSL-like Lipase/Acylhydrolase superfamily protein
AT1G20120 .1	-	-	Y	402	29	GDSL		GDSL-like Lipase/Acylhydrolase superfamily protein
AT1G74460 .1	-	-	Y	366	23	GDSL		GDSL-like Lipase/Acylhydrolase superfamily protein

AT2G42990.1	-	-	Y	350	26	GDSL		GDSL-like Lipase/Acylhydrolase superfamily protein
AT1G20132.1	-	-	Y	383	26	GDSL		GDSL-like Lipase/Acylhydrolase superfamily protein
AT1G71120.1	-	-	Y	362	26	GDSL	GLIP6	GDSL-motif lipase/hydrolase 6
AT3G48610.1	-	-	Y	520	30	GDSL		GDSL-like Lipase/Acylhydrolase superfamily protein
AT4G16230.1	-	-	Y	251	27	GDSL		GDSL-like Lipase/Acylhydrolase superfamily protein
AT5G63170.1	-	-	Y	338	29	GDSL		GDSL-like Lipase/Acylhydrolase superfamily protein
AT3G26430.1	-	-	Y	380	27	GDSL		GDSL-like Lipase/Acylhydrolase superfamily protein
AT2G26870.1	-	-	Y	514	28	GDSL	NPC2	Non-specific phospholipase C2
AT1G54790.2	-	-	Y	408	25	GDSL		GDSL-like Lipase/Acylhydrolase superfamily protein
AT3G26820.1	-	-	Y	634	23	GDSL		Esterase/lipase/thioesterase family protein
AT1G07230.1	-	-	Y	533	26	PLC	NPC1	Non-specific phospholipase C1
AT1G54010.1	-	-	Y	386	27	GDSL		GDSL-like Lipase/Acylhydrolase superfamily protein
AT1G75890.1	-	-	Y	379	25	GDSL		GDSL-like Lipase/Acylhydrolase superfamily protein
AT1G23500.1	-	-	Y	345	26	GDSL		GDSL-like Lipase/Acylhydrolase superfamily protein
AT3G26610.1	-		Y	470	25	Pectin lyase		Pectin lyase-like superfamily protein
AT1G02460.1	-	Y	Y	491	29	Pectin lyase		Pectin lyase-like superfamily protein

AT4G20040.1	-	-	Y	483	25	Pectin lyase		Pectin lyase-like superfamily protein
AT1G56710.1	-	-	Y	434	29	Pectin lyase		Pectin lyase-like superfamily protein
AT4G01890.1	-	-	Y	468	29	Pectin lyase		Pectin lyase-like superfamily protein
AT3G54920.1	-	-	Y	501	29	Pectin lyase	PMR6	Pectin lyase-like superfamily protein
AT5G48140.1	-	-	Y	395	25	Pectin lyase		Pectin lyase-like superfamily protein
AT3G07830.1	-	-	Y	397	25	Pectin lyase		Pectin lyase-like superfamily protein
AT4G32380.1	-	-	Y	354	31	Pectin lyase		Pectin lyase-like superfamily protein
AT5G04310.1	-	-	Y	518	27	Pectin lyase		Pectin lyase-like superfamily protein
AT3G16850.1	-	-	Y	455	29	Pectin lyase		Pectin lyase-like superfamily protein
AT2G36700.1	-	-	Y	333	25	Pectin lyase		Pectin lyase-like superfamily protein
AT5G47500.1	-	-	Y	362	28	Pectin lyase		Pectin lyase-like superfamily protein
AT1G69940.1	-	-	Y	361	28	Pectin lyase	PPME1	Pectin lyase-like superfamily protein
AT5G07410.1	-	-	Y	361	28	Pectin lyase		Pectin lyase-like superfamily protein
AT3G36659.1	-	Y	Y	264	45	PMEI		Plant invertase/pectin methylesterase inhibitor superfamily protein
AT3G47400.1	-	Y	Y	594	31	PMEI		Plant invertase/pectin methylesterase inhibitor superfamily
AT1G02550.1	-	Y	Y	242	34	PMEI		Plant invertase/pectin methylesterase inhibitor superfamily protein
AT5G46940.1	-	-	Y	176	32	PMEI		Plant invertase/pectin methylesterase inhibitor superfamily protein
AT1G09460.1	-	-	Y	330	48	X8		Carbohydrate-binding X8 domain superfamily protein
AT1G13830.1	-	-	Y	197	43	X8		Carbohydrate-binding X8 domain superfamily protein

AT2G03505.1	-	-	Y	168	37	X8		Carbohydrate-binding X8 domain superfamily protein
AT2G30933.1	-	-	Y	227	41	X8		Carbohydrate-binding X8 domain superfamily protein
AT1G79480.1	-	-	Y	397	49	X8		Carbohydrate-binding X8 domain superfamily protein
AT1G26450.1	-	-	Y	197	46	X8		Carbohydrate-binding X8 domain superfamily protein
AT1G29380.1	-	-	Y	315	41	X8		Carbohydrate-binding X8 domain superfamily protein
AT3G55260.1	-	-	Y	541	27	GH	HEXO1	beta-hexosaminidase 1
AT1G02360.1	-	-	Y	272	30	GH		Chitinase family protein
AT4G19820.1	-	-	Y	366	32	GH		Glycosyl hydrolase family protein with chitinase insertion domain
AT5G01930.1	-	-	Y	448	26	GH	MAN6	Glycosyl hydrolase superfamily protein
AT1G66250.1	-	-	Y	505	33	GH17		O-Glycosyl hydrolases family 17 protein
AT4G29360.1	-	Y	Y	534	34	GH17		O-Glycosyl hydrolases family 17 protein
AT2G16230.1	-	-	Y	503	31	GH17		O-Glycosyl hydrolases family 17 protein
AT5G67460.1	-	-	Y	380	35	GH17		O-Glycosyl hydrolases family 17 protein
AT3G55430.1	-	-	Y	449	30	GH17		O-Glycosyl hydrolases family 17 protein
AT3G23770.1	-	-	Y	476	31	GH17		O-Glycosyl hydrolases family 17 protein
AT2G19440.1	-	-	Y	478	25	GH17		O-Glycosyl hydrolases family 17 protein
AT4G14080.1	-	-	Y	478	30	GH17		O-Glycosyl hydrolases family 17 protein
AT4G26830.1	-	-	Y	455	31	GH17		O-Glycosyl hydrolases family 17 protein
AT1G11820.2	-	-	Y	511	33	GH17		O-Glycosyl hydrolases family 17 protein
AT2G01630.1	-	-	Y	501	34	GH17		O-Glycosyl hydrolases family 17 protein
AT2G39640.1	-	-	Y	549	30	GH17		glycosyl hydrolase family 17 protein |

AT4G17180 .1	-	-	Y	475	25	GH17		O-Glycosyl hydrolases family 17 protein
AT3G13560 .1	-	-	Y	505	34	GH17		O-Glycosyl hydrolases family 17 protein
AT3G07320 .1	-	-	Y	460	29	GH17		O-Glycosyl hydrolases family 17 protein
AT4G11050 .1	-	Y	Y	626	29	GH17	GH9C3	glycosyl hydrolase 9C3
AT5G42720 .1	-	-	Y	438	35	GH17		O-Glycosyl hydrolases family 17 protein
AT1G09790 .1	-	-	Y	454	26	COBL	COBL6	COBRA-like protein 6 precursor
AT5G60920 .1	-	-	Y	456	31	COBL		COBRA-like extracellular glycosyl-phosphatidyl inositol-anchored protein family
AT5G60950 .1	-	-	Y	204	31	COBL	COBL5	COBRA-like protein 5 precursor
AT3G29810 .1	-	-	Y	441	29	COBL	COBL2	COBRA-like protein 2 precursor
AT3G16860 .1	-	-	Y	653	31	COBL	COBL8	COBRA-like protein 8 precursor
AT4G16120 .1	-	-	Y	661	31	COBL	COBL7	COBRA-like protein7 precursor
AT5G49270 .1	-	Y	Y	663	29	COBL		COBRA-like protein precursor
AT2G41400 .1	-	-	Y	150	33	Ole E		Pollen Ole e 1 allergen and extensin family protein
AT2G47530 .1	-	-	Y	184	36	Ole E		Pollen Ole e 1 allergen and extensin family protein
AT2G34700 .1	-	-	Y	175	37	Ole E		Pollen Ole e 1 allergen and extensin family protein
AT3G16670 .1	-	-	Y	154	31	Ole E		Pollen Ole e 1 allergen and extensin family protein
AT5G15780 .1	-	-	Y	401	43	Ole E		Pollen Ole e 1 allergen and extensin family protein
AT2G16630 .1	-	-	Y	359	35	Ole E		Pollen Ole e 1 allergen and extensin family protein
AT3G16660 .1	-	-	Y	180	27	Ole E		Pollen Ole e 1 allergen and extensin family protein
AT5G13140 .1	-	-	Y	267	32	Ole E		Pollen Ole e 1 allergen and extensin family protein

AT2G33790.1	Y	Y	Y	239	35	Ole E	AGP30I	Pollen Ole e 1 allergen and extensin family protein
AT1G28290.1	Y	Y	Y	359	44	Ole E	AGP31I	Pollen Ole e 1 allergen and extensin family protein
AT4G15200.1	-	Y	Y	764	34	Formin	FH3	Actin-binding FH2 (formin homology 2) family protein
AT1G24150.1	-	-	Y	725	31	Formin	FH4	Actin-binding FH2 (formin homology 2) family protein
AT3G05470.1	-	Y	Y	884	27	Formin		Actin-binding FH2 (formin homology 2) family protein
AT5G54650.1	-	-	Y	900	31	Formin	FH5	Actin-binding FH2 (formin homology 2) family protein
AT5G48360.1	-	Y	Y	782	33	Formin		Actin-binding FH2 (formin homology 2) family protein
AT1G70140.1	-	-	Y	760	31	Formin	FH8	Actin-binding FH2 (formin homology 2) family protein
AT5G67470.1	-	-	Y	899	36	Formin	FH6	Actin-binding FH2 (formin homology 2) family protein
AT3G07540.1	-	Y	Y	841	36	Formin		Actin-binding FH2 (formin homology 2) family protein
AT2G43800.1	-	-	Y	894	35	Formin		Actin-binding FH2 (formin homology 2) family protein
AT2G20300.1	-	-	Y	744	30	RLK	ALE2	Protein kinase superfamily protein
AT2G23300.1	-	-	Y	773	29	RLK		Leucine-rich repeat protein kinase family protein
AT5G56890.1	-	-	Y	1113	40	RLK		Protein kinase superfamily protein
AT1G10850.1	-	-	Y	663	30	RLK		Leucine-rich repeat protein kinase family protein
AT4G36180.1	-	-	Y	1136	27	RLK		Leucine-rich receptor-like protein kinase family protein
AT3G19300.1	-	Y	Y	663	29	RLK		Protein kinase superfamily protein
AT5G03140.1	-	-	Y	711	32	RLK		Concanavalin A-like lectin protein kinase family protein
AT5G58540.1	-	Y	Y	484	31	RLK		Protein kinase superfamily protein
AT5G11020.1	-	-	Y	433	27	RLK		Protein kinase superfamily protein
AT5G67280.1	-	-	Y	751	29	RLK		Receptor-like kinase

AT2G20850.1	-	Y	Y	775	28	RLK	SRF1	STRUBBELIG-receptor family 1
AT4G03390.1	-	-	Y	776	29	RLK	SRF3	STRUBBELIG-receptor family 3
AT2G13800.1	-	-	Y	601	25	RLK	SERK5/BAK8	Somatic embryogenesis receptor-like kinase 5
AT1G63570.1	-	Y	Y	284	39	RLK		Receptor-like protein kinase-related family protein
AT4G18640.1	-	Y	Y	678	27	RLK	MRH1	Leucine-rich repeat protein kinase family protein
AT4G39110.1	-	-	Y	878	28	RLK		Malectin/receptor-like protein kinase family protein
AT3G56100.1	-	-	Y	719	30	RLK	MRLK/IMK3	Meristematic receptor-like kinase
AT5G39020.1	-	-	Y	813	24	RLK		Malectin/receptor-like protein kinase family protein
AT5G28680.1	-	-	Y	858	27	RLK	ANX2	Malectin/receptor-like protein kinase family protein
AT5G01550.1	-	-	Y	688	28	RLK		Lectin receptor kinase
AT5G38990.1	-	-	Y	880	26	RLK		Malectin/receptor-like protein kinase family protein
AT2G21480.1	-	-	Y	871	28	RLK		Malectin/receptor-like protein kinase family protein
AT4G38830.1	-	-	Y	665	25	RLK	CRK26	Cysteine-rich RLK (RECEPTOR-like protein kinase) 26
AT2G40270.1	-	-	Y	489	31	RLK		Protein kinase family protein
AT1G70690.1	-	-	Y	299	28	RLK	HWI1/PDLP5	Receptor-like protein kinase-related family protein
AT2G31880.1	-	-	Y	641	25	RLK	SOBIR1/EVR	Leucine-rich repeat protein kinase family protein
AT5G54380.1	-	-	Y	855	30	RLK	THE1	protein kinase family protein
AT3G08680	-	Y	Y	641	30	RLK		Leucine-rich repeat protein kinase family protein
AT1G34300.1	-	-	Y	829	29	RLK		Lectin protein kinase family protein
AT1G51940.1	-	-	Y	651	27	RLK-LysM		Protein kinase family protein / peptidoglycan-binding LysM domain-containing protein
AT2G25440.1	-	-	Y	671	25	RLP	RLP20	Receptor like protein 20

AT3G22030.1	-	-	Y	253	25	RLP		Receptor protein kinase-related
AT1G28340.1	-	-	Y	626	28	RLP	RLP4	Receptor like protein 4
AT1G34290.1	-	-	Y	266	22	RLP	RLP5	Receptor like protein 5
AT2G42800.1	-	-	Y	462	30	RLP	RLP29	Receptor like protein 29
AT1G63550.1	-	-	Y	324	37	RLP		Receptor-like protein kinase-related family protein
AT5G41280.1	-	-	Y	286	31	RLP		Receptor-like protein kinase-related family protein
AT4G18760.1	-	Y	Y	431	36	RLP-LRR	RLP51	Receptor like protein 51
AT4G28380.1	-	-	Y	391	30	RLP-LRR		Leucine-rich repeat (LRR) family protein
AT5G25550.1		Y	Y	433	28	RLP-LRR		Leucine-rich repeat (LRR) family protein
AT3G24480.1	-	-	Y	494	31	RLP-LRR		Leucine-rich repeat (LRR) family protein
AT3G19320.1	-	-	Y	493	30	RLP-LRR		Leucine-rich repeat (LRR) family protein
AT1G21880.2	-	-	Y	416	37	RLP-LysM		LysM domain protein
AT1G77630.1	-	Y	Y	423	38	RLP-LysM		LysM domain protein
AT1G71980.1	-	Y	Y	448	39	Protease		Protease-associated (PA) RING/U-box zinc finger family protein
AT4G34980.1	-	-	Y	764	33	Protease	SLP2	Subtilisin-like serine protease 2
AT1G35630.1	-	-	Y	318	27	Protease		Protease-associated (PA) RING/U-box zinc finger family protein
AT4G38220.2	-	-	Y	433	28	Protease		Peptidase M20/M25/M40 family protein
AT1G09750.1	-	-	Y	449	37	Protease		Eukaryotic aspartyl protease family protein
AT5G07030.1	-	-	Y	455	36	Protease		Eukaryotic aspartyl protease family protein
AT5G43100.1	-	Y	Y	631	27	Protease		Eukaryotic aspartyl protease family protein
AT4G21326.1	-	-	Y	754	31	Protease	SBT3.12	Subtilase 3.12
AT3G51330.1	-	Y	Y	529	30	Protease		Eukaryotic aspartyl protease family protein
AT3G51350.1	-	Y	Y	528	30	Protease		Eukaryotic aspartyl protease family protein

AT4G20990.1	-	-	Y	267	24	Carbonic anhydrase		Alpha carbonic anhydrase 4
AT5G04180.1	-	Y	Y	277	25	Carbonic anhydrase	ACA3	Alpha carbonic anhydrase 3
AT5G56330.1	Y	Y	Y	350	40	Carbonic anhydrase	AGP33I	Alpha carbonic anhydrase 2
AT5G21105.1	-		Y	588	27	Chimeric		Plant L-ascorbate oxidase
AT4G19410.1	-	-	Y	391	29	Chimeric		Pectinacetylesterase family protein
AT5G23870.3	-	-	Y	451	28	Chimeric		Pectinacetylesterase family protein
AT3G17070.1	-	-	Y	339	32	Chimeric		Peroxidase family protein
AT2G18980.1	-	-	Y	323	29	Chimeric		Peroxidase superfamily protein
AT2G43290.1	-	-	Y	215	21	Chimeric		Calcium-binding EF-hand family protein
AT2G42840.1	-	-	Y	306	50	Chimeric	PDF1	protodermal factor 1
AT5G14920.1	-	-	Y	275	54	Chimeric		Gibberellin-regulated family protein
AT1G12090.1	-	-	Y	137	36	Chimeric	ELP	extensin-like protein
AT3G28790.1	-	-	Y	608	47	Chimeric		Protein of unknown function (DUF1216)
AT3G52480.1	-	-	Y	209	36	Chimeric		NA
AT5G27710.1	-	-	Y	335	28	Chimeric		NA
AT3G29300.1	-	-	Y	213	43	Chimeric		NA
AT2G30700.1	-	-	Y	480	32	Chimeric		NA
AT3G11640.1	-	-	Y	186	26	Chimeric		NA
AT1G65720.1	-	-	Y	180	33	Chimeric		NA
AT1G78460.1	-	-	Y	219	32	Chimeric		SOUL heme-binding family protein
AT5G64720.1	-	-	Y	155	31	Chimeric		Protein of unknown function (DUF1278)
AT4G39340.1	-	-	Y	127	34	Chimeric		Protein of unknown function (DUF1278)
AT5G62630.1	-	-	Y	696	31	Chimeric	HIPL2	hipl2 protein precursor
AT4G38080.1	-	-	Y	128	50	Chimeric		hydroxyproline-rich glycoprotein family protein
AT5G60650.1	-	-	Y	147	38	Chimeric		NA
AT2G28410.1	-	-	Y	115	38	Chimeric		NA
AT2G22510.1	-	-	Y	124	55	Chimeric		hydroxyproline-rich glycoprotein family protein

AT5G20630.1	-	-	Y	211	32	Chimeric	GLP3	germin 3
AT5G62630.1	-	-	Y	696	31	Chimeric	HIPL2	hipl2 protein precursor
AT3G45245.1	-	-	Y	116	25	Chimeric		ECA1 gametogenesis related family protein
AT2G34870.1	-	-	Y	116	48	Chimeric	MEE26	hydroxyproline-rich glycoprotein family protein
AT5G42560.1	-	-	Y	296	33	Chimeric	HVA22	Abscisic acid-responsive (TB2/DP1, HVA22) family protein
AT5G66816.1	-	-	Y	101	25	Chimeric		NA
AT5G26070.1	-	-	Y	102	46	Chimeric		hydroxyproline-rich glycoprotein family protein
AT3G18050.1	-	-	Y	335	32	Chimeric		NA
AT2G36695.1	-	-	Y	117	33	Chimeric		NA
AT1G24267.2	-	-	Y	344	31	Chimeric		Protein of unknown function (DUF1664)
AT5G48210.1	-	-	Y	103	29	Chimeric		Protein of unknown function (DUF1278)
AT1G61688.1	-	-	Y	105	33	Chimeric		Defensin-like (DEFL) family protein
AT3G44430.1	-	-	Y	206	33	Chimeric		NA
AT5G56369.1	-	-	Y	93	32	Chimeric		Defensin-like (DEFL) family protein
AT5G42785.1	-	-	Y	115	26	Chimeric		NA
AT4G20420.1	-	-	Y	142	42	Chimeric		Tapetum specific protein TAP35/TAP44
AT3G26110.1	-	-	Y	128	43	Chimeric		Anther-specific protein agp1-like
AT3G22053.1	-	-	Y	149	29	Chimeric		NA
AT3G54040.1	-	-	Y	183	35	Chimeric		PAR1 protein
AT4G28160.1	-	-	Y	111	43	Chimeric		hydroxyproline-rich glycoprotein family protein
AT4G22010.1	-	-	Y	541	27	Chimeric	SKS4	SKU5 similar 4
AT5G09480.1	-	-	Y	144	37	Chimeric		hydroxyproline-rich glycoprotein family protein
AT1G12845.1	-	-	Y	119	37	Chimeric		NA
AT2G17230.1	-	-	Y	363	31	Chimeric	EXL5	EXORDIUM like 5
AT5G51105.1	-	-	Y	123	37	Chimeric		Protein of unknown function (DUF1278)

AT1G41830.1	-	-	Y	542	26	Chimeric	SKS6	SKU5-similar 6
AT5G67400.1	-	-	Y	329	28	Chimeric	RHS19	root hair specific 19
AT5G52390.1	-	-	Y	195	29	Chimeric		PAR1 protein
AT3G47300.1	-	-	Y	209	29	Chimeric		SELT-like protein precursor
AT2G14900.1	-	-	Y	108	29	Chimeric		Gibberellin-regulated family protein
AT4G10457.1	-	-	Y	92	21	Chimeric	SCRL1	SCR-like 1
AT4G13263.1	-	-	Y	149	25	Chimeric		Protein of unknown function (DUF784)
AT2G21110.1	-	-	Y	186	29	Chimeric		Disease resistance-responsive (dirigent-like protein) family protein
AT1G76160.1	-	-	Y	541	27	Chimeric	SKS5	SKU5 similar 5
AT2G40530.1	-	-	Y	105	34	Chimeric		NA
AT1G21850.1	-	-	Y	551	28	Chimeric	SKS8	SKU5 similar 8
AT2G28790.1	-	-	Y	249	31	Chimeric		Pathogenesis-related thaumatin superfamily protein
AT4G22900.1	-	-	Y	343	25	Chimeric		Protein of unknown function (DUF1191)
AT1G21680.1	-	-	Y	706	31	Chimeric		DPP6 N-terminal domain-like protein
AT3G18715.1	-	-	Y	93	28	Chimeric	IDL4	inflorescence deficient in abscission (IDA)-like 4
AT3G45252.1	-	-	Y	96	28	Chimeric		ECA1 gametogenesis related family protein
AT3G27410.1	-	-	Y	132	41	Chimeric		NA
AT4G26880.1	-	-	Y	152	28	Chimeric		Stigma-specific Stig1 family protein
AT5G66920.1	-	-	Y	546	28	Chimeric	SKS17	SKU5 similar 17
AT1G21670.1	-	-	Y	703	33	Chimeric		NA
AT4G24350.1	-	-	Y	336	25	Chimeric		Phosphorylase superfamily protein
AT3G47070.1	-	-	Y	100	29	Chimeric		NA
AT2G04031.1	-	-	Y	151	27	Chimeric		Protein of unknown function (DUF784)
AT2G21740.1	-	-	Y	125	32	Chimeric		Protein of unknown function (DUF1278)

AT5G34581.1	-	-	Y	161	27	Chimeric		hydroxyproline-rich glycoprotein family protein
AT3G14860.2	-	-	Y	493	31	Chimeric		NHL domain-containing protein
AT5G44570.2	-	-	Y	140	43	Chimeric		NA
AT1G16950.1	-	-	Y	93	33	Chimeric		NA
AT3G01960.1	-	-	Y	138	30	Chimeric		NA
AT5G35405.1	-	-	Y	149	22	Chimeric		Protein of unknown function (DUF784)
AT5G60630.1	-	-	Y	139	32	Chimeric		NA
AT2G20870.1	-	-	Y	140	36	Chimeric		cell wall protein precursor, putative
AT5G44580.1	-	-	Y	98	31	Chimeric		NA
AT1G34800.1	-	-	Y	92	27	Chimeric		Plant thionin family protein
AT1G34805.1	-	-	Y	92	27	Chimeric		Plant thionin family protein
AT1G34815.1	-	-	Y	92	27	Chimeric		Plant thionin family protein
AT1G34930.1	-	-	Y	92	25	Chimeric		Plant thionin family protein
AT1G34860.1	-	-	Y	92	27	Chimeric		Plant thionin family protein
AT1G34795.1	-	-	Y	92	27	Chimeric		Plant thionin family protein
AT1G34820.1	-	-	Y	92	27	Chimeric		Plant thionin family protein
AT1G35035.1	-	-	Y	92	28	Chimeric		Plant thionin family protein
AT2G28671.1	-	Y	Y	299	53	Chimeric		NA
AT1G03820.1	Y	Y	Y	222	25	Chimeric	AGP28I	
AT1G61900.1	-	-	Y	433	31	Chimeric		NA
AT3G07390.1	-	-	Y	273	38	Chimeric		auxin-responsive family protein
AT3G20520.1	-	-	Y	729	29	Chimeric	SVL3	
AT4G25240.1	-	-	Y	589	26	Chimeric	SKU5L1	
AT4G26690.1	-	-	Y	759	33	Chimeric	SHV3 PLC-like	
AT4G28100.1	-	-	Y	304	32	Chimeric	Chimeric	NA
AT5G55480.1	-	-	Y	766	32	Chimeric	SVL1	
AT1G69295.1	-	-	Y	222	47	Chimeric	PDCB4	plasmodesmata callose-binding protein 4
AT1G18650.1	-	-	Y	184	40	Chimeric	PDCB3	plasmodesmata callose-binding protein 3

AT5G61130.1	-	-	Y	201	42	Chimeric	PDCB1	plasmodesmata callose-binding protein
AT1G78820.1	-	-	Y	455	26	Chimeric		D-mannose binding lectin protein with Apple-like carbohydrate-binding domain
AT3G06260.1	-	-	Y	351	24	Chimeric	GATL4	galacturonosyltransferase-like 4
AT5G01050.1	-	-	Y	586	29	Chimeric		Laccase/Diphenol oxidase family protein
AT3G44990.1	-	-	Y	293	29	Chimeric	XTR8	Xyloglucan endo-transglycosylase-related 8
AT5G07520.1	-	-	Y	228	44	Gly-rich	GRP18	Glycine-rich protein family
AT1G30795.1	-	-	Y	109	46	Gly-rich		Glycine-rich protein family
AT2G15340.1	-	-	Y	119	37	Gly-rich		Glycine-rich protein family
AT3G29033.1	-	-	Y	167	24	Gly-rich		Glycine-rich protein family
AT1G63057.1	-	-	Y	116	28	His-rich		Hist-rich
AT1G63055.1	-	-	Y	109	27	His-rich		Hist-rich
AT3G19020.1	-	-	Y	956	46	LRR-EXT		Leucine-rich repeat (LRR) family protein
AT1G12040.1	-	-	Y	744	46	LRR-EXT	LRX1	leucine-rich repeat/extensin 1
AT4G13340.1	-	-	Y	760	47	LRR-EXT		Leucine-rich repeat (LRR) family protein
AT1G49490.1	-	-	Y	847	44	LRR-EXT		Leucine-rich repeat (LRR) family protein
AT4G22470.1	Y	Y	Y	375	51	EXT	HAE4	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein
AT1G54970.1	-	-	Y	335	46	EXT-AGP		
AT3G62680.1	-	-	Y	313	46	EXT-AGP		
AT1G44191.1	-	-	Y	359	57	EXT-AGP		
AT3G50580.1	Y	Y	Y	250	46	EXT-AGP	HAE2	NA
AT5G11990.1	-	-	Y	181	52	EXT-AGP		NA
AT3G22120.1	-	-	Y	334	56	EXT-AGP	CWLP	cell wall-plasma membrane linker protein

Table II. AGPs identified in the *M. truncatula* genome.

AGI	Ma 2017	Moore 2018	AA length	PAST %	AGP class	Annotation
Medtr2g094170.1	Y	Y	125	47	Classical	Transmembrane protein, putative
Medtr4g459420.1	Y	Y	144	56	Classical	Transmembrane protein, putative
Medtr4g059520.1	Y	Y	144	56	Classical	Transmembrane protein, putative
Medtr4g093780.1	Y	Y	181	71	Classical	Transmembrane protein, putative
Medtr4g127180.1	Y	Y	146	66	Classical	Transmembrane protein, putative
Medtr8g069283.1	Y	Y	100	66	Classical	Transmembrane protein, putative
Medtr8g069315.1	Y	Y	101	68	Classical	Transmembrane protein, putative
Medtr7g032620.1	Y	Y	127	60	Classical	Transmembrane protein, putative
Medtr3g049390.1	Y	Y	116	50	Classical	Transmembrane protein, putative
Medtr3g099520.1	Y	Y	128	75	Classical	Transmembrane protein, putative
Medtr4g032750.1	-	Y	230	54	Classical	Transmembrane protein, putative
Medtr4g093810.1	Y	Y	98	72	Classical	Transmembrane protein, putative
Medtr4g093790.1	Y	Y	100	68	Classical	Transmembrane protein, putative
Medtr4g093825.1	Y	Y	103	68	Classical	Transmembrane protein, putative
Medtr4g093835.1	Y	Y	101	70	Classical	Transmembrane protein, putative
Medtr5g013285.1	Y	Y	100	62	Classical	Transmembrane protein, putative
Medtr6g021940.1	Y	Y	124	70	Classical	Transmembrane protein, putative
Medtr6g039360.1	Y	Y	168	72	Classical	Transmembrane protein, putative
Medtr6g039690.1	Y	Y	171	74	Classical	Transmembrane protein, putative

Medtr0902s0010.1	Y	Y	185	74	Classical	Transmembrane protein, putative
Medtr3g007900.1	Y	Y	136	53	Classical	GASAGASTSnakin HC chr312363561234999 20130731 ; MatureChain: 19-154
Medtr5g077360.1	Y	Y	122	50	Classical	Leguminosin group485 secreted peptide HC chr53302810033027104 20130731 ; MatureChain: 26-147
Medtr5g077715.1	Y	Y	175	46	Classical	Leguminosin group485 secreted peptide HC chr53316768233166106 20130731 ; MatureChain: 26-200
Medtr5g095990.1	Y	Y	131	54	Classical	GASAGASTSnakin HC chr54196929641967991 20130731 ; MatureChain: 19-149
Medtr8g022490.1	Y	Y	152	50	Classical	Leguminosin group485 secreted peptide HC chr879670357967849 20130731 ; MatureChain: 26-177
Medtr8g022550.1	Y	Y	149	52	Classical	Leguminosin group485 secreted peptide HC chr879857837986582 20130731 ; MatureChain: 26-174
Medtr2g009150.1	-	Y	94	37	Classical	Transmembrane protein, putative
Medtr2g009275.1	Y	Y	74	65	Classical	Transmembrane protein, putative
Medtr2g009290.1	Y	Y	71	62	Classical	Transmembrane protein, putative
Medtr7g083830.1	Y	Y	83	56	Classical	Transmembrane protein, putative
Medtr2g450250.1	-	Y	57	51	Classical	Transmembrane protein, putative
Medtr8g089615.1	-	Y	82	42	Classical	Transmembrane protein, putative

Medtr6g045633.1	-	Y	75	42	Classical	Leguminosin prolinerich group669 secreted peptide HC chr61653075116529919 20130731 ; MatureChain: 21-95
Medtr3g111910.1	Y	Y	201	49	Classical	Transmembrane protein, putative
Medtr5g004980.1	Y	Y	200	71	Classical	Transmembrane protein, putative HC chr5230877229070 20130731 ; MatureChain: 24-223
Medtr5g032670.1	Y	Y	123	82	Classical	Hypothetical protein HC chr51407464514075917 20130731 ; MatureChain: 23-145
Medtr1g022340.1	-	Y	147	49	Classical	Transmembrane protein, putative LC chr1:7067015-7065822 20130731
Medtr1g112220.1	-	Y	81	56	Classical	Transmembrane protein, putative HC chr1:50753076-50753390 20130731
Medtr2g020370.1	-	Y	123	53	Classical	Transmembrane protein, putative HC chr2:6775699-6774908 20130731
Medtr2g026875.1	-	Y	155	70	Classical	Transmembrane protein, putative LC chr2:9781196-9781735 20130731
Medtr3g460760.1	-	Y	191	72	Classical	Hypothetical protein HC chr3:23867366-23868679 20130731
Medtr7g032400.1	-	Y	76	60	Classical	Transmembrane protein, putative HC chr7:11381022-11380648 20130731
Medtr7g032430.1	-	Y	50	64	Classical	Transmembrane protein, putative HC chr7:11396461-11396026 20130731
Medtr7g032470.1	-	Y	76	60	Classical	Transmembrane protein, putative HC chr7:11404486-11404117 20130731

Medtr1g016850.1	Y	Y	37	63	AG peptide	Transmembrane protein, putative HC chr145280724527427 20130731 ; MatureChain: 25-61
Medtr1g052120.1	Y	Y	39	54	AG peptide	Transmembrane protein, putative HC chr12100717521006594 20130731 ; MatureChain: 23-61
Medtr3g028270.1	Y	Y	42	50	AG peptide	Transmembrane protein, putative HC chr389651328965667 20130731 ; MatureChain: 26-67
Medtr5g008670.1	Y	Y	46	37	AG peptide	Transmembrane protein, putative LC chr518657151865512 20130731 ; MatureChain: 22-67
Medtr5g035610.1	Y	Y	43	33	AG peptide	Transmembrane protein, putative HC chr5:15501113-15501772 20130731
Medtr5g035640.1	Y	Y	53	38	AG peptide	Transmembrane protein, putative HC chr51550847615508987 20130731 ; MatureChain: 24-76
Medtr6g029260.1	Y	Y	33	52	AG peptide	Transmembrane protein, putative HC chr61006924610068704 20130731 ; MatureChain: 27-59
Medtr6g029330.1	Y	Y	33	52	AG peptide	Transmembrane protein, putative HC chr61010693610106129 20130731 ; MatureChain: 27-59
Medtr7g085780.1	Y	Y	33	58	AG peptide	Transmembrane protein, putative HC chr73325275233253550 20130731 ; MatureChain: 27-59

Medtr3g104110.1	Y	Y	38	61	AG peptide	Transmembrane protein, putative LC chr34797057247970739 20130731 ; MatureChain: 18-55
Medtr1g028610.1	Y	Y	43	47	AG peptide	arabinogalactan peptide like protein HC chr197046419703281 20130731 ; MatureChain: 20-62
Medtr3g090830.1	Y	Y	36	39	AG peptide	arabinogalactan peptide like protein HC chr34123493141234172 20130731 ; MatureChain: 28-63
Medtr6g086350.1	-	Y	38	29	AG peptide	arabinogalactan protein LC chr63244022432440033 20130731 ; MatureChain: 26-63
Medtr6g086365.1	-	Y	63	24	AG peptide	arabinogalactan protein LC chr63246133432459573 20130731 ; MatureChain: 26-88
Medtr6g086390.1	Y	Y	38	32	AG peptide	arabinogalactan peptide like protein HC chr63246422932463703 20130731 ; MatureChain: 27-64
Medtr6g090535.1	Y	Y	44	39	AG peptide	arabinogalactan protein HC chr63441166534410320 20130731 ; MatureChain: 20-63
Medtr8g099245.1	Y	Y	44	44	AG peptide	arabinogalactan peptide like protein HC chr84176556641763694 20130731 ; MatureChain: 32-75
Medtr8g083480.1	Y	Y	44	48	AG peptide	arabinogalactan protein HC chr84360039243600914 20130731 ; MatureChain: 26-69

Medtr3g464630.1	-	Y	48	36	AG peptide	Transmembrane protein, putative LC chr32599262725993229 20130731 ; Matu
Medtr3g037730.1	-	Y	56	29	AG peptide	Clavata3ESR (CLE) gene family member MtCLE11 HC chr31387447713873747 20130731 ; MatureChain: 28-83
Medtr3g027772.1	Y	Y	43	45	AG peptide	Transmembrane protein, putative HC chr3:8755526-8755723 20130731
Medtr3g028260.1	Y	Y	43	42	AG peptide	Transmembrane protein, putative HC chr3:8962413-8963001 20130731
Medtr3g028340.1	Y	Y	43	47	AG peptide	Transmembrane protein, putative HC chr3:8988222-8988674 20130731
Medtr3g028450.1	Y	Y	43	40	AG peptide	Transmembrane protein, putative LC chr3:9015269-9015961 20130731
Medtr3g028460.1	Y	Y	43	45	AG peptide	Transmembrane protein, putative HC chr3:9019158-9019743 20130731
Medtr3g028465.1	Y	Y	43	45	AG peptide	Transmembrane protein, putative HC chr3:9023312-9023905 20130731
Medtr3g028480.1	Y	Y	43	42	AG peptide	Transmembrane protein, putative HC chr3:9026022-9026583 20130731
Medtr3g028495.1	Y	Y	43	45	AG peptide	Transmembrane protein, putative HC chr3:9034496-9034963 20130731
Medtr3g028540.1	Y	Y	43	45	AG peptide	Transmembrane protein, putative HC chr3:9101047-9101784 20130731
Medtr3g028550.1	Y	Y	43	45	AG peptide	Transmembrane protein, putative HC chr3:9108686-9108956 20130731
Medtr5g008650.1	Y	Y	50	32	AG peptide	Transmembrane protein, putative HC chr5:1861708-1861505 20130731

Medtr7g092350.1	Y	Y	51	55	AG peptide	Transmembrane protein, putative HC chr7:36592637-36592855 20130731
Medtr8g036500.1	Y	Y	52	41	AG peptide	Transmembrane protein, putative LC chr8:13503881-13503663 20130731
Medtr5g035650.1	-	Y	45	40	AG peptide	Transmembrane protein, putative
Medtr4g095280.1	Y	Y	186	71	Lys Rich	Hypothetical protein
Medtr5g016250.1	Y	Y	203	68	Lys Rich	Hypothetical protein
Medtr8g069925.1	Y	Y	599	63	Lys Rich	Hypothetical protein
Medtr0002s0900.1	Y	Y	232	63	Lys Rich	Hypothetical protein
Medtr2g093980.1	Y	Y	261	39	FLA	Fasciclin domain protein HC chr2:40048746-40047457 20130731
Medtr2g018340.1	Y	Y	199	33	FLA	Fasciclin domain protein HC chr2:5778670-5779378 20130731
Medtr2g437700.1	Y	Y	408	39	FLA	Fasciclin-like arabinogalactan protein HC chr2:14853943-14855880 20130731
Medtr4g053380.1	Y	Y	274	38	FLA	Fasciclin-like arabinogalactan protein HC chr4:19402736-19404065 20130731
Medtr4g059790.1	Y	Y	271	38	FLA	Fasciclin-like arabinogalactan protein HC chr4:22059038-22060252 20130731
Medtr4g059840.1	Y	Y	266	37	FLA	Fasciclin-like arabinogalactan protein HC chr4:22076175-22074615 20130731
Medtr4g084160.1	Y	Y	225	28	FLA	Fasciclin domain protein HC chr4:32804307-32803564 20130731
Medtr5g098420.1	Y	Y	396	30	FLA	Fasciclin-like arabinogalactan protein HC chr5:43093083-43091217 20130731
Medtr1g009910.1	Y	Y	264	38	FLA	Fasciclin-like arabinogalactan protein LC chr116131591612264 20130731 ; MatureChain: 24-287

Medtr1g009920.1	Y	Y	204	38	FLA	Fasciclin domain protein HC chr116225951620447 20130731 ; MatureChain: 24-227
Medtr1g009940.1	Y	Y	264	38	FLA	Fasciclin domain protein HC chr116225951620447 20130731 ; MatureChain: 24-227
Medtr1g009950.1	Y	Y	245	42	FLA	Fasciclin domain protein HC chr116377891636871 20130731 ; MatureChain: 24-268
Medtr2g007870.1	Y	Y	331	36	FLA	Fasciclin domain protein HC chr211802911179140 20130731 ; MatureChain: 24-354
Medtr2g017970.1	Y	Y	257	40	FLA	Fasciclin domain protein HC chr252657325267208 20130731 ; MatureChain: 24-280
Medtr2g017960.1	Y	Y	254	39	FLA	Fasciclin domain protein HC chr252687625270188 20130731 ; MatureChain: 24-277
Medtr2g017950.1	Y	Y	267	37	FLA	Fasciclin domain protein HC chr252731195274452 20130731 ; MatureChain: 29-295
Medtr2g087810.1	Y	Y	423	27	FLA	Fasciclin-like arabinogalactan protein 16 precursor HC chr23695627336953488 20130731 ; MatureChain: 24-446
Medtr2g090765.1	Y	Y	223	35	FLA	Fasciclin-like arabinogalactan protein HC chr23891996238918307 20130731 ; MatureChain: 27-249
Medtr2g093970.1	Y	Y	268	37	FLA	Fasciclin domain protein HC chr24004583240044669 20130731 ; MatureChain: 26-293

Medtr2g093990.1	Y	Y	224	35	FLA	Fasciclin-like arabinogalactan protein HC chr24005111540049995 20130731 ; MatureChain: 27-250
Medtr3g460780.1	Y	Y	392	34	FLA	Fasciclin-like arabinogalactan protein HC chr32387913923880737 20130731 ; MatureChain: 30-421
Medtr4g050170.1	Y	Y	436	28	FLA	Fasciclin-like arabinogalactan protein 16 precursor HC chr41774750217752074 20130731 ; MatureChain: 31-466
Medtr4g059720.1	Y	Y	224	34	FLA	Fasciclin-like arabinogalactan protein HC chr42203457022035620 20130731 ; MatureChain: 27-250
Medtr4g099150.1	Y	Y	227	31	FLA	Fasciclin domain protein LC chr44106286741062127 20130731 ; MatureChain: 20-246
Medtr5g098060.1	Y	Y	241	42	FLA	Fasciclin-like arabinogalactan protein HC chr54293328142935396 20130731 ; MatureChain: 22-262
Medtr7g075453.1	Y	Y	397	44	FLA	Fasciclin-like arabinogalactan protein HC chr72825181728250054 20130731 ; MatureChain: 22-418
Medtr7g101080.1	Y	Y	229	37	FLA	Fasciclin-like arabinogalactan protein HC chr74078363740782599 20130731 ; MatureChain: 28-256

Medtr7g106760.3		Y	425	29	FLA	Fasciclin-like arabinogalactan protein 16 precursor HC chr74348176543477708 20130731 ; MatureChain: 24-448
Medtr7g106760.1	Y	Y	425	29	FLA	Fasciclin-like arabinogalactan protein 16 precursor HC chr74348151543477779 20130731 ; MatureChain: 24-448
Medtr7g106760.2		Y	425	29	FLA	Fasciclin-like arabinogalactan protein 16 precursor HC chr74348176543478342 20130731 ; MatureChain: 24-448
Medtr8g087450.1	Y	Y	270	39	FLA	Fasciclin domain protein HC chr83614781636149210 20130731 ; MatureChain: 23-292
Medtr8g087460.1	Y	Y	288	41	FLA	Fasciclin domain protein HC chr83615068136151957 20130731 ; MatureChain: 29-316
Medtr8g087470.1	Y	Y	288	42	FLA	Fasciclin domain protein HC chr83615407036155160 20130731 ; MatureChain: 29-316
Medtr8g086360.1	Y	Y	171	24	PAG	Plastocyanin-like domain protein LC chr8:35820481-35819459 20130731
Medtr4g081100.1	Y	Y	194	38	PAG	Plastocyanin-like domain protein HC chr4:31394809-31394041 20130731
Medtr5g006040.1	Y	Y	160	35	PAG	Plastocyanin-like domain protein HC chr5:739373-738212 20130731
Medtr4g066110.1	Y	Y	148	30	PAG	Plastocyanin-like domain protein LC chr4:24910220-24910812 20130731
Medtr4g067200.1	Y	Y	148	30	PAG	Plastocyanin-like domain protein LC chr4:25360406-25359814 20130731

Medtr1g077790.1	Y	Y	167	35	PAG	Plastocyanin-like domain protein HC chr1:34753995-34753085 20130731
Medtr2g088990.1	Y	Y	207	42	PAG	Plastocyanin-like domain protein HC chr2:37537716-37538597 20130731
Medtr2g101300.1	Y	Y	261	39	PAG	Plastocyanin-like domain protein HC chr2:43547176-43547831 20130731
Medtr3g099570.1	Y	Y	162	43	PAG	Blue copper-like protein HC chr3:45643508-45642850 20130731
Medtr1g014120.1	Y	Y	163	30	PAG	Plastocyaninlike domain protein HC chr131153353114127 20130731 ; MatureChain: 22-184
Medtr1g105130.1	-	Y	133	25	PAG	Plastocyaninlike domain protein HC chr12498462124985495 20130731 ; MatureChain: 23-155
Medtr1g105120.1	-	Y	133	25	PAG	Plastocyaninlike domain protein HC chr12499049624991335 20130731 ; MatureChain: 23-155
Medtr1g090420.1	Y	Y	195	34	PAG	Plastocyaninlike domain protein LC chr14050754140508971 20130731 ; MatureChain: 26-220
Medtr1g104800.1	Y	Y	163	38	PAG	Plastocyaninlike domain protein HC chr14721856647220241 20130731 ; MatureChain: 23-185
Medtr2g083250.1	Y	Y	238	43	PAG	Plastocyaninlike domain protein HC chr23495210234954240 20130731 ; MatureChain: 24-261

Medtr3g099580.1	Y	Y	163	47	PAG	Plastocyaninlike domain protein HC chr34565094245649884 20130731 ; MatureChain: 24-186
Medtr3g099680.1	Y	Y	197	49	PAG	Plastocyaninlike domain protein HC chr34570279145703579 20130731 ; MatureChain: 24-220
Medtr3g099980.1	Y	Y	195	49	PAG	Plastocyaninlike domain protein HC chr34588663445887385 20130731 ; MatureChain: 24-218
Medtr4g078410.1	Y	Y	165	27	PAG	Plastocyaninlike domain protein HC chr43031536530314687 20130731 ; MatureChain: 23-187
Medtr4g081100.1	-	Y	193	38	PAG	Plastocyaninlike domain protein HC chr43139480931394041 20130731 ; MatureChain: 30-222
Medtr4g114870.1	Y	Y	345	51	PAG	Plastocyaninlike domain protein HC chr44733265247334953 20130731 ; MatureChain: 26-370
Medtr4g124280.1	Y	Y	216	33	PAG	Plastocyaninlike domain protein LC chr45131906551320408 20130731 ; MatureChain: 23-238
Medtr4g130800.1	Y	Y	244	47	PAG	Plastocyaninlike domain protein HC chr45450413654505468 20130731 ; MatureChain: 27-270

Medtr5g088990.1	Y	Y	210	26	PAG	Plastocyaninlike domain protein HC chr53864315238645471 20130731 ; MatureChain: 31-240
Medtr6g013170.1	Y	Y	149	22	PAG	Plastocyaninlike domain protein HC chr641454894142683 20130731 ; MatureChain: 27-175
Medtr6g083240.1	Y	Y	152	38	PAG	Plastocyaninlike domain protein HC chr63115840431157789 20130731 ; MatureChain: 25-176
Medtr7g086090.1	Y	Y	220	43	PAG	Plastocyaninlike domain protein LC chr73339715433395401 20130731 ; MatureChain: 24-243
Medtr7g086100.1	Y	Y	208	46	PAG	Plastocyaninlike domain protein LC chr73340020733399140 20130731 ; MatureChain: 24-231
Medtr7g086140.1	Y	Y	138	28	PAG	Plastocyaninlike domain protein HC chr73341104033410034 20130731 ; MatureChain: 24-161
Medtr7g086160.1	Y	Y	139	29	PAG	Plastocyaninlike domain protein HC chr73341690233415916 20130731 ; MatureChain: 24-162
Medtr7g086190.1	Y	Y	138	28	PAG	Plastocyaninlike domain protein HC chr73342496133423956 20130731 ; MatureChain: 24-161

Medtr7g086220.1	Y	Y	138	27	PAG	Plastocyaninlike domain protein HC chr73343699733436074 20130731 ; MatureChain: 23-160
Medtr7g086230.1	Y	Y	153	32	PAG	Plastocyaninlike domain protein LC chr73344242233441805 20130731 ; MatureChain: 24-176
Medtr7g086280.1	Y	Y	153	32	PAG	Plastocyaninlike domain protein LC chr73345267233452025 20130731 ; MatureChain: 24-176
Medtr8g007020.1	Y	Y	166	40	PAG	Plastocyaninlike domain protein HC chr811801011181060 20130731 ; MatureChain: 22-187
Medtr8g007035.1	Y	Y	178	43	PAG	Plastocyaninlike domain protein HC chr811869001188335 20130731 ; MatureChain: 22-199
Medtr8g463180.1	Y	Y	282	50	PAG	Plastocyaninlike domain protein HC chr82221446722212599 20130731 ; MatureChain: 22-303
Medtr8g094990.1	Y	Y	255	42	PAG	Plastocyaninlike domain protein LC chr83969178239690137 20130731 ; MatureChain: 23-277
Medtr8g095013.1	Y	Y	277	47	PAG	Plastocyaninlike domain protein LC chr83970373939705150 20130731 ; MatureChain: 23-299

Medtr8g095020.1	Y	Y	295	46	PAG	Plastocyaninlike domain protein LC chr83971034339711774 20130731 ; MatureChain: 21-315
Medtr0334s0010.1	Y	Y	175	48	PAG	Plastocyaninlike domain protein HC scaffold033434686498 20130731 ; MatureChain: 26-200
Medtr1g098580.1	Y	Y	216	40	PAG	Cupredoxin superfamily protein, putative HC chr14439406944391772 20130731 ; MatureChain: 22-237
Medtr1g098580.2	-	Y	190	39	PAG	Cupredoxin superfamily protein, putative HC chr14439390144392417 20130731 ; MatureChain: 22-211
Medtr1g098580.3	-	Y	212	42	PAG	Cupredoxin superfamily protein, putative HC chr14439402344391916 20130731 ; MatureChain: 22-233
Medtr2g090575.1	Y	Y	338	50	ENOD	Early nodulin-like protein LC chr23871224338710650 20130731 ; MatureChain: 23-360
Medtr2g090580.1	Y	Y	176	35	ENOD	Early nodulin-like protein HC chr23872088638719533 20130731 ; MatureChain: 25-200
Medtr3g092170.1	Y	Y	257	37	ENOD	Early nodulin-like protein LC chr34209547142094282 20130731 ; MatureChain: 21-277
Medtr3g105930.1	Y	Y	156	27	ENOD	Early nodulin-like protein HC chr34885456148853915 20130731 ; MatureChain: 25-180

Medtr4g130780.1	Y	Y	158	25	ENOD	ENOD16 HC chr45449608254497288 20130731 ; MatureChain: 25-182
Medtr7g090170.1		Y	124	27	ENOD	Early nodulin-like protein HC chr73539578535395287 20130731 ; MatureChain: 25-148
Medtr1g071720.1	Y	Y	164	42	LTP	Lipid transfer protein HC chr13185669131855319 20130731 ; MatureChain: 27-190
Medtr1g071720.2	-	Y	163	42	LTP	Lipid transfer protein HC chr13185669131855319 20130731 ; MatureChain: 27-189
Medtr1g103420.1	Y	Y	191	52	LTP	Lipid transfer protein HC chr14681764546816110 20130731 ; MatureChain: 27-217
Medtr1g103450.1	Y	Y	130	36	LTP	Lipid transfer protein HC chr14682442246825180 20130731 ; MatureChain: 25-154
Medtr2g008580.1	Y	Y	183	46	LTP	Lipid transfer protein HC chr215402851539002 20130731 ; MatureChain: 25-207
Medtr3g079210.1	Y	Y	187	42	LTP	Lipid transfer protein HC chr33563024835627567 20130731 ; MatureChain: 20-206
Medtr3g085210.1	Y	Y	168	41	LTP	Lipid transfer protein HC chr33851674438512453 20130731 ; MatureChain: 25-192
Medtr3g106740.1	Y	Y	106	36	LTP	Lipid transfer protein HC chr34932881649329721 20130731 ; MatureChain: 25-130

Medtr4g101780.2	Y	Y	125	43	LTP	Lipid transfer protein HC chr44210071642109572 20130731 ; MatureChain: 24-148
Medtr4g101780.1	-	Y	118	39	LTP	Lipid transfer protein HC chr44210071642102691 20130731 ; MatureChain: 24-141
Medtr4g109250.1	Y	Y	157	37	LTP	Lipid transfer protein HC chr44534662845345071 20130731 ; MatureChain: 21-177
Medtr4g115360.1	-	Y	207	42	LTP	Lipid transfer protein HC chr44764781047644558 20130731 ; MatureChain: 28-234
Medtr4g115360.2	-	Y	207	42	LTP	Lipid transfer protein HC chr44764772747644963 20130731 ; MatureChain: 28-234
Medtr5g070360.1	Y	Y	168	49	LTP	Lipid transfer protein HC chr52980437029802224 20130731 ; MatureChain: 23-190
Medtr7g073100.1	Y	Y	122	28	LTP	Lipid transfer protein HC chr72729616927296615 20130731 ; MatureChain: 27-148
Medtr8g446830.1	Y	Y	170	31	LTP	Lipid transfer protein HC chr81832348118325978 20130731 ; MatureChain: 26-195
Medtr1g012710.1	Y	Y	127	45	LTP	Protease inhibitorseed storageLTP family protein HC chr126088332607695 20130731 ; MatureChain: 25-151
Medtr1g097720.1	-	Y	87	28	LTP	Lipid transfer protein HC chr1:44059772-44059512 20130731
Medtr1g101320.1	-	Y	125	29	LTP	Lipid transfer protein HC chr1:45765339-45765713 20130731

Medtr1g101360.1	-	Y	98	34	LTP	Lipid transfer protein (LIM1/2/3/MEN-8) HC chr1:45783929-45783546 20130731
Medtr1g103490.1	-	Y	157	36	LTP	Lipid transfer protein HC chr1:46837798-46839766 20130731
Medtr2g008560.1	Y	Y	173	44	LTP	Lipid transfer protein HC chr2:1536456-1535394 20130731
Medtr3g046350.1	-	Y	118	32	LTP	Lipid transfer protein HC chr3:15314030-15314383 20130731
Medtr3g046443.1	-	Y	128	29	LTP	Lipid transfer protein HC chr3:15334099-15334482 20130731
Medtr3g046515.1	-	Y	121	31	LTP	Lipid transfer protein HC chr3:15378643-15379005 20130731
Medtr3g046530.1	-	Y	123	31	LTP	Lipid transfer protein HC chr3:15389565-15390140 20130731
Medtr3g046540.1	-	Y	123	31	LTP	Lipid transfer protein HC chr3:15393902-15394477 20130731
Medtr3g046580.1	-	Y	401	31	LTP	Lipid transfer protein HC chr3:15406938-15410926 20130731
Medtr3g046590.1	-	Y	123	30	LTP	Lipid transfer protein HC chr3:15414469-15415045 20130731
Medtr3g056640.1	-	Y	123	30	LTP	Lipid transfer protein HC chr3:22662869-22662286 20130731
Medtr3g464110.1	Y	Y	174	30	LTP	Lipid transfer protein HC chr3:25757731-25758252 20130731
Medtr3g083760.1	-	Y	204	30	LTP	Lipid transfer protein HC chr3:37789330-37787885 20130731
Medtr3g106750.1	Y	Y	130	34	LTP	Lipid transfer protein HC chr3:49331065-49331877 20130731

Medtr4g027800.1	-	Y	117	39	LTP	Lipid transfer protein HC chr4:9812808-9811604 20130731
Medtr4g027840.1	-	Y	140	38	LTP	Lipid transfer protein HC chr4:9831935-9832433 20130731
Medtr4g069210.1	-	Y	210	24	LTP	Lipid transfer protein HC chr4:26015723-26017178 20130731
Medtr4g077180.1	-	Y	115	34	LTP	Lipid transfer protein HC chr4:29568981-29569791 20130731
Medtr4g101280.1	-	Y	144	37	LTP	Lipid transfer protein HC chr4:41833903-41833046 20130731
Medtr4g101310.1	-	Y	135	33	LTP	protease inhibitor/seed storage/LTP family protein HC chr4:41841657-41842329 20130731
Medtr4g101330.1	Y	Y	170	37	LTP	Lipid transfer protein HC chr4:41847596-41846675 20130731
Medtr4g113405.1	Y	Y	155	31	LTP	Lipid transfer protein HC chr4:46617359-46618354 20130731
Medtr5g006940.1	Y	Y	213	50	LTP	Lipid transfer protein HC chr5:1143135-1141543 20130731
Medtr5g006950.1	Y	Y	169	45	LTP	Lipid transfer protein HC chr5:1146399-1147714 20130731
Medtr5g011980.2	Y	Y	157	27	LTP	Lipid transfer protein HC chr5:3537211-3534586 20130731
Medtr7g072810.1	-	Y	126	29	LTP	Lipid transfer protein HC chr7:27158009-27158386 20130731
Medtr7g072900.1	-	Y	118	36	LTP	Lipid transfer protein HC chr7:27220239-27220592 20130731
Medtr7g072930.1	-	Y	150	31	LTP	Lipid transfer protein HC chr7:27227863-27228312 20130731

Medtr7g072960.1	Y	Y	134	30	LTP	Non-specificLipid-transfer protein, putative LC chr7:27235957-27236400 20130731
Medtr7g072993.1	-	Y	138	33	LTP	Lipid transfer protein HC chr7:27249081-27249494 20130731
Medtr7g073130.1	-	Y	128	33	LTP	Lipid transfer protein HC chr7:27304401-27304784 20130731
Medtr7g073150.1	Y	Y	134	35	LTP	Lipid transfer protein HC chr7:27312621-27313022 20130731
Medtr7g073170.1	-	Y	120	28	LTP	Lipid transfer protein HC chr7:27322859-27323614 20130731
Medtr7g082590.1	-	Y	96	28	LTP	Lipid transfer protein HC chr7:31659374-31658860 20130731
Medtr7g094650.1	-	Y	172	23	LTP	Lipid transfer protein related HC chr7:37733712-37734227 20130731
Medtr7g095230.1	Y	Y	185	42	LTP	Lipid transfer protein HC chr7:38109826-38108564 20130731
Medtr0071s0070.1	-	Y	123	29	LTP	Lipid transfer protein HC scaffold0071:47776-47200 20130731
Medtr1g012630.1	-	Y	189	60	LTP	Lipid transfer protein HC chr1:2578840-2578208 20130731
Medtr1g012690.1	-	Y	131	48	LTP	Lipid transfer protein HC chr1:2592421-2591957 20130731
Medtr1g012700.1	-	Y	141	49	LTP	Lipid transfer protein HC chr1:2600229-2599711 20130731
Medtr2g026775.1	-	Y	735	67	LTP	Lipid transfer protein HC chr2:9742223-9744709 20130731
Medtr5g011960.1	-	Y	133	31	LTP	Lipid transfer protein HC chr5:3531965-3531118 20130731

Medtr7g083580.1	-	Y	204	49	LTP	Lipid transfer protein HC chr7:32163030-32163979 20130731
Medtr3g040680.1	-	Y	499	27	PLC	Nonspecific phospholipase C4 HC chr31434311514337780 20130731 ; MatureChain: 32-530
Medtr3g435540.1		Y	344	25	GDSL	GDSL-like lipaseacylhydrolase HC chr31166912111665026 20130731 ; MatureChain: 21-364
Medtr4g108690.1	Y	Y	344	26	GDSL	GDSL-like lipaseacylhydrolase HC chr44501274445015362 20130731 ; MatureChain: 25-368
Medtr5g009100.1	-	Y	356	25	GDSL	GDSL-like lipaseacylhydrolase HC chr520936292091194 20130731 ; MatureChain: 27-382
Medtr5g078210.1	-	Y	332	26	GDSL	GDSL-like lipaseacylhydrolase HC chr53342332333421258 20130731 ; MatureChain: 22-353
Medtr8g087870.1	-	Y	328	25	GDSL	GDSL-like lipaseacylhydrolase HC chr83635812436355685 20130731 ; MatureChain: 24-351
Medtr2g058670.1	Y	Y	400	32	Pectin lyase	Polygalacturonase 11c, putative HC chr2:24244330-24242502 20130731
Medtr2g058840.1	Y	Y	398	32	Pectin lyase	Polygalacturonase 11c, putative HC chr2:24297949-24299598 20130731
Medtr1g086390.1	Y	Y	453	30	Pectin lyase	Polygalacturonaseglycoside hydrolase family protein HC chr13865976538662894 20130731 ; MatureChain: 25-477
Medtr2g032710.1	-	Y	440	28	Pectin lyase	Polygalacturonase plantlike protein HC chr21232230012315408 20130731 ; MatureChain: 24-463

Medtr2g032710.2	-	Y	440	28	Pectin lyase	Polygalacturonase plantlike protein HC chr21231930112315408 20130731 ; MatureChain: 24-463
Medtr2g035760.1	-	Y	473	28	Pectin lyase	Polygalacturonase QRT3like protein HC chr21515891615161418 20130731 ; MatureChain: 25-497
Medtr2g035790.1	Y	Y	465	24	Pectin lyase	Polygalacturonase QRT3like protein HC chr21517861015181077 20130731 ; MatureChain: 24-488
Medtr2g095080.1	-	Y	201	30	Pectin lyase	Ripening related protein family HC chr24061779540616664 20130731 ; MatureChain: 22-222
Medtr2g095130.1	-	Y	193	29	Pectin lyase	Ripening related protein family HC chr24062835540627563 20130731 ; MatureChain: 28-220
Medtr1g086880.1	-	Y	230	30	PMEI	Plant invertasepectin methylesterase inhibitor HC chr13889779338898557 20130731 ; MatureChain: 25-254
Medtr1g086890.1	-	Y	223	31	PMEI	Plant invertasepectin methylesterase inhibitor HC chr13889995538900695 20130731 ; MatureChain: 24-246
Medtr5g081840.1	-	Y	154	26	PMEI	Pectinesterase inhibitor domain protein LC chr53507993735079313 20130731 ; MatureChain: 22-175

Medtr8g042910.1	Y	Y	590	31	PMEI	Pectinesterase inhibitor HC chr81661053016607577 20130731 ; MatureChain: 27-616
Medtr8g093460.1	-	Y	543	31	PMEI	Pectinesterasepectinesterase inhibitor HC chr83904513539042219 20130731 ; MatureChain: 19-561
Medtr1g084820.2	-	Y	550	43	X8	Carbohydrate-binding X8 domain protein HC chr13775444737757434 20130731 ; MatureChain: 22-571
Medtr1g084820.1	Y	Y	612	43	X8	Carbohydrate-binding X8 domain protein HC chr13775439137757434 20130731 ; MatureChain: 22-633
Medtr1g116870.1	Y	Y	335	50	X8	Carbohydrate-binding X8 domain protein HC chr15278632252784446 20130731 ; MatureChain: 28-362
Medtr2g049670.1	-	Y	257	55	X8	Carbohydrate-binding X8 domain protein HC chr2:22307554-22304641 20130731
Medtr7g111600.1	-	Y	282	48	X8	Carbohydrate-binding X8 domain protein HC chr7:45815016-45812250 20130731
Medtr8g085390.1	-	Y	165	44	X8	Carbohydrate-binding X8
Medtr3g082080.1	-	Y	217	42	X8	Carbohydrate-binding X8
Medtr5g044530.1	-	Y	238	45	X8	Carbohydrate-binding X8
Medtr3g048280.1	-	Y	188	42	X8	Carbohydrate-binding X8
Medtr1g022160.5	-	Y	397	34	GH	Glucan endo1,3betaglucosidaselike protein HC chr168969886892247 20130731 ; MatureChain: 21-417

Medtr1g022160.1	Y	Y	461	34	GH	Glucan endo1,3betaglucosidaselike protein HC chr168969576893217 20130731 ; MatureChain: 21-481
Medtr1g022160.4	-	Y	466	34	GH	Glucan endo1,3betaglucosidaselike protein HC chr168969886892247 20130731 ; MatureChain: 21-486
Medtr1g022160.2	-	Y	397	34	GH	Glucan endo1,3betaglucosidaselike protein HC chr168969886892247 20130731 ; MatureChain: 21-417
Medtr1g022160.3	-	Y	459	34	GH	Glucan endo1,3betaglucosidaselike protein HC chr168969886892247 20130731 ; MatureChain: 21-479
Medtr3g065460.1	-	Y	473	31	GH	Glucan endo1,3betaglucosidaselike protein HC chr32959016729594105 20130731 ; MatureChain: 19-491
Medtr4g116990.1	-	Y	364	34	GH	Glycoside hydrolase family 18 protein HC chr44845622948454178 20130731 ; MatureChain: 22-385
Medtr6g079630.1	-	Y	282	26	GH	Chitinase HC chr62905852029059610 20130731 ; MatureChain: 25-306

Medtr6g083270.1	Y	Y	435	29	GH	Chitinase Hevein PR4 Wheatwin2 HC chr63117533831171217 20130731 ; MatureChain: 23-457
Medtr8g074335.1	-	Y	287	28	GH	Chitinase (Class Ib) Hevein HC chr83143287031435038 20130731 ; MatureChain: 23-309
Medtr8g074350.1	-	Y	297	31	GH	Chitinase (Class Ib) Hevein HC chr83144272231445255 20130731 ; MatureChain: 24-320
Medtr1g055415.1	Y	Y	322	40	Ole E	Pollen Ole e I family allergens HC chr12454144824539849 20130731 ; MatureChain: 22-343
Medtr4g017240.1	Y	Y	326	43	Ole E	Pollen Ole e I family allergens HC chr453417595343041 20130731 ; MatureChain: 22-347
Medtr3g012020.1	Y	Y	372	42	Ole E	Pollen Ole e I family allergens HC chr331762503174857 20130731 ; MatureChain: 22-393
Medtr4g066680.1	Y	Y	317	42	Ole E	Pollen Ole e I family allergens HC chr42522232025223534 20130731 ; MatureChain: 22-338
Medtr4g088735.1	Y	Y	255	42	Ole E	Pollen Ole e I family allergens HC chr43532759835325086 20130731 ; MatureChain: 22-276
Medtr4g088740.1	Y	Y	381	40	Ole E	Pollen Ole e I family allergens HC chr43533594335334462 20130731 ; MatureChain: 22-402
Medtr4g124400.1	Y	Y	189	47	Ole E	Pollen Ole e I family allergens HC chr45145743051458050 20130731 ; MatureChain: 18-206

Medtr5g047890.1	Y	Y	323	39	Ole E	Pollen Ole e I family allergens HC chr52096382020965177 20130731 ; MatureChain: 20-342
Medtr6g042490.1	Y	Y	233	37	Ole E	Pollen Ole e I family allergens HC chr61467528214674127 20130731 ; MatureChain: 23-255
Medtr7g016950.1	-	Y	445	41	Ole E	Pollen protein Ole E Ilike protein HC chr753588465361469 20130731 ; MatureChain: 21-465
Medtr8g092230.1	Y	Y	275	37	Ole E	Pollen protein Ole E Ilike protein LC chr83854271538541582 20130731 ; MatureChain: 22-296
Medtr3g089570.1	-	Y	242	49	Ole E	Pollen Ole e I family allergens HC chr31629957216301945 20130731 ; MatureChain: 31-272
Medtr4g046767.1	-	Y	257	35	Ole E	Pollen Ole e I family allergens HC chr4:16591718-16592799 20130731
Medtr4g087890.1	-	Y	870	33	Formin	Formin -like 2 domain protein HC chr4:34465056-34468637 20130731
Medtr4g095780.1	-	Y	838	34	Formin	Formin -like 2 domain protein HC chr4:39940525-39944800 20130731
Medtr1g083260.1	Y	Y	884	29	Formin	Formin-like 2 domain protein HC chr13705554137051517 20130731 ; MatureChain: 26-909
Medtr4g081410.2	-	Y	965	35	Formin	Formin-like 2 domain protein HC chr43156074231567076 20130731 ; MatureChain: 26-990
Medtr4g081410.1	Y	Y	965	35	Formin	Formin-like 2 domain protein HC chr43156145031566634 20130731 ; MatureChain: 20-984

Medtr8g027995.1	-	Y	1049	38	Formin	Formin-like 2 domain protein HC chr81032825710323250 20130731 ; MatureChain: 23-1071
Medtr1g021630.1	Y	Y	656	24	RLK	Cysteine-rich receptor kinase-like protein LC chr165073336502386 20130731 ; MatureChain: 21-676
Medtr1g021635.1	-	Y	653	24	RLK	Cysteine-rich receptor kinase-like protein LC chr165176436514070 20130731 ; MatureChain: 22-674
Medtr1g069340.1	-	Y	732	29	RLK	Receptor-like SerineThreoninekinase ALE2 HC chr12995341529947213 20130731 ; MatureChain: 25-756
Medtr1g105615.1	-	Y	479	27	RLK	Cysteine-rich receptor kinase-like protein HC chr14760424747607685 20130731 ; MatureChain: 28-506
Medtr1g105615.2	-	Y	479	27	RLK	Cysteine-rich receptor kinase-like protein HC chr14760424247607699 20130731 ; MatureChain: 28-506
Medtr1g105650.1	Y	Y	632	29	RLK	Cysteine-rich receptor kinase-like protein HC chr14763442647637645 20130731 ; MatureChain: 30-661
Medtr2g016620.1	Y	Y	413	34	RLK	LRR Receptor-like kinase HC chr251246475122806 20130731 ; MatureChain: 25-437 -- SKETCHY
Medtr2g039290.3	-	Y	1206	40	RLK	Receptor-like SerineThreoninekinase ALE2 HC chr21722321317211976 20130731 ; MatureChain: 20-1225

Medtr2g039290.1	-	Y	1428	43	RLK	Receptor-like SerineThreoninekinase ALE2 HC chr21722321317211980 20130731 ; MatureChain: 20-1447
Medtr2g039290.2	-	Y	1440	43	RLK	Receptor-like SerineThreoninekinase ALE2 HC chr21722321317211976 20130731 ; MatureChain: 20-1459
Medtr3g075440.1	Y	Y	603	24	RLK	LRR Receptor-like kinase family protein HC chr33434281834340379 20130731 ; MatureChain: 26-628
Medtr3g093710.1	Y	Y	609	28	RLK	Receptor-like kinase HC chr34281500242818320 20130731 ; MatureChain: 27-635
Medtr3g093930.1	-	Y	955	26	RLK	Leucine-rich Receptor-like kinase family protein HC chr34291616742919917 20130731 ; MatureChain: 35-989
Medtr4g074080.1	-	Y	647	25	RLK	Receptor-like kinase HC chr42815490728162503 20130731 ; MatureChain: 28-674
Medtr4g085480.1	-	Y	432	26	RLK	SerineThreonine kinase, planttype protein, putative HC chr43334491033346941 20130731 ; MatureChain: 29-460
Medtr4g113100.1	-	Y	630	30	RLK	LRR Receptor-like kinase HC chr44647588646479891 20130731 ; MatureChain: 26-655
Medtr4g126270.1	-	Y	957	35	RLK	Receptor-like SerineThreoninekinase ALE2like protein, putative HC chr45251125652521195 20130731 ; MatureChain: 24-980

Medtr5g042440.1	-	Y	602	22	RLK	LysM-domain Receptor-like kinase HC chr51865454718659757 20130731 ; MatureChain: 31-632
Medtr5g068210.1	Y	Y	579	26	RLK	LRR Receptor-like kinase HC chr52883982328842750 20130731 ; MatureChain: 26-604
Medtr7g082310.1	-	Y	416	28	RLK	LRR Receptor-like kinase LC chr73155048131552209 20130731 ; MatureChain: 24-439
Medtr7g115740.2	-	Y	567	24	RLK	Lectin receptor kinase HC chr74787018447872190 20130731 ; MatureChain: 26-592
Medtr7g115740.1	-	Y	643	24	RLK	Lectin receptor kinase HC chr74787018447872190 20130731 ; MatureChain: 26-668
Medtr8g016330.1	-	Y	839	30	RLK	Receptorlike SerineThreoninekinase ALE2 HC chr855110115505857 20130731 ; MatureChain: 18-856
Medtr8g072590.1	-	Y	638	23	RLK	SerineThreonine kinase domain protein LC chr83068323930686420 20130731 ; MatureChain: 22-659
Medtr0341s0020.1	Y	Y	643	25	RLK	Cysteinerich receptor kinase-like protein LC scaffold034190124034 20130731 ; MatureChain: 21-663
Medtr2g012670.1	-	Y	686	29	RLK	Strubbelig receptor family 3 protein HC chr2:3236861-3230480 20130731
Medtr3g071480.1	-	Y	623	28	RLK	LRR receptor-like kinase HC chr3:32070026-32066014 20130731

Medtr4g046113.1	-	Y	815	31	RLK	strubbelig receptor family protein HC chr4:16201157-16193578 20130731
Medtr3g072410.1	-	Y	391	37	RLP	LysM domain GPIanchored protein HC chr33255362932557165 20130731 ; MatureChain: 22-412
Medtr7g104360.1	-	Y	401	28	PPA	Purple acid phosphatase superfamily protein HC chr74229032342293494 20130731 ; MatureChain: 27-427
Medtr1g039270.1	-	Y	450	27	Protease	Aspartic protease in GUARD CELLlike protein HC chr11451668814514095 20130731 ; MatureChain: 20-469
Medtr2g016180.1	Y	Y	464	26	Protease	Papain family cysteine protease HC chr248997504901997 20130731 ; MatureChain: 27-490
Medtr2g016270.1	-	Y	449	26	Protease	Papain family cysteine protease HC chr249471444949388 20130731 ; MatureChain: 27-475
Medtr4g084020.2	-	Y	428	25	Protease	Trypsin family protein HC chr43270788232704601 20130731 ; MatureChain: 29-456
Medtr8g075010.1	-	Y	504	30	Protease	Eukaryotic aspartyl protease family protein HC chr83170414231707373 20130731 ; MatureChain: 24-527
Medtr7g113740.1	-	Y	136	23	Protease	Signal peptidase complex subunit 3Blike protein HC chr74687855146875326 20130731 ; MatureChain: 32-167

Medtr4g036590.1	Y	Y	234	47	Hist-rich	Transmembrane protein, putative LC chr41324280013243567 20130731 ; MatureChain: 22-255
Medtr3g012570.1	Y	Y	276	50	Hist-rich	Hypothetical protein LC chr335494983548605 20130731 ; MatureChain: 22-297
Medtr4g019600.1	-	Y	131	37	Hist-rich	Transmembrane protein, putative LC chr461712696171930 20130731 ; MatureChain: 22-152
Medtr6g016930.1	-	Y	138	25	Leguminosin	Leguminosin group486 secreted peptide HC chr666130336613533 20130731 ; MatureChain: 29-166
Medtr6g016970.1	-	Y	116	25	Leguminosin	Leguminosin group486 secreted peptide HC chr666317136632147 20130731 ; MatureChain: 29-144
Medtr6g089320.1	-	Y	138	24	Leguminosin	Leguminosin group486 secreted peptide HC chr62282355522823055 20130731 ; MatureChain: 29-166
Medtr1g007880.2	-	Y	322	32	Chimeric	ABIL1like protein HC chr1609028606935 20130731 ; MatureChain: 21-342
Medtr1g021638.1	Y	Y	292	28	Chimeric	salt stress responseantifungal domain protein LC chr165212756522643 20130731 ; MatureChain: 22-313
Medtr2g042145.1	-	Y	92	28	Chimeric	Thionin related (TAP1) HC chr21839638118396835 20130731 ; MatureChain: 22-113

Medtr2g042163.1	-	Y	92	29	Chimeric	Thionin related (TAP1) HC chr21840230018402754 20130731 ; MatureChain: 22-113
Medtr0341s0010.1	Y	Y	287	29	Chimeric	salt stress responseantifungal domain protein LC scaffold03411358435 20130731 ; MatureChain: 21-307
Medtr0721s0020.1	-	Y	129	30	Chimeric	Transmembrane protein, putative LC scaffold072163305643 20130731 ; MatureChain: 28-156
Medtr5g024983.1	-	Y	99	29	Chimeric	Transmembrane protein, putative HC chr51010507610108942 20130731 ; MatureChain: 24-122
Medtr3g028560.1	-	Y	165	30	Chimeric	Transmembrane protein, putative LC chr391091959112249 20130731 ; MatureChain: 21-185
Medtr3g064040.1	-	Y	164	38	Chimeric	Transmembrane protein, putative HC chr32879461928795716 20130731 ; MatureChain: 28-191
Medtr2g090660.1	Y	Y	219	24	Chimeric	Hypothetical protein HC chr23881868938817390 20130731 ; MatureChain: 25-243
Medtr1g041410.1	-	Y	192	29	Chimeric	neurogenic locus notchlike protein HC chr11549015515486147 20130731 ; MatureChain: 23-214
Medtr2g005930.1	-	Y	443	30	Chimeric	DUF1682 family protein HC chr2384861381814 20130731 ; MatureChain: 26-468

Medtr3g095330.1	-	Y	114	31	Chimeric	ECA1 gametogenesis related family HC chr34355892943559330 20130731 ; MatureChain: 20-133
Medtr3g108760.1	Y	Y	240	45	Chimeric	surface protein, putative HC chr35023456550235467 20130731 ; MatureChain: 21-260
Medtr4g134220.1	-	Y	314	31	Chimeric	nicotiana tabacum ORF protein HC chr45619120856189400 20130731 ; MatureChain: 20-333
Medtr0002s0150.1	Y	Y	574	29	Chimeric	NAD(P)Hquinone oxidoreductase subunit H HC scaffold0002165003167020 20130731 ; MatureChain: 23-596
Medtr3g028560.1	-	Y	165	30	Chimeric	Transmembrane protein, putative LC chr391091959112249 20130731 ; MatureChain: 21-185
Medtr3g064040.1	-	Y	164	38	Chimeric	Transmembrane protein, putative HC chr32879461928795716 20130731 ; MatureChain: 28-191
Medtr5g081820.1	-	Y	203	25	Chimeric	Transmembrane protein, putative LC chr53507489035073090 20130731 ; MatureChain: 22-224
Medtr3g106510.1	-	Y	553	30	Chimeric	Transmembrane protein, putative HC chr34922409749228278 20130731 ; MatureChain: 23-575
Medtr3g030640.1	-	Y	245	30	Chimeric	ribonuclease T2 family protein HC chr397132629712341 20130731 ; MatureChain: 24-268

Medtr4g107400.1	-	Y	672	31	Chimeric	hipl2 protein precursor HC chr44441514444409998 20130731 ; MatureChain: 26-697
Medtr5g022810.1	-	Y	208	21	Chimeric	ribonuclease T2 family protein LC chr590374259038278 20130731 ; MatureChain: 22-229
Medtr2g030825.1	Y	Y	574	48	Chimeric	Late embryogenesis abundant (LEA)like protein LC chr21165398211656243 20130731 ; MatureChain: 22-595
Medtr2g030845.1	Y	Y	575	48	Chimeric	Late embryogenesis abundant (LEA)like protein LC chr21166775111669978 20130731 ; MatureChain: 22-596
Medtr2g079030.1	-	Y	522	36	Chimeric	MATE subfamily protein LC chr23309434233097785 20130731 ; MatureChain: 23-544
Medtr3g062840.1	-	Y	286	26	Chimeric	plantF17O147 protein HC chr32840098128399252 20130731 ; MatureChain: 26-311
Medtr1g006460.1	-	Y	204	55	Chimeric	Hypothetical protein LC chr12147148221472159 20130731 ; MatureChain: 22-225
Medtr5g043870.1	-	Y	207	59	Chimeric	Hypothetical protein LC chr51926285919262173 20130731 ; MatureChain: 22-228
Medtr6g007913.1	-	Y	119	32	Chimeric	GASA/GAST/Snakin HC chr6:2139621-2140861 20130731
Medtr1g088970.1	-	Y	140	45	Chimeric	Transmembrane protein, putative HC chr1:39956709-39955720 20130731
Medtr7g073280.1	-	Y	343	29	Chimeric	Transmembrane protein, putative HC chr7:27366686-27373900 20130731

Medtr1g069235.1	-	Y	128	43	Chimeric	photosystem I reaction center subunit IV A HC chr1:29816886-29819066 20130731
Medtr3g064510.1	-	Y	267	30	Chimeric	expansin A1 HC chr3:29068579-29070728 20130731
Medtr7g014880.1	-	Y	166	33	Chimeric	RALF HC chr7:4453529-4456208 20130731
Medtr4g084950.1	-	Y	136	42	Chimeric	Transmembrane protein, putative

Primer	Sequence
MtFH3-F	ATGGGAATTGAGAGATATATGGTT
MtFH3-R	TGTTGTTTTATTACTATGGTTATCTTC
AtFH3-F	CACCATGGGGAGATTGAGATTAGCG
AtFH3-R	CATCAAGATATCATCCTTCTTCTC
ATAle2-F	CACCATGCGGAACTTTGCGATGCTT
ATAle2-R	CGTTCTAAAACCTATTCCCTGG

Table III. Primers used in this study.

A Small Family of Arabinogalactan Peptides Necessary for Symbiosis in *Medicago truncatula*

William M. Moore[1,2,3] and Henrik V. Scheller[1,2,3]
[1]Joint BioEnergy Institute, Emeryville, California 94608
[2]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720
[3]Department of Plant and Microbial Biology, University of California, Berkeley, California 94720

Introduction

Arabinogalactan proteins (AGPs) are a diverse group of cell surface glycoproteins conserved in plants that are involved in a myriad of cellular and physiological processes including cell expansion (1, 2), proliferation (3), differentiation (4), and programmed cell death (5). Collectively they are characterized by extensive protein glycosylation with hydroxyproline *O*-linked arabinogalactan (AG) polysaccharides, which have a high degree of structural complexity and diversity with respect to branching and sugar composition. AGPs can be cross-linked into the plant cell wall by their carbohydrate moieties (6), tethered to the plasma membrane by a glycosylphosphatidylinositol (GPI) lipid anchor (7–9), or released as diffusible factors abundant in plant cell exudates (10–12). Many AGPs contain GPI anchors, which may allow them to mediate signaling between the plasma membrane and cell wall, organize into membrane microdomains, act as co-receptors, or as cleavable signals susceptible to phospholipase C (13).

Individual AGP-encoding genes often have developmentally regulated, cell-specific, or inducible expression patterns in plants (14–16). While AGP function is not well understood, however mounting evidence points to roles in positional cell signaling for many of these glycoproteins. Indeed, a number of AGPs purified from cell culture media have morphogenic properties linked to cell identity and fate determination. A prime example of this is xylogen, a chimeric AGP that promotes xylem cell differentiation *in vitro* (17). Xylogen proteins localize to the polar cell-cell junctions of differentiating vessel elements and *Arabidopsis thaliana* xylogen mutants have discontinuous and incomplete xylem (4). Therefore, xylogen is believed to mediate cell-cell induction of vessel member identity during vascular development. Similar positional signaling roles have been reported for AGPs during somatic embryogenesis. Somatic embryos develop through an initial asymmetric cell division that gives rise to a mother cell and an embryogenic daughter cell. Monoclonal antibody JIM8 recognizes an AGP epitope secreted by the mother cell that is required as a factor for the embryogenic differentiation of the daughter cell *in vitro* (18), which provides additional evidence to support the role of AGPs in autonomous and non-autonomous cell-cell signaling pathways. Lastly, free oligosaccharides released from ovule-specific AGPs have been shown to act as diffusible signals necessary for pollen tube attraction and targeting (19, 20). Activity of the purified AG oligosaccharide was dependent on the terminal rhamnose and glucuronic acid residues (21), indicating that AG sugar composition and structure has intrinsic signaling properties in plants.

While AGPs integrate a variety of developmental cues, they also mediate response to environmental factors (22) and have been implicated in a variety of plant-microbe interactions (23). The *A. thaliana agp17* mutant is resistant to *Agrobacterium tumifaciens* transformation, which have poor adherence to the plant cell surface (24). AGPs from *Pisum sativa* induce a novel type of polarized attachment by *Sinorhizobium meliloti in vitro*, while pectin and xylan do not (25). AGPs have similar effects in promoting *Bacillus subtilis* biofilm formation in polarized arrays (26). This indicates that bacteria specifically interact with AGP epitopes present in plant cell surfaces. AGPs are enriched in the secretory border-like cells of the root cap and are abundant in root exudates (27). AGPs secreted in root exudates have previously been shown to act as chemotactic attractants for *Aphanomyces* zoospores that prevent spore germination and promote cyst formation (28). This suggests that secreted AGPs may play a role in attracting, repelling, or modulating microbial communities in the rhizosphere.

AGPs have also been linked to variety of plant-microbial endosymbiosis encompassing a wide range of microorganisms and plant hosts. Monoclonal antibodies recognizing AGP carbohydrate epitopes indicate that AGPs accumulate in the plant-derived symbiotic membranes that encapsulate the endosymbiont. This has been demonstrated in *Rhizobium* symbiosis on *Pisum sativum* (29), *Frankia* symbiosis on *Alnus* spp (30), *Nostoc* symbiosis on *Gunnera manicata* and *Blasia pusilla* (31), and the arbuscular mycorrhizal (AM) symbiosis with a variety of plant hosts (32, 33). These membranes have a distinct identity based on their protein composition and create dynamic interfaces between the plant and microorganism that function in the bidirectional exchange of nutrients and information (34–37). Moreover, these compartments control the growth and development of the endosymbiont and are essential for their persistence inside plant cells(38, 39).

The accumulation of AGPs in symbiotic plant-microbial interfaces across a wide range of endosymbiosis has suggested that these glycoproteins may be important for mediating interactions between plant cells and endosymbionts. However, the identity of these AGPs and their corresponding specificity for symbiosis has remained unknown, largely due to the research challenges that accompany these enigmatic glycoproteins (40). Here have used the bioinformatic tool, AGPredictor, to identify AGP-encoding genes in the *Medicago truncatula* genome, which has been used a model plant species for the molecular study of *Rhizobium*-legume symbiosis and AM symbiosis. Cross-referencing predicted AGPs to the *M. truncatula* Genome Expression Atlas identified a small family of SYMBIOSIS-ASSOCIATED ARABINOGALACTAN PEPTIDES (SAPs) differentially expressed in *S. meliloti* infected root nodules and roots colonized by the AM fungus *Rhizphagus irregularis.* Here we show that SAPs localize to symbiotic membranes and are necessary for the differentiation of *S. meliloti* and *R. irregularis* within these compartments, indicating that SAPs are essential components of these interfaces.

Materials and Methods

Plant Material and Hairy Root Transformation

M. truncatula Gaertn. cv Jemalong A17 (Noble Foundation) was used for all experiments. Hairy root transformation was conducted according Floss *et al.* (41) using

Rhizobium rhizogenes (formerly *Agrobacterium rhizogenes*) strain Arqua-1. Transgenic roots were selected on Faehrus medium containing kanamycin (25 mg/ml) for two weeks and adventitious roots emanating above the wound site were removed prior to transplanting into sterile sand. Transformants were allowed to recover for one week prior to inoculation with either *S. meliloti* or *R. irregularis.* All plants were grown under 16 h photo period at 22°C and fertilized twice a week according to the microbial inoculant used.

Nodulation Experiments

For nodulation experiments roots were flood inoculated with 5 ml (OD 0.1) resuspension cultures of wild-type *S. meliloti* strain 1021 (ATCC 51124). For live cell imaging the *S. meliloti* 1021 strain expressing mCherry from Haney and Long (42) was used. Plants were fertilized twice a week with half strength Murashige and Skoog (MS) media without nitrogen (PhytoTechnology Laboratories). Nodules were assayed at 3 wpi.

Mycorrhizal Experiments

Axenic *R. irregularis* spores (DAOM 181602) were purchased from Premier Tech (Rivière-du-Loup, Québec, Canada) and used for all experiments. Approximately 1200 spores were pre-germinated overnight in dilute root exudate at room temperature with gentle rocking and applied directly to roots during transplanting. Plants were grown in sterile sand and fertilized twice a week with half-strength Hoaglands Basal Salt media modified with 20 µM phosphate (43). Root mycorrhization with *R. irregularis* was examined at 4 wpi by confocal microscopy.

Cloning

Single RNAi constructs targeted a 400 bp region from each of the three SAPs, which encompassed the entire coding sequence including part of the 5’ and 3’ UTR. SAPs share little sequence similarity and target sequences were determined to be specific based on BLAST search. Targets were amplified by PCR from cDNA prepared from AMF colonized *M. truncatula* roots and cloned into Gateway (GW) entry vector pENTR1A using XbaI and XhoI restriction sites. Resulting entry clones were recombined by LR reaction into GW RNAi destination vector pHELLSGATE12 (44). The *SAP1+2+3* triple RNAi target sequence was generated by DNA synthesis and was 600 bp in length, consisting of three 200 bp sequences from each SAP concatenated together. The *SAP1+2-RNAi* construct was 400 bp and amplified from the *SAP1+2+3* template sequence. Both sequences were subcloned into pENTR1A and recombined into pHELLSGATE12 as previously described.

For SAP-promoter GUS-reporter constructs a 2 kb genomic sequence including the 5’UTR from each peptide was cloned into GW TOPO-entry vector pCR8 (Invitrogen). Resulting entry clones were recombined by LR reaction into GUS expression vector pGWB3 (45) to generate pSAP:GUS expression constructs. SAP localization constructs using GFP and mCherry were assembled by In-Fusion cloning into the pCR8 vector backbone using sequences amplified from genomic DNA. All localization constructs used the native *SAP* promoter and terminator sequences corresponding to each peptide. Resulting entry clones were recombined into the plant transformation vector pGWB1 by LR reaction.

For western blot *AtAGP9* was subcloned without stop codon into GW entry vector pENTR1A and recombined by LR reaction into pEARLYGATE101 to generate *35S:AtAGP9-YFP-HA* expression constructs. The full *SP-mCherry-SAP2* coding sequenc was subcloned from the native localization construct into pGWB17 to generate *35S:SP-mCherry-SAP2* expression constructs. All constructs in this study were confirmed by sequencing and primers are listed in Table II.

RNA Extraction and qPCR

Total RNA was extracted from root using Trizol (Thermo Fisher), treated with Turbo DNAseI (Thermo Fisher), and further purified using an RNEAsy column (Qiagen). Resulting cDNA was prepared using iScript cDNA synthesis kit (BioRad) using 500 ng of total RNA. *SAP* qRT-PCR primers were determined to be gene-specific by BLAST search. *SAPS* do not have introns, so it was not possible to place primers at exon-exon junctions. Gene expression was quantified using the $\Delta\Delta C_T$ method relative to the housekeeping gene EF-1α.

Western blot

Plant tissues were ground in liquid nitrogen and total protein was extracted by boiling in 2X SDS-PAGE buffer for 10 minutes. Proteins were run on a precast TGX-mini protein gel with 16-8% gradient in SDS-PAGE buffer. Proteins were transferred to PVDF membranes, blocked overnight in 3% BSA TBST, probed for one hour with either anti-HA or anti-mCherry mouse primary antibody (1:5,000), followed by anti-mouse horse radish peroxidase-conjugated secondary antibody (1:10,000), with subsequent washes with 3% BSA TBST in between. Signal was detected by chemiluminescence on an Amersham imager using SuperSignal™ luminol HRP substrate (Thermo Fisher). AtAGP9-YFP-HA and mCherry-SAP2 fusion proteins were transiently expressed in *Nicotiana benthamiana* by leaf infiltration and collected 4 dpi. The mCherry-SAP2 fusion protein was also expressed under its native promoter in *M. truncatula* hairy roots colonized *R. irregularis* and collected 4 wpi.

Acetylene Reduction Assay

Nitrogen fixation was assayed in silenced roots 4 weeks post inoculation with *S. meliloti*. Chimeric plants were gently uprooted from sand and placed on a strip of Whatman filter paper (pre-wet with 1/2x MS without nitrogen) inside a 25 ml culture tube topped with a rubber septum. Acetylene (1 ml) was injected into the culture tube, which was placed in a growth chamber for 6 h prior to sampling. Gas samples were measured using gas chromatography flame ionization detection on a Shimadzu GC-2014.

Histochemistry

Nodules were fixed and embedded in Technovit 7100 plastic resin according to the manufacturer’s directions. A Reichardt microtome and glass knife was used to cut longitudinal nodule sections approximately 6 µm thick, which were subsequently stained with toluidine blue-O and imaged under bright field with a Leica DM6B fluorescence microscope. Routine observations of nodules were made using fresh hand sections stained with Syto13 for nucleic acids (Life Technologies) and imaged by confocal microscopy. AM colonized roots were stained with wheat germ agglutinin AlexaFluor 488 (Thermo Fisher) as previously described. GUS staining of SAP-reporter was done as previously described (46).

Confocal and Light Microscopy

Electron Microscopy

Nodules were directly harvested into 0.1 M cacodylate buffer (pH 7.2) containing 2% (v/v) EM-grade glutaraldehyde and fixed overnight at 4°C. The tissue was rinsed 3× with 0.1 M cacodylate buffer (pH 7.2), stained for 1 h with 1% (w/v) osmium tetraoxide, rinsed 3× with buffer, and followed by an additional 3 rinses with distilled water. Nodules were dehydrated in an acetone gradient (35/50/70/80/95/100% v/v) and bisected to aid infiltration. Nodules were infiltrated with acetone:Epon resin at 2:1, 1:1, and 1:2 dilution ratios for 1 h each followed by pure Epon. The resin was freshly changed after 1 h and allowed to infiltrate overnight with gentle rocking. Samples were then infiltrated with pure resin plus accelerator for two hours and embedded in Pelco molds, which were left in a 65°C oven to polymerize for two days. Ultra-thin 70 nm ultrathin sections were collected on grids using a diamond knife and Reichert microtome. Grids were stained with 2% (v/v) aqueous uranyl acetate for 5 min, rinsed with distilled water 5×, stained with lead citrate for 5 min, and rinsed another 5× with distilled water using a Pelco Grid Staining System. Grids were imaged using a Technai 1200 electron microscope. At least three nodules were imaged per construct.

Results

Identifying *M. truncatula* AGPs Expressed in AM Symbiosis and Nodulation

Previous publications have shown that AGPs aggregate at symbiotic plant-microbial interfaces through the use of carbohydrate-directed monoclonal antibodies (30–33, 47, 48). However, the identity of these proteins and their specificity for these symbioses is not known. To investigate this we developed a new bioinformatic search tool to identify AGP-encoding genes in the *M. truncatula* genome (Chapter 1), which is a model plant for the molecular study of both arbuscular mycorrhizal and rhizobia-legume symbioses. All predicted AGP encoding genes were cross-referenced to the *Medicago truncatula* Gene Expression Atlas (MtGEA) (49, 50), which returned 191 Affymetrix probe set IDs that were used for global expression profiling in *M. truncatula* roots. Microarray data from roots treated with either beneficial microorganisms (*S. meliloti* and *R. irregularis*) or pathogenic organisms (*Ralstonia solanacearum*, *Macrophomina phaseolina*, *Phymatotrichopsis omnivora*) was used to generate a heatmap of AGP expression plotted as log 2 fold-change relative to uninoculated experimental controls (Fig. 1). Using this method we found two clusters of AGPs exclusively expressed in either AM roots or root nodules (Fig. 1, Table I). AM-inducible AGPs include five plastocyanin-like AGPs including Blue Copper-binding Protein 1 (BCP1) and BCP2, which have historically been used as markers for AM colonization, but have not previously been reported as AGP. We also identified one chitinase-like AGP, one lipid transfer protein AGP and two small AGP peptides. One of these peptides

(Medtr5g035640) was previously described as ARBUSCULAR MYCORRHIZAL AGP 1 (AMA1), which was identified in a cDNA library prepared from AM colonized roots. Nodule-inducible AGPs included three plastocyanin-like AGPs (including ENOD16 and ENOD20), one chitinase-like AGP, one lipid transfer protein AGP, and one AG peptide – similar to the AGP encoding genes expressed in AM symbiosis. We also identified two pectin lyase-like AGP, one LRR AGP, and two classical AGPs expressed in root nodules.

The Symbiosis-association AG Peptide (SAP) Gene Family

Interestingly, the 3 small AGP peptides we identified form a small family of tandemly duplicated genes on chromosome 5 (Fig. 2 *A*) but are differentially expressed during AM symbiosis and nodulation (Fig. 1; Fig. 2 *C*). AMA1 is the founding member of this family, however, we collectively refer to this group as Symbiosis-associated AGP Peptides (SAP1, Medtr5g035640; SAP2, Medtr5g035650; SAP3, Medtr5g35610) to better reflect their expression pattern. All three SAP peptides have a conserved *N*-terminal secretion signal, a *C*-terminal GPI-anchor signal, and are predicted to be between 15-29 amino acids in length after signal peptide cleavage (Fig. 2 *B*). Even though SAPs occur as linear tandem gene duplications they share very little sequence homology with less than 30% identity conserved at the amino acid level and less than 18% at the nucleotide level. In total only 5 amino acids are conserved between the mature peptide sequences. These include two non-contiguous proline residues predicted to be glycosylated with AG, a serine residue that serves as the GPI anchor addition site, and two interspersed glycine residues near the *N*-terminus (Fig. 2 *B*).

SAP Expression Patterning in *M. truncatula* roots

SAP peptides were identified based on their symbiosis-induced expression patterns using publicly available transcriptomic data. To validate this we first used qPCR to measure SAP gene expression in *M. truncatula* roots 4 weeks post inoculation (4 wpi) with *R. irregularis* or *S. meliloti* (Fig 2 *C*). Results show that SAP1 and SAP2 were only expressed in roots after inoculation with *R. irregularis* and are specific for AM symbiosis. However, SAP2 was expressed at a lower level. SAP3 was strongly upregulated during nodulation with *S. meliloti* but was also expressed in uninoculated roots and was not transcriptionally responsive to either *R. irregularis* inoculation or nutrient deficiency.

We further investigated the spatial expression pattern of these peptides by expressing *SAP*-promoter *GUS*-reporter constructs in transgenic hairy roots. Roots expressing *SAP1* reporter constructs inoculated with *R. irregularis* had dark GUS staining specific to cortical root cells containing arbuscules (Fig. 3 *B*). *SAP2* reporter constructs also showed AM-dependent expression in the inner root cortex, however, due to the weak staining it was unable to determine if it was arbuscule-specific based on staining alone (Fig. 3 *D*). Staining was not observed in mock-inoculated roots (Fig. 3 *A, C*). Roots expressing *SAP3* reporter constructs had strong GUS staining in root nodules (Fig. 4 *B*). Nodule cross-sections showed that staining predominantly occurred in the transition zone and early nitrogen fixation zone associated with bacterial growth and differentiation (Fig. 4 *E*). Furthermore, staining was specific to cells infected with *S. meliloti* and did not occur in the uninfected cells (Fig. 4 *E*). Roots expressing *SAP3* reporter constructs also showed GUS staining in AM colonized tissue but was not restricted to arbuscule-containing cells of the inner root cortex (Fig. 4 *D*). GUS staining was also present in mock-inoculated control roots, however the spatial distribution was unexpectedly different in response to nitrogen or phosphate starvation (Fig. 4 *A, C*), which had not previously been resolved by qPCR measurements of whole root systems (Fig 2 *C*).

SAP3 is Required for Nodule Development and Function

The expression pattern of SAP3 in root nodules suggested that this peptide might function in nodule development. To investigate this, we took a reverse-genetics approaching using RNAi-mediated gene silencing to knockdown *SAP3* expression in transgenic roots. Composite plants expressing either *SAP3-RNAi* or *GUS-RNAi* control constructs were grown without nitrogen and assayed 3 wpi with *S. meliloti*. Plants expressing *GUS-RNAi* constructs had green shoots and developed bright pink root nodules indicative of active symbiotic nitrogen fixation (Fig. 5 *A, C, E*). In contrast the nodules formed on *SAP3-RNAi* roots remained small, white, and did not accumulate the pink leghemoglobin seen in control nodules (Fig. 5 *D*). At 4 wpi, the shoots of *SAP3-RNAi* composite plants became highly chlorotic, a characteristic sign of nitrogen starvation, which suggested nitrogen fixation was impaired (Fig. 5 *A*). We measured nitrogen fixation efficiency by acetylene reduction assay and confirmed that symbiotic nitrogen fixation is strongly reduced in *SAP3-RNAi* nodules (Fig. 5 *I*). *SAP3* expression was measured by qPCR and confirmed that the transcript is efficiently silenced by the *SAP3-RNAi* construct in nodulated roots (Fig. 5 *G*). Functional nodules rarely ever formed on *SAP3-RNAi* root systems, however, the total number of nodules formed was unaffected (Fig. 5 *H*). This indicates that *SAP3* is required for nodule development and function, but not for early signaling events necessary for nodule initiation.

We further investigated the developmental defects in *SAP3-RNAi* nodules using a combination of light microscopy, confocal microscopy, and transmission electron microscopy. Plastic embedded longitudinal sections stained with Toluidine Blue-O show that *SAP3-RNAi* nodules have fewer infected cells relative to control nodules and the symbiosomes within these cells appear poorly developed with abnormal spaces occurring between the bacteria (Fig 6). *SAP3-RNAi* nodules also accumulated compound starch in the adjacent uninfected cells (Fig. 6). This suggested that *S. meliloti* might be impaired in growth and differentiation within the symbiosome compartment. We investigated this further using live cell imaging of *S. meliloti* expressing mCherry in fresh hand-sectioned nodules. Control nodules were tightly packed with fluorescent elongated bacteroids (Fig. 7). In *SAP3-RNAi* nodules *S. meliloti* rapidly lost fluorescence upon cell elongation and bacteroid differentiation. Starch accumulation was more apparent in fresh hand sections, which completely filled the adjacent uninfected cells (Fig. 8). Transmission electron microscopy revealed large spaces between the symbiosome membrane and *S. meliloti*, suggesting that the bacteria are not able to interact with these interfaces (Fig. 9).

SAP3 Localization in Root Nodules

SAP3 was localized in root nodules by expressing a GFP-SAP3 fusion protein under control of the native SAP3 promoter and terminator sequence. The GFP coding sequence was inserted between the SAP3 secretion signal and the mature peptide sequence to ensure proper targeting and propeptide processing. Roots expressing pSAP3:SP-GFP-SAP3-GPI constructs were inoculated with *S. meliloti* and nodules were examined by confocal laser scanning microscopy at 3 wpi (Fig. 10). The GFP-SAP3 fusion protein localized to symbiosome membranes in the transition and early nitrogen fixation zones of the nodule (Fig. 10 *A, B*) but appears to be less present in mature bacteroids (Fig. 10 *C*). Therefore, SAP3 labeling appeared to be transient and associated with the growth and differentiation of bacteria in young symbiosomes. Unexpectedly a large amount of GFP signal was also detected in the uninfected cells of the nodule (Fig 10 *A, C*). This was surprising because previous experiments with GUS staining had indicated that SAP3 is only expressed in cells containing bacteria. The GFP signal accumulated in uninfected cells appeared to be cytoplasmic and the shadow of plant nuclei can be can be seen in these cells (Fig. 10 *A*). We also examined SAP3 localization in older nodules at 6wpi and found that the GFP signal continued to accumulate in the cytoplasm of cells as symbiosomes undergo programmed senescence (Fig 10 *D, E*).

SAP3 is Mobilized from the Symbiosome Compartment

Based on the discrepancies between SAP3 promoter activity and protein localization we hypothesized that SAP3 might be released as a cell mobile signal. To investigate this further we expressed a variety of GFP fusion constructs under control of the SAP3 promoter to test the localization of SAP3 in root nodules (Fig. 11). First we tested whether GFP placement influenced the localization pattern of SAP3 by moving the GFP sequence between the mature peptide and the GPI-anchor signal. GFP placement did not have any affect on SAP3 localization, which transiently labeled symbiosome membranes and continued to accumulate in uninfected cells. Small starch granules were occasionally observed in the uninfected cells containing cytoplasmic GFP signal (Fig. 10 *C*, Fig. 11). Next we tested whether this localization pattern was specific to the SAP3 peptide by expressing a GFP fusion construct with the SAP3 *N*-terminal secretion signal and *C*-terminal GPI-anchor signal, but without mature SAP3 peptide sequence. Surprisingly, The GFP signal in these nodules strongly accumulated in mature symbiosome membranes and was not detected in uninfected cells (Fig. 11). This suggests that the mature SAP3 peptide sequence is required for the movement of GFP out of the symbiosome compartment. Since SAP3 is a GPI-anchored peripheral membrane protein, we tested whether membrane disassociation was also required for SAP3 mobility by replacing the GPI anchor sequence with a single pass transmembrane domain. SP-GFP-SAP3-TM constructs remained in the symbiosome membrane and were not detected in the uninfected cells (Fig. 12). Interestingly, this also triggered heavy starch accumulation similar to *SAP3-RNAi* nodules. This indicates that the release of SAP3 from the symbiosome membrane is necessary for signal mobility and implies that retention of SAP3 negatively impacts nodulation.

Localization of SAP1, SAP2, and SAP3 in AM Colonized Roots

We used a similar cloning strategy to localize SAP peptides in AM colonized roots. The mCherry coding sequence was inserted between the *N*-terminal secretion signal and mature peptide sequence for SAP1 and SAP2 to generate SP-mCherry-SAP-GPI fusion proteins. Roots expressing mCherry-SAP constructs were baited with *R.*

irregularis and imaged 4 wpi. Results show that SAP1 and SAP2 specifically localized to periarbuscular membranes (Fig. 13). Signal was detected at very early stages of arbuscule branching, which suggest that SAP1 and SAP2 play an early role in arbuscule development (Fig. 13 *C*). It was difficult to localize SAP3 relative to regions of AM colonization so we made co-transformed roots using SAP2-mCherry as a marker for arbuscule development. SAP3-GFP was expressed in cells developing arbuscules in the inner root cortex and localized to the plasma membrane (Fig. 14 *A, B*). Weak GFP signal also emanated from the periarbuscular membrane and co-localized with SAP2 (Fig. 14 *B*). The periarbuscular compartment is highly acidified by H^+-ATPases which make it difficult to localize GFP labeled proteins to this space.

SAPs 1, 2, and 3 are Required for AM symbiosis

Co-localization of SAPs to periarbuscular membranes suggested that these peptides might act redundantly in AM symbiosis. We investigated this using a combinatorial RNAi approach to knockdown multiple transcripts in hairy roots and screen for AM-related morphological phenotypes. Preliminary data from these experiments indicated that single *SAP-RNAi* (data not shown) and double *SAP1+2-RNAi* constructs had no effect on AM morphology (Fig. 15 *A, B*). However, triple *SAP1+2+3-RNAi* constructs resulted in stunted arbuscule formation and reduced root colonization intensity (Fig. 15 *C-F*). This data suggests that the SAPs likely have redundant functions, which appear to be important for arbuscule development. However, more in-depth investigation is necessary.

Analysis of SAP2 by Western Blot

SAP peptides contain non-contiguous hydroxyproline motifs that are predicted to be glycosylated with arabinogalactan. We investigated possible post-translational modification of SAP2 by analyzing mCherry fusion proteins by SDS-PAGE and western blot. SAP2 was expressed under either its native promoter in *M. truncatula* roots colonized with AMF or under the constitutive *35S* promoter in *N. benthamiana* leaves. A known AGP from *A. thaliana* (AGP9-YFP-HA) was expressed in *N. benthamiana* and used as a relative control for comparison. The AtAGP9-HA-YFP and SAP2-mCherry fusion proteins have a predicted molecular weight (MW) of 45 and 28 kDa, respectively, based on their primary amino acid sequences. 35s:AtAGP9-HA-YFP expressed in tobacco produced two high molecular weight bands, which is characteristic of AGPs, at approximately 90 and 55 kDa, far above the predicted MW (Fig. 16). Expression of *35S:SP-mCherry-SAP2-GPI* in *N. benthamiana* produced only a single band corresponding to predicted MW of the fusion protein, indicating that post-translation modification did not occur. However, when the same fusion protein is expressed under its native promoter in *M. truncatula* roots colonized by AMF, two high molecular weight bands are produced above 55 and 70 kDa similar to the AtAGP9-YFP-HA construct (Fig. 16). This suggests that post-translation modification of SAP2 could either be specific to *M. truncatula* or require a symbiotic tissue-dependent context.

Discussion

Previous studies had shown that AGP carbohydrate epitopes localize to symbiotic membranes in a variety of different plant-microbial endosymbioses. This suggests that AGPs may have an important and shared function in these interfaces, yet their identity and corresponding specificity for endosymbiosis has remained unknown due to the research challenges that accompany these enigmatic glycoproteins. We addressed this question using the bioinformatic search tool, AGPredictor, to identify AGP-encoding genes in the *M. truncatula* genome (Chapter 1) and cross-referenced this list to community expression data. Through this approach we identified two groups of AGPs discretely expressed in either AM symbiosis or nodulation. Both groups contained similar types of AGPs, including a large number of plastocyanin-like AGPs and a small family of AG peptides that we have named SYMBIOSIS-ASSOCIATED ARABINOGALACTAN PEPTIDES (SAPs). SAPs were chosen for further study due to their small size, which may allow them to act as diffusible signals.

The SAP family is composed of three tandemly duplicated genes on chromosome 5 that are differentially expressed in AM symbiosis and root nodules. This reflects a pattern of gene duplication and neofunctionalization consistent with the evolutionary narrative of the rhizobium-legume symbiosis arising through expansion of AM symbiosis-specific gene families (51). SAPs share little sequence homology but have conserved post-translational modification sites for AG glycosylation and GPI-anchoring, indicating these are the relevant features of these peptides. Analysis of SAP2 fusion proteins by SDS-PAGE and western blot suggests that post-translational modification does occur but is either species or tissue specific.

RNAi-mediated knockdown of SAP3 strongly impaired nodule development and nitrogen fixation, resulting in a Nod$^+$/Fix$^-$ phenotype. Symbiosomes in SAP3-RNAi nodules were poorly differentiated with large spaces between the bacteria and the symbiosome membrane, which could suggest that *S. meliloti* fail to interact with these membrane interfaces. This was correlated with a massive accumulation of starch in uninfected cells. However, it is unclear whether this was the cause or consequence of the impaired bacterial development.

The *SAP3* promoter was only active in cells containing *S. meliloti* and SAP3-GFP fusion proteins localized to differentiating symbiosome membranes, but also occurred in the cytoplasm of uninfected cells. Insertion of the GFP coding sequence between either the *N*-terminal secretion signal or *C*-terminal GPI signal did not affect the localization pattern of the SAP3-GFP fusion protein. However, GFP expressed with the SAP3 secretion and GPI signal sequences, but without the SAP3 peptide, remained within the symbiosome and did not move to the uninfected cells. This indicates that the mature 14-amino acid SAP3 peptide is necessary for the mobility of GFP out of the symbiosome compartment.

We tested whether release from the symbiosome membrane was necessary for GFP mobility by replacing the GPI signal sequence with a transmembrane domain. This caused the SAP3-GFP fusion protein to be retained in the symbiosome membrane indicating that membrane-release is also required for SAP3 mobility. Surprisingly, retention of SAP3 on the symbiosome membrane triggered starch accumulation, suggesting that removal of SAP3 from the membrane is necessary for normal development. GPI-anchors can be enzymatically cleaved by phosphatidylinositol-specific phospholipase C (PI-PLC) activity, which allows for the targeted release of proteins from the membrane surface. Interestingly, the *M. truncatula* mutant *deficient nodule formation 2* (*dnf2*) encodes a nonfunctional nodule-specific PI-PLC that has a strikingly similar nodule phenotype to the SAP3-RNAi nodules, which are characterized by poorly differentiated symbiosomes with large spaces between the bacteria and the symbiosome membrane (52). Very little is known about endogenous plant PI-PLC specificity*,* however, mammalian PI-PLCs have been demonstrated to possess preferential selectivity for individual GPI-anchored proteins (53). Therefore, based on the distinct phenotype shared between *dnf2* and *SAP3-RNAi* nodules, we hypothesize that *DNF2* and *SAP3* act in a genetic pathway where DNF2-mediated release of SAP3 from the symbiosome membrane is required for bacterial growth and differentiation. This is partially supported by the observation that retention of SAP3 in the symbiosome using a transmembrane domain also triggers heavy start accumulation.

The role *SAPs* play in AM symbiosis is less clear. SAP1 and SAP2 specifically localize to the periarbuscular membrane and are present at early developmental stages prior to arbuscule branching. SAP3 is also induced by AM symbiosis and appeared to have an overlapping localization pattern with SAP1 and SAP2 but was also present in the plasma membrane of surrounding uncolonized cells. Initial attempts at screening single and double *SAP1+2-RNAi* constructs did not have a visible effect on root colonization or morphology of the symbiosis. However, *SAP1+2+3-RNAi* constructs resulted in impaired arbuscule formation and sparse root colonization, which suggest that SAPs function redundantly in AM symbiosis. Arbuscules in *SAP1+2+3-RNAi* roots were stunted and failed to develop the highly branched haustoria characteristic of a functional AM symbiosis. This is analogous to the *SAP3-RNAi* nodulation phenotype, where in both cases the endosymbiont fails to grow and differentiate within symbiotic membrane compartment. While we were able to observe morphological phenotypes associated with the *SAP1+2+3-RNAi* construct, this data is still considered preliminary because we have yet to quantify these results. The phenotype was reproducible, but the severity was variable. We attribute this not only to the proposed redundancy of these peptides but also the poor performance of *35S* promoter-driven RNAi constructs in AM roots. It was recently shown that the *35S* promoter is silenced in *M. truncatula* roots during AM colonization and is not expressed in the inner root cortex (54). This has been shown to mask the RNAi phenotype of the arbuscule-specific gene *MIG1*, which has a strong arbuscule-related RNAi phenotype when silenced under the *A. thaliana* ubiquitin 3 promoter that is not seen using *35s*-driven RNAi constructs (55). We have developed *pAtUBQ3:SAP-RNAi* constructs that will be used in future experiments.

SAP functional redundancy in AM symbiosis is a perfect example that highlights the utility of the AGPredictor software over other programs. Schultz et al (2008) identified *SAP1* as an AM-specific transcript using amino acid composition bias (%PAST) but failed to identify *SAP2* and *SAP3* (56). *SAP* peptides have low sequence homology and are not BLAST searchable, which is likely why subsequent papers were never published in this area. Recently, Ma et al. (2017) identified *SAP1* and *SAP3* using the noncontiguous hydroxyproline hypothesis, but still failed to identify *SAP2* (57). We find this somewhat counter-intuitive because SAP2 and SAP3 have the same number of proline residues capable of becoming glycosylated. Furthermore, we provide some evidence to support that *SAP2* contains post-translational modifications consistent with a known AGP based on the altered migration pattern of these proteins on SDS-PAGE reducing gels. This supports our assessment that deeper-reaching predictive tools like AGPredictor are necessary to identify and functionally characterize AGPs in plants.

In summary, we have identified a novel family of AGP peptides that localize to symbiotic membranes and are required for the growth and differentiation of *S. meliloti* and *R. irregularis* within these compartments. However, the mechanism underlying SAP function remains elusive and we can only speculate based the literature and the phenotypic data at hand. Many possibilities exist, however we hypothesize that SAPs may possibly function in stimulating nutrient release, suppressing defense response, or promoting the differentiation of plant cells engaged in symbiosis.

As previously discussed, some AGPs have been characterized as secreted signals that promote plant cell identity and differentiation. In this regard perhaps SAPs promote the acquisition of “symbiotic-identity” in the plant host cell necessary for the cellular reprogramming that controls the development of the symbiosis. Mobility of SAP3 from infected to uninfected cells in root nodules could suggest a role in tissue patterning between these two distinct cell types. The appearance of SAP3-GFP in the cytoplasm of uninfected cells was not expected, however in the literature there is precedent for the uptake of extracellular AGPs into the plant cytoplasm during pollen self-incompatibility (58–60). In AM symbiosis we did not observe anything to suggest that SAPs are mobilized between cells, but we speculate that similar roles in tissue patterning could be achieved by the partially overlapping localization pattern between SAPs 1 and 2 vs SAP3 in arbuscule-containing cells and adjacent uninfected cells.

The *SAP3-RNAi* nodulation phenotype, which is characterized by heavy starch accumulation in uninfected cells and large spaces between symbiosome membranes, is correlated with the localization pattern of the SAP3-GFP protein and could suggest that SAP3 may play a role in stimulating starch release from these cells. Carbon metabolism predominately occurs in the uninfected cells where accumulated sugars are converted to carboxylic acids by glycolytic enzymes and released to the growing bacteria in the adjacent infected cells (61). Only uninfected cells, as opposed to infected cells, are able to actively take up sucrose and glucose from the apoplast (62). Uninfected cells also have higher levels of sucrose synthase and alkaline invertase activity involved in sucrose cleavage (63–66). Furthermore, low oxygen tension in infected cells prevents mitochondrial respiration from efficiently providing the bacteria with carbon (67). Sugars released in uninfected cells can either be converted to malate for bacterial nutrition or stored as starch based on the growing demand and nitrogen-fixing status of the bacteria. It is possible that the release of SAP3 from the symbiosome membrane and accumulation in uninfected cells may act as a signal approximating the relative infection level and developmental status of the infected cell, thereby influencing starch release. Conceptually this is similar to the function of the arabinosylated CLE44 peptide, which is secreted by the nodule meristem but detected in leaves as a long distance signal used to approximate the number of nodules in the root system (68). Interestingly, AGPs have previously been reported to play a role in gibberellic acid (GA)-induced starch release during barley seed germination (69). GA is synthesized by embryo epithelial cells and transcriptionally activates α-amylase secretion from aleurone cells into the starchy seed endosperm (70). At least one aleurone-specific AGP was identified and disruption of membrane bound AGPs in aleurone cell protoplasts with β-glucosyl Yariv reagent inhibited GA-induced transcription of the α-amylase promoter (69). This provides an example where an AGP may participate in non-autonomous cell signaling that regulates the release of starch in neighboring cells. Based on the available evidence we hypothesize that SAP3 could stimulate the release of carbon necessary for the growth of the bacteria within in the symbiosome compartment. However, due to the reciprocal nature of nutrient exchange it is not possible to definitively tell, because inhibiting nitrogen fixation also prevents starch release. If starch accumulation were the cause of the *SAP3-RNAi* phenotype, as opposed to the consequence, then we would expect the phenotype to be suppressed by over expression of hexokinase or phosphoglucomutase. Future work will be necessary investigate this in greater detail.

Bibliography

1. Willats WG, Knox JP (1996) A role for arabinogalactan-proteins in plant cell expansion: evidence from studies on the interaction of beta-glucosyl Yariv reagent with seedlings of Arabidopsis thaliana. *Plant J* 9(6):919–925.
2. Shi H (2003) The Arabidopsis SOS5 Locus Encodes a Putative Cell Surface Adhesion Protein and Is Required for Normal Cell Expansion. *THE PLANT CELL ONLINE* 15(1):19–32.
3. Serpe MD, Nothnagel EA (1994) Effects of Yariv phenylglycosides onRosa cell suspensions: Evidence for the involvement of arabinogalactan-proteins in cell proliferation. *Planta* 193(4):542–550.
4. Motose H, Sugiyama M, Fukuda H (2004) A proteoglycan mediates inductive interaction during plant vascular development. *Nature* 429(6994):873–878.
5. Gao M, Showalter AM (1999) Yariv reagent treatment induces programmed cell death in Arabidopsis cell cultures and implicates arabinogalactan protein involvement. *Plant J* 19(3):321–331.
6. Tan L, et al. (2013) An Arabidopsis cell wall proteoglycan consists of pectin and arabinoxylan covalently linked to an arabinogalactan protein. *Plant Cell* 25(1):270–287.
7. Schultz CJ, Ferguson KL, Lahnstein J, Bacic A (2004) Post-translational modifications of arabinogalactan-peptides of Arabidopsis thaliana. Endoplasmic reticulum and glycosylphosphatidylinositol-anchor signal cleavage sites and hydroxylation of proline. *J Biol Chem* 279(44):45503–45511.
8. Oxley D, Bacic A (1999) Structure of the glycosylphosphatidylinositol anchor of an arabinogalactan protein from Pyrus communis suspension-cultured cells. *Proc Natl Acad Sci USA* 96(25):14246–14251.
9. Svetek J, Yadav MP, Nothnagel EA (1999) Presence of a glycosylphosphatidylinositol lipid anchor on rose arabinogalactan proteins. *J Biol Chem* 274(21):14724–14733.
10. Huang Y, et al. (2016) Nanospherical arabinogalactan proteins are a key component of the high-strength adhesive secreted by English ivy. *Proc Natl Acad Sci USA* 113(23):E3193-202.
11. Smallwood M, Yates EA, Willats WGT, Martin H, Knox JP (1996) Immunochemical comparison of membrane-associated and secreted arabinogalactan-proteins in rice and carrot. *Planta* 198(3):452–459.
12. Baldwin TC, McCann MC, Roberts K (1993) A Novel Hydroxyproline-Deficient Arabinogalactan Protein Secreted by Suspension-Cultured Cells of Daucus carota

(Purification and Partial Characterization). *Plant Physiol* 103(1):115–123.

13. Schultz C, Gilson P, Oxley D, Youl J, Bacic A (1998) GPI-anchors on arabinogalactan-proteins: implications for signalling in plants. *Trends Plant Sci* 3(11):426–431.
14. Nguema-Ona E, Coimbra S, Vicré-Gibouin M, Mollet J-C, Driouich A (2012) Arabinogalactan proteins in root and pollen-tube cells: distribution and functional aspects. *Ann Bot* 110(2):383–404.
15. van Hengel AJ, Barber C, Roberts K (2004) The expression patterns of arabinogalactan-protein AtAGP30 and GLABRA2 reveal a role for abscisic acid in the early stages of root epidermal patterning. *Plant J* 39(1):70–83.
16. Pereira AM, et al. (2014) Differential expression patterns of arabinogalactan proteins in Arabidopsis thaliana reproductive tissues. *J Exp Bot* 65(18):5459–5471.
17. Motose H, Sugiyama M, Fukuda H (2001) An arabinogalactan protein(s) is a key component of a fraction that mediates local intercellular communication involved in tracheary element differentiation of zinnia mesophyll cells. *Plant Cell Physiol* 42(2):129–137.
18. McCabe PF, Valentine TA, Forsberg LS, Pennell RI (1997) Soluble Signals from Cells Identified at the Cell Wall Establish a Developmental Pathway in Carrot. *Plant Cell* 9(12):2225–2241.
19. Mizukami AG, et al. (2016) The AMOR Arabinogalactan Sugar Chain Induces Pollen-Tube Competency to Respond to Ovular Guidance. *Curr Biol* 26(8):1091–1097.
20. Dresselhaus T, Coimbra S (2016) Plant reproduction: AMOR enables males to respond to female signals. *Curr Biol* 26(8):R321-3.
21. Jiao J, et al. (2017) Structure-Activity Relation of AMOR Sugar Molecule That Activates Pollen-Tubes for Ovular Guidance. *Plant Physiol* 173(1):354–363.
22. Lamport DTA, Kieliszewski MJ, Showalter AM (2006) Salt stress upregulates periplasmic arabinogalactan proteins: using salt stress to analyse AGP function. *New Phytol* 169(3):479–492.
23. Nguema-Ona E, Vicré-Gibouin M, Cannesan M-A, Driouich A (2013) Arabinogalactan proteins in root-microbe interactions. *Trends Plant Sci* 18(8):440–449.
24. Gaspar YM, et al. (2004) Characterization of the Arabidopsis lysine-rich arabinogalactan-protein AtAGP17 mutant (rat1) that results in a decreased efficiency of agrobacterium transformation. *Plant Physiol* 135(4):2162–2171.
25. Xie F, Williams A, Edwards A, Downie JA (2012) A plant arabinogalactan-like glycoprotein promotes a novel type of polar surface attachment by Rhizobium leguminosarum. *Mol Plant Microbe Interact* 25(2):250–258.
26. Beauregard PB, Chai Y, Vlamakis H, Losick R, Kolter R (2013) Bacillus subtilis biofilm induction by plant polysaccharides. *Proc Natl Acad Sci USA* 110(17):E1621-30.
27. Koroney AS, et al. (2016) Root exudate of Solanum tuberosum is enriched in galactose-containing molecules and impacts the growth of Pectobacterium atrosepticum. *Ann Bot*.
28. Cannesan MA, et al. (2012) Effect of arabinogalactan proteins from the root caps of pea and Brassica napus on Aphanomyces euteiches zoospore chemotaxis and germination. *Plant Physiol* 159(4):1658–1670.
29. Tsyganova AV, et al. (2009) Distribution of legume arabinogalactan protein-extensin (AGPE) glycoproteins in symbiotically defective pea mutants with abnormal infection threads. *Cell Tissue Biol* 3(1):93–102.

30. Berry AM, et al. (2002) Arabinogalactan proteins are expressed at the symbiotic interface in root nodules of Alnus spp. *New Phytol* 155(3):469–479.
31. Jackson O, Taylor O, Adams DG, Knox JP (2012) Arabinogalactan proteins occur in the free-living cyanobacterium genus Nostoc and in plant-Nostoc symbioses. *Mol Plant Microbe Interact* 25(10):1338–1349.
32. Balestrini R, Hahn MG, Faccio A, Mendgen K, Bonfante P (1996) Differential localization of carbohydrate epitopes in plant cell walls in the presence and absence of arbuscular mycorrhizal fungi. *Plant Physiol* 111(1):203–213.
33. Bonfante-Fasolo, et al. Immunocytochemical location of hydroxyproline rich glycoproteins at the interface between a mycorrhizal fungus and its host plants.
34. Balestrini R, Bonfante P (2005) The interface compartment in arbuscular mycorrhizae: A special type of plant cell wall? *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology* 139(1):8–15.
35. Bonfante P, Genre A (2010) Mechanisms underlying beneficial plant-fungus interactions in mycorrhizal symbiosis. *Nat Commun* 1:48.
36. Clarke VC, Loughlin PC, Day DA, Smith PMC (2014) Transport processes of the legume symbiosome membrane. *Front Plant Sci* 5:699.
37. Javot H, Penmetsa RV, Terzaghi N, Cook DR, Harrison MJ (2007) A Medicago truncatula phosphate transporter indispensable for the arbuscular mycorrhizal symbiosis. *Proc Natl Acad Sci USA* 104(5):1720–1725.
38. Wang D, et al. (2010) A nodule-specific protein secretory pathway required for nitrogen-fixing symbiosis. *Science* 327(5969):1126–1129.
39. Ivanov S, et al. (2012) Rhizobium-legume symbiosis shares an exocytotic pathway required for arbuscule formation. *Proc Natl Acad Sci USA* 109(21):8316–8321.
40. Tan L, et al. (2012) Arabinogalactan-proteins and the research challenges for these enigmatic plant cell surface proteoglycans. *Front Plant Sci* 3:140.
41. Floss DS, Schmitz AM, Starker CG, Gantt JS, Harrison MJ (2013) Gene silencing in Medicago truncatula roots using RNAi. *Methods Mol Biol* 1069:163–177.
42. Haney CH, Long SR (2010) Plant flotillins are required for infection by nitrogen-fixing bacteria. *Proc Natl Acad Sci USA* 107(1):478–483.
43. Hoagland DR, Arnon DI (1950) The water-culture method for growing plants without soil. *Circular California Agricultural Experiment Station*.
44. Helliwell C, Waterhouse P (2003) Constructs and methods for high-throughput gene silencing in plants. *Methods* 30(4):289–295.
45. Nakagawa T, et al. (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J Biosci Bioeng* 104(1):34–41.
46. Kawaharada Y, et al. (2017) Differential regulation of the Epr3 receptor coordinates membrane-restricted rhizobial colonization of root nodule primordia. *Nat Commun* 8:14534.
47. Perotto S, Vandenbosch KA, Butcher GW, Brewin NJ (1991) Molecular composition and development of the plant with the peribacteroid membrane of pea root nodules. *Development*.
48. Perotto S, Donovan N, Drobak BK (1995) Differential expression of a glycosyl inositol phospholipid antigen on the peribacteroid membrane during pea nodule development. *MPMI-Molecular Plant …*.

49. He J, et al. (2009) The Medicago truncatula gene expression atlas web server. *BMC Bioinformatics* 10:441.
50. Benedito VA, et al. (2008) A gene expression atlas of the model legume Medicago truncatula. *Plant J* 55(3):504–513.
51. Young ND, et al. (2011) The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* 480(7378):520–524.
52. Bourcy M, et al. (2013) Medicago truncatula DNF2 is a PI-PLC-XD-containing protein required for bacteroid persistence and prevention of nodule early senescence and defense-like reactions. *New Phytol* 197(4):1250–1261.
53. Müller A, Klöppel C, Smith-Valentine M, Van Houten J, Simon M (2012) Selective and programmed cleavage of GPI-anchored proteins from the surface membrane by phospholipase C. *Biochim Biophys Acta* 1818(1):117–124.
54. Devers EA, Teply J, Reinert A, Gaude N, Krajinski F (2013) An endogenous artificial microRNA system for unraveling the function of root endosymbioses related genes in Medicago truncatula. *BMC Plant Biol* 13:82.
55. Heck C, et al. (2016) Symbiotic Fungi Control Plant Root Cortex Development through the Novel GRAS Transcription Factor MIG1. *Curr Biol* 26(20):2770–2778.
56. Schultz CJ, Harrison MJ (2008) Novel plant and fungal AGP-like proteins in the Medicago truncatula-Glomus intraradices arbuscular mycorrhizal symbiosis. *Mycorrhiza* 18(8):403–412.
57. Ma Y, et al. (2017) Bioinformatics prediction and evolution analysis of arabinogalactan proteins in the plant kingdom. *Front Plant Sci* 8:66.
58. Pistil K A style-specific 120-kDa glycoprotein enters pollen tubes of Nicotiana alata in vivo.
59. Goldraij A, et al. (2006) Compartmentalization of S-RNase and HT-B degradation in self-incompatible Nicotiana. *Nature* 439(7078):805–810.
60. Lee CB, Kim S, McClure B (2009) A pollen protein, NaPCCP, that binds pistil arabinogalactan proteins also binds phosphatidylinositol 3-phosphate and associates with the pollen tube endomembrane system. *Plant Physiol* 149(2):791–802.
61. White J, Prell J, James EK, Poole P (2007) Nutrient sharing between symbionts. *Plant Physiol* 144(2):604–614.
62. Peiter E, Schubert S (2003) Sugar uptake and proton release by protoplasts from the infected zone of Vicia faba L. nodules: evidence against apoplastic sugar supply of infected cells. *J Exp Bot* 54(388):1691–1700.
63. Flemetakis E, et al. (2006) Spatial and temporal organization of sucrose metabolism in Lotus japonicus nitrogen-fixing nodules suggests a role for the elusive alkaline/neutral invertase. *Plant Mol Biol* 62(1–2):53–69.
64. Gordon AJ, Minchin FR, James CL, Komina O (1999) Sucrose synthase in legume nodules is essential for nitrogen fixation. *Plant Physiol* 120(3):867–878.
65. Kavroulakis N, Flemetakis E, Aivalakis G, Katinakis P (2000) Carbon metabolism in developing soybean root nodules: the role of carbonic anhydrase. *Mol Plant Microbe Interact* 13(1):14–22.
66. Singh R (Punjab AU Ludhiana (Inde)Department of Biochemistry), Karamdeep L, Bhullar SS, Gupta KA (1994) Metabolism of free sugars in relation to the activities of enzymes involved in sucrose metabolism and nitrogen assimilation in the developing nodules of chickpea. *Plant Physiology and Biochemistry (France)*.

67. Day DA (Australian NU Canberra (Australia)Division of Biochemistry and Molecular Biology), Copeland L (1991) Carbon metabolism and compartmentation in nitrogen-fixing legume nodules [symbiotic nitrogen-fixation, dicarboxylic acids, oxygen-limitation, nodule structure]. *Plant Physiology and Biochemistry (France)*.
68. Okamoto S, Shinohara H, Mori T, Matsubayashi Y, Kawaguchi M (2013) Root-derived CLE glycopeptides control nodulation by direct binding to HAR1 receptor kinase. *Nat Commun* 4:2191.
69. Suzuki Y, Kitagawa M, Knox JP, Yamaguchi I (2002) A role for arabinogalactan proteins in gibberellin-induced alpha-amylase production in barley aleurone cells. *Plant J* 29(6):733–741.
70. Kaneko M, Itoh H, Ueguchi-Tanaka M, Ashikari M, Matsuoka M (2002) The alpha-amylase induction in endosperm during rice seed germination is caused by gibberellin synthesized in epithelium. *Plant Physiol* 128(4):1264–1270.

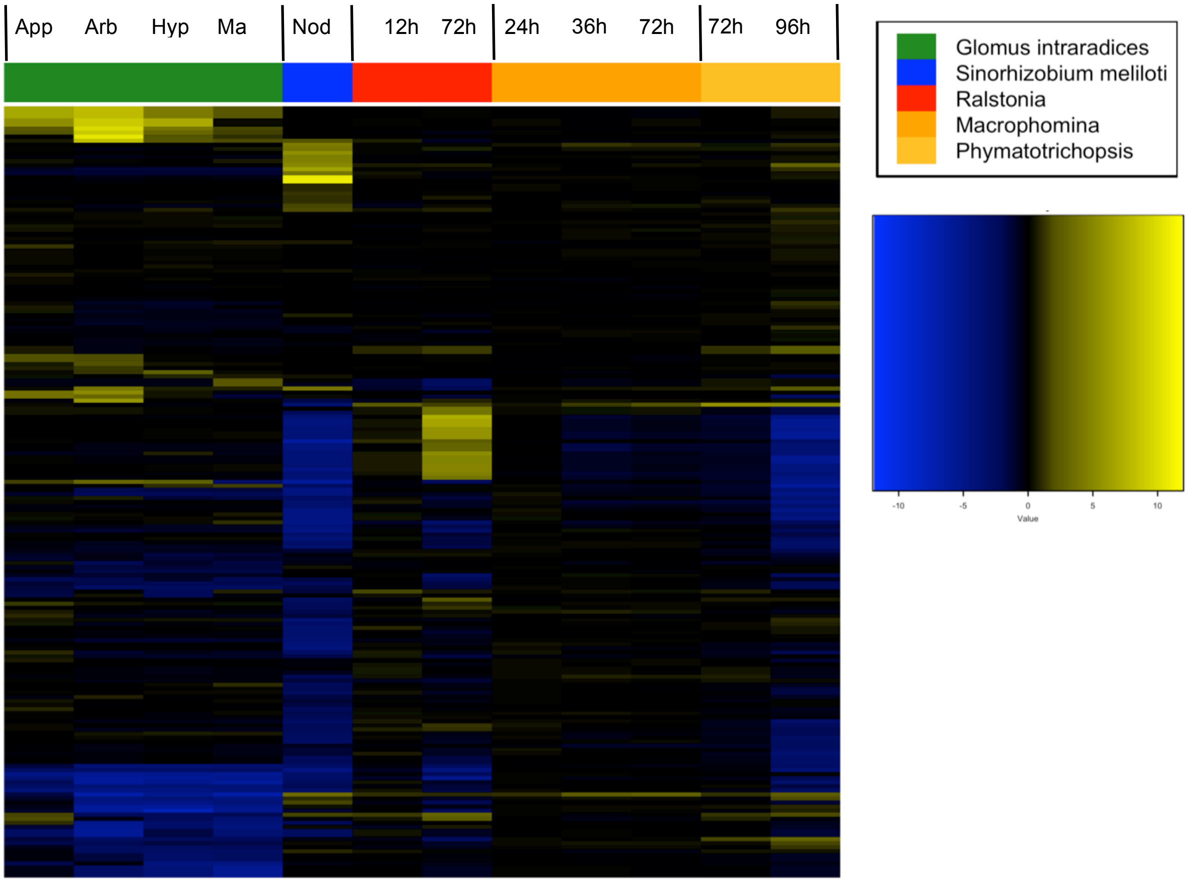
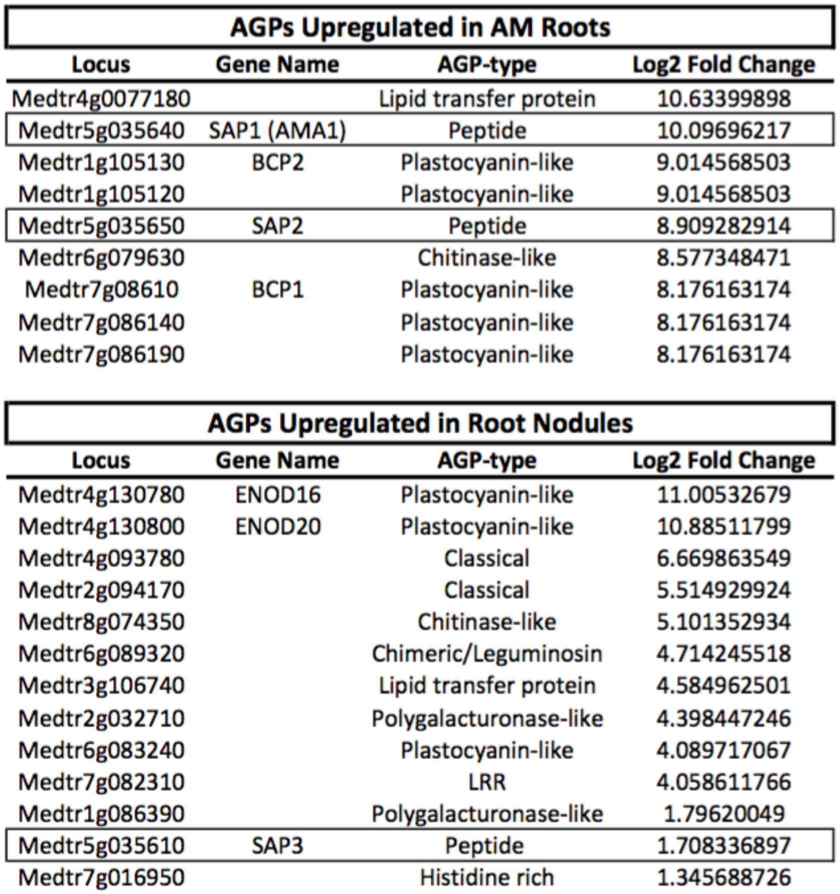


Figure 1. Heat map of AGP expression in *M. truncatula* roots in response to inoculation with either symbiotic or pathogenic microorganisms. Data were taken from the Medicago Gene Expression Atlas and plotted as log 2-fold change in expression relative to the uninoculated control for each experiment. Primary data as follows: laser capture microdissection of different cell types colonized by the AM fungus *Rhizophagus irregularis* (formerly *Glomus* intraradices) (Hodgekamp *et al.* 2013); App, cortical cells underneath appresoria; Arb, cortical cells containing arbuscules; Hyp, cortical cells containing hyphae; Mat, cortical cells containing mature structures i.e. vesicles/arbuscules. RNA-seq of whole root nodules (Nod) infected with *Sinorhizobium meliloti* vs roots (Roux *et al*. 2014). Roots infected with *Ralstonia solanacearum* at 12 h and 72 h time points (Balzergue *et al.* 2008, unpublished). Roots infected with *Macrophomina phaseolina* at 24, 36, and 72 h time points (Mah *et al.* 2012). Roots infected with *Phymatotrichopsis omnivora* at 72 h and 96 h time points (Uppalapati *et al.* 2009).



AGPs Upregulated in AM Roots

Locus	Gene Name	AGP-type	Log2 Fold Change
Medtr4g0077180		Lipid transfer protein	10.63399898
Medtr5g035640	SAP1 (AMA1)	Peptide	10.09696217
Medtr1g105130	BCP2	Plastocyanin-like	9.014568503
Medtr1g105120		Plastocyanin-like	9.014568503
Medtr5g035650	SAP2	Peptide	8.909282914
Medtr6g079630		Chitinase-like	8.577348471
Medtr7g08610	BCP1	Plastocyanin-like	8.176163174
Medtr7g086140		Plastocyanin-like	8.176163174
Medtr7g086190		Plastocyanin-like	8.176163174

AGPs Upregulated in Root Nodules

Locus	Gene Name	AGP-type	Log2 Fold Change
Medtr4g130780	ENOD16	Plastocyanin-like	11.00532679
Medtr4g130800	ENOD20	Plastocyanin-like	10.88511799
Medtr4g093780		Classical	6.669863549
Medtr2g094170		Classical	5.514929924
Medtr8g074350		Chitinase-like	5.101352934
Medtr6g089320		Chimeric/Leguminosin	4.714245518
Medtr3g106740		Lipid transfer protein	4.584962501
Medtr2g032710		Polygalacturonase-like	4.398447246
Medtr6g083240		Plastocyanin-like	4.089717067
Medtr7g082310		LRR	4.058611766
Medtr1g086390		Polygalacturonase-like	1.79620049
Medtr5g035610	SAP3	Peptide	1.708336897
Medtr7g016950		Histidine rich	1.345688726

Table I. Predicted *M. truncatula* AGPs identified as being upregulated in during AM symbiosis or root nodulation.

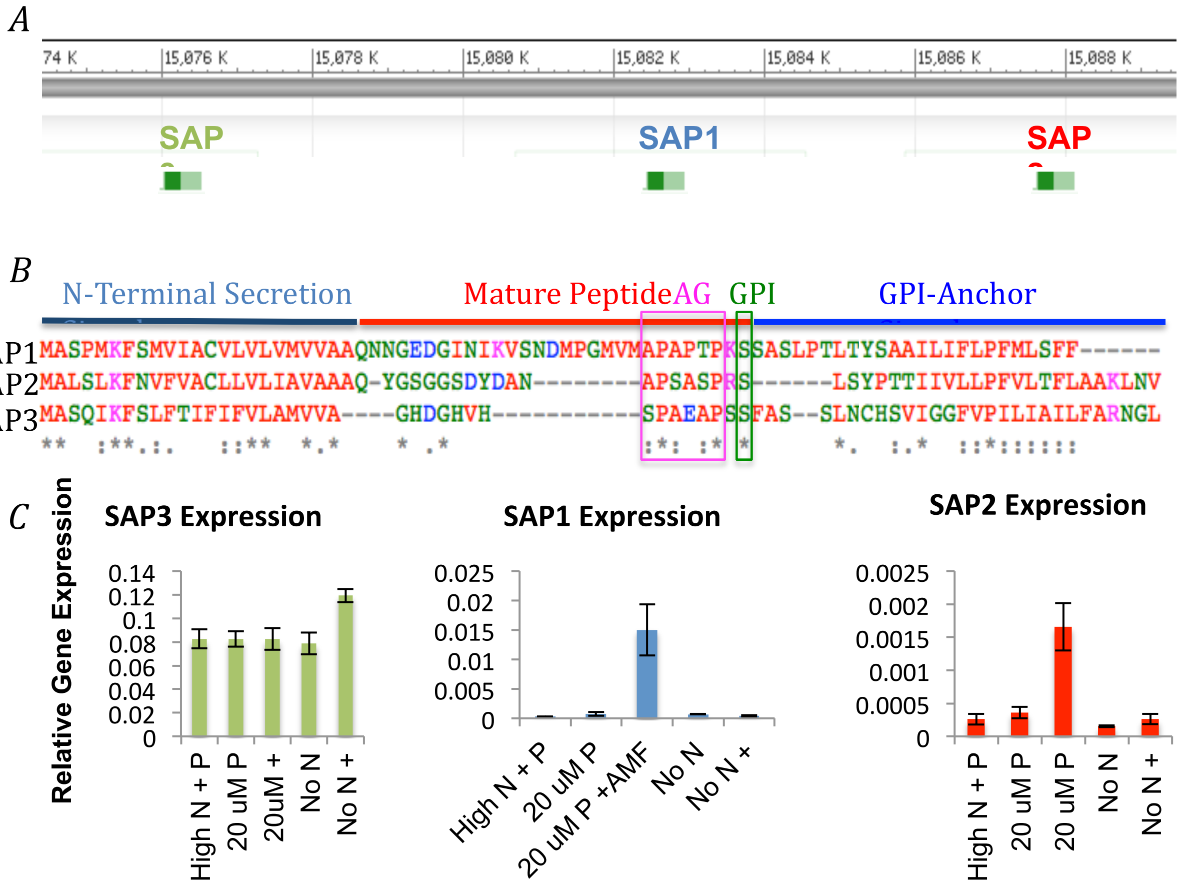


Fig 2. *SAPs* arise from tandem gene duplications but are differentially expressed in roots. (*A*) Jbrowse screen shot of *M. truncatula* chromosome 5 showing the genomic location of *SAPs*. Dark and light green colors indicate the coding sequence and untranslated regions of SAP transcripts. (*B*) MUSCLE multiple protein alignment of *SAPs.* (*C*) *SAP* expression measure by qPCR relative to EF1-α in *M. truncatula* roots during nodulation, AM symbiosis, and nutrient stress. All roots were sampled at 4 wpi (*n* = 5).

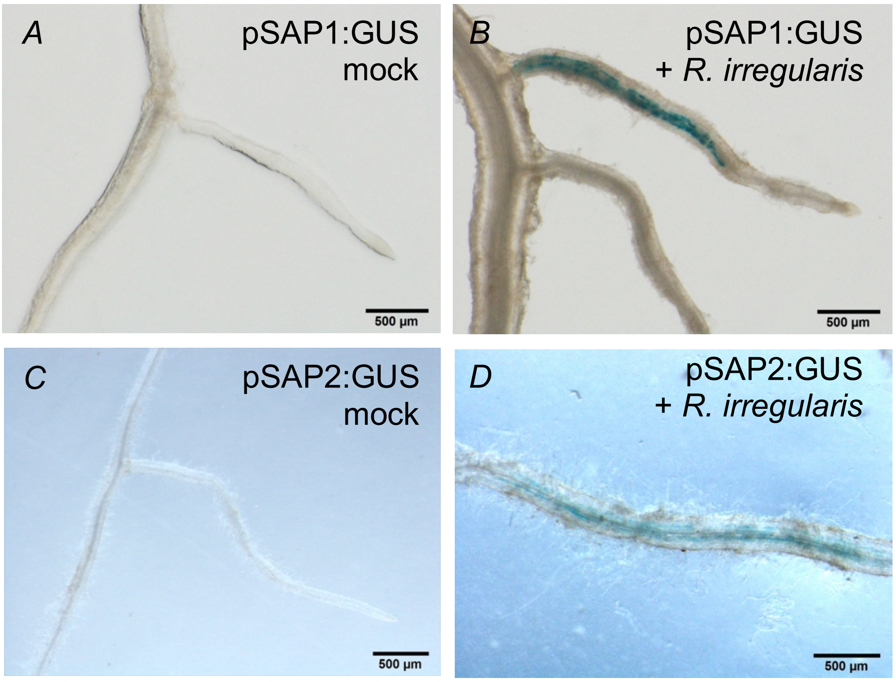


Fig 3. GUS staining of *M. truncatula* roots expressing either *SAP1* or *SAP2* reporter constructs. Roots were stained 4 wpi with *R. irregularis.*

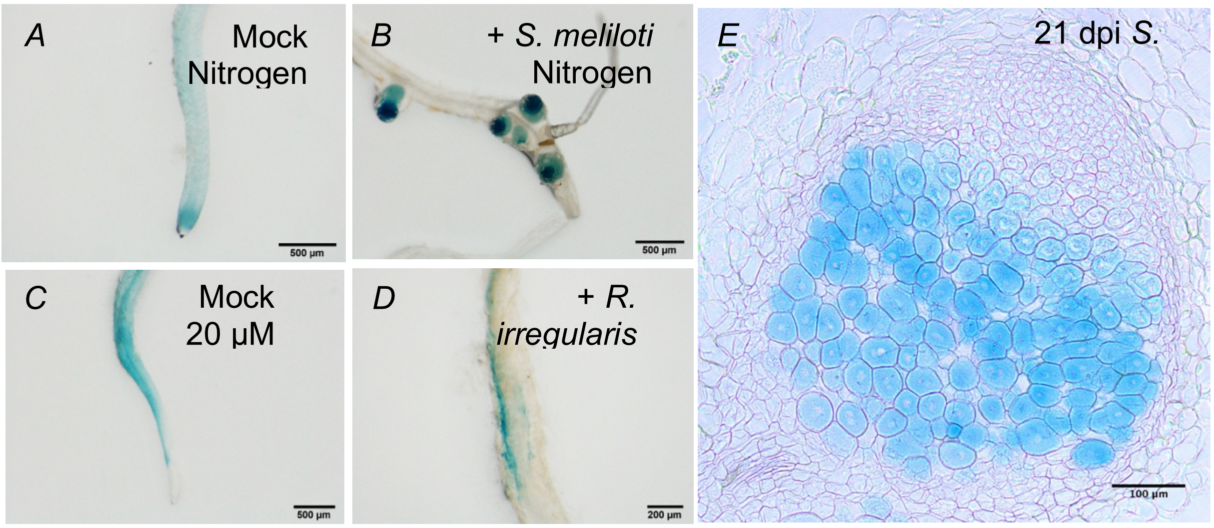


Figure 4. GUS staining of *M. truncatula* roots expressing *SAP3* reporter constructs. Roots were inoculated with either *S. meliloti* or *R. irregularis* and stained at either 3 wpi or 4 wpi, respectively. Results indicate *SAP3* has a different spatiotemporal expression pattern in symbiotic tissues and mock treated roots in response to nutrient stress.

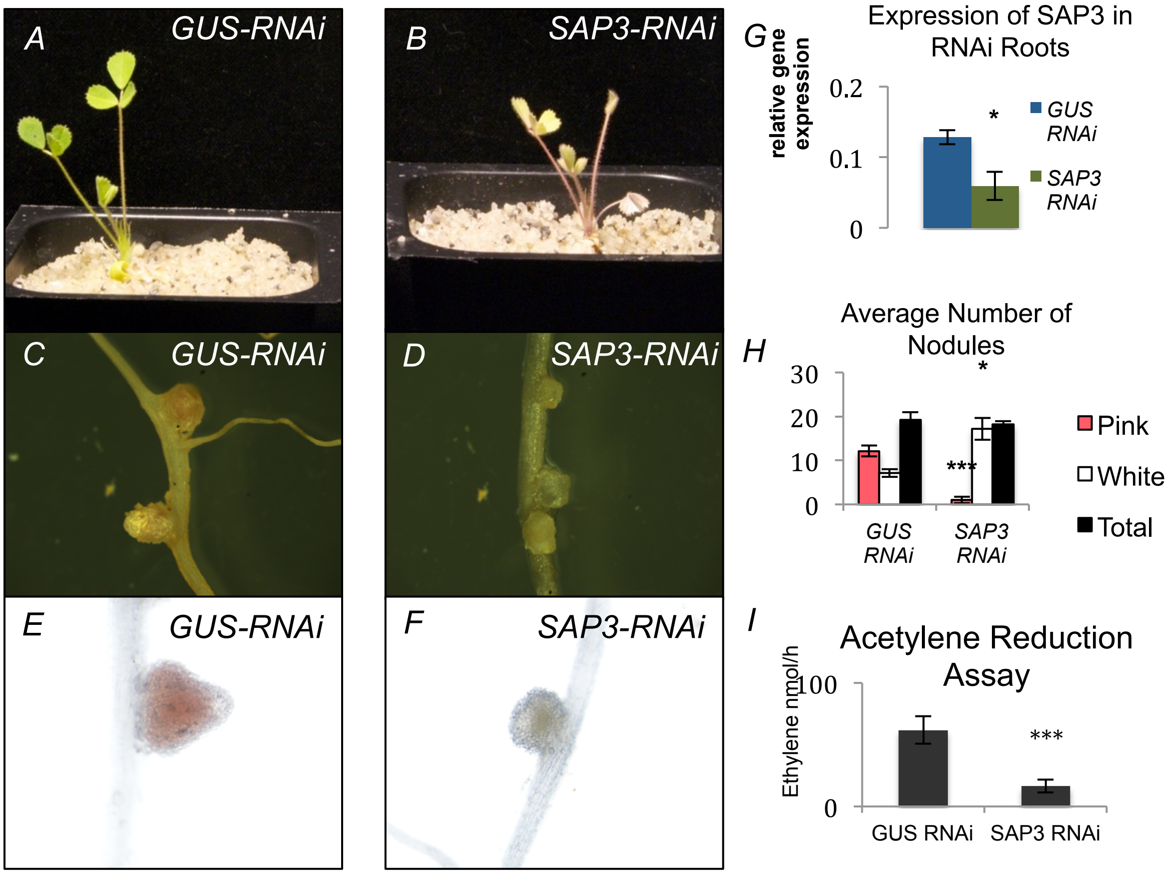


Fig 5. *SAP3* silencing impairs nodule development and function. (*A, B*) Shoots and (*C-F*) nodules of RNAi plants 3 wpi (*A-D*) and 4 wpi (*E-F*) with *S. meliloti* grown under nitrogen deprivation. (*G*) *SAP3* expression measured by qPCR relative EF1-α in nodulated RNAi roots (*n* = 4). (*H*) Average number of nodules produced per root system quantified based on the visible appearance of leghemoglobin (pink) (*n* = 5). (*I*) Nitrogen fixation efficiency measured by acetylene reduction assay (*n* = 4). Significant differences are indicated based on Student's T-test (* > 0.05, ** >0.01, ** >0.001). Each data is from an independent, representative, and reproducible experiment.

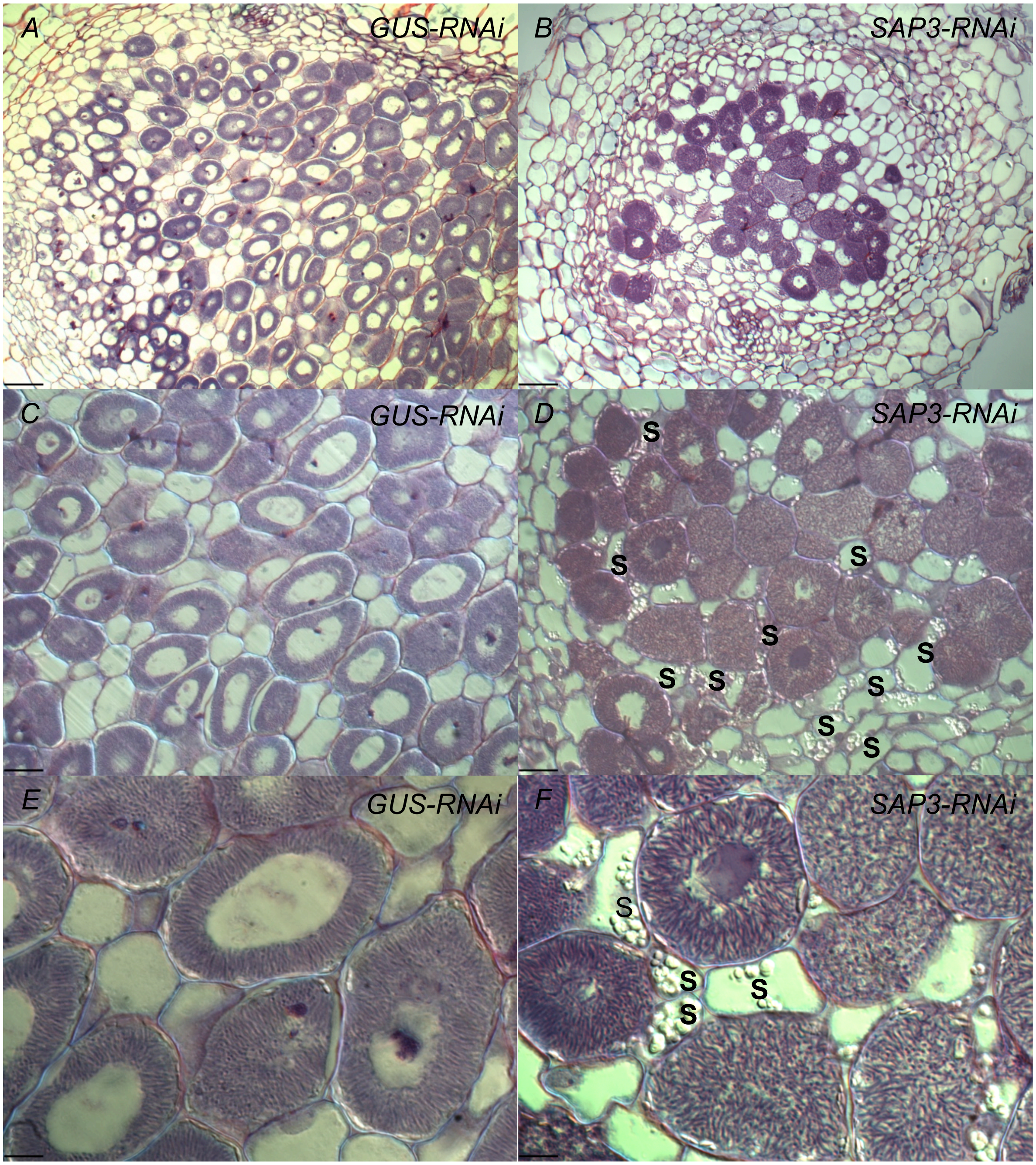


Fig. 6 Semi-thin sections of *SAP3-RNAi* and *GUS-RNAi* root nodules 3 wpi stained with Toludine Blue O. Scale bars = 50 µm (*A, B*), 25 µm (*C, D*), 10 µm (*E, F*).

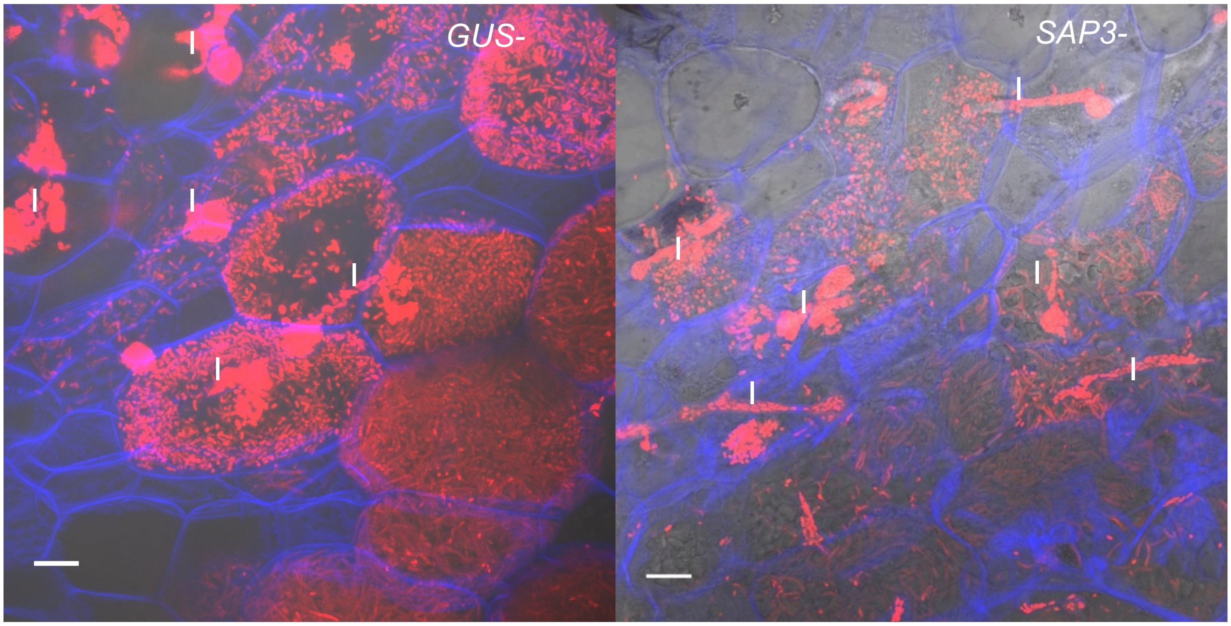


Fig 7. Confocal live cell imaging of *S. meliloti* expressing mCherry (red) in *SAP3-RNAi* root nodules stained with Calcofluor white (blue) 3wpi. (*A*) Image shows *S. meliloti* being released from infection threads (IT) into plant cells and differentiating into elongated bacteroids. (*B*) In *SAP3-RNAi* nodules *S. meliloti* release from ITs but lose fluorescence as the bacteria begin to differentiate. Scale bar = 10 µm.

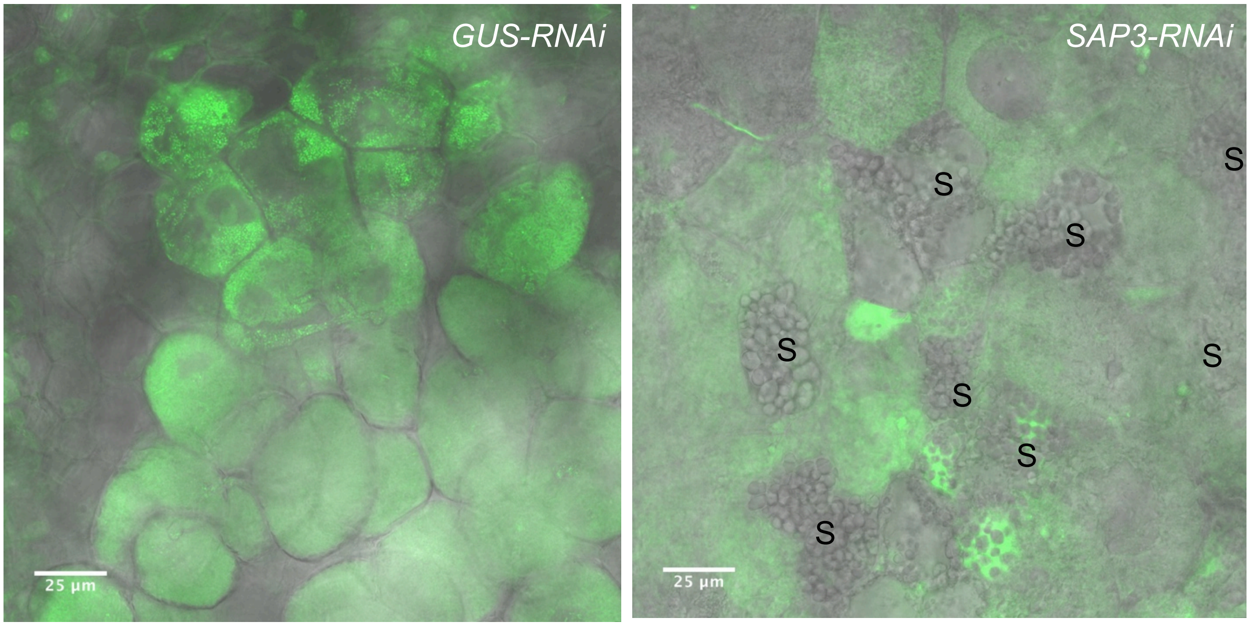


Fig 8. *SAP3-RNAi* nodules strongly accumulate starch in uninfected cells. Image show fresh nodule hand sections stained with Syto 13 for nucleic acids to label colonized plant cells. S, indicates heavy starch accumulation in *SAP3-RNAi* nodules.

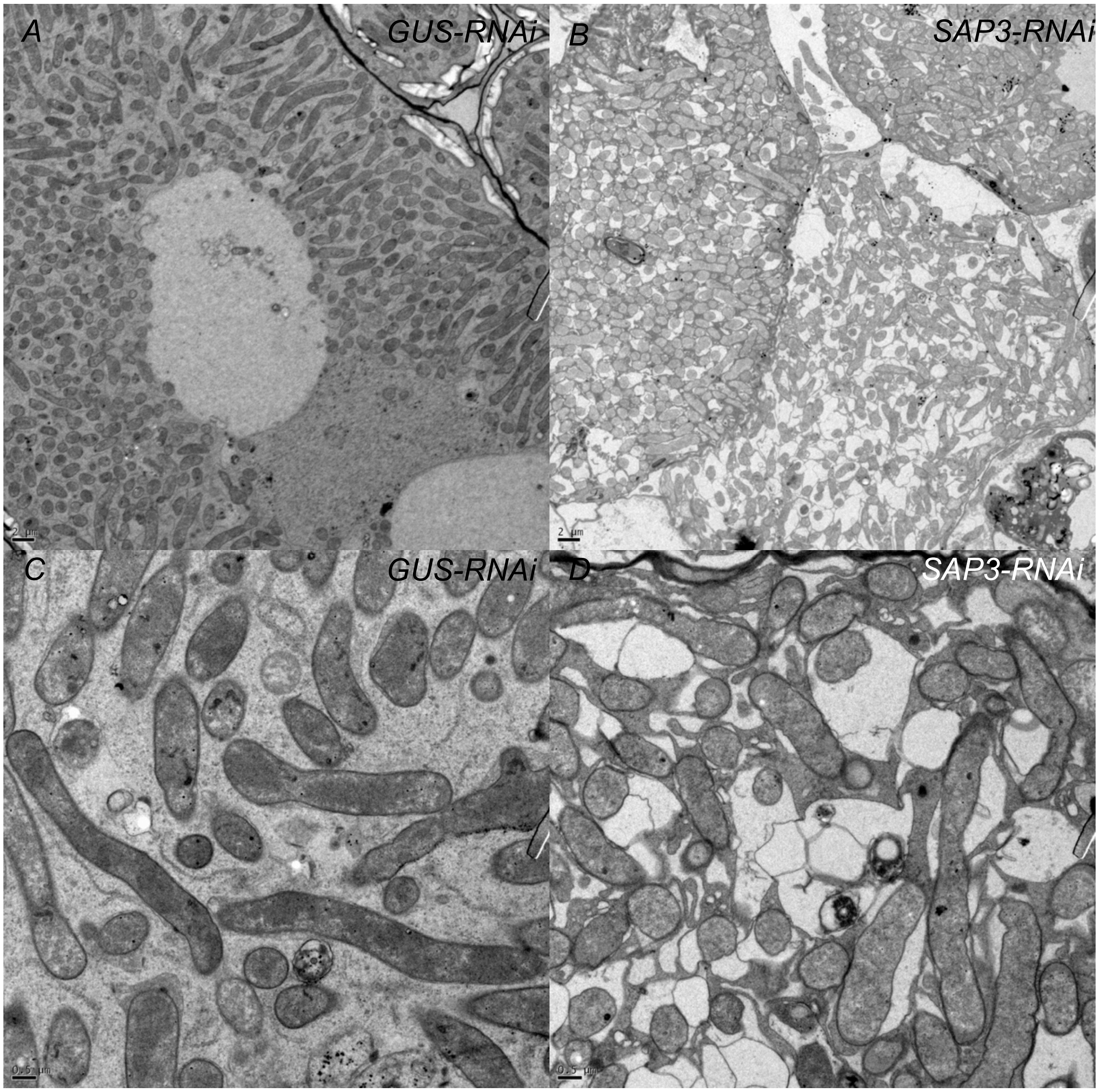


Fig 9. TEM images of *SAP3-RNAi* nodules 3 wpi with *S. meliloti.* Images show large spaces form between*S. meliloti* and the symbiosome membrane in *SAP3-RNAi* (*B, D*) nodules leading to poor bacterial growth and development.

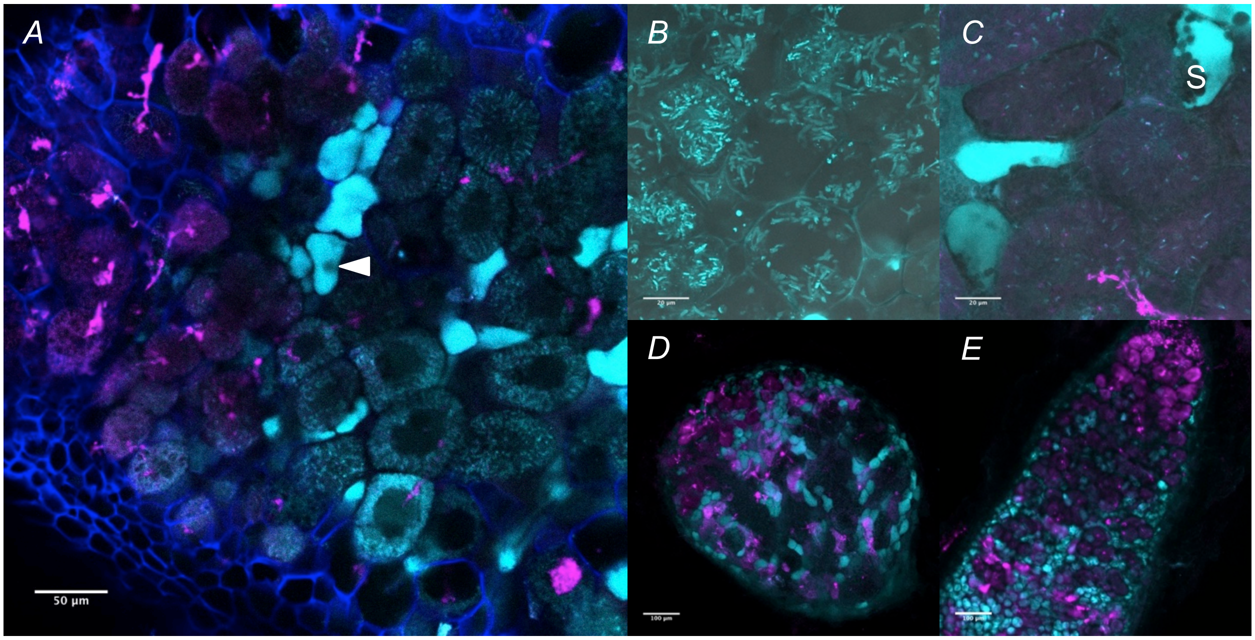


Fig 10. Localization of pSAP3:SP-GFP-SAP3 (cyan) in root nodules infected with *S. meliloti* expressing mCherry (magenta). (*A*) Nodule 3 wpi. Arrow head indicates the shadow of a nucleus in an uninfected cell. Plant cell walls were stained with Calcofluor white. (*B*) Young differentiating symbiosomes contain strong GFP-SAP3 signal (mCherry signal not imaged). Light plasmolysis with 0.8 M sucrose show that SAP3 signal is also present on the plasma membrane. (*C*) GFP-SAP3 signal is absent from mature symbiosomes and strongly accumulates inunifected cells, occasionally observed with starch, S. (*D*) SAP3-GFP expressing nodule 3 wpi. (*E*) SAP3-GFP expression nodule 6 wpi. Scale bars = 50 µm (*A*), 20 µm (*B, C*), and 100 µm (*D, E*).

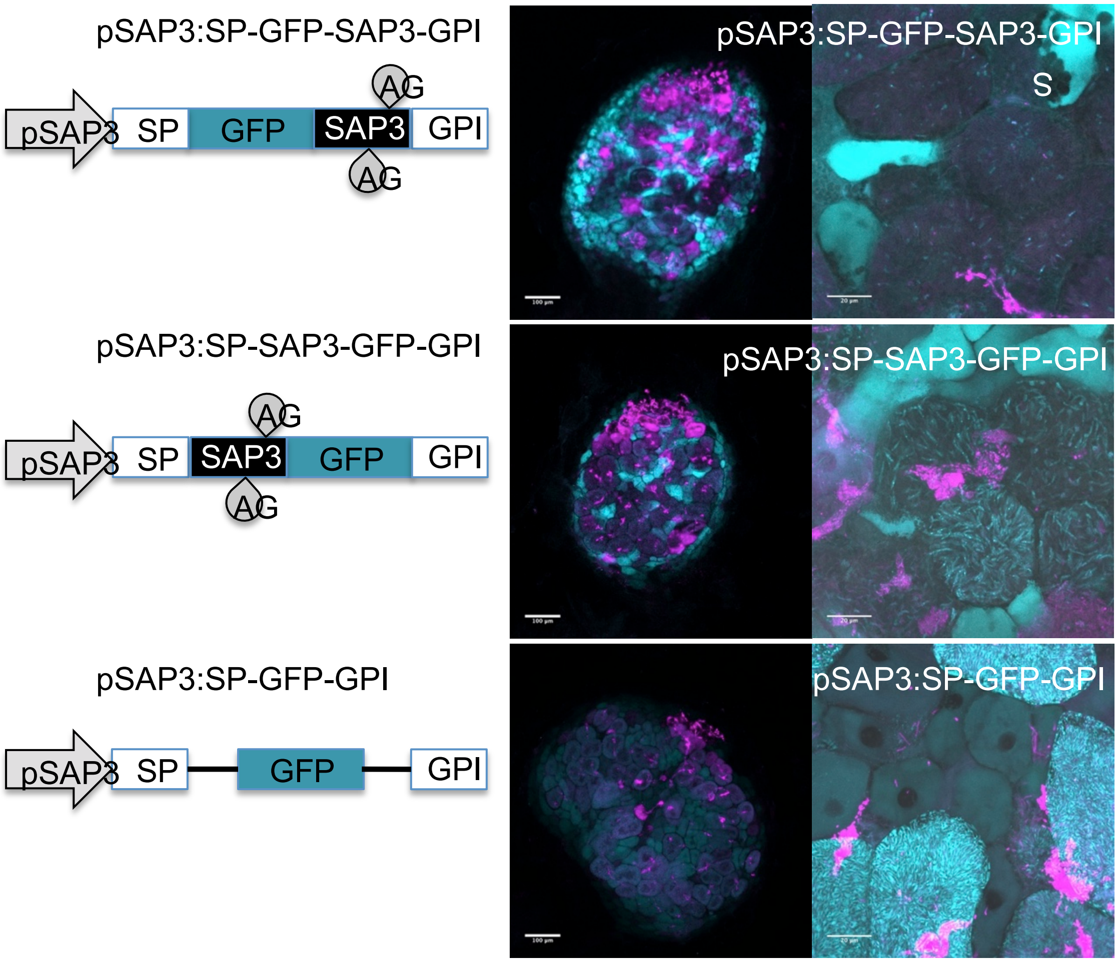


Fig 11. Localization of iterative SAP3 localization constructs in nodules infected with *S. meliloti* expressing mCherry. Merged images acquired by confocal microscopy are shown. Green/Red fluorescence represented as cyan/magenta. Scale bars = 100 µm and 20 µm.

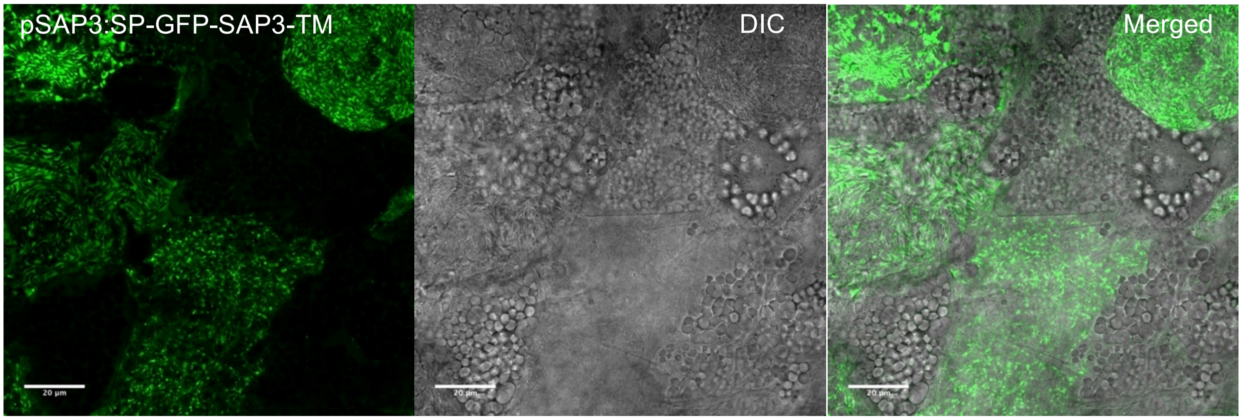


Fig 12. Localization of pSAP3:SP-GFP-SAP3-TM constructs in nodules 3 wpi with *S. meliloti.* GFP signal is retained in the symbiosome membrane and large amounts of starch accumulate in uninfected cells. Scale bar = 20 µm.

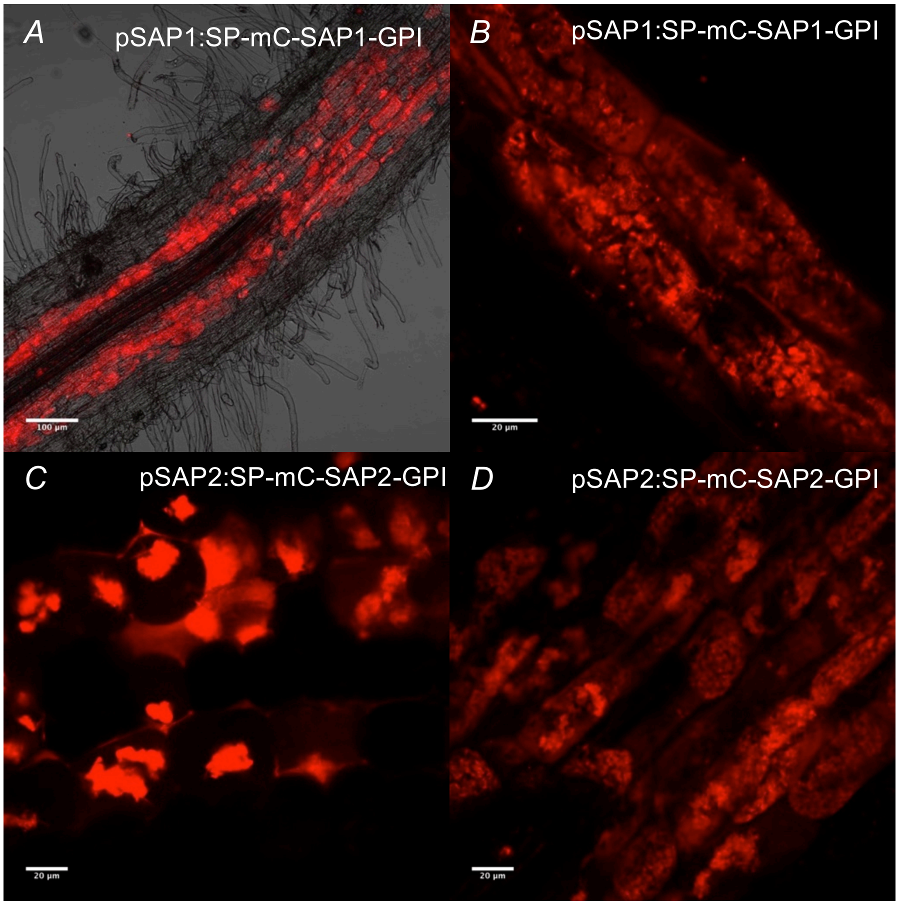


Fig 13. Localization of SAP1 and SAP2 in *M. truncatula* roots 4 wpi with *R. irregularis*. (*A, B*) SP-mCherry-SAP1-GPI localizes to the periarbuscular membrane. (*C, D*) SP-mCherry-SAP2-GPI also localizes to the periarbuscular membrane and can be observed in young arbuscules prior to branching (*C*). Scale bars = 100 µm (*A*) and 20 µm (*B-D*).

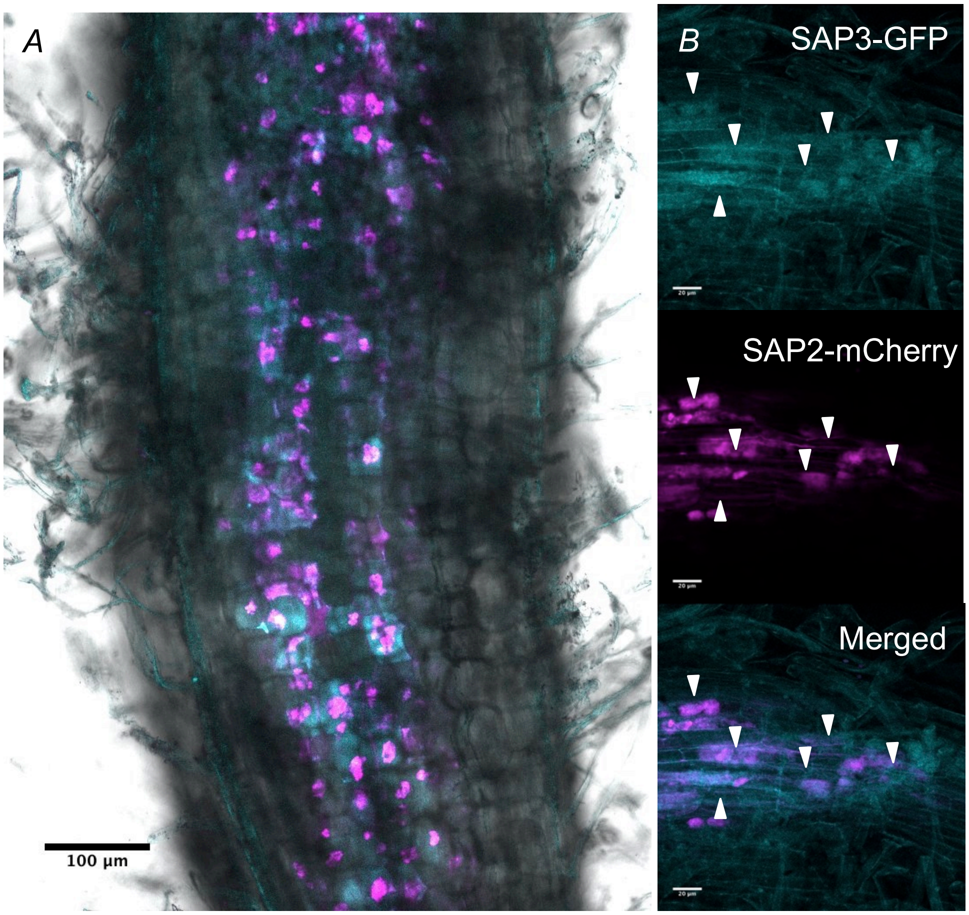


Fig 14. Co-localization of SAP3 and SAP2 in *M. truncatula* roots 4 wpi with *R. irregularis*. (*A*) Spatial localization pattern of SAP3 is correlated with arbuscule development. (*B*) Overlapping localization patterns of SP-GFP-SAP3-GPI and SP-mCherry-SAP2-GPI constructs. Scale bar = 20µm.

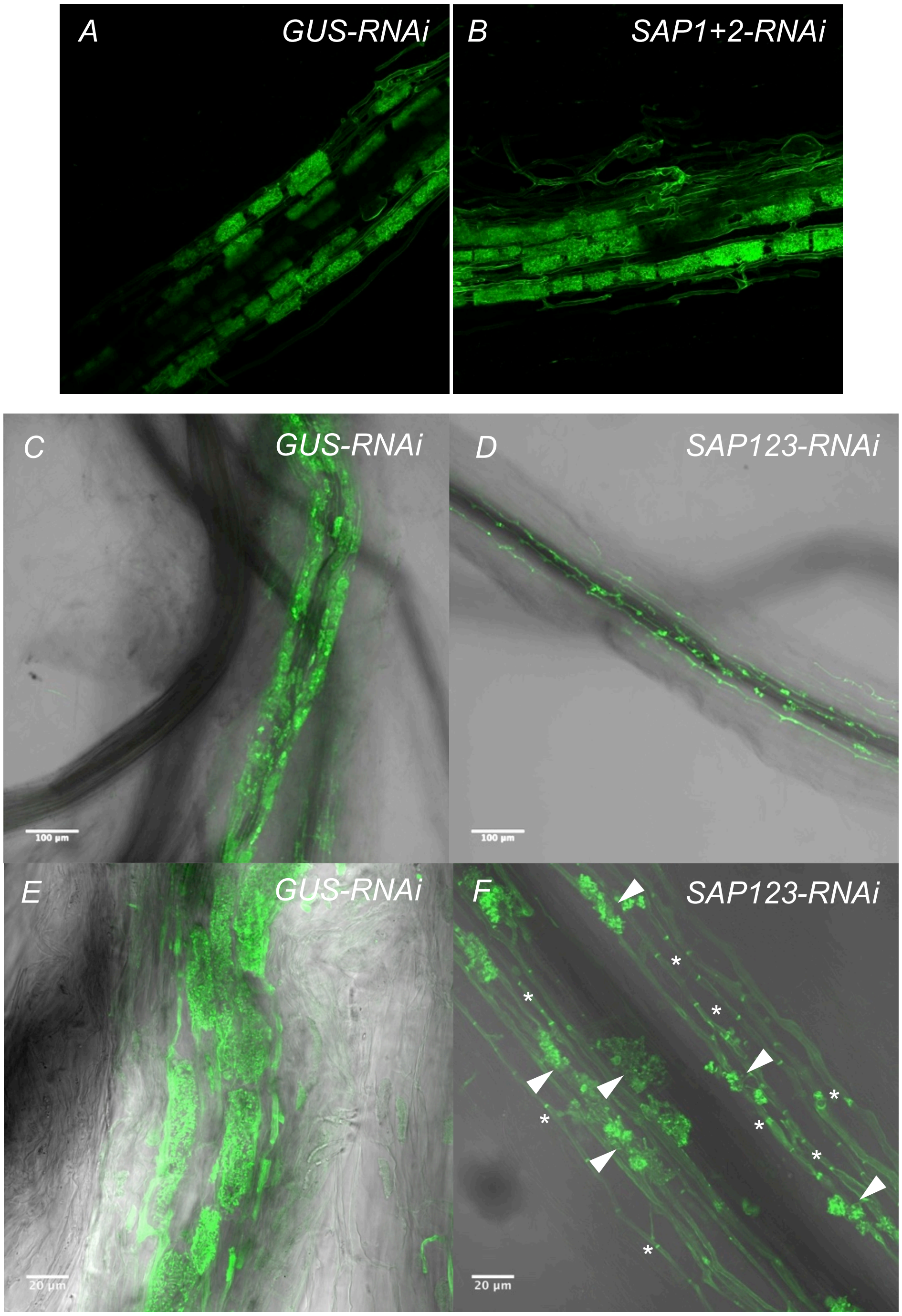


Fig 15. Preliminary phenotypes of *R. irregularis* colonized roots expressing *SAP-RNAi* constructs. (*A, B*) No mycorrhizal phenotype was observed associated with expression of the *SAP1+2-RNAi* construct relative *GUS-RNAi* control roots. (*C, D*) Expression of the *SAP1+2+3-RNAi* construct resulted in sparse root colonization and stunted arbuscule formation relative to *GUS-RNAi* control roots. Stunted arbuscules (arrows) and septate hyphae (asterisks) in *SAP1+2+3-RNAi* roots are indicated.

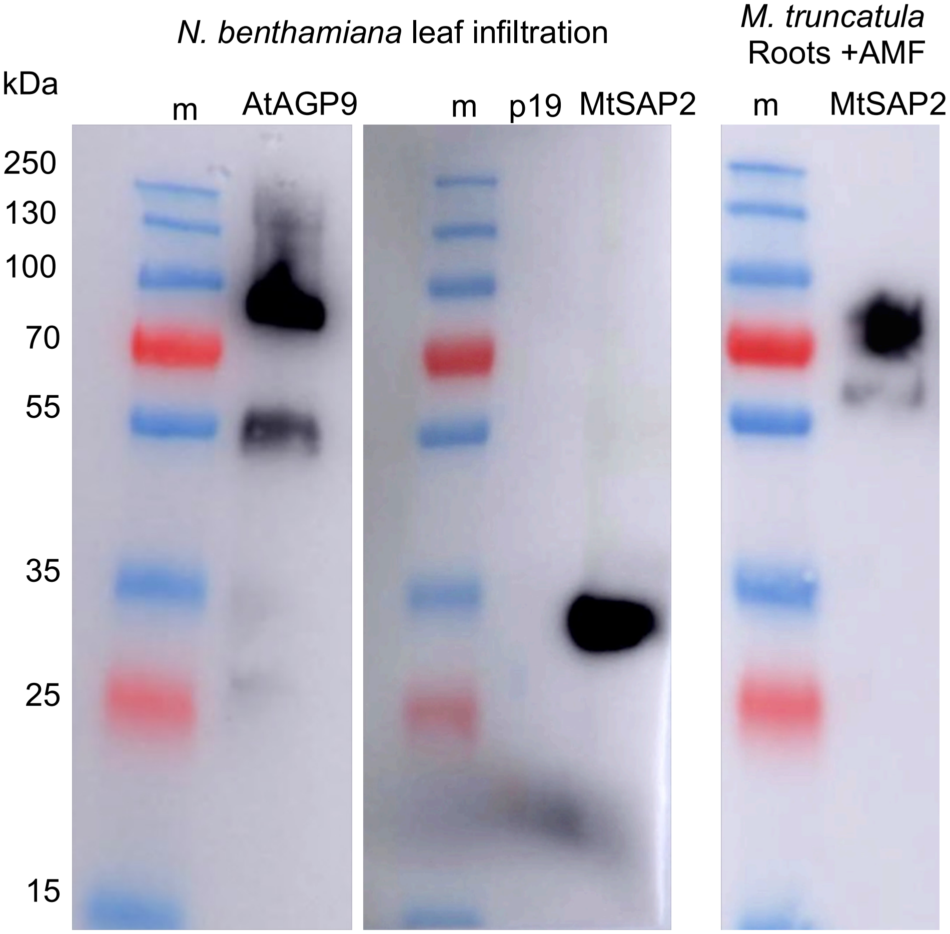


Fig 16. Western blot of AtAGP9-YFP-HA and SP-mCherry-SAP2-GPI proteins separated by reducing SDS-PAGE. PVDF membranes were probed with anti-HA and anti-mCherry primary antibodies (1:2,500) produced in mouse (Invitrogen) followed by (1:10,000) anti-mouse IgG HRP secondary antibody (Invitrogen).The mCherry-SAP2 fusion protein is predicted to be aprroximately 28 kDa, while AtAGP9-YFP-HA is predicted to be ~ 45 kDa. P19 was used as a negative control. PageRule Plus Prestained Protein Ladder (Invitrogen) was used as MW standard.

Table I. Primers used in this study

Primer number	Primer name	Primer sequence	Construct
WM001	SAP1-RNAi-F	CCATCTAGACAATAAAATGGCTTCACCAATG	RNAi primers
WM002	SAP2-RNAi-F	CCATCTAGACAACAATGGCTTTATCACTCAAG	
WM003	SAP3-RNAi-F	CCATCTAGACTACACCTTCAATAAGCTGATC	
WM004	SAP12-RNAi-F	CCATCTAGACAATGGAGAAGATGGGATAAATAT	
WM005	SAP123-RNAi-F	CCATCTAGACAATGGAGAAGATGGGATAAATAT	
WM006	SAP1-RNAi-R	GCATGAATTCACAACAACTTAAAGGCTAGTCC	
WM007	SAP2-RNAi-R	GCATGAATTCGACAAGACAAAGAGGGAAGCTT	
WM008	SAP3-RNAi-R	GCATGAATTCACAACATAGGAAGGTCAAAACG	
WM009	SAP12-RNAi-F	GCATGAATTCAGTAAGCACAAAAGGAAGAAGAAC	
WM010	SAP123-RNAi-F	GCATGAATTCTTAATACAACAATTCAATCAAACAACAT	
WM011	SAP1proGus-F	TATACTGCAGAAATAAATCACTTTGGACAAACTA	Promoter (GUS) Primers
WM012	SAP1proGus-R	ACTCTCGAGTTTATTGATTTCACTTGGTGATTTA	
WM013	SAP2proGus-F	TATACTGCAGACTTGTTAATTGTATGAAACTAGTA	
WM014	SAP2proGus-R	ACTCTCGAGTGTTGAATATATGCACTTGATTATA	
WM015	SAP3proGus-F	TATACTGCAGGCAAGCTTTTCAGCTATAAGC	
WM016	SAP3proGus-R	ACTCTCGAGTTTTGTTATGTTGAACTGATCAG	
WM017	qPCR EF-1α F	TGACAGGCGATCTGGTAAGG	qPCR Primers
WM018	qPCR EF-1α R	TCAGCGAAGGTCTCAACCAC	
WM019	qPCR SAP1 F	GATGGGATAAATATCAAAGTGTC	
WM020	qPCR SAP1 R	GCAGATGATTTTGGTGTTGG	
WM021	qPCR SAP2 F	CAACAATGGCTTTATCACTCAAG	
WM022	qPCR SAP2 R	TAGTCAGAACCACCTGATCC	
WM023	qPCR SAP3 F	CAGTAATTGGTGGATTTGTTCC	
WM024	qPCR SAP3 R	GGAACAAATCCACCAATTACTG	
WM025	SAP1pro-F	TACAAAAAAGCAGGCTCCGAATTATGCGTCGTAGCTGAACACTGTG	pSAP1: SP-

WM026	SAP1pro-R	TGCTCACCATAGCTGCAACAACCATAACCAAAACC	mCherry -SAP1-GPI, pCR8
WM027	SAP1mC-F	TGTTGCAGCTATGGTGAGCAAGGGCGAG	
WM028	SAP1mC-R	CATTGTTTTGCTTGTACAGCTCGTCCATGC	
WM029	SAP1term-F	GCTGTACAAGCAAAACAATGGAGAAGATGGG	
WM030	SAP1term-R	CAAGAAAGCTGGGTCGAATTATCACCATCAATCTACATCTTTTTATAT	
WM031	SAP2pro-F	TACAAAAAAGCAGGCTCCGAATTGTAATGTTATAGCAACATTTAC	pSAP2: SP-mCherry -SAP2-GPI, pCR8
WM032	SAP2pro-R	TGCTCACCATAGCAGCAGCAACTGCAATTAAAAC	
WM033	SAP2mC-F	TGCTGCTGCTATGGTGAGCAAGGGCGAG	
WM034	SAP2mC-R	TCCATATTGCTTGTACAGCTCGTCCATGC	
WM035	SAP2term-F	GCTGTACAAGCAATATGGATCAGGTGGTTCTG	
WM036	SAP2term-R	CAAGAAAGCTGGGTCGAATTGATATATGCTCTCGAAGGG	
WM037	SAP3pro-F	TACAAAAAAGCAGGCTCCGAATTGGGACTCATTAAGCAGAGATCAAAATTG	pSAP3:-SP-GFP-SAP3-GPI, pCR8
WM038	SAP3pro-R	TGCTCACCATACCGGCAACCACCATGGC	
WM039	SAP3gfp-F	GGTTGCCGGTATGGTGAGCAAGGGCGAG	
WM040	SAP3gfp-R	GCCCGTCGTGCTTGTACAGCTCGTCCATG	
WM041	SAP3term-F	GCTGTACAAGCACGACGGGCATGTTCAC	
WM042	SAP3term-R	CAAGAAAGCTGGGTCGAATTTGCCTTATCCTAATGTGAGGG	
WM043	SAP3pro-F	TACAAAAAAGCAGGCTCCGAATTGGGACTCATTAAGCAGAGATCAAAATTG	pSAP3: SP-SAP3-GFP-GPI, pCR8
WM044	SAP3pro-R	TGCTCACCATAGAAGGTGCTTCCGCCGG	
WM045	SAP3gfp-F	AGCACCTTCTATGGTGAGCAAGGGCGAG	
WM046	SAP3gfp-R	TTGCAAAGCTGCTCTTGTACAGCTCGTCCATG	
WM047	SAP3term-F	GTACAAGAGCAGCTTTGCAAGCAGCCTC	
WM048	SAP3term-R	CAAGAAAGCTGGGTCGAATTTGCCTTATCCTAATGTGAGG	
WM049	SAP3pro-F	TACAAAAAAGCAGGCTCCGAATTGGGACTCATTAAGCAGAGATCAAAATTG	pSAP3: SP-GFP-

WM050	SAP3pro-R	TGCTCACCATACCGGCAACCACCATGGC	GPI, pCR8
WM051	SAP3gfp-F	GGTTGCCGGTATGGTGAGCAAGGGCGAG	
WM052	SAP3gfp-R	TTGCAAAGCTGCTCTTGTACAGCTCGTCCATG	
WM053	SAP3term-F	GTACAAGAGCAGCTTTGCAAGCAGCCTC	
WM054	SAP3term-R	CAAGAAAGCTGGGTCGAATTTGCCTTATCCTAATGTGAGG	
WM055	SAP3tmd-R	CTCCCGCAATAGAAGGTGCTTCCGCCGGAG	pSAP3: SP-GFP-SAP3-TMD, pCR8
WM056	TMD-F	AGCACCTTCTATTGCGGGAGGAGTTGCTGCA	
WM057	TMD-R	GGCTGCTTGCTCAGTCCTGCGGCTTTTTCCTTCGC	
WM058	SAP3tmd-F	GCAGGACTGAGCAAGCAGCCTCAATTGTCACTC	

Reprogramming of GIPC Glycosylation is Critical for the Function of Symbiotic Membranes in *Medicago truncatula*

William M. Moore[1,2,3], Candace Chan[1,2], Toshiki Ishikawa[4], Emilie A. Rennie[1,2], Heidi M.-L. Wipf[3], Veronica Benites[1,2], Maki Kawai-Yamada[4], Jenny C. Mortimer[1,2], and Henrik V. Scheller[1,2,3]
[1]Joint BioEnergy Institute, Emeryville, California 94608
[2]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720
[3]Department of Plant and Microbial Biology, University of California, Berkeley, California 94720
[4]Graduate School of Science and Engineering, Saitama University, Japan

Abstract

Glycosyl inositol phosphorylceramides (GIPCs) are abundant components of the plant plasma membrane outer leaflet that are believed to function in membrane organization and signal transduction. Here we identified GLUCOSAMINE INOSITOL PHOSPHORYLCERAMIDE TRANSFERASE1 (GINT1) as a glycosyltransferase enzyme highly expressed in *Medicago truncatula* root nodules and demonstrate that this enzyme functions in the synthesis of glucosamine-decorated GIPC structures *in planta. GINT1* promoter characterization indicates that *GINT1* expression is developmentally regulated root nodules and mycorrhized roots associated with the development of perimicrobial membranes. Knockdown of the *GINT1* mRNA transcript did not affect overall root growth but strongly impaired nodulation and arbuscular mycorrhizal (AM) symbiosis, resulting in the senescence of symbiosomes and arbuscules. Our results indicate that while *M. truncatula* root sphingolipidome primarily consist of hexose-decorated GIPCs, symbiosis-induced reprogramming of GIPC glycosylation to contain glucosamine-decorated GIPCs is critical for the development, function, and persistence of plant-perimicrobial membranes during nodulation and AM symbiosis.

Introduction

Central to the development and function of plant endosymbiosis is the synthesis of a specialized host-derived membrane that serves as an interface between the plant and the microorganism. These membranes bear a symbiotic identity distinct from the plasma membrane based on protein composition (1, 2) and are imperative for facilitating the bidirectional exchange of nutrients and information (3-7). Moreover, symbiotic membranes help compartmentalize the microorganism and are essential for the intracellular accommodation and persistence of the endosymbiont inside plant cells (8, 9).

Development of the plant-microbial interface is highly coordinated with the intracellular accommodation and differentiation of the endosymbiont. Rhizobia and arbuscular mycorrhizal fungi enter plant cells through host-initiated asymmetric invaginations of the plasma membrane (10-12) that grow by means of targeted vesicle secretion and membrane fusion (13,14). While these symbioses culminate in vastly different morphological structures, they share a symbiosis-dedicated secretory pathway that drives the rapid growth of membrane around the endosymbiont (14-17).

Immunolabeling of symbiotic tissues with glycan-directed monoclonal antibodies has indicated that glycolipids and glycoproteins aggregate in peribacteroid and periarbuscular membranes (18-22) and it has been widely hypothesized that this interfacial glycocalyx could play an important role in mediating interactions between the host plant and microorganism. One of these antibodies, monoclonal antibody JIM18, has been shown to recognize a glucosamine, inositol, and phosphate-containing glycolipid developmentally regulated in young *Pisum sativum* (pea) symbiosome membranes (23). The JIM18 antigen was resistant to mild alkaline hydrolysis and, therefore, likely to be a glycosyl inositol phosphorylceramide (GIPC), which is a class of lipids resistant to alkaline hydrolysis unlike other phospholipids (24).

GIPCs are highly glycosylated sphingolipids found in plants and fungi that are believed to be the functional analogs of gangliosides in animal cells (25). GIPCs are the predominant sphingolipid found in plant tissues and are major components of the plasma membrane accounting for 25-40 mol% of total plasma membrane lipids (26, 27). GIPCs contribute to the asymmetry of the plasma membrane by preferentially accumulating in the outer leaflet of the phospholipid bilayer where they are believed to be involved in membrane organization, a variety of cell-surface interactions and cell-signaling processes (25, 28, 29), though exact mechanisms are not yet known. Due to the abundance of GIPCs in detergent-insoluble membranes (28), it has been suggested that they may contribute to the assembly of protein complexes in membrane microdomains, or so-called lipid rafts (30). Other evidence points to roles in protein anchoring, crosslinking between the cell wall and plasma membrane (31), and signaling (32-36) – none of which are mutually exclusive processes. Recently, proteins secreted by necrotrophic plant pathogens have been shown to specifically target the carbohydrate head group of GIPCs (37) indicating that microorganisms directly interact with these plant cell surface epitopes.

Plant GIPC head groups have a core structure consisting of an inositol phosphorylceramide (IPC) linked to an α-glucuronic acid (GlcA-IPC) followed by either a hexose (Hex-GlcA-IPC) or hexosamine/*N*-acetyl-hexosamine (HexN(Ac)-GlcA-IPC) residue, which can be further glycosylated to contain additional sugars (25, 38). GIPC head group composition is developmentally regulated in plant tissues and structurally diverse across plant species (38-41). Very little is known about the function of the carbohydrate head group. However, *Arabidopsis thaliana* mutants defective in GIPC head group biosynthesis clearly illustrate the importance of these structures in plant growth, development, and defense response (32-35, 42).

Here we have identified *Medicago truncatula* GLUCOSAMINE INOSITOL PHOSPHORYLCERAMIDE TRANSFERASE1 (GINT1) as a glycosyltransferase enzyme highly expressed in roots during nodulation and mycorrhizal symbiosis. We provide evidence that this enzyme functions in the synthesis of HexN(Ac)-GlcA-IPC structures *in planta*, consistent with the recently reported function of the *A. thaliana* and *Oryza sativa* GINT1 orthologs (43). Characterization of the *MtGINT1* promoter revealed dynamic spatiotemporal expression patterns in *M. truncatula* roots during both nodulation and AM symbiosis. RNAi-mediated silencing of *MtGINT1* did not affect overall root growth but significantly impaired nodulation and AM symbiosis leading to the senescence of symbiosomes and arbuscules. Our results show that while the *M. truncatula* root sphingolipidome primarily consists of Hex-GIPCs, local production of

HexN(Ac)-GIPC species within a symbiotic cell-specific context is necessary for the development, function, and persistence of plant-microbial membrane interfaces in *M. truncatula* roots.

Materials and Methods

Plant Material and Root Transformation

Medicago truncatula Gaertn. cv Jemalong A17 (Noble Foundation) was used for all experiments. Hairy root transformation was conducted according to Floss *et al.* (44) using *Rhizobium rhizogenes* strain Arqua1. Chimeric plants were grown in cones containing sterile sand under a 16 h photoperiod and 22 °C ambient temperature. Plants were allowed to recover for one week prior to inoculation with either *Sinorhizobium meliloti* or *Rhizophagus irregularis*. In normal growth experiments plants were fertilized twice a week with half-strength MS media.

Nodulation Experiments

Wild-type *Sinorhizobium meliloti* strain 1021 (ATCC 51124) was used for all experiments except when indicated. *S. meliloti* 1021 expressing mCherry was provided as a generous gift from Dr. Sharon Long (45) and used for live cell imaging experiments. Roots were flood inoculated with *S. meliloti* grown in liquid Tryptone Yeast extract (TY) media and re-suspended in 5 ml (OD 0.1 half-strength MS media without nitrogen (PhytoTechnology Laboratories). Plants were fertilized twice a week and nodulated root systems were assayed at 3 wpi and 4 wpi.

Mycorrhizal Experiments

Rhizophagus irregularis spores (DAOM 181602) were purchased from Premier Tech (Rivière-du-Loup, Québec, Canada) and used for all experiments. Approximately 1200 spores were pre-germinated overnight in dilute root exudate at room temperature with gentle rocking and applied directly to roots during transplanting. Plants were grown in sterile sand and fertilized twice a week with half strength Hoaglands Basal Salt media modified with 20 µM phosphate (46). Root colonization 5 wpi with *R. irregularis* was quantified by the gridline intersect method (47) at 100X magnification on a Leica DM6B microscope (Leica Microsystems, Heerbrugge, Switzerland). Root fragments were sampled at random from five independently transformed root systems per construct and scored with gridlines set at 200 µm intervals. A total of 480 lines and over a thousand root-line intersects were scored for each root system.

Cloning

RNAi constructs targeted a 400-base pair region beginning at the start codon of the *MtGINT1* coding sequence that was determined to be gene-specific by BLAST search against the *M. truncatula* genome. The sequence was cloned into the Gateway entry vector pENTR1A (Thermo Fisher Scientifc) using XbaI/XhoI restriction sites and recombined into pHELLSGATE12 RNAi vector (48) by LR reaction. A 423 bp sequence targeting the *E. coli* GUS gene was used as a control. pHELLSGATE12 RNAi vectors were used for all nodulation experiments. Mycorrhization and GIPC experiments used the RNAi vector pRNAi-GG (49) that we modified to contain the *pAtUBQ10:DsRed:tNOS* fluorescent marker cassette from expression vector p921 (50).

The fluorescent marker cassette from p921 was amplified by PCR and recombined into SpeI-linearized RNAi-GG plasmid by In-Fusion cloning (Invitrogen) to generate the pRNAi-GGred vector. In this case the same 400-bp sequence corresponding to *GINT1* was directly cloned into pRNAi-GGred using BsaI restriction sites. For *GINT1* promoter reporter constructs 2 kb of genomic sequence upstream from the start codon was cloned into TA-TOPO GW entry vector pCR8 (Thermo Fisher Scientific) and confirmed by sequencing. Entry clones were recombined by LR reaction into GUS-expression vector pGWB3 and YFP-expression vector pGWB40 (51) for GUS staining and fluorescence microscopy. For overexpression of *MtGINT1* in *M. truncatula* roots, the full-length *GINT1* coding sequence was cloned directly into p921 using SpeI and MluI restriction sites to generate *AtUBQ3:MtGINT1*. For overexpression of *MtGINT1* in *A. thaliana*, the coding sequence without stop codon was cloned into TA-TOPO pCR8 and recombined by LR reaction into pGWB17 to generate 35s:MtGINT1-3xMYC. All constructs were confirmed by sequencing and primers are listed (SI Table I).

Histochemistry

GUS staining of *M. truncatula* root systems with X-Gluc was conducted as previously described (52). All bright-field images were collected on a Leica DM6 microscope.

Confocal Laser Scanning Microscopy

For live cell imaging of root nodules *S. meliloti* strain expressing mCherry was used and plant cell walls were stained with Calcofluor White. For live dead staining experiments wild-type *S. meliloti* was used and nodules were stained in 5 µM Syto9 and 30 µM propidium iodine for 20 min at room temperature. In both cases fresh nodule hand sections were made and immediately imaged. Mycorrhizal roots colonized by *R. irregularis* were visualized by staining fungal chitin with wheat germ agglutinin Alexa Fluor 488 (Thermo Fischer Scientific) at 0.5 µg/ml in 80 µM PIPES buffer (pH 7.0). Mycorrhized roots were fixed in 50% (v/v) ethanol, cleared in 20% (w/v) KOH at 65C for two days, neutralized, and stained with WGA-488 for at least 3 days at 4°C prior to imaging. Images were captured on a Zeiss LSM 710 laser scanning microscope (Carl Zeiss, Oberkochen, Germany).

Electron Microscopy

Nodules were collected into 0.1 M cacodylate buffer (pH 7.2) containing 2% (v/v) EM-grade glutaraldehyde and fixed overnight at 4°C. Nodules were rinsed 3× with 0.1 M cacodylate buffer (pH 7.2), stained for 1 h with 1% (w/v) osmium tetraoxide, and rinsed 3× with buffer, followed by an additional 3 rinses with distilled water. Nodules were then dehydrated in an acetone gradient (35/50/70/80/95/100/100% v/v) and bisected longitudinally to aid resin infiltration. Acetone-Epon resin was infiltrated at 2:1, 1:1, and 1:2 dilution ratios for 1 h each, followed by pure Epon, which was freshly changed after one hour and allowed to infiltrate overnight with gentle rocking. Samples were infiltrated with pure resin plus accelerator for two hours, embedded in Pelco molds, and polymerized in a 65°C oven for two days. Ultra-thin 70 nm thick sections were collected on to grids using a diamond knife and Reichert microtome. Grids were stained using a Pelco Grid Staining System with 2% (v/v) aqueous uranyl acetate for 5 min, lead citrate for 5 min, and rinsed at least 5× with distilled water in between. Grids were imaged using a Technai 1200 electron microscope.

RNA Extraction and qPCR

Total RNA was extracted from roots using Trizol (Thermo Fisher), treated with Turbo DNAseI (Thermo Fisher) according to manufacturer's directions, and further purifed using an RNEAsy column (Qiagen). cDNA was prepared using a cDNA synthesis kit (BioRad) from 500 ng of total RNA. *GINT1* qRT-PCR primers (SI Table I) were chosen to span the junction between the third and fourth exon outside the RNAi-targeted region of the transcript. Gene expression was quantified using the $\Delta\Delta C_T$ method relative to the housekeeping gene EF-1α.

Acetylene Reduction Assay

Nitrogen fixation was assayed in silenced roots 4 wpi with *S. meliloti*. Chimeric plants were gently uprooted from sand and placed on a strip of Whatman filter paper (pre-wet with 1/2x MS without nitrogen) inside a 25 ml culture tube topped with a rubber septum. Acetylene (1 ml) was injected into the culture tube, which was placed in a growth chamber for 6 h prior to sampling. Gas samples were measured using gas chromatography flame ionization detection on a Shimadzu GC-2014.

Salicylic Acid Measurements

Salicylic acid (SA) metabolite extraction and measurement was done as previously described (34, 35), but HCl hydrolysis was omitted in order to measure active SA within the greater pool of SA-glucoside.

Sphingolipidomics

Lyophilized plant tissues (5 to 10 mg) were homogenized in 450 µl of methanol/1-butanol (1:2, v/v) and subjected to heat-denaturation of enzymes at 80°C for 10 min. The homogenate was mixed with 300 µl of 1 N KOH and further incubated at 60°C for 30 min to eliminate glycerolipids. The mixture was acidified with 1.5 ml of 0.4 N HCl and extracted with additional 1 ml 1-butanol. After vigorous shaking and centrifugation, the upper 1-butanol layer was collected and evaporated. The residue was dissolved in 150 µl of THF/methanol/water (2:1:2, v/v/v) containing 0.1% formic acid. GIPC composition was analyzed by LC-MS/MS according to previous reports (34, 53).

Results

MtGINT1 Belongs to Glycosyltransferase Family 64 and is Expressed in Symbiotic Tissues

MtGINT1 (Medtr1g090920) belongs to a small three-member clade of GT family 64 enzymes (Fig. 1 *A*) (54) and was identified based on several recent transcriptomic studies (55-58) as being highly upregulated in root nodules and roots colonized by mycorrhizal fungi (SI Fig. 1). Two members from this family have recently been characterized in *A. thaliana* and *O. sativa* as GTs involved in synthesizing the polar glycan head group of GIPCs. *AtGMT1* (34) and *AtGINT1/OsGINT1* (43) function as mannosyl- and glucosaminyl- transferases, respectively, and represent a major fork in

the GIPC glycan biosynthesis pathway that produces distinct classes of GIPC (Fig. 1 *C*). Transcriptome profiling of GT family 64 in *M. truncatula* roots using the Symbimics database (55) show that *MtGINT1* is highly upregulated in root nodules with only small transcriptional changes occurring in *MtGMT1* or *MtEPCL1* expression (Fig. 1 *B*).
The *MtGINT1* promoter was used to drive expression of GUS or YFP reporter constructs to further investigate the spatiotemporal of *MtGINT1* in hairy roots. Roots expressing YFP reporter constructs confirmed that *MtGINT1* is highly expressed in root nodules (Fig. 1 *D*). GUS staining in nodules was primarily restricted to the infection zone and interzone (Fig. 1 *E*, arrow), with little observable staining below in the nitrogen fixation zone. This is in agreement with RNA-seq data acquired from different developmental zones of the nodule (58, 59) (SI Fig. 1). Roots inoculated with *R. irregularis* showed staining in the inner root cortex associated with the spatial domain of AM colonization, which was detected by staining with WGA-488 for fungal chitin (Fig. 1 *F, G*). This is consistent with the arbuscule-induced expression of *MtGINT1* reported in laser capture micro-dissection microarray studies (57, 58) (SI Fig. 1). This implies that HexN(Ac)-type GIPCs, perhaps as opposed to Hex-type GIPCs, could play an important role in facilitating these two beneficial endosymbioses.

MtGINT1 **Silencing Impairs Nodulation**

RNAi-mediated gene silencing was used to investigate *MtGINT1* function in hairy roots. Reduction of *MtGINT1* transcript levels did not affect overall root growth under normal growth conditions (SI Fig. 2) but strongly impaired nodulation resulting in small white nodules and chlorotic leaves consistent with a loss of symbiotic nitrogen fixation (Fig. 2 *A-F*). We tested nitrogen fixation efficiency using the acetylene reduction assay and confirmed that nitrogen fixation is substantially impaired in *MtGINT1*-RNAi nodules (Fig. 2 *H*). Since nodules were able to form but remained small, white, and incapable of fixing nitrogen (nod+/fix-) we conclude that GINT1 function is required at early stages of nodule development but not required for nodule initiation.

We further investigated the *MtGINT1*-RNAi phenotype using live cell imaging of *S. meliloti* expressing mCherry in nodules 3 wpi. Control nodules were filled with elongating bacteroids (Fig. 2 *I*); however, *GINT1*-RNAi nodules had apparent defects in infection thread release and symbiosome development (Fig. 2 *J*). Release of bacteria from infection threads was disorganized in the infection zone and bacteria appeared to senesce within differentiating symbiosomes (Fig. 2 *J, L*; SI Fig. 3). Transmission electron microscopy revealed that large amounts of membrane debris aggregated at infection thread tips and *S. meliloti* was improperly released into disorganized clouds of nascent membrane, which appeared uncoordinated and non-adherent to the bacteria (Fig. 2 *L*). Prior to infection thread release *S. meliloti* were noted to contain distinctly thick layers of exopolysaccharide not seen in control nodules (Fig. 2 *L*), suggesting impaired interaction with the infection thread membrane. Below the infection zone *S. meliloti* were highly degraded with most cells being completely collapsed and a large amount of presumably lytic compartments accumulated alongside the senescing bacteria (Fig. 2 *N*). Live-dead staining using nucleic acid stains Syto 9 (live) and propidium iodide (dead) further supported that the majority of plant cells in *GINT1*-RNAi nodules contain senescent bacteria (SI Fig. 4).

***MtGINT1* Silencing Does Not Increase Salicylic Acid Content**

Since *A. thaliana* GIPC glycosylation mutants *iput1*, *gmt1*, and *gonst1*, have phenotypes that include ectopic cell death, H_2O_2 production, and salicylic acid (SA) accumulation (34-35), and links between sphingolipid metabolism and SA have been reported(59, 60), we questioned whether the cell death phenotype observed in *GINT1*-RNAi nodules could be a result of SA accumulation. SA content in RNAi roots was measured using LC-MS/MS but no significant difference between *GINT1*-RNAi roots and *GUS*-RNAi control roots was found (Fig. 2 *G*). In nodulation experiments *GINT1*-RNAi roots actually had significantly less SA than *GUS*-RNAi control roots, indicating that the bacterial senescence in *GINT1* silenced nodules is not likely a result of an SA-mediated defense response.

MtGINT1 Functions as a GIPC HexN(Ac)-Transferase *in planta*

A. thaliana and *O. sativa* GINT1 orthologs have recently been characterized as GIPC *N*-acetyl-glucosaminyltransferases (43). To further test this function in *M. truncatula,* we took a loss-of-function/gain-of-function approach using LC-MS/MS to measure GIPC content in hairy roots expressing either *MtGINT1*-RNAi or *MtGINT1*-overexpression constructs. LC-MS/MS data from control roots showed that the *M. truncatula* root sphingolipidome contained both Hex-GlcA-IPC and HexN(Ac)-GlcA-IPC structures, but predominantly consisted of the Hex variety, which constituted three quarters of the total GIPC population (SI Fig. 5 *A*). Both Hex and HexN(Ac)-type GIPCs use the same pool of ceramide bases (SI Fig. 5 *B*) indicating that the only difference between these GIPC structures is the single monosaccharide substitution in the carbohydrate head group. Expression of the *MtGINT1*-RNAi construct reduced total HexN(Ac)-GlcA-IPC content by 65% relative to *GUS*-RNAi control roots (Fig. 3 *A*) and impacted all HexN(Ac)-containing GIPC species – while roots carrying *AtUBQ3:MtGINT1* overexpression constructs had a 160% relative increase in total HexN-GlcA-IPC content (Fig. 3 *B*). Overall changes in Hex-GlcA-IPC content were not significantly different in either experiment (Fig. 3 *C, D*) and supports that *MtGINT1* expression specifically affects the HexN(Ac) GIPC population. To garner more support for *MtGINT1* function as a HexN(Ac) GIPC transferase we ectopically overexpressed *MtGINT1* in *A. thaliana*, which completely lack HexN(Ac) GIPC structures in vegetative tissues. Heterologous expression of MtGINT1 protein in *A. thaliana* resulted in non-native HexN(Ac)-decorated GIPCs in leaf tissue that were not detected in wild-type plants (Fig. 3 *E, F*). These results support that MtGINT1 functions in the glycosylation of HexN(Ac)-decorated GIPCs similar to the *O. sativa* and *A. thaliana* GINT1 orthologs.

***MtGINT1*-RNAi Roots have Increased Arbuscule Senescence**

Since the *GINT1* promoter is spatially active in both nodules and AM colonized roots, and *GINT1* silencing imparts a strong nodulation phenotype, we were interested to determine if GINT1 activity is also required to sustain AM symbiosis. To test this we inoculated silenced root systems with *R. irregularis* and allowed the symbiosis to form for 5 weeks under phosphate-limited conditions. *GINT1*-silenced roots were found to support less AM colonization and fewer arbuscules relative to *GUS*-RNAi control roots (Fig. 4 *E*) and an increase in fungal senescence was observed (Fig. 4 *B, D*). While

GUS-RNAi control roots contained abundant highly branched arbuscules (Fig. 4 *A, C*), most of the arbuscules formed in *GINT1*-RNAi roots appeared stunted and septate with septa also occurring throughout the intercellular mycelium. We measured expression of the arbuscule-specific phosphate transporter *PT4* (61) by qRT-PCR as an independent measure of arbuscule abundance and symbiotic function and found that *PT4* expression is significantly reduced in *GINT1*-silenced roots (Fig. 4 *F, G*).

Discussion

Plant endosymbiosis relies heavily on dynamic cell surface interactions between the microorganism and the plant plasma membrane that culminate in the development of expansive cellular membrane-interfaces. Previous work by Perotto *et al.* (23) identified a glucosamine-containing GIPC epitope that is developmentally regulated in young pea symbiosome membranes, suggesting that plant glycosphingolipids may play an important role in this process. Consistent with this observation we have identified the GT *MtGINT1* as highly upregulated in *M. truncatula* root nodules and provide evidence that this enzyme functions in the synthesis of HexN(Ac)-GlcA-IPC structures *in planta*. While the *M. truncatula* root sphingolipidome predominantly consists of Hex-GlcA-IPC structures, analysis of the *MtGINT1* promoter indicates that local synthesis of HexN(Ac)-GlcA-IPC structures occurs in symbiotic tissues. Knockdown of the *MtGINT1* mRNA transcript does not affect normal root growth but strongly impairs symbiosis, resulting in the compromised development and rapid senescence of symbiosomes and arbuscules. SA accumulation is not increased in *MtGINT1* silenced roots, therefore the symbiosis-defective phenotypes appear to be the direct consequence of altered GIPC glycan structure, and not indirect effects of SA-induced hypersensitive response. Taken together, our data indicate that symbiosis-induced reprogramming of GIPC glycosylation by *MtGINT1* to contain HexN(Ac)-GlcA-IPC structures is essential for the development, function, and persistence of plant perimicrobial membranes in *M. truncatula* roots.

While more experimentation is needed to determine the role GIPCs play in symbiotic membranes, it is likely that they have important functions in membrane organization and cell signaling. GIPCs are believed to assist in the assembly of membrane protein microdomains (26, 30) and there is evidence to support that such organization occurs during symbiosis (62). While altering membrane topology is expected to have pleiotropic effects, we speculate that part of the *MtGINT1*-RNAi nodulation phenotype could be explained by compromised exopolysaccharide receptor 3 (*Epr3*) signaling. *Epr*3 is a LysM-containing receptor-like kinase localized to the infection thread membrane that binds to rhizobial EPS and functions in the targeted membrane-restricted release of rhizobia into plant cells (52, 63). We propose that impaired *Epr3* signaling by membrane perturbation could result in the excessive secretion of EPS and uncoordinated release of rhizobia observed in *MtGINT1*-silenced nodules. Though more work is needed to understand protein dynamics within this altered membrane environment.

Not mutually exclusive from this is the possibility is that GIPCs are involved in important membrane-pectin crosslinking in symbiotic membranes. GIPCs have previously been implicated in forming boron bridges with the pectic polysaccharide rhamnogalacturonan II (32) and related cell adhesion phenotypes have been reported for the *gmt1* and *gint1* mutants in *A. thaliana* and *O. sativa* (43, 64). Boron deficiency impairs nodule infection (65) and the presence of RGII, glycolipids, and glycoporoteins epitopes in symbiosome membranes in boron-dependent (66). Pectin remodeling at infection thread tips is necessary for bacterial release into plant cells (67) and pectic epitopes are present in the periplasmic space of young arbuscules and symbiosomes (66, 68). Therefore, we cannot rule out the possibility that membrane crosslinking to nascent pectin networks by GIPCs could be important for the rapid coordinated growth of perimicrobial membranes or their adhesion to the microorganism.

However, perhaps the most intriguing questions are why do developmentally regulated changes in GIPC glycosylation occur and what is the functional significance of the glucosamine glycan structure in symbiosis? The spatial regulation of the *MtGINT1* promoter and strong symbiosis phenotypes observed in *MtGINT1*-RNAi roots suggest a unique and important role for glucosamine-decorated GIPCs that hexose GIPC structures cannot fulfill. This notion is supported by the fact that the rice OsGINT1 protein is only partially able to rescue the *A. thaliana gmt1* mutant, which indicates that the sugar composition of the glycan head group is important for GIPC function (43) and suggests that Hex-type and HexN(Ac)-type GIPC structures are not completely interchangeable. In *M. truncatula* it is tempting to speculate whether changes in GIPC glycosylation could reflect inherent signaling properties associated with HexN(Ac) decorated GIPC structures in symbiosis. Previous groups have proposed that glycolipids and glycoproteins released from developing symbiosome membranes may function in signaling (23, 66) and other glucosamine-containing molecules like lipochitooligosaccharides, chitooligosaccharides, and glucosamine derivatives play prominent roles in symbiosis signaling (69-72). It is also tempting to speculate whether these GIPCs impart a common membrane topology shared in symbiosis or have structural features recognized by the endosymbiont similar to plant-pathogen interactions.

Lastly, while we have provided evidence to support that *MtGINT1* function is imperative for symbiosis in *M. truncatula*, it remains unclear how well conserved this phenomenon is across plant species and their shared interactions with AMF and other endophytes. GIPC head groups have a high degree of structural diversity in glycosylation patterning across plant species; therefore it's reasonable to question whether similar changes in GIPC head group composition occur during symbiosis. In *M. truncatula* it is still possible that larger GIPC glycan structures exist, as these more highly glycosylated GIPCs are not well detected by LC-MS/MS methods. This opens up many new and exciting questions to be further investigated regarding the structure-function relationship of GIPCs in plant membranes and related interactions with microorganisms.

Acknowledgements

This work conducted by the Joint BioEnergy Institute was supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy. Part of the work was supported by an NSF graduate fellowship to WMM (DGE 1106400), and by JSPS KAKENHI to TI (17K15411) and MKY (26292190). We thank Dr. Mi Yeon Lee for assistance with plant growth.

Author Contributions

WMM, JCM and HVS designed the research. WMM, CC, TI, MKY, EAR and HW conducted experiments and analyzed data. WMM and HVS wrote the manuscript. All authors discussed results and approved the final manuscript.

References

1. Pumplin N, Harrison MJ (2009) Live-Cell Imaging Reveals Periarbuscular Membrane Domains and Organelle Location in *Medicago truncatula* Roots during Arbuscular Mycorrhizal Symbiosis. *Plant Physiol* 151(2):809–819.

2. Catalano CM, Lane WS, Sherrier DJ (2004) Biochemical characterization of symbiosome membrane proteins from *Medicago truncatula* root nodules. *Electrophoresis* 25(3):519–531.

3. Zhang Q, Blaylock LA, Harrison MJ (2010) Two *Medicago truncatula* Half-ABC Transporters Are Essential for Arbuscule Development in Arbuscular Mycorrhizal Symbiosis. *Plant Cell* 22(5):1483–1497.

4. Harrison MJ, Dewbre GR, Liu J (2002) A Phosphate Transporter from *Medicago truncatula* Involved in the Acquisition of Phosphate Released by Arbuscular Mycorrhizal Fungi. *Plant Cell* 14(10):2413–2429.

5. Guether M, et al. (2009) A Mycorrhizal-Specific Ammonium Transporter from *Lotus japonicus* Acquires Nitrogen Released by Arbuscular Mycorrhizal Fungi. *Plant Physiol* 150(1):73–83.

6. Clarke VC, Loughlin PC, Day DA, Smith PM (2014) Transport processes of the legume symbiosome membrane. *Front Plant Sci* 5:699.

7. Vincill ED, Szczyglowski, Roberts DM (2005) GmN70 and LjN70. Anion Transporters of the Symbiosome Membrane of Nodules with a Transport Preference for Nitrate. *Plant Physiol* 137(4):1435–1444.

8. Bapaume L, Reinhardt D (2012) How membranes shape plant symbioses: signaling and transport in nodulation and arbuscular mycorrhiza. *Front Plant Sci* 3:223.

9. Rich MK, Schorderet M, Reinhardt D (2014) The role of the cell wall compartment in mutualistic symbioses of plants. *Front Plant Sci* 1:699.

10. Genre A, Chabaud M, Timmers T, Bonfante P, Barker DG (2005) Arbuscular Mycorrhizal Fungi Elicit a Novel Intracellular Apparatus in Medicago truncatula Root Epidermal Cells before Infection. *Plant Cell* 17(12):3489–3499.

11. Genre A, Chabaud M, Faccio A, Barker DG, Bonfante P (2008) Prepenetration Apparatus Assembly Precedes and Predicts the Colonization Patterns of

Arbuscular Mycorrhizal Fungi within the Root Cortex of Both Medicago truncatula and Daucus carota. *The Plant Cell* 20(5):1407–1420.

12. Fournier J, et al. (2008) Mechanism of Infection Thread Elongation in Root Hairs of Medicago truncatula and Dynamic Interplay with Associated Rhizobial Colonization. *Plant Physiol* 148(4):1985–1995.
13. Genre A, et al. (2011) Multiple Exocytotic Markers Accumulate at the Sites of Perifungal Membrane Biogenesis in Arbuscular Mycorrhizas. *Plant and Cell Physiol* 53(1):244–255.
14. Ivanov S, et al. (2012) Rhizobium-legume symbiosis shares an exocytotic pathway required for arbuscule formation. *Proc Natl Acad Sci USA* 109(21):8316–8321.
15. Zhang X, Pumplin N, Ivanov S, Harrison MJ (2015) EXO70I Is Required for Development of a Sub- domain of the Periarbuscular Membrane during Arbuscular Mycorrhizal Symbiosis. *Current Biol* 25(16):2189–2195.
16. Huisman R, et al. (2016) A symbiosis-dedicated SYNTAXIN OF PLANTS 13II isoform controls the formation of a stable host-microbe interface in symbiosis. *New Phytologist* 211(4):1338–1351.
17. Gavrin A, Kulikova O, Bisseling T, Fedorova EE (2017) Interface Symbiotic Membrane Formation in Root Nodules of *Medicago truncatula*: the Role of Synaptotagmins MtSyt1, MtSyt2 and MtSyt3. *Front Plant Sci* 8(43):249.
18. Bonfante P (2001) At the interface between mycorrhizal fungi and plants: the structural organization of cell wall, plasma membrane and cytoskeleton. In Mycota, IX Fungal Associations (Springer-Verlag, Berlin), 45–91.
19. Bonfante-Fasolo P, Tamagnone L, Peretto R (1991) Immunocytochemical location of hydroxyproline rich glycoproteins at the interface between a mycorrhizal fungus and its host plants. *Protoplasma* 165:127.
20. Balestrini R, Bonfante P (2014) Cell wall remodeling in mycorrhizal symbiosis: a way towards biotrophism. *Front Plant Sci* 5:237.
21. Balestrini R, Romera C, Puigdomenech P, Bonfante P (1994) Location of a cell-wall hydroxyproline-rich glycoprotein, cellulose and β-1, 3-glucans in apical and differentiated regions of maize mycorrhizal roots. *Planta* 195(2):201–209.
22. Balestrini R, Hahn MG, Faccio A, Mendgen K, Bonfante P (1996) Differential Localization of Carbohydrate Epitopes in Plant Cell Walls in the Presence and Absence of Arbuscular Mycorrhizal Fungi. *Plant Physiol* 111(1):203–213.
23. Perotto S, Donovan N, Drøbak B, Brewin N (1995) Differential expression of a

glycosyl inositol phospholipid antigen on the peribacteroid membrane during pea nodule development. *Molecular Plant Microbe Interactions* 8(4): 560-568.

24. Brockerhoff H (1963) Breakdown of phospholipids in mild alkaline hydrolysis. *J Lipid Res* 4:96–99.

25. Gronnier J, Germain V, Gouguet P, Cacas J-L, Mongrand S (2016) GIPC: Glycosyl Inositol Phospho Ceramides, the major sphingolipids on earth. *Plant Signaling Behavior* 11(4):1–7.

26. Cacas J-L, et al. (2015) Revisiting Plant Plasma Membrane Lipids in Tobacco: A Focus on Sphingolipids. *Plant Physiol* 170(1):367–384.

27. Sperling P, Franke S, Lüthje S, Heinz E (2005) Are glucocerebrosides the predominant sphingolipids in plant plasma membranes? *Plant Physiol and Biochem* 43(12):1031–1038.

28. Michaelson LV, Napier JA, Molino D, Faure J-D (2016) Plant sphingolipids their importance on cellular organization and adaption. *Biochimica et Biophysica Acta* 1861:1329–1335.

29. Markham JE, Lynch DV, Napier JA, Dunn TM, Cahoon EB (2013) Plant sphingolipids: function follows form. *Current Opinion in Plant Biology* 16(3):350–357.

30. Cacas J-L, et al. (2012) Lipids in plant membrane rafts. *Progress in Lipid Research* 51(3):272–299.

31. Voxeur A, Fry SC (2014) Glycosylinositol phosphorylceramides from Rosacell cultures are boron-bridged in the plasma membrane and form complexes with rhamnogalacturonan II. *Plant J* 79(1):139–149.

32. Rennie EA, Ebert B, Miles GP, Cahoon RE, Christiansen KM, Stonebloom S, Khatab H, Twell D, Petzold CJ, Adams PD, *et al.* (2014) Identification of a sphingolipid α-glucuronosyltransferase that is essential for pollen function in Arabidopsis. *Plant Cell* 26: 3314–3325.

33. Mortimer JC, Yu X, Albrecht S, Sicilia F, Huichalaf M, Ampuero D, Michaelson LV, Murphy AM, Matsunaga T, Kurz S, *et al.* (2013) Abnormal glycosphingolipid mannosylation triggers salicylic acid–mediated responses in Arabidopsis. *Plant Cell* 25: 1881–1894.

34. Fang L, Ishikawa T, Rennie EA, Murawska GM, Lao J, Yan J, Tsai AY-L, Baidoo EEK, Xu J, Keasling JD, *et al.*(2017) Loss of inositol phosphorylceramide sphingolipid mannosylation induces plant immune responses and reduces cellulose content in Arabidopsis. *Plant Cell* 28: 2991–3004

35. Tartaglio V, Rennie EA, Cahoon R, Wang G, Baidoo E, Mortimer JC, Cahoon EB,

Scheller HV. 2017. Glycosylation of inositol phosphorylceramide sphingolipids is required for normal growth and reproduction in Arabidopsis. *Plant J* 89: 278–290.

36. Tanaka T, et al. (2013) Identification of a sphingolipid-specific phospholipase D activity associated with the generation of phytoceramide-1-phosphate in cabbage leaves. *FEBS J* 280(16):3797–3809.

37. Lenarčič T, et al. (2017) Eudicot plant-specific sphingolipids determine host selectivity of microbial NLP cytolysins. *Science* 358(6369):1431–1434.

38. Cacas J-L, et al. (2013) Biochemical survey of the polar head of plant glycosylinositolphosphoceramides unravels broad diversity. *Phytochemistry* 96:191–200.

39. Blaas N, Humpf H-U (2013) Structural Profiling and Quantitation of Glycosyl Inositol Phosphoceramides in Plants with Fourier Transform Mass Spectrometry. *J Agric Food Chem* 61(18):4257–4269.

40. Tellier FDR, Maia-Grondard A, Schmitz-Afonso I, Faure J-D (2014) Comparative plant sphingolipidomics reveals specific lipids in seeds and oil. *Phytochemistry* 103:50–58.

41. Buré C, et al. (2011) Fast screening of highly glycosylated plant sphingolipids by tandem mass spectrometry. *Rapid Commun Mass Spectrom* 25(20):3131–3145.

42. Wang W, et al. (2008) An Inositolphosphorylceramide Synthase Is Involved in Regulation of Plant Programmed Cell Death Associated with Defense in Arabidopsis. *Plant Cell* 20(11):3163–3179.

43. Ishikawa T, Fang L, Rennie E, Sechet J, Yan J, Jing B, Moore W, Cahoon EB, Scheller HV, Kawai-Yamada M, Mortimer JC. GINT1 is a GIPC GlcNAc glycosyltransferase. *Plant Physiology*, in press (accepted 5/1/18).

44. Floss DS, Schmitz AM, Starker CG, Gantt JS, Harrison MJ. (2013) Gene Silencing in Medicago Truncatula Roots Using RNAi. In *Legume Genomics: Methods and Protocols* (Humana Press, Totowa, NJ) 1069: 163-177

45. Haney CH, Long SR (2010) Plant flotillins are required for infection by nitrogen-fixing bacteria. *Proc Natl Acad Sci USA* 107(1):478–483.

46. Hoagland DR, Arnon, D (1938) The water culture method for growing plants without soil. UC College of Agriculture, Ag. Exp. Station, Berkeley, CA. Circular. 347, 1-39.

47. McGongigle TP, Miller HM, Evans DG, Fairchild GL, Swan JA (1990) A new method which give an objective measure colonization of roots by vesicular-arbuscular mycorrhizal fungi. *New Phytologist* 115: 495-501.

48. Helliwell CA, Wesley V, Wielopolska AJ, Waterhouse PM (2002) High-throughput

vectors for efficient gene silencing in plants. *Functional Plant Biol* 29(10): 1217-1225.

49. Yan P, et al. (2012) High-Throughput Construction of Intron-Containing Hairpin RNA Vectors for RNAi in Plants. *PLoS ONE* 7(5):e38186.

50. Devers EA, Teply J, Reinert A, Gaude N, Krajinski F (2013) An endogenous artificial microRNA system for unraveling the function of root endosymbiosesrelated genes in Medicago truncatula. *BMC Plant Biology* 13:82.

51. Nakagawa T, Kurose T, Hino T, Tanaka K, Kawamukai M, Niwa Y, Toyooka K, Matsuoka K, Jinbo T, Kimura T (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J of Bioscience and Bioengineering* 104(1): 34-41.

52. Kawaharada Y, et al. (2017) Differential regulation of the Epr3 receptor coordinates membrane-restricted rhizobial colonization of root nodule primordia. *Nature Communications* 8:1–11.

53. Ishikawa T, Ito Y, Kawai-Yamada M (2016) Molecular characterization and targeted quantitative profiling of the sphingolipidome in rice. *Plant J* 88: 681–693.

54. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42: 490-495.

55. Roux B, et al. (2014) An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant J* 77(6):817–837.

56. Limpens E, Moling S, Hooiveld G, Pereira PA, Bisseling T, Becker JD, Küster H (2013) Cell- and tissue-specific transcriptome analyses of Medicago truncatula root nodules. *PloS One* 8(5): e64377

57. Gaude N, Bortfield Sm Duensing N, Lohse M, Krajinski F (2012) Arbuscule-containing and non-colonized cortical cells of mycorrhizal roots undergo extensive and specific reprogramming during arbuscular mycorrhizal development. *The Plant Journal* 69(3):510-528.

58. HodgeKamp C, Küster H (2013) A roadmap of cell-type specific gene expression during sequential stages of the arbuscular mycorrhiza symbiosis. *BMC Genomics* 14:306.

59. Magnin-Robert M, et al. (2015) Modifications of sphingolipid content affect tolerance to hemibiotrophic and necrotrophic pathogens by modulating plant defense responses in *Arabidopsis*. *Plant Physiol* 169(3):2255-2274.

60. Shi C, et al. (2015) A systematic simulation of the effect of salicylic acid on

sphingolipid metabolism. *Front Plant Sci* 6(4):1241.

61. Javot H, Penmetsa RV, Terzaghi N, Cook DR, Harrison MJ (2007) A Medicago truncatula phosphate transporter indispensable for the arbuscular mycorrhizal symbiosis. *Proc Natl Acad Sci USA* 104(5):1720-1725.

62. Ott T (2017) ScienceDirectMembrane nanodomains and microdomains in plant–microbe interactions. *Current Opinion in Plant Biology* 40:82–88.

63. Kawaharada Y, et al. (2015) Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature* 523(7560):308–312.

64. Singh SK, Eland C, Harholt J, Scheller HV, Marchant A (2005) Cell adhesion in *Arabidopsis thaliana* is mediated by ECTOPICALLY PARTING CELLS 1 0 a glycosyltransferase (GT64) related to the animal exostosins. *Plant J* 43: 387-397.

65. Bolaños L, Brewin NJ, Bonilla I (1996) Effects of boron on *Rhizobium*-legume cell-surface interactions and nodule development. *Plant Physiology* 110: 1249–1256.

66. Redondo-Nieto M, Pulido L, Reguer M, Bonilla I, Bolaños L (2007) Developmentally regulated membrane glycoproteins sharing antigenicity with rhamnogalacaturonan II are not detected in nodulated boron deficient *Pisum sativum*. *Plant, Cell, and Env* 30: 1436-1443.

67. Xie F, et al. (2012) Legume pectate lyase required for root infection by rhizobia. *Proc Natl Acad Sci USA* 109(2):633–638.

68. Bonfante-Fasolo P, Vian B, Perotto S, Faccio A, Knox JP (1990) Cellulose and pectin localization in roots of mycorrhizalAllium porrum: labelling continuity between host cell wall and interfacial material. *Planta* 180(4):537

69. Geurts R, Bisseling T. (2002) Rhizobium nod factor perception and signaling. *Plant Cell* 14:239-249.

70. GENRE A, et al. (2013) Short-chain chitin oligomers from arbuscular mycorrhizal fungi trigger nuclear Ca 2+spiking in Medicago truncatularoots and their production is enhanced by strigolactone. *New Phytologist* 198(1):190–202.

71. Nadal M, et al. (2017) An N-acetylglucosamine transporter requiredfor arbuscular mycorrhizal symbiosesin rice and maize. *Nature Plants*:1–7.

72. Lefebvre B (2017) A new role for N-acetylglucosamine. *Nature Plants* 3:1–2.

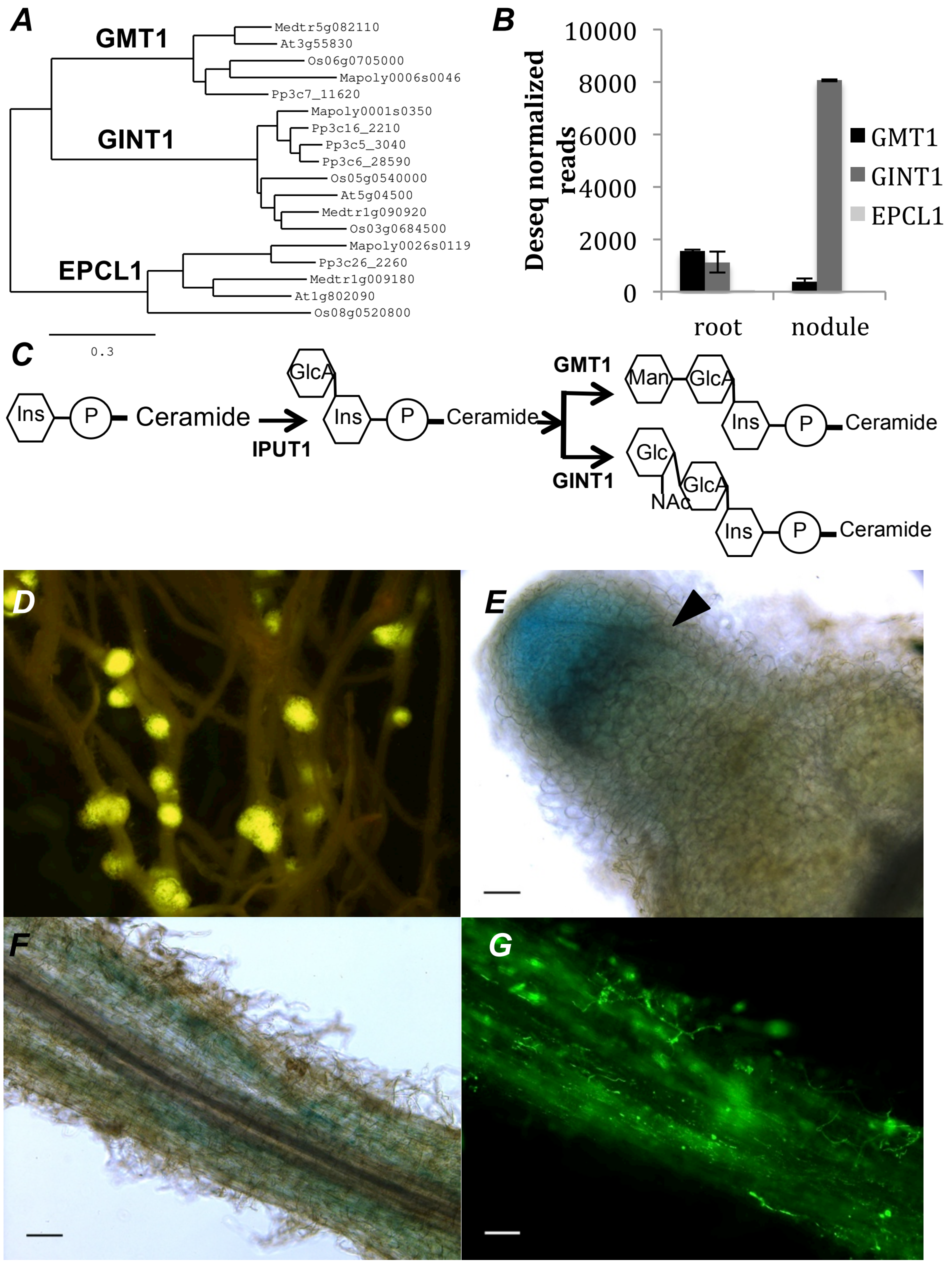


Fig. 1. MtGINT1 belongs to GT family 64 and is expressed in *M. truncatula* root nodules and mycorrhized roots. (*A*) Maximum likelyhood tree of GT family 64 in selected plant species (Pp, *Physcomitrell patens; MaPoly, Marchantia polymorpha*; At, *Arabidopsis thaliana*; Medtr, *Medicago truncatula*; Os, *Oryza sativa*). (*B*) RNA-seq profiling of *M. truncatula* GT64 in roots and root nodules from Symbimics database (Roux *et al.,* 2014) (*C*) GIPC head group biosynthesis pathway (Ins, inositol; GlcA, glucuronic acid; Man, mannose; GlcN(Ac), *N*-acetyl-glucosamine). (*D*) *M. truncatula* hairy roots expressing MtGINT1 promoter YFP-reporter construct 3 wpi *S. meliloti*. (*E*, *F*) X-gluc staining of hairy roots expressing MtGINT1 promoter GUS reporter constructs 3 wpi with *S. meliloti* (*E*) and 4 wpi with *R. irregularis* (*F*). (*G*) Fluorescence microscopy image of the same specimen as *F* stained with WGA-488 for fungal chitin (Scale bar =100 µm).

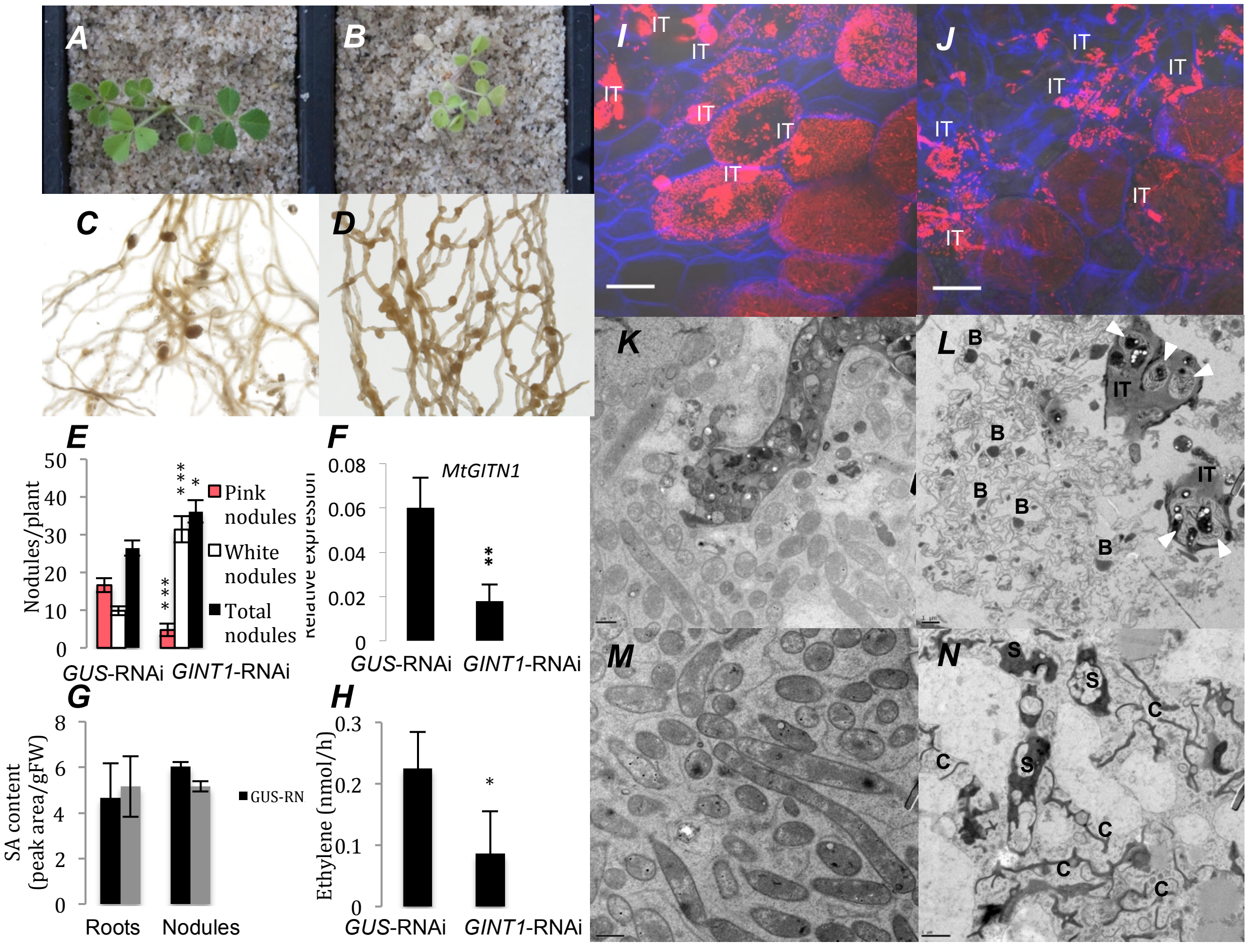
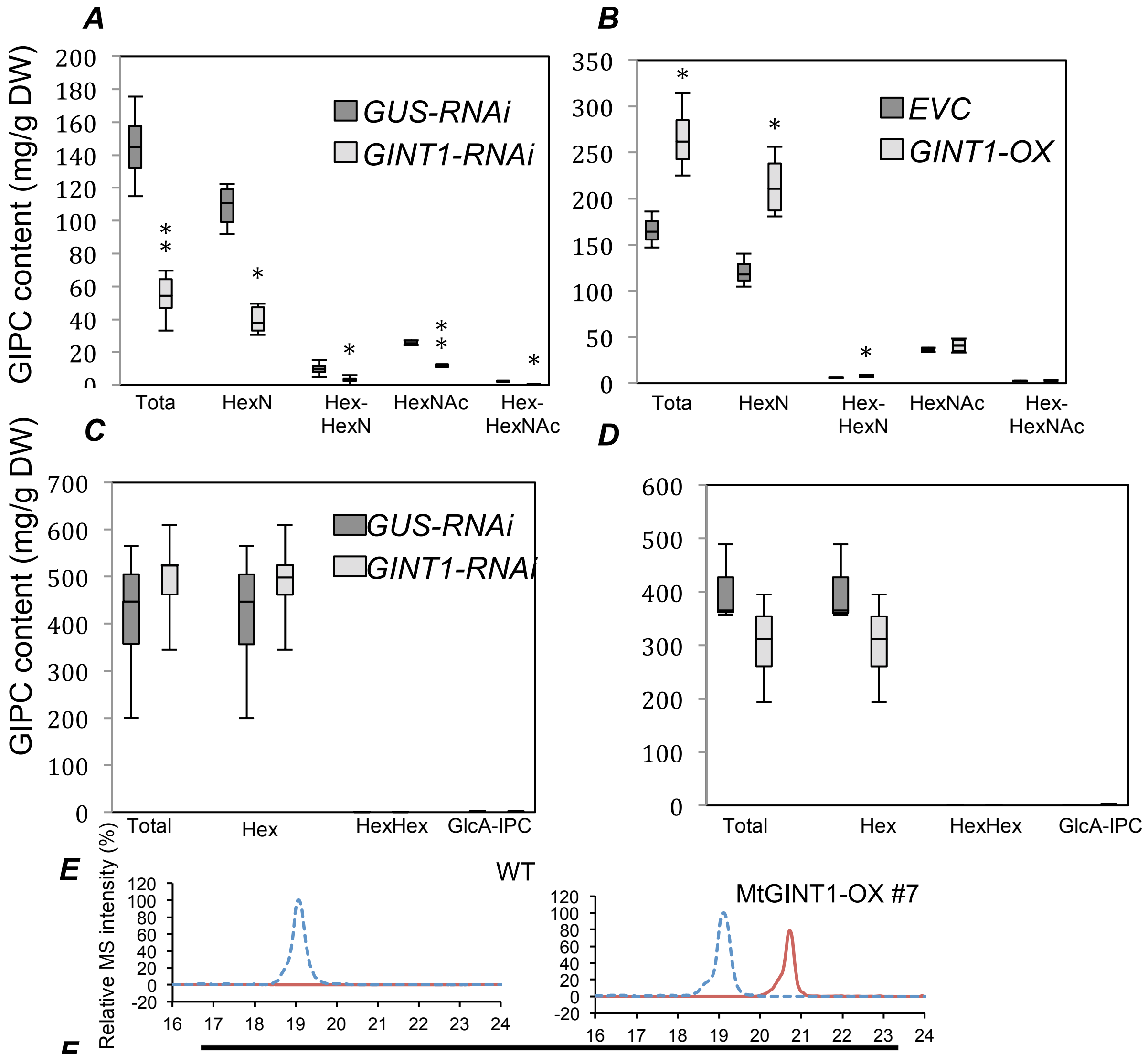


Fig. 2. *MtGINT1*-silencing impairs nodulation resulting in altered membrane dynamics and bacterial senescence. (*A*-D) Shoots and nodulated roots from chimeric plants expressing *GUS*-RNAi (*A,* C) and *MtGINT1*-RNAi (*B, D)* constructs. (*E*) Average number of nodules per RNAi root system ± SEM. (*F*) *MtGINT1* transcript abundance measured relative to EF-1α by qRT-PCR ± SD (*n*=4). (*G*) SA content measured by LC-MS/MS in nodulated and no inoculum control roots 3 wpi (*n*=3) ± SD. (*H*) Nitrogen fixation efficiency measured by acetylene reduction assay (*n*=3) of nodulated roots 3 wpi ± SD. Significant differences are indicated (Student's t-test, *p*-value: *< 0.05, **<0.005, ***<0.001). (*I, J*) Confocal live-cell imaging of *S. meliloti* expressing mCherry in *GUS*-RNAi nodules (*I*) and *MtGINT1*-RNAi nodules (*J*) 3 wpi, scale bar = 20 µm. Plant cell walls were stained with Calcofluor White. (*K, L*) TEM images of infection threads releasing *S. meliloti* in *GUS*-RNAi (*K*) and *MtGINT1*-RNAi (*L*) nodules 3 wpi. White arrows point to excessive EPS secreted by *S. meliloti* while in infection threads of *MtGINT1*-RNAi nodules (B, bacteria; IT, infection thread). (*M, N*) TEM images of *S. meliloti* in the distal nitrogen fixation zone, below infection thread release, in *GUS*-RNAi (*M*) and *MtGINT1*-RNAi (*N*) nodules 3 wpi (S, senescent bacteria; C, collapsed bacteria). Scale bar = 1 µm.



F

Genotype	Total HexNAc	HexHexN	HexHexNAc
Wt	0	0	0
MtGINT1-OX #3	35.5±0.5	28.4±0.8	7.1±1.3
MtGINT1-OX #7	34.3±3.4	26.9±2.3	7.4±1.1
MtGINT1-OX #9	5.5±1.4	4.1±1.0	1.3±0.4

Fig. 3. *MtGINT1* functions as a HexN(Ac) GIPC transferase *in planta.* (*A, B*) HexN(Ac)-GIPC content in *MtGINT1*-RNAi (*A*) and *MtGINT1*-OX (*B*) roots compared to control roots measured by LC-MS/MS. (*C, D*) Hex-GIPC content in *MtGINT1*-RNAi (*C*) and *MtGINT1*-OX (*D*) roots compared to control roots (*GINT1*-RNAi *n*=6*; GUS*-RNAi*, n*=4*; MtGINT1*-OX*, n*=4*; EVC*-OX*, n*=3). (*E*) LC-MS/MS chromatograms targeting GIPCs with the Hex (blue dashed line, *m*/*z* 1260.8 > 662.6) or HexN (red solid line, *m*/*z* 1259.8 > 662.6) head group on t18:1-h24:1 ceramide backbone. An asterisk indicates the GIPC peak. (*F*) HexN(Ac)-GIPC content of *A. thalia*na leaves from wild-type and plants expressing *35s:MtGINT1* constructs (average 6 plants per line ± SD). Significant differences determined by Student's T-test are indicated (*p*-value: *< 0.05, **<0.005).

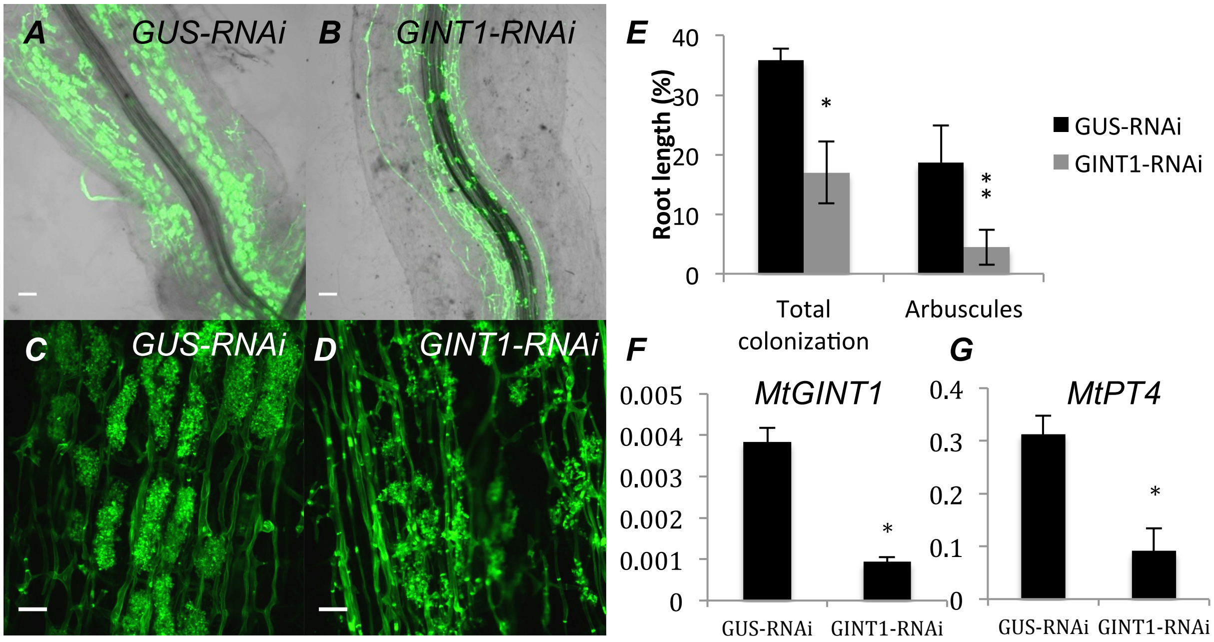
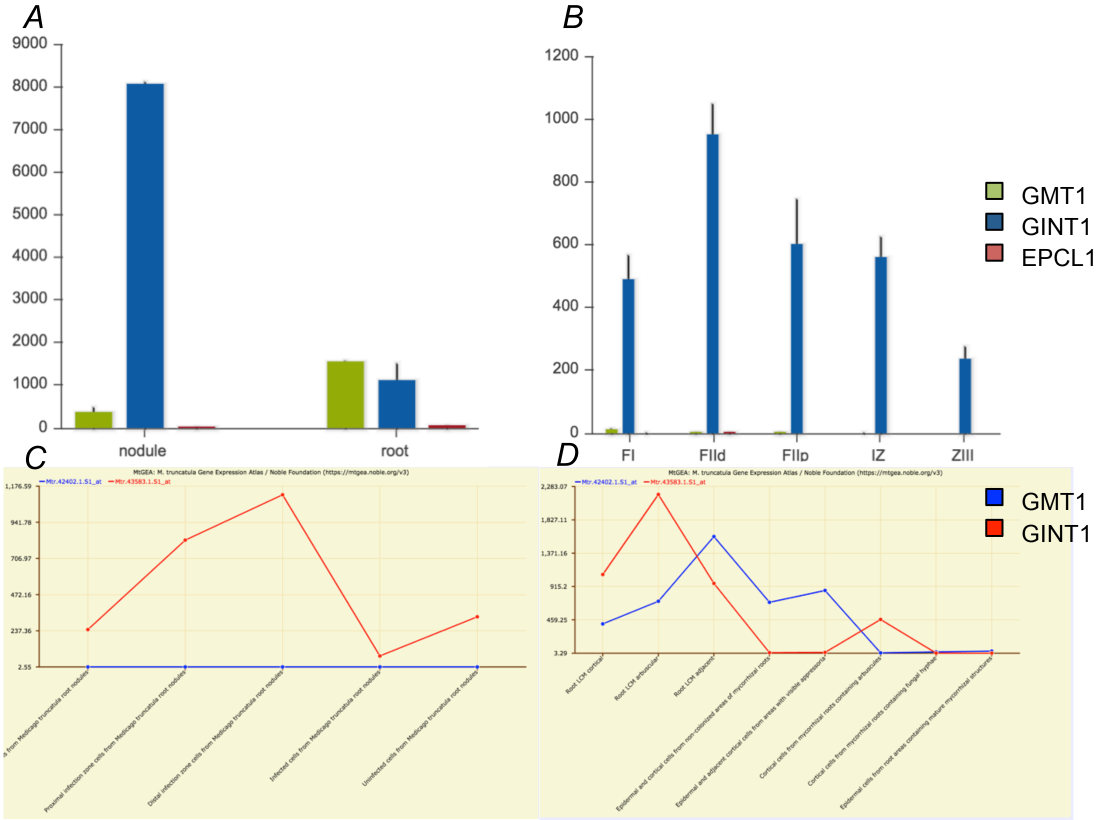
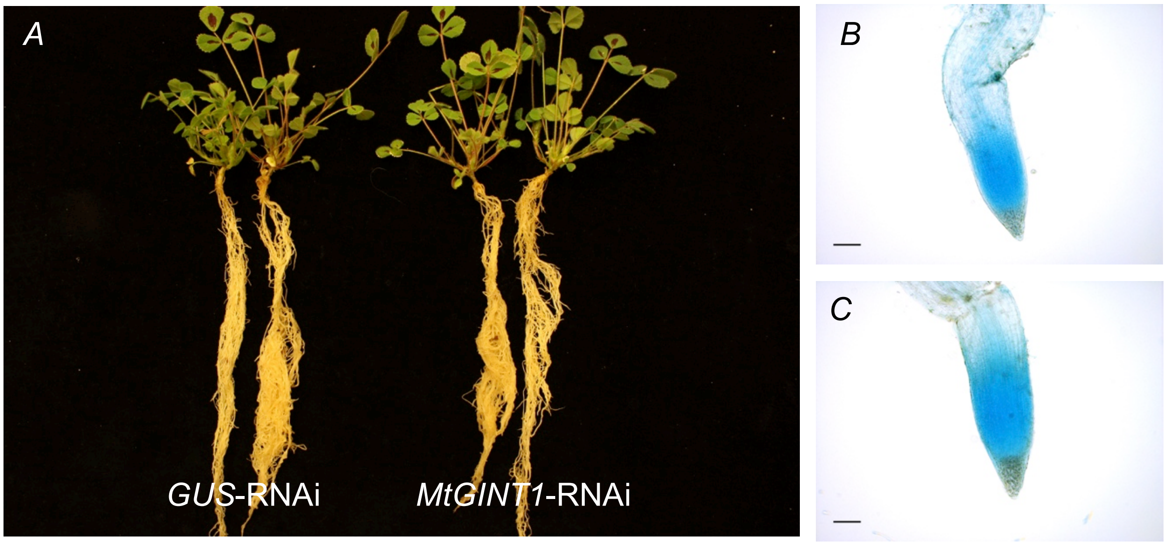


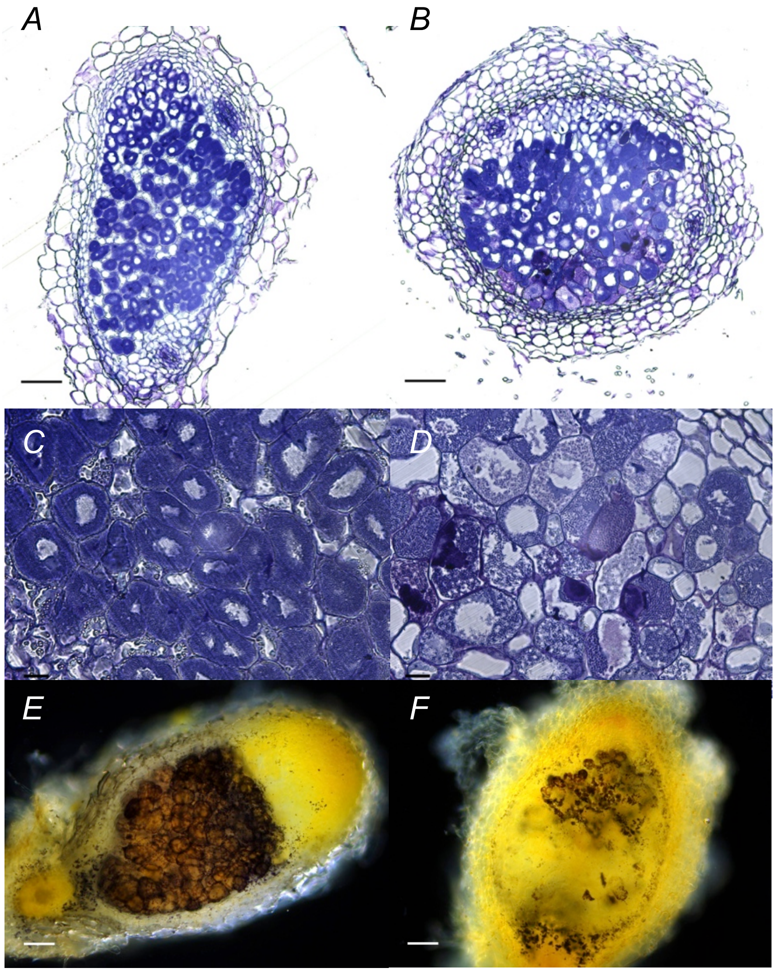
Fig. 4. AM colonization is impaired in *MtGINT1*-silenced roots. Plants were grown for 5 weeks with *R. irregularis* under limited phosphate conditions. (*A-D*) Mycorrhized *GUS*-RNAi (*A, C*) and *MtGINT1*-RNAi (*B, D*) roots stained with WGA-488 for fungal chitin (*A, B,* scale bar = 50 µm; *C, D* scale bar = 20 µm. (*E*) Mycorrhizal colonization based on gridline intersect (*n* = 5) represented as the average ± SEM. (*F, G*) *MtGINT1* and *MtPT4* expression measured relative to EF-1α by qRT-PCR (*n* = 4) ± SD. Significant differences between *MtGINT1*-RNAi and control roots are indicated by Student's t-test, (*p*-value: *< 0.05, **<0.005).



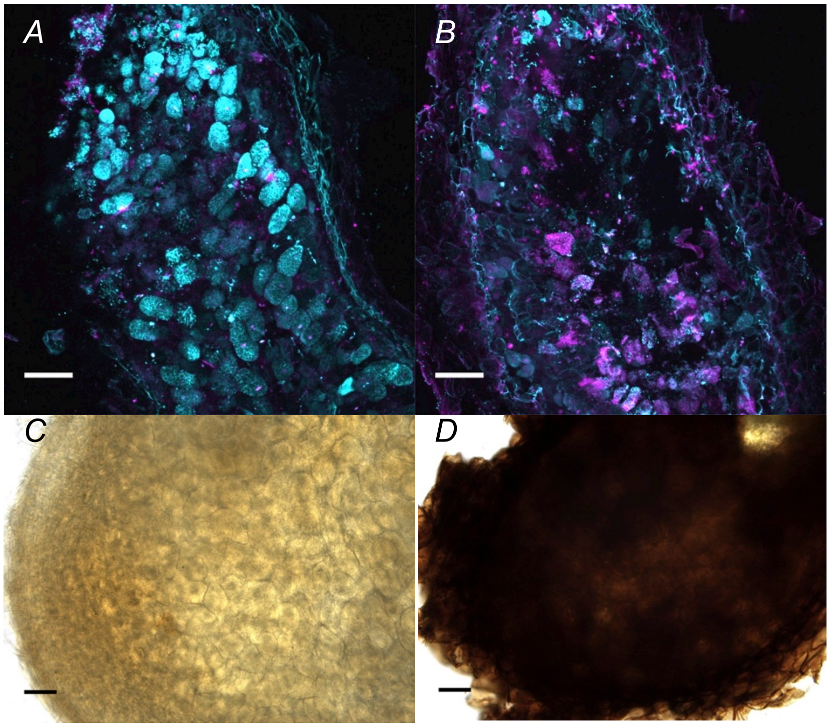
SI Fig 1. Laser capture microdissection transcriptome profiling of GT family 64 in Medicago roots during nodulation and AM symbiosis based on publicly accessible databases: Symbimics Database (1) (*A, B*) and the Medicago Gene Expression Atlas (2) (*C, D*). (*A, B*) Deseq-normalized RNA-seq reads in root nodules vs roots (*A*) and different developmental zones of the root nodule (*B*) collected by laser-capture microdissection (LCM) (1) (FII, meristem; FIId-FIIp, distal and proximal infection zone; IZ, interzine; ZIII, nitrogen fixation zone. (*C*) Microarray data from different zones of the root nodule (2, 3). (*D*) Microarray data from different root cell types colonized by *R. irregularis* (2, 4, 5).



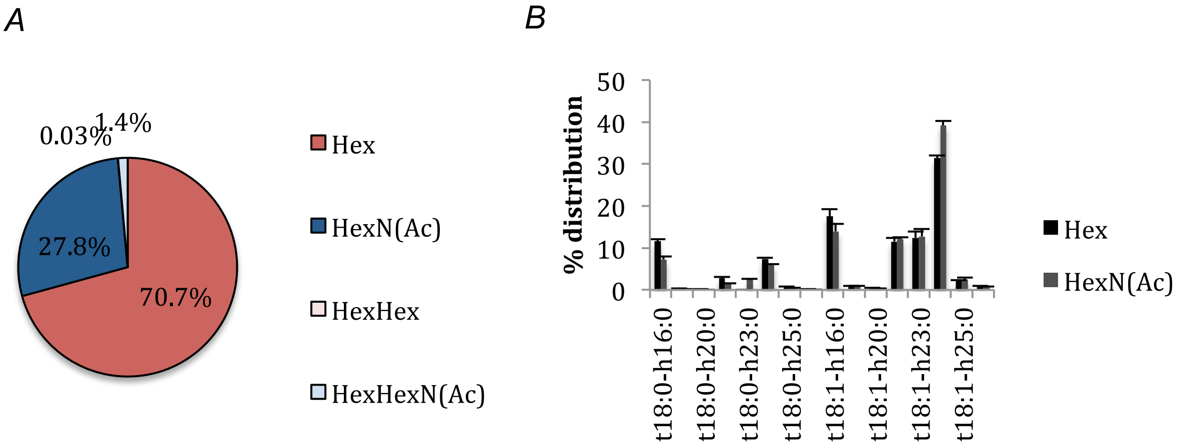
SI Fig 3. *GINT1*-silencing does not affect root growth. (*A*) Images of Chimeric RNAi plants grown for 4-weeks and fertilized with 1/2X MS. (*B*, *C*) X-gluc staining in root tips expressing *MtGINT1* promoter GUS-reporter constructs in uninoculated (*B*) and mycorrhized root systems (*C*).



SI Fig 3. *MtGINT1*-RNAi nodulation phenotype 3 wpi with *S. meliloti*. (*A*, *B*) 6 µm thick plastic embedded sections of *GUS*-RNAi (*A*, *C*) and *MtGINT1*-RNAi (*B*, *D*) nodules stained with Toludine Blue O. Symbiosomes appear degraded in *MtGINT1* silenced nodules (*D*). Scale bar = 100 µm (*A, B*), 20 µm (*C, D*). (*E, F*) Nodules stained with Lugol reagent to detect starch (dark brown) in *GUS*-RNAi (*E*) and *MtGINT1*-RNAi (*F*) nodules. Starch distribution is altered in the infection zone of *MtGINT1* silenced nodules. Scale bar = 100 µm.



SI Fig 4. Cell death and H_2O_2 accumulation *MtGINT1*-silenced nodules. (*A, B*) Live-dead staining using Syto 9 (live; cyan) and propidium iodide (dead; magenta) of *GUS*-RNAi (*A*) and *MtGINT1*-RNAi nodules (*B*), scale bar = 100 µm. (*C, D*) DAB staining to detect H_2O_2 (dark brown) in *GUS*-RNAi (*C*) and *MtGINT1*-RNAi nodules (*D*), scale bar = 50 µm.



SI Figure 5. GIPC composition of control roots. (*A*) Distribution of GIPC by glycan head group. (*B*) Distribution of Hex GIPC and HexN(Ac) GIPCs by ceramide class.

Primer name	Sequence
GINT1-RNAi-F	**CCATCTAGA**ATGGGTTCCGGTCAG
GINT1-RNAi-R	**GCATGAATTC**CACAAGTCACCACA
GUS-RNAi-F	**CCATCTAGA**CTTACGTGGCAAAGG
GUS-RNAi-R	**GCATGAATTC**GAACATTACATTGA
GINT1-RNAiGG-F	**ACCAGGTCTCAGGAG**ATGGGTTC
GINT1-RNAiGG-R	**ACCAGGTCTCATCGT**CACAAGTCA
GUS-RNAiGG-F	**ACCAGGTCTCAGGAG**CTTACGTG
GUS-RNAiGG-R	**ACCAGGTCTCATCGT**GAACATTAC
GINT1-PRO-F	GCAGATTGTGGAGTTTAGATAAC
GINT1-PRO-R	GAACATCTTCCTTAACTTGGTGT
GINT1-CDS-F	ATGGGTTCCGGTCAGATTAG
GINT1-CDS-R	CTATACATCCCAACCATCTCT
GINT1-CDSns-R	TACATCCCAACCATCTCTCC
GINT1-qPCR-F	GTATTTGAGCACAATGGCCAG
GINT1-qPCR-R	GATAAAGGAATCAATAAGGGGC
EF1α-qPCR-F	TGACAGGCGATCTGGTAAGG
EF1α-qPCR-R	TCAGCGAAGGTCTCAACCAC
PT4-qPCR-F	GGATTCTTTTGCACGTTCTTGG
PT4-qPCR-R	CCTGTCATTTGGTGTTGCAGTG

SI Table I. Primers used in this study.

Summary and future directions

Plants form a variety of endosymbioses with bacteria and fungi that promote plant growth and fitness. Critical to the development and function of these beneficial relationships are the synthesis of specialized plant-derived membranes that serve as interfaces for nutrient and information exchange between the microorganism and the plant cell. Early pioneering work by Bonfante, Ballestrini, Perotto, and others in the 1990's and early 2000's showed that antibodies recognizing glycoproteins, glycolipids, and cell wall epitopes label the interfacial material associated with periarbuscular and symbiosome membranes (1–4). Among these epitopes were arabinogalactan proteins (AGPs), glycosyl inositolphosphorylceramides (GIPCs), and pectic domains rhamnogalacturonan I and II, which reflect the heterogeneity and structural complexity of this material. These epitopes are developmentally regulated and actively deposited within the symbiotic compartment, which has lead to the wide spread hypothesis that this interfacial glycocalyx might play an important role in symbiosis. However, this has not been investigated at the molecular or genetic level.

The underlying motivation for this dissertation was to identify genetic determinants of the symbiosis-associated glycocalyx and assess their function in nodulation and AM symbiosis. AGPs were interesting targets because they are involved in cell signaling and have previously been implicated in four different and diverse plant-microbial symbioses (1, 3, 5–7), but have not been described in depth outside of *Arabidopsis thaliana* (8). To address this we developed new bioinformatics tools that we used to identify predicted AGP-encoding genes in the *Medicago truncatula* genome (Chapter 1). These were then cross-referenced to the *Medicago* Gene Expression Atlas (9, 10), which identified two groups of AGPs expressed in either root nodules or AM colonized roots, and uncovered previously unknown redundancy in both the plastocyanin-like AGPs and AGP peptides (Chapter 2). We found that the AGP peptides form a small family of tandemly duplicated genes, which we have named *SYMBIOSIS-ASSCOIATED ARABINOGLACTAN PEPTIDES* (*SAPs*) that are differentially expressed in root nodules and AM colonized roots. SAPs localized to symbiotic membranes and knockdown of *SAP* expression using RNAi-mediated gene silencing impaired the growth and differentiation of *S. meliloti* and *R. irregularis* within these compartments. We also identified a glycosyltransferase gene highly expressed in root nodules and AM colonized roots as *GIPC GLUCOSAMINE TRANSFERASE 1* (*GINT1*) and showed that the corresponding enzyme functions in the synthesis of HexN(Ac) decorated GIPCs *in planta* (Chapter 3). Knockdown of *GINT1* expression impaired the development of symbiotic membranes resulting in the senescence of symbiosomes and arbuscules. Taken together these results provide genetic evidence to support that specific AGPs and GIPCs are necessary for symbiosis, consistent with the initial observations and hypotheses made by Bonfante and Perotto (1–4).

While we were able to show that specific components of the glycocalyx are important for endosymbiosis, the underlying mechanisms in which they function remain obscure. In the case of AGPs and GIPCs a multitude of explanations are possible. GIPCs have previously been associated with lipid rafts in detergent insoluble membranes (11) and shown to form boron cross-linked bridges with RG-II (12). Future studies could investigate the boron-dependence of the *GINT1-RNAi* phenotype or ligand-induced recruitment of receptor-like kinase LYK3 and REMORIN proteins into raft-associated nanodomains (13, 14). The original observations made by Perotto *et al.* (4) had noted that the JIM18 epitope localized to young symbiosome membranes but was absent from the membrane surrounding mature bacteroids. The JIM18 epitope is likely a GIPC and hence these observations suggested that GIPCs might be developmentally released from the symbiosome membrane. Similarly, we found that SAP3 is a GPI-anchored protein that is developmentally released from symbiosome membranes during bacteroid differentiation. GIPC and GPI glycolipids share an inositol-phosphate linkage that can be cleaved by phosphatydilinositol phospholipase C or D enzymes (15, 16). Thus, a general shedding of GIPCs and GPI-anchored proteins may occur during bacteroid differentiation. It is also possible that SAP3 might be GIPC-anchored as opposed to GPI-anchored (17), since we did not investigate this at the biochemical level. Interestingly, we found that *SAP3-RNAi* nodules had a phenotype similar to that reported for the nodule-specific phospholipase C mutant *dnf2* (18), which may provide a putative mechanism for the release of SAP3 from the symbiosome membrane. The biochemical activity, specificity, and endogenous substrate of *DNF2* are not currently known. Therefore future studies should investigate the relationship between *SAP3* and *DNF2* to determine if these proteins operate in the same genetic pathway. The *dnf2* phenotype was previously shown to be a function of defense response, which had suggested that DNF2 might function in local suppression of defense signaling during nodulation (19). If this is true and SAP3 is a target of DNF2, then this could suggest that the SAPs have immunogenic properties. This hypothesis could be tested by measuring expression of defense related genes in *SAP3-RNAi* and *SAP3-ox* roots in responses to defense elicitors like flg22, chitin, and ulvan.

Lastly, in the course of identifying predicted AGPs in *M. truncatula* we found a large number of proteins with convincing noncontiguous proline motifs not previously reported to be AGP. Our assessment was later supported by the work of Ma *et al.* (2017) that identified many of the same proteins and came to a similar conclusion (20), which suggests that the number and diversity of AGP-encoding genes has been previously underestimated. Among these proteins we found the type-I formins as the most intriguing new class of AGP. Type-I formins are transmembrane proteins that localize to the plasma membrane and have cytosolic domains involved in actin nucleation and microtubule organization (21, 22). We found that many type-I formins have putative AGP and extension-type protein glycosylation sites, which may help explain previous observations that had linked AGP glycans to cytoskeleton organization (23–25). This has major implications for cell wall biosynthesis where formins may organize cortical microtubules to guide cellulose synthase inside the cell, while providing a glycosylated scaffold for cellulose deposition outside the cell. Extensins are well known for having self-assembling properties *in vitro* that create filamentous networks (26) and self-assembling properties have also been reported for some AGPs (27). In this regard it is possible that formins might function as hubs that integrally interconnect the intracellular cytoskeleton with an extracellular *glycoskeleton* network of self-assembling glycoproteins. This potentially opens up a new and exciting area of research and we hope to see future work investigating the functional consequences of formin glycosylation through the use of mutants complemented with proteins either lacking glycosylation sites or containing scrambled glycosylation codes.

Bibliography

1. Balestrini R, Hahn MG, Faccio A, Mendgen K, Bonfante P (1996) Differential localization of carbohydrate epitopes in plant cell walls in the presence and absence of arbuscular mycorrhizal fungi. *Plant Physiol* 111(1):203–213.
2. Bonfante-Fasolo, et al. Immunocytochemical location of hydroxyproline rich glycoproteins at the interface between a mycorrhizal fungus and its host plants.
3. Perotto S, Vandenbosch KA, Butcher GW, Brewin NJ (1991) Molecular composition and development of the plant with the peribacteroid membrane of pea root nodules. *Development*.
4. Perotto S, Donovan N, Drobak BK (1995) Differential expression of a glycosyl inositol phospholipid antigen on the peribacteroid membrane during pea nodule development. *MPMI-Molecular Plant ...*.
5. Jackson O, Taylor O, Adams DG, Knox JP (2012) Arabinogalactan proteins occur in the free-living cyanobacterium genus Nostoc and in plant-Nostoc symbioses. *Mol Plant Microbe Interact* 25(10):1338–1349.
6. Berry AM, et al. (2002) Arabinogalactan proteins are expressed at the symbiotic interface in root nodules of Alnus spp. *New Phytol* 155(3):469–479.
7. Bradley, et al. Isolation of monoclonal antibodies reacting with peribacteroid membranes and other components of pea root nodules containing Rhizobium leguminosarum.
8. Tan L, et al. (2012) Arabinogalactan-proteins and the research challenges for these enigmatic plant cell surface proteoglycans. *Front Plant Sci* 3:140.
9. He J, et al. (2009) The Medicago truncatula gene expression atlas web server. *BMC Bioinformatics* 10:441.
10. Benedito VA, et al. (2008) A gene expression atlas of the model legume Medicago truncatula. *Plant J* 55(3):504–513.
11. Cacas J-L, et al. (2016) Revisiting plant plasma membrane lipids in tobacco: A focus on sphingolipids. *Plant Physiol* 170(1):367–384.
12. Voxeur A, Fry SC (2014) Glycosylinositol phosphorylceramides from Rosa cell cultures are boron-bridged in the plasma membrane and form complexes with rhamnogalacturonan II. *Plant J* 79(1):139–149.
13. Haney CH, et al. (2011) Symbiotic rhizobia bacteria trigger a change in localization and dynamics of the Medicago truncatula receptor kinase LYK3. *Plant Cell* 23(7):2774–2787.
14. Liang P, et al. (2018) Symbiotic root infections in Medicago truncatula require remorin-mediated receptor stabilization in membrane nanodomains. *Proc Natl Acad Sci USA*.
15. Müller A, Klöppel C, Smith-Valentine M, Van Houten J, Simon M (2012) Selective and programmed cleavage of GPI-anchored proteins from the surface membrane by phospholipase C. *Biochim Biophys Acta* 1818(1):117–124.
16. Tanaka T, et al. (2013) Identification of a sphingolipid-specific phospholipase D activity associated with the generation of phytoceramide-1-phosphate in cabbage leaves. *FEBS J* 280(16):3797–3809.
17. Svetek J, Yadav MP, Nothnagel EA (1999) Presence of a glycosylphosphatidylinositol lipid anchor on rose arabinogalactan proteins. *J Biol Chem* 274(21):14724–14733.
18. Bourcy M, et al. (2013) Medicago truncatula DNF2 is a PI-PLC-XD-containing protein required for bacteroid persistence and prevention of nodule early senescence and

defense-like reactions. *New Phytol* 197(4):1250–1261.
19. Berrabah F, et al. (2014) Growth conditions determine the DNF2 requirement for symbiosis. *PLoS One* 9(3):e91866.
20. Ma Y, et al. (2017) Bioinformatics prediction and evolution analysis of arabinogalactan proteins in the plant kingdom. *Front Plant Sci* 8:66.
21. van Gisbergen PAC, Bezanilla M (2013) Plant formins: membrane anchors for actin polymerization. *Trends Cell Biol* 23(5):227–233.
22. Deeks MJ, et al. (2010) The plant formin AtFH4 interacts with both actin and microtubules, and contains a newly identified microtubule-binding domain. *J Cell Sci* 123(Pt 8):1209–1215.
23. Sardar HS, Yang J, Showalter AM (2006) Molecular interactions of arabinogalactan proteins with cortical microtubules and F-actin in Bright Yellow-2 tobacco cultured cells. *Plant Physiol* 142(4):1469–1479.
24. Nguema-Ona E, Bannigan A, Chevalier L, Baskin TI, Driouich A (2007) Disruption of arabinogalactan proteins disorganizes cortical microtubules in the root of Arabidopsis thaliana. *Plant J* 52(2):240–251.
25. Driouich A, Baskin TI (2008) Intercourse between cell wall and cytoplasm exemplified by arabinogalactan proteins and cortical microtubules. *Am J Bot* 95(12):1491–1497.
26. Cannon MC, et al. (2008) Self-assembly of the plant cell wall requires an extensin scaffold. *Proc Natl Acad Sci USA* 105(6):2226–2231.
27. Hijazi M, et al. (2014) Arabinogalactan protein 31 (AGP31), a putative network-forming protein in Arabidopsis thaliana cell walls? *Ann Bot* 114(6):1087–1097.

Appendix I. Known AGPs used for AGPredictor program development.

```
>AT1G03820.1 | Symbols:  | unknown protein; FUNCTIONS IN:
molecular_function unknown; INVOLVED IN: biological_process
unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 21 plant
structures; EXPRESSED DURING: 13 growth stages; Has 1345 Blast
hits to 1122 proteins in 102 species: Archae - 2; Bacteria - 28;
Metazoa - 28; Fungi - 30; Plants - 109; Viruses - 0; Other
Eukaryotes - 1148 (source: NCBI BLink). | chr1:960008-960676
REVERSE LENGTH=222
MASFALKPIFCFIAVFCFIVHNVEAREGKLFFSKFTHIDRPNNKDVALSPAPAPGLAQANGRLG
NGSFGPGSGMIPQTKESWPSSSTTTDEEFEKLMATF
DEEKNTKLPEAFEEEEESEDSEDLNEPKDKYNNNNNNNGYTYTTNNYNDNGRGYGNEEEKQGMS
DTRVMENGKYFYDTRGRNSENTPSRGYENARGNDHT
NEFETMEEYYKSLEGSQEEYEP
>AT1G03870.1 | Symbols: FLA9 | FASCICLIN-like arabinoogalactan 9
| chr1:982625-983368 REVERSE LENGTH=247
MATTRLTLAPLLLIAAVLLATKATAQPAAPAPEPAGPINLTAILEKGGQFTTFIHLLNITQVGS
QVNIQVNSSSEGMTVFAPTDNAFQNLKPGTLNQLSP
DDQVKLILYHVSPKYYSMDDLLSVSNPVRTQASGRDNGVYGLNFTGQTNQINVSTGYVETRISN
SLRQQRPLAVYVVDMVLLPGEMFGEHKLSPIAPAPK
SKSGGVTDDSGSTKKAASPSDKSGSGEKKVGLGFGLGLIVLCLKFLF
>AT1G15190.1 | Symbols:  | Fasciclin-like arabinogalactan family
protein | chr1:5227275-5228021 FORWARD LENGTH=248
MAKISSASCFRAIFLGALIIILCLPHPSTGVPLEELERAIAILRVRGRALFANAIITSDLLFDLL
SDESLTLFAPTDSMLFDLDMTHSLPFYVSTLRLHSV
PLRLSLSGLRSLPNSSSLPTLLPSHRLLLTKHSSSNDSIFLDGVQLLIPGLFDGQHIAVHGLAD
LLPLTAPSSPNRLVEDSTALAKSPWFLGSRFSPAPE
PYFAFMDLSPAESPSVEEVSPSPSWGEGEEDFIVGDEGGPLDGRNNGF
>AT1G24520.1 | Symbols: BCP1 | homolog of Brassica campestris
pollen protein 1 | chr1:8688699-8689058 FORWARD LENGTH=119
MGRQNIVVVVALVFMAIIGLAAAASSPSPSASPSKAPAASKTDHVEAPVTDDQIGTTDDDAAPT
PGDGDVAVAGPLGSDSSYDNAPTGSADSAKSGAAAL
GVSAVVVGVTSIAGSFLLL
>AT1G28290.1 | Symbols: AGP31 | arabinogalactan protein 31 |
chr1:9889331-9890843 REVERSE LENGTH=359
MGFIGKSVLVSLVALWCFTSSVFTEEVNHKTQTPSLAPAPAPYHHGHHHPHPPHHHHPHPHPHP
HPPAKSPVKPPVKAPVSPPAKPPVKPPVYPPTKAPV
KPPTKPPVKPPVSPPAKPPVKPPVYPPTKAPVKPPTKPPVKPPVYPPTKAPVKPPTKPPVKPPV
YPPTKAPVKPPTKPPVKPPVSPPAKPPVKPPVYPPT
KAPVKPPVSPPTKPPVTPPVYPPKFNRSLVAVRGTVYCKSCKYAAFNTLLGAKPIEGATVKLVC
KSKKNITAETTTDKNGYFLLLAPKTVTNFGFRGCRV
YLVKSKDYKCSKVSKLFGGDVGAELKPEKKLGKSTVVVNKLVYGLFNVGPFAFNPSCPK
>AT1G31250.1 | Symbols:  | proline-rich family protein |
chr1:11166658-11167500 REVERSE LENGTH=165
MANQYFRMAFLLCLFLSLSYQSIRIEAREGKACIGNCYGSSAPPVPPNASLSIPPNSSPSVKLT
PPYASPSVKLTPPYASPSVRPAGTTPNASPSVKLTP
```

```
PYASPSMRPAGTPNASPSVKLTPPYASPSVRPTGTTPNASPSLTPPNPSPSEKFIPPNASPFIH
T
>AT1G35230.1 | Symbols: AGP5 | arabinogalactan protein 5 |
chr1:12917184-12917585 FORWARD LENGTH=133
MASKSVVVFLFLALVASSVVAQAPGPAPTISPLPATPTPSQSPRATAPAPSPSANPPPSAPTTA
PPVSQPPTESPPAPPTSTSPSGAPGTNVPSGEAGPA
QSPLSGSPNAAAVSRVSLVGTFAGVAVIAALLL
>AT1G36150.1 | Symbols:  | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr1:13528257-13529470 FORWARD LENGTH=256
MKPSFVLLSIVLLLSSSLSDAADFGSPSQPPSMAPTPQPSNSTDCSSVIYSMVDCLSFLTVGST
DPSPTKTCCVGVKTVLNYSPKCLCSALESSREMGFV
LDDTKALAMPKICNVPIDPNCDVSTPAASTPVSPPVESPTTSPSSAKSPAITPSSPAVSHSPPP
VRHSSPPVSHSSPPVSHSSPPTSRSSPAVSHSSPVV
AASSPVKAVSSSTASSPRAASPSPSPSPSISSSGILLVSKLFIAVVMVSSFLYILA
>AT1G51915.1 | Symbols:  | cryptdin protein-related |
chr1:19293587-19293865 FORWARD LENGTH=67
MATERFSTMLISVLVLALVLSPILPCQATRAHLDAETRMLRRVCPSCVCCAPAPRGACCPCRCP
KNP
>AT1G55330.1 | Symbols: AGP21, ATAGP21 | arabinogalactan protein
21 | chr1:20648634-20648810 REVERSE LENGTH=58
MEAMKMKMMVFIMVVAVAFSAATAATVEAPAPSPTSDAAMFVPALFASVVALASGFIF
>AT1G63530.1 | Symbols:  | BEST Arabidopsis thaliana protein
match is: hydroxyproline-rich glycoprotein family protein
(TAIR:AT1G63540.1); Has 10212 Blast hits to 4024 proteins in 434
species: Archae - 1; Bacteria - 1259; Metazoa - 3608; Fungi -
2247; Plants - 291; Viruses - 90; Other Eukaryotes - 2716
(source: NCBI BLink). | chr1:23563654-23565236 FORWARD
LENGTH=499
MRVDFSEPEYRNCYHCRDSGILNTQESEPNNVIGSSITSVPVSSGPFQALGHASAAIGSTSTSS
PVQSSSSFSFGSAPAATTSVGVSSTSTPTPISFLPF
GYVPPPPPSSSLSGPAQNVPWSISAYSTAGRGVEWAWARPDVAVYPASSSSSTSTVVFGTTPV
SVSSLFGPSQDFPKTTFACCSSTPAPTRCECSFAKP
FFGSSPASSSSNLFGPNPSTTTASVFGTSATSGSSLFGPAQHPLSYDSIGGSAFAGVKTSSPQP
LGFNGATTLLPSSHEGQNISSFKPYVDKSHDELRWE
YYKKEAKGGSFPAAHASPIGSKPNAVFSPSTVSTSLFASPSVPDHHQRTNFSTTLGNFTGPAFG
FFPAAAGGATTFSSTGFGQPNAPFPGFDSQFPSAFS
SFSTTYPAVSSSVQRPHETGAVSSPAFGCTACGATSSSSASDHFTFNGATTPPGLFFPTTSSGP
MMFGTTLAAQGTTPALQTYPVQGFILLPFAAMSLQ
>AT1G63540.1 | Symbols:  | hydroxyproline-rich glycoprotein
family protein | chr1:23567495-23569495 FORWARD LENGTH=635
MKIDFSEPECSNCYQFTNPDLLNTPESEQSNVIGSSITSVPVNDGPVPPLELDSAAVVSTSTSS
PVQALGHDSGATVSTSTSSPVQIFSSPFSFGSAHAA
ITPVSSGPAPSPTFGEPRILLATSGSGASATSTSSTSSPLHSSSPFSFGSAPAAITSVSSGPAQ
SPASSPRLWIDRFATSSSASATSSSSTSSPFHSSSL
LGFAPAVTSVSSAPTPACGPTQAFGQPTQAFGLSMFGSTPRFEITGFPFQASASRNSPSPSFGP
AHNCGKPAFGSPFGNNVAFARPDVGISPVASSSTST
```

EIFGATPASLFSPFGPMQAPVQASASSTSTFPPFGCVPPSPSSGSSLFNSAFGSLPAPSSSNFF
GQSSSNLLGQNPSTTGVGYLPGSPLNSSFPGFGVGY
LPGSSSNLFRSNPPNFGGGSIGAGPQHFGFNGDASVLPSTPFSLSPAFSSNTNTGSYPFASHEW
SRPTEQGSMNPGYAPTHEGDNSSGWSFPTSKGNIYI
SISASKPYLHKSHEELRWEDYKQGDKGGPFPAAPASTIGSRPNAAFSPPTVSPPAHGCTACGAT
SSSSASRHFTFNGATSPPSAATTPPGLFFPSTGFAP
MMFGTNLAVQGTSPALQAYPVQGYILLPFAAMTLQ
>AT1G68725.1 | Symbols: AGP19, ATAGP19 | arabinogalactan protein 19 | chr1:25809298-25810130 FORWARD LENGTH=248
MESNSIIWSLLLASALISSFSVNAQGPAASPVTSTTTAPPPTTAAPPTTAAPPPTTTTPPVSAA
QPPASPVTPPPAVTPTSPPAPKVAPVISPATPPPQP
PQSPPASAPTVSPPPVSPPPAPTSPPPTPASPPPAPASPPPAPASPPPAPVSPPPVQAPSPISL
PPAPAPAPTKHKRKHKHKRHHHAPAPAPIPPSPPSP
PVLTDPQDTAPAPSPNTNGGNALNQLKGRAVMWLNTGLVILFLLAMTA
>AT2G04780.1 | Symbols: FLA7 | FASCICLIN-like arabinoogalactan 7 | chr2:1677488-1678252 FORWARD LENGTH=254
MAKMQLSIFIAVVALIVCSASAKTASPPAPVLPPTPAPAPAPENVNLTELLSVAGPFHTFLDYL
LSTGVIETFQNQANNTEEGITIFVPKDDAFKAQKNP
PLSNLTKDQLKQLVLFHALPHYYSLSEFKNLSQSGPVSTFAGGQYSLKFTDVSGTVRIDSLWTR
TKVSSSVFSTDPVAVYQVNRVLLPEAIFGTDVPPMP
APAPAPIVSAPSDSPSVADSEGASSPKSSHKNSGQKLLLAPISMVISGLVALFL
>AT2G14890.1 | Symbols: AGP9 | arabinogalactan protein 9 | chr2:6399679-6400755 FORWARD LENGTH=191
MARSFAIAVICIVLIAGVTGQAPTSPPTATPAPPTPTTPPPAATPPPVSAPPPVTTSPPPVTTA
PPPANPPPPVSSPPPASPPPATPPPVASPPPPVASP
PPATPPPVATPPPAPLASPPAQVPAPAPTTKPDSPSPSPSSSPPLPSSDAPGPSTDSISPAPSP
TDVNDQNGASKMVSSLVFGSVLVWFMI
>AT2G20520.1 | Symbols: FLA6 | FASCICLIN-like arabinogalactan 6 | chr2:8840663-8841406 FORWARD LENGTH=247
MSSSLFSYVVLLIFLFTIPYIQSQPTAPAPTTEKSPINLTAILEAGHQFTTLIQLLNTTQVGFQ
VSVQLNSSDQGMTIFAPTDNAFNKLKPGTLNSLTYQ
QQIQLMLYHIIPKYYSLSDLLLASNPVRTQATGQDGGVFGLNFTGQAQSNQVNVSTGVVETRIN
NALRQQFPLAVYYVVDSVLLPEELFGTKTTPTGAPAP
KSSTSSSDADSPAADDEHKSAGSSVKRTSLGIVVSFALFCCSVIYIA
>AT2G22470.1 | Symbols: AGP2, ATAGP2 | arabinogalactan protein 2 | chr2:9538400-9538795 REVERSE LENGTH=131
MNSKAMQALIFLGFLATSCLAQAPAPAPTTVTPPPTALPPVTAETPSPIASPPVPVNEPTPAPT
TSPTTSPVASPPQTDAPAPGPSAGLTPTSSPAPGPD
GAADAPSAAWANKAFLVGTAVAGALYAVVLA
>AT2G23130.1 | Symbols: AGP17, ATAGP17 | arabinogalactan protein 17 | chr2:9844411-9845186 FORWARD LENGTH=185
MTRNILLTVTLICIVFITVGGQSPATAPIHSPSTSPHKPKPTSPAISPAAPTPESTEAPAKTPV
EAPVEAPPSPTPASTPQISPPAPSPEADTPSAPEIA
PSADVPAPALTKHKKKTKKHKTAPAPGPASELLSPPAPPGEAPGPGPSDAFSPAADDQSGAQRI
SVVIQMVGAAAIAWSLLVLAF
>AT2G23990.2 | Symbols: ENODL11, AtENODL11 | early nodulin-like protein 11 | chr2:10206835-10207911 REVERSE LENGTH=226

MVSLISIVSVVFLLFTTFYHFGEARIINVGGSLDAWKVPESPNHSLNHWAESVRFQVGDALCSF
VMMVKIRMLVIVGYTFMFKYDSKIDSVLQVTKENYE
KCNTQKPLEEHKDGYTTVKLDVSGPYYFISGAPSGNCAKGEKVTVVVQSPNHPKPGPAAVTPTL
PPKPSTTPAAPAPAPPTPSPKSSTSTMAPAPAPAKS
SAVGLVAGNGIFWASTLVAVIGLAFA
>AT2G24450.1 | Symbols: FLA3 | FASCICLIN-like arabinogalactan protein 3 precursor | chr2:10393019-10393861 REVERSE LENGTH=280
MGLKVSSSLLCLTILLAVSSIVSAVNITRVLEKYPEFSTMTELLAKTELTPIINKRQTITVLAL
NNDAIGSISGRPEEEVKNILMNHVVLDYFDELKLKA
LKEKSTLLTTLYQSTGLGQQQNGFLNCTKSNGKIYFGSGVKGAPQTAEYITTVFRNPYNLSVVQ
ISMPIVAPGLGSPVKVPPPPPMSSPPAPSPKKGAAT
PAPAPADEGDYADAPPGLAPETAPASAPSESDSPAPAPDKSGKKKMAAADEAEPPSSASNTGLS
FGAVLVLGFVASFVGF
>AT2G25060.1 | Symbols: ENODL14, AtENODL14 | early nodulin-like protein 14 | chr2:10662308-10662930 FORWARD LENGTH=182
MFLSASMASSSLHVAIFSLIFLFSLAAANEVTVGGKSGDWKIPPSSSYSFTEWAQKARFKVGDF
IVFRYESGKDSVLEVTKEAYNSCNTTNPLANYTDGE
TKVKLDRSGPFYFISGANGHCEKGQKLSLVVISPRHSVISPAPSPVEFEDGPALAPAPISGSVR
LGGCYVVLGLVLGLCAWF
>AT2G26720.1 | Symbols: | Cupredoxin superfamily protein | chr2:11384782-11385402 FORWARD LENGTH=206
MALIKNNIFFTSLLIFVTLFGVAVGGTVHKVGNTKGWTMIGGDYEAWASSRVFQVGDTLVFAYN
KDYHDVTEVTHNDFEMCESSKPLRRYKTGSDSISLT
KPGLQHFICGVPGHCKKGQKLQIHVLPASLGHVAVPVPGPVRSQSSSSSPSPSPLVDPPVNNAP
QYQMGPTPASHSAASADFIFTFSFDLTLIDLCTFFI
LFFILV
>AT2G28440.1 | Symbols: | proline-rich family protein | chr2:12161226-12162032 FORWARD LENGTH=268
MAKKLCFIVMLSICLLIFDFAGAQEESPSPAAVSPGREPSTDSPLSPSSSPEEDSPLSPSSSPE
EDSPLPPSSSPEEDSPLAPSSSPEVDSPLAPSSSPE
VDSPQPPSSSPEADSPLPPSSSPEANSPQSPASSPKPESLADSPSPPPPPPQPESPSSPSYPEP
APVPAPSDDDSDDDPEPETEYFPSPAPSPELGMAQD
IKASDAAGEELNDERGEDYGMSGLEKAGIAIGTILGVGAIVIGALVYKKRRDNMTRARYTYFTE
GEFL
>AT2G31050.1 | Symbols: | Cupredoxin superfamily protein | chr2:13212150-13212752 FORWARD LENGTH=200
MALIKSNAFFTSLLILVALFGISVGGTVHKVGDSDGWTIMSVNYETWASTITFQVGDSLVFKYN
KDFHDVTEVTHNDYEMCEPSKPLARYETGSDIVILT
KPGLQHFICGFPGHCDMGQKLQIHVLPASLGPVAAPVPGPVRPPSSFSSPSQSPLAESPVNHAP
VQYQMGPSPAPHSAASNSNVWIGLCFLPLLSLLILV
>AT2G32300.1 | Symbols: UCC1 | uclacyanin 1 | chr2:13722510-13723464 FORWARD LENGTH=261
MASREMLIIISVLATTLIGLTVATDHTIGGPSGWTVGASLRTWAAGQTFAVGDNLVFSYPAAFH
DVVEVTKPEFDSCQAVKPLITFANGNSLVPLTTPGK
RYFICGMPGHCSQGMKLEVNVVPTATVAPTAPLPNTVPSLNAPSPSSVLPIQPLLPLNPVPVLS
PSSSTPLPSSSLPLIPPLSPALSPATAAGTSLPLFP
GSPGSSSSTTSTKTVGTFPSSTTGTTADLAGADSPPADSSSAAKTLVLGFGFMVAMMLHLF

```
>AT2G33790.1 | Symbols: ATAGP30, AGP30 | arabinogalactan protein
30 | chr2:14293862-14295237 REVERSE LENGTH=239
MGIIGKSVSLTLFALLCFTSSVFTLGVNQPGSSDPFHSLPQHLPLPPIKLPTLPPAKAPIKLPA
YPPAKAPIKLPTLPPAKAPIKLPTLPPIKPPVLPPV
YPPKYNKTLVAVRGVVYCKACKYAGVNNVQGAKPVKDAVVRLVCKNKKNSISETKTDKNGYFML
LAPKTVTNYDIKGCRAFLVKSPDTKCSKVSSLHDGG
KGSVLKPVLKPGFSSTIMRWFKYSVYNVGPFAFEPTCPK
>AT2G35860.1 | Symbols: FLA16 | FASCICLIN-like arabinogalactan
protein 16 precursor | chr2:15059859-15061810 FORWARD LENGTH=445
MDSSYGATKFLLLLFLTTSIATALPDNKPVPGQINSNSVLVALLDSHYTELAELVEKALLLQTL
EEAVGKHNITIFAPRNDALERNLDPLFKSFLLEPRN
LKSLQSLLMFHILPKRITSPQWPSLSHHHRTLSNDHLHLTVDVNTLKVDSAEIIRPDDVIRPDG
IIHGIERLLIPRSVQEDFNRRRSLRSISAVIPEGAP
EVDPRTHRLKKPSPAVPAGAPPVLPIYDAMSPGPSLAPAPAPGPGGPRGHFNGDAQVKDFIHTL
LHYGGYNEMADILVNLTSLATEMGRLVSEGYVLTVL
APNDEAMAKLTTDQLSEPGAPEQIMYYHIIPEYQTEESMYNAVRRFGKVKYDSLRFPHKVLAQE
ADGSVKFGHGDGSAYLFDPDIYTDGRISVQGIDGVL
FPKEETPATEIKPAAPVVKKVSKSRRGKLMEVACRMMGSRFIPCQ
>AT2G41905.1 | Symbols:  | BEST Arabidopsis thaliana protein
match is: arabinogalactan protein 23 (TAIR:AT3G57690.1); Has
35333 Blast hits to 34131 proteins in 2444 species: Archae -
798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531;
Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink). |
chr2:17495766-17495951 FORWARD LENGTH=61
MEMKKIACGVVFAAASMTAVMAAEVGAPAPGPAASGASVAVPALGSLVGASLVSLFAYYLN
>AT2G45000.1 | Symbols: EMB2766 | structural constituent of
nuclear pore | chr2:18564156-18567632 FORWARD LENGTH=739
MSGFPFGQSNSVGGFSFGSSSATNSSSASSTTSPLSFSFNQSSNPSSTGFGFGSSVSSTPASST
TPSFGFGASSTPSFGFGSSASSSTPSFGFGSSASVT
PASTTPSFGFGTAASSSAPAPSLFGSSTTNASSAAPGSSPFGFVTSSASSTATPSSSLFGAPAS
SAATPSSSPFGAAPASGSTPLFGSSPSLFSAPSSAS
ASNSSLFGASSSAATSTSPLFGAPSSATGATPSFSVASSAPGSSSSIFGATGSSPSFSVASSAS
GSSPSIFGATGSSPFFGSSSSAGSTPSLFASSSSGA
TTSSPSPFGVSTFNSSSTSNTSNASASPFSASTGFSFLKSTASSTTSSTTPSAPPQTASSSSSF
SFGTSANSGFNLSTGSSAAPASSTSGAVFSIATTTT
TSSSTPAATSAPASSAPASTMAFPSFGVTSSATNTTPASSAATFSTTGFGLASSTPATGSTNSF
TGFAVPKTSTPASSSQPQTTSPAFSFSLPSSTSTTA
PATSSATTTQTTLVVPSSSGTSTAVAPVAGSPKLPSEITGKTVEEIIKEWNTELQERTGRFRKQ
ANAIAEWDKRILQNRDVLLRLEIEVAKVVETQSSLE
RQLELIETHQQEVDKALQSMEEEAERIYNDERKSLLDDEAASTRDAMYEQSELVERELEHMTEQ
IRSIIQSVNANQGGELEAIDGMSPLDVVVRILNNQL
SSLMWIDEKAEEFSSRIQKIALQGSGGDRELMAPKHWMS
>AT2G45470.1 | Symbols: FLA8, AGP8 | FASCICLIN-like
arabinogalactan protein 8 | chr2:18742797-18744059 REVERSE
LENGTH=420
MAASQTFSLLAFTFSLLAFASTVSSHNITQILADSPDYSSFNSYLSQTKLADEINSRTTITVLV
LNNGAMSALAGKHPLSVIKSALSLLVLLDYYDPQKL
```

```
HKISKGTTLSTTLYQTTGNAPGNLGFVNITDLKGGKVGFGSAASGSKLDSSYTKSVKQIPYNIS
ILEIDAPIIAPGVLTAPAPSASLSNITGLLEKAGCK
TFANLLVSSGVLKTYESAVEKGLTVFAPSDEAFKAEGVPDLTKLTQAEVVSLLEYHALAEYKPK
GSLKTNKNNISTLATNGAGKFDLTTSTSGDEVILHT
GVAPSRLADTVLDATPVVIFTVDNVLLPAELFGKSKSPSPAPAPEPVTAPTPSPADAPSPTAAS
PPAPPTDESPESAPSDSPTGSANSKSANAAVGVSTP
SLFTALVTIAAIAVSVSLCS
>AT2G46330.1 | Symbols: AGP16, ATAGP16 | arabinogalactan protein
16 | chr2:19018730-19019108 REVERSE LENGTH=73
MASRNSVTGFALFSFVFAVILSLAGAQSLAPAPAPTSDGTSIDQGIAYLLMVVALVLTYLIHPL
DASSSYSFF
>AT2G47930.1 | Symbols: AGP26, ATAGP26 | arabinogalactan protein
26 | chr2:19617219-19617629 REVERSE LENGTH=136
MSVSLFTAFTVLSLCLHTSTSEFQLSTISAAPSFLPEAPSSFSASTPAMSPDTSPLFPTPGSSE
MSPSPSESSIMPTIPSSLSPPNPDAVTPDPLLEVSP
VGSPLPASSSVCLVSSQLSSLLLVLLMLLLAFCSFF
>AT3G01700.1 | Symbols: AGP11, ATAGP11 | arabinogalactan protein
11 | chr3:258981-259391 FORWARD LENGTH=136
MARLFVVVALLALAVGTVFAADAPSAAPTASPTKSPTKAPAAAPKSSAAAPKASSPVAEEPTPE
DDYSAASPSDSAEAPTVSSPPAPTPEADGPSSDGPS
SDGPAAAESPKSGATTNVKLSIAGTVAAAGFFIFSL
>AT3G01730.1 | Symbols:  | unknown protein; Has 8 Blast hits to
8 proteins in 2 species: Archae - 0; Bacteria - 0; Metazoa - 0;
Fungi - 0; Plants - 8; Viruses - 0; Other Eukaryotes - 0
(source: NCBI BLink). | chr3:266628-266891 FORWARD LENGTH=87
MRAFSLMICVLVLATLIMVAESKSSYSGNDETLTPPPPMAPSPSPTGTDGGHFGEPKTLLPPPN
AATFTTCPLLITSTLISALAFVF
>AT3G06360.1 | Symbols: AGP27, ATAGP27 | arabinogalactan protein
27 | chr3:1928976-1929353 FORWARD LENGTH=125
MASSILLTLITFIFLSSLSLSSPTTNTIPSSQTISPSEEKISPEIAPLLPSPAVSSTQTIPSSS
TLPEPENDDVSADPDPAFAPSASPPASSLASLSSQA
PGVFIYFVFAAVYCFSLRLLAVSAI
>AT3G11700.1 | Symbols: FLA18 | FASCICLIN-like arabinogalactan
protein 18 precursor | chr3:3698992-3700971 FORWARD LENGTH=462
MDRCIYGCSVITIFFSFFFLLNASALESGHHNITGSGQINSNSVLVALLDSRYTELAELVEKAL
LLQTLEDAVGRHNITIFAPRNEALERDLDPDFKRFL
LQPGNLKSLQTLLLSHIIPKRVGSNQWPEENSGRVKHVTLGHDQVLHLSKLKGTNGKRLVNSAV
ITRPDDLTRPDGLIHGIERLLIPRSVQEDFNRRRNL
RSISAVLPEGAPEIDPRTNRLKKSATAVSVPAGSPPVLPIESAMAPGPSLAPAPAPGPGGAHKH
FNGDAQVKDFIHTLLHYGGYNEMADILVNLTSLATE
MGRLVSEGYVLTVLAPNDEAMGKLTTDQLSEPGAPEQIMYYHIIPEYQTEESMYNSVRRFGKVK
YETLRFPHKVGAKEADGSVKFGSGDRSAYLFDPDIY
TDGRISVQGIDGVLFPEEKEEETVKKPTGPVKKVVQPRRGKLLEVACSMLGAIGKDSYLSRC
>AT3G12660.1 | Symbols: FLA14 | FASCICLIN-like arabinogalactan
protein 14 precursor | chr3:4019060-4019827 FORWARD LENGTH=255
MSSSLTIFFFFFASTFLYTSSNSFNITNILNEHDDFSNFNQLLSETQLASTINKRQTITVLVVS
NGALSSLSGQPTSVIKKILSLHIVLDYYDQKKLKNL
```

```
SKKTVLLTTLFQSSGLARGQQGFLNATVMKNGDVAFGSAVPGSSLDAQLQDTVAALPFNISVLH
ISSAIMIDVKGDNAPTASPLSPVSSPPRPAESPNDD
GQDFDEPPSSAPGAAADEPSENAGSANGVSRNDSQPAFAFTLLMSFIWWFMARLR
>AT3G13520.1 | Symbols: AGP12, ATAGP12 | arabinogalactan protein
12 | chr3:4409087-4409269 FORWARD LENGTH=60
MESMKMKLIVVLMVAIVAFSAVGNVAAQTEAPAPSPTSDAAMFVPALFASVAALASGFLF
>AT3G20570.1 | Symbols: ENODL9, AtENODL9 | early nodulin-like
protein 9 | chr3:7186754-7187453 REVERSE LENGTH=203
MARNLKSMMLCGFGLLCFLMIVDRAYAREFTVGGATGWTVPSGSQVYSQWAEQSRFQIGDSLLF
VYQSNQDSVLQVTRDAYDSCNTDSPTAKFADGKTSV
TLNHSGPYYFISGNKDNCKKNEKLVVIVMADRSGNKNTASSPPSPAPAPSGESAPSPPVSGTFE
MTPAPTPTTSEDTPNSAASSLSFVAALLGAALASTL
FLH
>AT3G20865.1 | Symbols: AGP40 | arabinogalactan protein 40 |
chr3:7308699-7308887 REVERSE LENGTH=62
MEMKNIFVALFISAVLVSSVSAATMESPAPSPGASSASTVAFPVVGSIVAASLSAFLALLLQ
>AT3G22070.1 | Symbols:  | proline-rich family protein |
chr3:7775530-7776066 REVERSE LENGTH=178
MLCYVGKATKIFIFIVTVVFVIGLVVAFGVLRRHSHHCSGDYCSSSTDPSSSSSSSSSTSPFI
TPFPNPNPNPNPNPPVLGSSPPSPTDSSSSTSISPN
PPAPIVNPNPPPPSTPNPPPEFSPPPPDLDTTTAPPPPSTDIPIPPPPPAPVSASPPLTPPSSV
VTSPAPVHAKLVND
>AT3G45230.1 | Symbols:  | hydroxyproline-rich glycoprotein
family protein | chr3:16569123-16569650 FORWARD LENGTH=175
MKLEFIIVAMMLSLVLVSGEILTKSSPAPSPDLADSPLIHASPPSKLGSHNSPAESPIEYSSPP
EPETEHSPSPSPANSPSVSPPLPNDSQSPSSSASPS
PSPEASDVNHSDITGIEGEKLPSGSGGGMSGGKKVGVAFGAIAAVCVVGVAGFVYKKRQENIRR
SRYGYAAREIL
>AT3G46550.1 | Symbols: SOS5 | Fasciclin-like arabinogalactan
family protein | chr3:17136612-17137874 REVERSE LENGTH=420
MANVISISHFTLLALPYLLLLLSSTAAAINVTAVLSSFPNLSSFSNLLVSSGIAAELSGRNSLT
LLAVPNSQFSSASLDLTRRLPPSALADLLRFHVLLQ
FLSDSDLRRIPPSGSAVTTLYEASGRTFFGSGSVNVTRDPASGSVTIGSPATKNVTVLKLLETK
PPNITVLTVDSLIVPTGIDITASETLTPPPTSTSLS
PPPAGINLTQILINGHNFNVALSLLVASGVITEFENDERGAGITVFVPTDSAFSDLPSNVNLQS
LPAEQKAFVLKFHVLHSYYTLGSLESITNPVQPTLA
TEEMGAGSYTLNISRVNGSIVTINSGVVLAVVTQTAFDQNPVSVFGVSKVLLPKELFPKSGQPV
ATAPPQEISLSPESSSEQPSRLVSPPREIVSSGAVK
RPLGFLVLWCWCIAFCYVLV
>AT3G52370.1 | Symbols: FLA15 | FASCICLIN-like arabinogalactan
protein 15 precursor | chr3:19417549-19419549 FORWARD LENGTH=436
MDDLSKLLFFLLLTISITTALPDKPGSGQINSNSVLVALLDSHYTELAELVEKALLLQTLEEAV
GQHNITIFAPRNDALEKNLDPEFKSFLLQPKNLKSL
QSLLMFHILPKRITSPQFSSAVVSHRTLSNDHLHFTNGKVNSAEITKPDDLTRPDGIIHGIERL
LIPRSVQEDFNRRRSLRSIAAVLPEGAPEVDPRTHR
LKKKPAPIPAGAPPVLPVYDAMSPGPSLAPAPAPGPGGPRHHFNGEAQVKDFIHTLLHYGGYNE
MADILVNLTSLATEMGRLVSEGYVLTVLAPNDEAMA
```

```
KLTTDQLSEPGAPEQIMYYHIIPEYQTEESMYNSVRRFGKIRYDSLRFPHKVEAQEADGSVKFG
HGDGSAYLFDPDIYTDGRISVQGIDGVLFPEEKTPV
EKKTGVPVVKKAPKPRRGKLMEVACTMLGSQFPTCQ
>AT3G57690.1 | Symbols: AGP23, ATAGP23 | arabinogalactan protein
23 | chr3:21384269-21384454 FORWARD LENGTH=61
MEMKKIACGVLFAAASMTAVMAAEEVGAPAPGPASAASAALPALGSLVGASLVSLFSYYLH
>AT3G60270.1 | Symbols:  | Cupredoxin superfamily protein |
chr3:22278029-22278762 REVERSE LENGTH=187
MGSTAATALLLLLLLVAVPAVFAVTFQVGDNDGWTIGVEYTSWVSEKTFRVGDTLEFKYGPSHS
VAVVNKADYDGCETSRPTQSFSDGDTKIDLTKVGAI
HFLCLTPGHCSLGMKLAVQVLAAVSLEPPPSPSAPSPSPSAPSPSPSAPSPSPSPGNAENLKNA
ASKGIMSYGKIVVTMVLMYGVLN
>AT3G60280.1 | Symbols: UCC3 | uclacyanin 3 | chr3:22279867-
22280633 REVERSE LENGTH=222
MGSTVAAALLLFLAAVPAVFAATFKVGDISGWTSNLDYTVWLTGKTFRVGDTLEFVYGLSHSVS
VVDKAGYDNCDSSGATQNFADGDTKIDLTTVGTMHF
LCPTFGHCKNGMKLAVPVLAAAPSPSTPSSPPSTPSTPSSPPSTPSTPSSPPSPPSPPSPSLPP
SSLPPSASPPTNGTPDSETLTPPPAPLPPSLSPNAA
SKGVMSYGIIGVTMILMYAVMT
>AT3G60900.1 | Symbols: FLA10 | FASCICLIN-like arabinogalactan-
protein 10 | chr3:22499573-22500841 REVERSE LENGTH=422
MATSRAFTLFAFTLSLLTVASTVSGHNITQILSDTPEYSSFNNYLSQTKLADEINSRTTITVLV
LNNGAMSSLAGKHPLSVVKNALSLLVLLDYYDPLKL
HQLSKGTTLTTTLYQTTGHALGNLGFVNVTDLKGGKVGFGSAAPGSKLDSSYTKSVKQIPYNIS
VLEINAPIIAPGILTAPAPSSAGVSNITGLLEKAGC
KTFANLLVSSGVIKTFESTVEKGLTVFAPSDEAFKARGVPDLTNLTQAEVVSLLEYHALAEYKP
KGSLKTNKDAISTLATNGAGKYDLTTSTSGDEVILH
TGVGPSRLADTVVDETPVVIFTVDNVLLPAELFGKSSSPAPAPEPVSAPTPTPAKSPSPVEAPS
PTAASPPAPPVDESSPEGAPSDSPTSSENSNAKNAA
FHVNAPALFTALVTIAATSLLL
>AT3G61640.1 | Symbols: AGP20, AtAGP20 | arabinogalactan protein
20 | chr3:22810283-22810629 REVERSE LENGTH=74
MASRNSVAVIALFAFVFAVISPFAGAQSLAPAPSPTSDGTSIDQGIAYLLMVVALVLTYLIHPL
DASSSSYTFF
>AT4G09030.1 | Symbols: AGP10, ATAGP10 | arabinogalactan protein
10 | chr4:5792249-5792632 FORWARD LENGTH=127
MASKSVVVLLFLALIASSAIAQAPGPAPTRSPLPSPAQPPRTAAPTPSITPTPTPTPSATPTAA
PVSPPAGSPLPSSASPPAPPTSLTPDGAPVAGPTGS
TPVDNNNAATLAAGSLAGFVFVASLLL
>AT4G12730.1 | Symbols: FLA2 | FASCICLIN-like arabinogalactan 2
| chr4:7491598-7492809 REVERSE LENGTH=403
MAYLRRAATALVLIFQLHLFLSLSNAHNITRILAKDPDFSTFNHYLSATHLADEINRRQTITVL
AVDNSAMSSILSNGYSLYQIRNILSLHVLVDYFGTK
KLHQITDGSTSTASMFQSTGSATGTSGYINITDIKGGKVAFGVQDDDSKLTAHYVKSVFEKPYN
ISVLHISQVLTSPEAEAPTASPSDLILTTILEKQGC
KAFSDILKSTGADKTFQDTVDGGLTVFCPSDSAVGKFMPKFKSLSPANKTALVLYHGMPVYQSL
QMLRSGNGAVNTLATEGNNKFDFTVQNDGEDVTLET
```

```
DVVTAKVMGTLKDQEPLIVYKIDKVLLPREIYKAVKTSAPAPKSSKKKPKNAEADADGPSADAP
SDDDVEVADDKNGAVSAMITRTSNVVTAIVGLCFGV
WLM
>AT4G16980.1 | Symbols:  | arabinogalactan-protein family |
chr4:9557087-9557581 FORWARD LENGTH=164
MASSFSSQAFFLLTLSMVLIPFSLAQAPMMAPSGSMSMPPMSSGGGSSVPPPVMSPMPMMTPPP
MPMTPPPMPMTPPPMPMAPPPMPMASPPMMPMTPST
SPSPLTVPDMPSPPMPSGMESSPSPGPMPPAMAASPDSGAFNVRNNVVTLSCVVGVVAAHFLLV
>AT4G26320.1 | Symbols: AGP13 | arabinogalactan protein 13 |
chr4:13317235-13317414 REVERSE LENGTH=59
MEAMKMRLFVAVLVAAMAFSAVQQAAAVEAPAPSPTSDASLAIPAFFASVATLAFGFLF
>AT4G27520.1 | Symbols: ENODL2, AtENODL2 | early nodulin-like
protein 2 | chr4:13750668-13751819 REVERSE LENGTH=349
MTFLKMKSLSFFFTILLSLSTLFTISNARKFNVGGSGAWVTNPPENYESWSGKNRFLVHDTLYF
SYAKGADSVLEVNKADYDACNTKNPIKRVDDGDSEI
SLDRYGPFYFISGNEDNCKKGQKLNVVVISARIPSTAQSPHAAAPGSSTPGSMTPPGGAHSPKS
SSPVSPTTSPPGSTTPPGGAHSPKSSSAVSPATSPP
GSMAPKSGSPVSPTTSPPAPPKSTSPVSPSSAPMTSPPAPMAPKSSSTIPPSSAPMTSPPGSMA
PKSSSPVSNSPTVSPSLAPGGSTSSSPSDSPSGSAM
GPSGDGPSAAGDISTPAGAPGQKKSSANGMTVMSITTVLSLVLTIFLSA
>AT4G28365.1 | Symbols: ENODL3, AtENODL3 | early nodulin-like
protein 3 | chr4:14033012-14033688 REVERSE LENGTH=199
MGLVMRFDLYLMFVMLMGLGFTISNGYKFYVGGKDGWVPTPSEDYSHWSHRNRFQVNDTLHFKY
AKGKDSVLEVTEQEYNTCNTTHPLTSLSDGDSLFLL
SHSGSYFFISGNSQNCLKGQKLAVKVLSTVHHSHSPRHTSPSPSPVHQELSSPGPSPGVEPSSD
SNSRVPAPGPATAPNSAGLVGPGMVVLVIMISSLF
>AT4G30590.1 | Symbols: ENODL12, AtENODL12 | early nodulin-like
protein 12 | chr4:14935760-14936469 REVERSE LENGTH=190
MGIIVPVLTLVFLLFAKVSHGASNPRVILVGGSVGSWKVPDSPNNTLNHWAENNRFKVGDFIVW
KYDMKVDSVLQVTKEDYESCNTANPLKQYNDGNTKV
ALDKSGPYFFISGAPGNCAKGEKITLVVLAERKSGGGSSSGDAPKVSPVSPTAQTPAPAPGPAA
AHNAAVGLKVASGWFLTAVVVGLAMA
>AT4G31370.1 | Symbols: FLA5 | FASCICLIN-like arabinogalactan
protein 5 precursor | chr4:15223838-15224674 REVERSE LENGTH=278
MGLKASLSLLSLTILLVFSKVVTANNITLAFQKYSKFSTMRDLFIKTKLIAAIDKYQTITVLAV
SNDAISSITNRSEVELRNILMTHVILDYYDELKLQG
MREKSIMLTTLYQTTGLGEQMNGFLNVSKSKGRVYFGSEVKNSPLNAEYVSTVYHNPYNLSIIQ
ITMPIVAPGLSLAIFPPPPPYVHVAPYPTPMDASVV
PAPGPAADDNSPDSAVPKTPPAPATDTPEADSPAPAPSADNEKIEAADKAKPSSSASKAGWSFD
VILLLAFLASFAGF
>AT4G31840.1 | Symbols: ENODL15, AtENODL15 | early nodulin-like
protein 15 | chr4:15401798-15402426 FORWARD LENGTH=177
MASSSLLVTIFLCISVFFFSSVNANEVTVGGKSGDWKIPPSSSFSFNEWAQKARFKVGDFIVFK
YEAGKDSVLQVTREAYEKCNTTSPKASYTDGNTKVK
LDQAGPVYFVSGTEGHCQKGQKLRLVVITPRNSAFSPGPSPSEFDGPAVAPTSGAAKLAGGFSV
VFGLVLGLWAFFF
>AT4G32490.1 | Symbols: ENODL4, AtENODL4 | early nodulin-like
protein 4 | chr4:15678811-15679556 REVERSE LENGTH=221
```

```
MVFVKMTDVYLMIVMLMGLGFSIELSNGHKFYVGGRDGWVLTPSEDYSHWSHRNRFQVNDTLYF
KYVKGKDSVLEVSEKEYNTCNTTHPLTSLSDGDSLF
LLSRSDPFFFVSGNSGSCLKGQKLAVTVMSTGHHSHTPRHPSPSPSPSASPVRKALLSPAPIPV
HKALSSPAPTPGVDPSHSEVLAPAPGPAAAVRNLAG
SVAPGVISLGLVLVIMISSMV
>AT4G37450.1 | Symbols: AGP18, ATAGP18 | arabinogalactan protein
18 | chr4:17605926-17606734 REVERSE LENGTH=209
MDRNFLLTVTLICIVVAGVGGQSPISSPTKSPTTPSAPTTSPTKSPAVTSPTTAPAKTPTASAS
SPVESPKSPAPVSESSPPPTPVPESSPPVPAPMVSS
PVSSPPVPAPVADSPPAPVAAPVADVPAPAPSKHKKTTKKSKKHQAAPAPAPELLGPPAPPTES
PGPNSDAFSPGPSADDQSGAASTRVLRNVAVGAVAT
AWAVLVMAF
>AT4G40090.1 | Symbols: AGP3 | arabinogalactan protein 3 |
chr4:18581085-18581504 REVERSE LENGTH=139
MALKTLQALIFLGLFAASCLAQAPAPAPITFLPPVESPSPVVTPTAEPPAPVASPPIPANEPTP
VPTTTPPTVSPPTTSPTTSPVASPPKPYALAPGPSGP
TPAPAPAPRADGPVADSALTNKAFLVSTVIAGALYAVLA
>AT5G03170.1 | Symbols: FLA11, ATFLA11 | FASCICLIN-like
arabinogalactan-protein 11 | chr5:752898-753638 REVERSE
LENGTH=246
MATSRTFIFSNLFIFFLVIATTYGQAPAPGPSGPTNITAILEKAGQFTLFIRLLKSTQASDQIN
TQLNSSSSNGLTVFAPTDNAFNSLKSGTLNSLSDQQ
KVQLVQFHVLPTLITMPQFQTVSNPLRTQAGDGQNGKFPLNITSSGNQVNITTGVVSATVANSV
YSDKQLAVYQVDQVLLPLAMFGSSVAPAPAPEKGGS
VSKGSASGGDDGGDSTDSSDAERTGFGFGIRITTVAAIAASSSLWI
>AT5G06390.1 | Symbols: FLA17 | FASCICLIN-like arabinogalactan
protein 17 precursor | chr5:1952939-1955047 FORWARD LENGTH=458
MDRRIYGGSAVIHLFLFFSVLIFSAASALSKNQSPSSGSGQINSNSVLVALLDSRYTELAELVE
KALLLQTLEDAVGRHNITIFAPRNEALERDLDPEFK
RFLLEPGNLKSLQTLLMFHIIPNRVGSNQWPSEESGRVKHHTLGNDQVRLSNGQGKKMVDLAEI
IRPDDLTRPDGLIHGIERLLIPRSVQEDFNRRRSLQ
SISAVLPEGAPEVDPRTNRLKKPAAPVPAGSPPALPIQSAMAPGPSLAPAPAPGPGGKQHHFDG
EAQVKDFIHTLLHYGGYNEMADILVNLTSLATEMGR
LVSEGYVLTVLAPNDEAMAKLTTDQLSEPGAPEQIVYYHIIPEYQTEESMYNSVRRFGKVKFDT
LRFPHKVAAKEADGSVKFGDGEKSAYLFDPDIYTDG
RISVQGIDGVLFPQEEEVVESVKKPVKKIVQPRRGKLLEVACSMLGAFGKDTYLSKCR
>AT5G06920.1 | Symbols: FLA21 | FASCICLIN-like arabinogalactan
protein 21 precursor | chr5:2142858-2143919 FORWARD LENGTH=353
MGCCSSDCFVYFILSIALAFMAISTTLRSPPDSEPTIPIAFSSSSPSLSLNASNTLRQSNFKAI
ATLLHISPEIFLSSSPNTTLFAIEDASFFNTSSLHP
LFLKQLLHYHTLPLMLSMDDLLKKPQGTCLPTLLHHKSVQISTVNQESRTAEVNHVRITHPDMF
LGDSLVIHGVIGPFSPLQPHSDHLIHTPLCQSDTTN
KTSNNEEVPVSIDWTRIVQLLSSNGFVPFAIGLHSVLNRIVNDHNHHKNLTGVTILATPNLVSL
SSASPFLYEVVRHHILVQRLTYKDFASMSDKATVKT
LDPYQDLTITRRNVNSSGGDFMISGVEIVDPDMFSSSNFVIHGISHTLEIPHV
>AT5G10430.1 | Symbols: AGP4, ATAGP4 | arabinogalactan protein 4
| chr5:3277828-3278235 REVERSE LENGTH=135
```

```
MGSKIVQVFLMLALFATSALAQAPAPTPTATPPPATPPPVATPPPVATPPPAATPAPATPPPAA
TPAPATTPPSVAPSPADVPTASPPAPEGPTVSPSSA
PGPSDASPAPSAAFSNKAFFAGTAFAAIMYAAVLA
>AT5G11740.1 | Symbols: AGP15, ATAGP15 | arabinogalactan protein
15 | chr5:3784318-3784503 FORWARD LENGTH=61
MAISKASIVVLMMVIISVVASAQSEAPAPSPTSGSSAISASFVSAGVAAVAALVFGSALRI
>AT5G12880.1 | Symbols:  | proline-rich family protein |
chr5:4068714-4068935 REVERSE LENGTH=73
MRRSWPTRLILLMVLTVVITMIAAAYGYSSVSSSKHKFPHYKYKAPSPPTTYSPYRYFSPPPVT
DSDSAAYVR
>AT5G14380.1 | Symbols: AGP6 | arabinogalactan protein 6 |
chr5:4636088-4636540 REVERSE LENGTH=150
MARQFVVLVLLTLTIATAFAADAPSASPKKSPSPTAAPTKAPTATTKAPSAPTKAPAAAPKSSS
ASSPKASSPAAEGPVPEDDYSASSPSDSAEAPTVSS
PPAPTPDSTSAADGPSDGPTAESPKSGAVTTAKFSVVGTVATVGFFFFSF
>AT5G18690.1 | Symbols: AGP25, ATAGP25 | arabinogalactan protein
25 | chr5:6233776-6234126 REVERSE LENGTH=116
MAFSFLNKLLIIFIFIFISLSSSSPTISLVQQLSPEIAPLLPSPGDALPSDDGSGTIPSSPSPP
DPDTNDGSYPDPLAFSPFASPPVSSPSPPPSLPSAG
VLLISLIISSASFLAL
>AT5G21160.3 | Symbols:  | LA RNA-binding protein |
chr5:7198817-7203879 REVERSE LENGTH=833
MMAETEGSVADDRELITREGGIGTKSPWKTTTSPVETIDAPVMGAHSWPALADAAQQPRPKNPP
APAPAPPSKNIPTSIPIPTPAVTGQAKSKGGGKANP
GHKNPSGRHSKPGPRSNQNGPPPPPYLVHAVPYHPPPFPPMVPLPHAAGPDFPYAPYPPYPVPV
PPVTESGNEKQVQASPLPPVLPAPQGDPGKPWPHQR
GFDPRNMPQGAGPRNFGRPPFMGPAPGFLVGPGPGFPGPVYYLPGPPPGAIRGPYPPRFAPYPV
NQGPPILSPEKLDLRDRVLKQVEYYFSDENLENDHY
LISLMDEEGWVPTKIIAGFKRVKAMTMDVDFIVYALGFSNSVEVQGDQIRKRDKWSDWIPASKK
STSAETIGDGDKDSPKSITSGDNFGNPSKGSSKPTV
SDFSSEGAQSSRTNNYKSGNLKSSADEKRNVEDLSNDFSNTFLLDEELDLEHRSPRKSGLSMSK
SIEYEDDDMAVDDQDIQKLVIVTQNSGKSDGAGIGG
TEAKNIPKELASTINDGLYYFEQELKKKRSGRRKNNSHLDTKDGKIKSGEGLNTKLGENSAAND
GGEEHGIITSRRKQNKGTHKHHTAHARRFFSSNIRN
NGNISESPPSSSIGFFFGSTPPDSHGPRLSKLSSSPQCTLSGSSPPVGSLPKSFPPFQHPSHQL
LEENGFKQEKYLKYRKRCLNERKKLGSGCSEEMNHL
YRFWSYFLRDTFVLSMYDDFQKFALEDAAGNYDYGLECLFRFYSYGLEKHFDEDLYKDFEKLSL
DFYHKGNLYGLEKYWAFHHYRGKEEPITKHPELEKL
LKEEFRSIDDFRAKETITNQKENKRKKNKKKMM
>AT5G24105.1 | Symbols: AGP41 | arabinogalactan protein 41 |
chr5:8152097-8152656 FORWARD LENGTH=63
MSGSRLFFGVSTIVSIIFAILLPMAHAQSAAPAPAPTSDGTTIDQGIAYVLMLVALVLTYLIH
>AT5G25090.1 | Symbols: ENODL13, AtENODL13 | early nodulin-like
protein 13 | chr5:8647117-8647755 REVERSE LENGTH=186
MAQRTLVATFFLIFFLLTNLVCSKEIIVGGKTSSWKIPSSPSESLNKWAESLRFRVGDTLVWKY
DEEKDSVLQVTKDAYINCNTTNPAANYSNGDTKVKL
ERSGPYFFISGSKSNCVEGEKLHIVVMSSRGGHTGGFFTGSSPSPAPSPALLGAPTVAPASGGS
ASSLTRQVGVLGFVGLLAIVLL
```

```
>AT5G26330.1 | Symbols:  | Cupredoxin superfamily protein |
chr5:9241614-9242635 REVERSE LENGTH=187
MAAIIVAALACIVVMLRLSEAAVYKVGDSAGWTTIANVDYKLWASTKTFHIGDTVLFEYNPQFH
NVMRVTHPMYRSCNTSKPISTFTTGNDSITLTNHGH
HFFFCGVPGHCLAGQKLDLHVLLPASSTPLSDPPTSSSSSPPSTTIPAAGVPGPSPSLAASLPS
MVTAQIVAVVTLLVSLAFTNFAS
>AT5G40730.1 | Symbols: AGP24, ATAGP24 | arabinogalactan protein
24 | chr5:16301153-16301362 FORWARD LENGTH=69
MMMMTKMFVQIAVVCLLATMAVVSAHEGHHHHAPAPAPGPASSSTVVSATNMFTVLAIAAVALV
VGSNH
>AT5G40940.1 | Symbols: FLA20 | putative fasciclin-like
arabinogalactan protein 20 | chr5:16406656-16407930 REVERSE
LENGTH=424
MNKKPSKYFPSLNRTDYKNCNFLCNGFSLKPNEKKSSKNASMASKLLTTFFLIFFVLDIDLVAT
SMTSVSSAVEVLSDSGYLSMGLTLKLANQDLNLEDW
QELTLFAPSDQSFSKFGQPSLLDMKYQLSPTRLPGETLRNLPNGAKIPTLRSNYSLTVTNSSRF
GGKTSINNVVVQDSPVFDDGYVVIYGSDEFFTSPTK
ISDDSSSSSSIPSTTSSTGSIPIPSSATQTPPSPNIASDSTRNLPNRSKPVNRFNIFESASRLL
MSRGFVIIATFLALQLEDNTSGNDTKITVFAPIDEA
IPNPTTKFSDYVTIFRGHVVSQLLLWKDLQKFAKEGSILQTVLKGYEIEISLSGDILLLNGVPL
IYPDLYVNDWIAVHGFNQMIVTKEKQVDVGDSITVL
NNGEQEEEGVHGEYSSELGDYGLH
>AT5G44130.1 | Symbols: FLA13 | FASCICLIN-like arabinogalactan
protein 13 precursor | chr5:17761128-17761871 FORWARD LENGTH=247
MATTPLLLLLLTAVFLSTEITAQRAAPAPGPAGPINITAILEKGGQFVTLIRLLNTTQIGNQIN
IQINSSSEGMTVLAPTDNAFQNLKPGTLNKLSPDDQ
VKLILYHVSPKFYTLEDLLSVSNPVRTQASGRDVGGVYGLNFTGQGNQVNVSTGVVETRLSTSL
RQERPLAVYVVDMVLLPEEMFGERKISPMAPPPKSK
SPDVSDDSESSKKAAAPSESEKSGSGEMNTGLGLGLGLVVLCLKFLL
>AT5G53250.1 | Symbols: AGP22, ATAGP22 | arabinogalactan protein
22 | chr5:21603715-21604007 FORWARD LENGTH=63
MASLKFPLEILAVFVIISVILLPIAQSHSSSPAPAPTSDGTSIDQGIAYVLMMVALALTYFIH
>AT5G53870.1 | Symbols: ENODL1, AtENODL1 | early nodulin-like
protein 1 | chr5:21870033-21871228 REVERSE LENGTH=370
MSAIMKSLCFSFLILASFATFFSVADAWRFNVGGNGAWVTNPQENYNTWAERNRFQVNDSLYFK
YAKGSDSVQQVMKADFDGCNVRNPIKNFENGESVVT
LDRSGAFYFISGNQDHCQKGQKLIVVVLAVRNQPSAPAHSPVPSVSPTQPPKSHSPVSPVAPAS
APSKSQPPRSSVSPAQPPKSSSPISHTPALSPSHAT
SHSPATPSPSPKSPSPVSHSPSHSPAHTPSHSPAHTPSHSPAHAPSHSPAHAPSHSPAHAPSHS
PAHSPSHSPATPKSPSPSSSPAQSPATPSPMTPQSP
SPVSSPSPDQSAAPSDQSTPLAPSPSETTPTADNITAPAPSPRTNSASGLAVTSVMSTLFSATF
TFLMFA
>AT5G55730.1 | Symbols: FLA1 | FASCICLIN-like arabinogalactan 1
| chr5:22558375-22560392 REVERSE LENGTH=424
MAKKMSSLIIIFNILLLLTTQTHAHNVTRLLANHPSFSSFSHFLTQTHLADEINRRRTITVCAV
DNAAMSALTSKGYTLSTLKNILSLHVLLDYFGTKKL
HQIRDGSALAATLFQATGAAPGTSGFVNITDLRGGKVGFGPDGGDLSSFFVKSIEEVPYNISII
QISRVLPSETAAAPTPAPAEMNLTGIMSAHGCKVFA
```

```
ETLLTNPGASKTYQESLEGGMTVFCPGDDAMKGFLPKYKNLTAPKKEAFLDFLAVPTYYSMAML
KSNNGPMNTLATDGANKFELTVQNDGEKVTLKTRIN
TVKIVDTLIDEQPLAIYATDKVLLPKELFKASAVEAPAPAPAPEDGDVADSPKAAKGKAKGKKK
KAAPSPDNDPFGDSDSPAEGPDGEADDATADDAGAV
RIIGGAKAGLVVSLLCLFASSWLL
>AT5G56330.1 | Symbols: ATACA8, ACA8 | alpha carbonic anhydrase
8 | chr5:22813768-22816162 FORWARD LENGTH=350
MKISSLGWVLVLIFISITIVSSAPAPKPPKPKPAPAPTPPKPKPTPAPTPPKPKPKPAPTPPKP
KPAPAPTPPKPKPAPAPTPPKPKPKPAPTPPNPKPT
PAPTPPKPKPAPAPAPTPAPKPKPAPKPAPGGEVEDETEFSYETKGNKGPAKWGTLDAEWKMCG
IGKMQSPIDLRDKNVVVSNKFGLLRSQYLPSNTTIK
NRGHDIMLKFKGGNKGIGVTIRGTRYQLQQLHWHSPSEHTINGKRFALEEHLVHESKDKRYAVV
AFLYNLGASDPFLFSLEKQLKKITDTHASEEHIRTV
SSKQVKLLRVAVHDASDSNARPLQAVNKRKVYLYKPKVKLMKKYCNISSY
>AT5G56540.1 | Symbols: AGP14, ATAGP14 | arabinogalactan protein
14 | chr5:22893243-22893425 FORWARD LENGTH=60
MEAMKMKLYVVVLVAVIAFSTVHQTVAAVDAPAPSPTSDASSFIPTFFASVAVMAFGFFF
>AT5G60490.1 | Symbols: FLA12 | FASCICLIN-like arabinogalactan-
protein 12 | chr5:24325916-24326665 REVERSE LENGTH=249
MEHSLIIILLFTVLLLLTTTPGILSQPSPAVAPAPPGPTNVTKILEKAGQFTVFIRLLKSTGVAN
QLYGQLNNSDNGITIFAPSDSSFTGLKAGTLNSLTD
EQQVELIQFHVIPSYVSSSNFQTISNPLRTQAGDSADGHFPLNVTTSGNTVNITSGVTNTTVSG
NVYSDGQLAVYQVDKVLLPQQVFDPRPPAPAPAPSV
SKSKKKKDDSDSSSDDSPADASFALRNVGSVCDAVSFCVMSVMLAWFYL
>AT5G64310.1 | Symbols: AGP1, ATAGP1 | arabinogalactan protein 1
| chr5:25722018-25722413 FORWARD LENGTH=131
MAFSKSLVFVLLAALLISSAVAQSPAPAPSNVGGRRISPAPSPKKMTAPAPAPEVSPSPSPAAA
LTPESSASPPSPPLADSPTADSPALSPSAISDSPTE
APGPAQGGAVSNKFASFGSVAVMLTAAVLVI
>AT5G65390.1 | Symbols: AGP7 | arabinogalactan protein 7 |
chr5:26128883-26129275 REVERSE LENGTH=130
MNSKIIEAFFIVALFTTSCLAQAPAPSPTTTVTPPPVATPPPAATPAPTTTPPPAVSPAPTSSP
PSSAPSPSSDAPTASPPAPEGPGVSPGELAPTPSDA
SAPPPNAALTNKAFVVGSLVAAIIYAVVLA
>AT1G55260.1 | Symbols:  | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr1:20614663-20616158 FORWARD LENGTH=227
MIPSSNQYVDTHRVTEDRAIYIHEHTKRLEDTFLSRENTTHRTMEKSTRTLFITIVITSMLLGF
GNSDLAQDREECTNQLIELSTCIPYVGGDAKAPTKD
CCAGFGQVIRKSEKCVCILVRDKDDPQLGIKINATLAAHLPSACHITAPNITDCISILHLPRNS
TLAKEFENLGRIEDNYNSTSPTQIHKDGTGGGKAEP
VKSNGWKEKSWLGVELLIYLLVSLIFF
>AT2G13820.1 | Symbols:  | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr2:5774295-5776279 REVERSE LENGTH=169
MAYATILMIFSVVALMSGERAHAAVDCSSLILNMADCLSFVTSGSTVVKPEGTCCSGLKTVVRT
GPECLCEAFKNSGSLGLTLDLSKAASLPSVCKVAAP
```

```
PSARCGLSVSGDPPATAPGLSPTAGAGAPALSSGANAATPVSSPRSSDASLLSVSFAFVIFMAL
ISSFY
>AT2G27130.1 | Symbols:  | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr2:11595379-11596395 FORWARD LENGTH=176
MLTTNTLAVLLLLFLSLCSGQSPPAPEPIAADGPSSPVNCLVSMLNVSDCFSYVQVGSNEIKPE
AACCPELAGMVQSSPECVCNLYGGGASPRFGVKLDK
QRAEQLSTICGVKAPSPSLCSVLGFPTISPAGSEDSSSGSEGSDKDKKNGAMTTKYCGVALNSL
ALLLLFTFLSLS
>AT2G44290.1 | Symbols:  | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr2:18305418-18306202 REVERSE LENGTH=205
MESRKIKVMATAIALIMVAMVVDAAGADKGKDKEECTAQLVGMATCLPYVQGKAKSPTPDCCSG
LKQVINSDMKCLCMIIQERNDPDLGLQVNVSLALAL
PSVCHATADITKCPALLHLDPNSPDAQVFYQLAKGLNETVSASAPTGSASEPTSMSSTPGSSAG
NNSGRTTSVPGTNHAQSFSKQWLGLEVVAHFFVIFY
IFILV
>AT2G44300.1 | Symbols:  | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr2:18307468-18308286 REVERSE LENGTH=204
MESRKINLMATAIALIVVAMVVAAADDKTKDKEECTEQLVGMATCLPYVQGQAKSPTPDCCSGL
KQVLNSNKKCLCVIIQDRNDPDLGLQINVSLALALP
SVCHAAADVTKCPALLHLDPNSPDAQVFYQLAKGLNKTGPASAPTGSSPGPISISPTSGSDDGN
NSGRTTSVPGRNHAQSFYKQWLGLEVVFHFFVIFYI
FILV
>AT2G48130.1 | Symbols:  | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr2:19685263-19685977 REVERSE LENGTH=183
MGYRRSYAITFVALVAALWSVTKAQPSSSCVSTLTTLSPCLSYITGNSTTPSQPCCSRLDSVIK
SSPQCICSAVNSPIPNIGLNINRTQALQLPNACNIQ
TPPLTQCNAATGPTAQPPAPSPTEKTPDVTLTPTSLPGARSGVGGGSKTVPSVGTGSSSRNVDP
LPLHFLMFAVLVVCTSSFL
>AT3G22600.1 | Symbols:  | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr3:8006711-8007397 REVERSE LENGTH=170
MKMEMGLVFLTVFMAVMSSTMVSAQSSCTNALISMSPCLNYITGNSTSPNQQCCNQLSRVVQSS
PDCLCQVLNGGGSQLGINVNQTQALGLPRACNVQTP
PVSRCNTGGGGGGSTSDSPAESPNSSGPGNGSKTVPVGEGDGPPSSDGSSIKFSFPLIAFFSAV
SYMAIF
>AT3G43720.1 | Symbols:  | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr3:15615549-15617099 REVERSE LENGTH=193
MSNVVVIAVVLIVASLTGHVSAQMDMSPSSGPSGAPDCMANLMNMTGCLSYVTVGEGGGAAKPD
KTCCPALAGLVESSPQCLCYLLSGDMAAQLGIKIDK
AKALKLPGVCGVITPDPSLCSLFGIPVGAPVAMGDEGASPAYAPGSMSGAESPGGFGSGPSASR
GSDAPSSAPYSLFLNLIIFPLAFAFYIFC
```

```
>AT4G08670.1 | Symbols:  | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr4:5536765-5538210 REVERSE LENGTH=208
MKQSLLLSFVLLLLSSSSLVTPIHARNKSNPAKSPVGAPAPGPSSSDCSTVIYSMMDCLGYLGV
GSNETKPEKSCCTGIETVLQYNPQCICAGLVSAGEM
GIELNSTRALATPKACKLSIAPPHCGIITSGATTPGASPVSPSAGAPTTSPSAAKSPETSATSP
SSDETPSMTAPSPSSSGTNILSVPALTIVFVIVSSV
AYISAFSN
>AT4G14815.1 | Symbols:  | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr4:8505122-8505760 FORWARD LENGTH=156
MKPRMCLILFIALMRVMSIVSAQSSCTNVLISMAPCLSFITQNTSLPSQQCCNQLAHVVRYSSE
CLCQVLDGGGSQLGINVNETQALALPKACHVETPPA
SRCHSGSSVNSHSEHGNGSKTVPREKSSSDGSIKFSFPLLAILFTASYITLIYAKY
>AT5G09370.1 | Symbols:  | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr5:2909450-2910475 REVERSE LENGTH=158
MAYFSTATSLLLLVLSVSSPYVHGASDCDTLVITLFPCLPFISIGGTADTPTASCCSSLKNILD
TKPICLCEGLKKAPLGIKLNVTKSATLPVACKLNAP
PVSACDSLPPASPPTANGQAPVWGSGWAPAPSPSKGNSLIPISGFSFVIVTALAMFRI
>AT5G64080.1 | Symbols:  | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr5:25645475-25646638 REVERSE LENGTH=182
MATHSSFTATTPLFLIVLLSLSSVSVLGASHHHATAPAPSVDCSTLILNMADCLSFVSSGGTVA
KPEGTCCSGLKTVLKADSQCLCEAFKSSASLGVTLN
ITKASTLPAACKLHAPSIATCGLSVAPSTAPGLAPGVAAAGPETAGFLAPNPSSGNDGSSLIPT
SFTTVLSAVLFVLFFSSA
>AT1G07460.1 | Symbols:  | Concanavalin A-like lectin family
protein | chr1:2290201-2290977 FORWARD LENGTH=258
MGQAFYGFPIAFKNSTNSSNSFSFSTSFVFSIDAPGHGLTFLISPSMDFTQAMPSQYLGLFNTT
NNGNSTNRILAVEFDTVKSNEFLDIDDNHVGIDVNG
LVSVESAPAAFFSNKQSKNISLKLSSKDPIRAWIEYNGVERLLNVTLATLDTSKPNFPLLSRQM
NLSEIFMEKMYVGFSASTGNITSNHDVLGWSFSREG
KPQDFDLKLLPSLSTPSPSDFDDLDPIPDPPSDSATVKPNHTKTMIIICTLAIMIFMI
>AT1G21090.1 | Symbols:  | Cupredoxin superfamily protein |
chr1:7384854-7386199 FORWARD LENGTH=242
MGCSQKHLTSMLFFYFFCFLSLFSRPSLSATFLVDGVSVWKSPTVHTGDSVIFRHKYGYDLYIF
RNKDAFNVCNFTQATLLTKPNSTSFTWYPSRTGSYY
FSFTNNTSLPKTCQLNQKLTVQVILAAASPPSQPPATAPVPVSEGGVISSPSSYPWPLGPREGS
AFSPGPSPSEITSVTVPGKDGVPFINSNPAVPLPTG
DVDSTSINPLPTSTNSAHQVMMMTLTVKLGLCCVAMFLFLVL
>AT1G21880.2 | Symbols: LYM1 | lysm domain GPI-anchored protein
1 precursor | chr1:7680689-7682526 FORWARD LENGTH=416
MKIPEKPIFLIFVSLILASSLTFTATAKSTIEPCSSNDTCNALLGYTLYTDLKVSEVASLFQVD
PISILLANAIDISYPDVENHILPSKLFLKIPITCSC
VDGIRKSVSTHYKTRPSDNLGSIADSVYGGLVSAEQIQEANSVNDPSLLDVGTSLVIPLPCACF
NGTDNSLPAVYLSYVVKEIDTLVGIARRYSTTITDL
```

```
MNVNAMGAPDVSSGDILAVPLSACASKFPRYASDFGLIVPNGSYALAAGHCVQCSCALGSRNLY
CEPASLAVSCSSMQCRNSNLMLGNITVQQTSAGCNV
TTCDYNGIANGTILTMLTRSLQPRCPGPQQFAPLLAPPDTVPRDVMYAPAPSPDFDGPGSIASS
PRSSMLPGGGILPGNPANGPAGSISTASASSVSYFF
ITFLISIASFSLALSS
>AT1G61900.1 | Symbols:  | unknown protein; FUNCTIONS IN:
molecular_function unknown; INVOLVED IN: biological_process
unknown; LOCATED IN: anchored to plasma membrane, plasma
membrane, anchored to membrane; EXPRESSED IN: 23 plant
structures; EXPRESSED DURING: 13 growth stages; BEST Arabidopsis
thaliana protein match is: unknown protein (TAIR:AT2G30700.1);
Has 65 Blast hits to 65 proteins in 12 species: Archae - 0;
Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 65; Viruses - 0;
Other Eukaryotes - 0 (source: NCBI BLink). | chr1:22882508-
22884722 REVERSE LENGTH=433
MTRRAEFEMGLFVILQSMFLISLCSSQKPEEFLPEISPDTSPQPFLPFIAPSPMVPYINSTMPK
LSGLCSLNFSASESLIQTTSHNCWTVFAPLLANVMC
CPQLDATLTIILGKASKETGLLALNRTQSKHCLSDLEQILVGKGASGQLNKICSIHSSNLTSSS
CPVINVDEFESTVDTAKLLLACEKIDPVKECCEEAC
QNAILDAATNISLKASETLTDNSDRINDCKNVVNRWLATKLDPSRVKETLRGLANCKINRVCPL
VFPHMKHIGGNCSNELSNQTGCCRAMESYVSHLQKQ
TLITNLQALDCATSLGTKLQKLNITKNIFSVCHISLKDFSLQVGNQESGCLLPSLPSDAIFDKD
TGISFTCDLNDNIPAPWPSSSLSSASTCKKPVRIPA
LPAAASSQPRLHDEGVTRLVIFVLSMLLVMLLS
>AT1G63550.1 | Symbols:  | Receptor-like protein kinase-related
family protein | chr1:23569786-23570890 FORWARD LENGTH=324
MARIIITLTIPLFYFFFFSLLSHQTMSQPDHIFTVCNPTNNFTQTSSYETNRDTLLASLRESSS
LGHYSNATEGLSPDTVHGMFLCRGDITTASCVDCVQ
TATTEIASNCTLNKRAVIYYDECMVRYSNVSFSSELEIVPSITIYSLRSAPNPTRFNQTLTEKF
SELIFNVSSSSLVPYFVEDQERVTQSEGSYDLDTMV
QCSPDLDIFNCTVCLRVAFFRISTCCGLPSYAKIFTPKCLLRFQTSVLLSPPPSPSAPPPRSPP
PKSSPPSSLPQTPSPPLVFTPPQNVPNPSGSFSFNV
LKGNVIFGRIVVTMTALVFALVDL
>AT1G66970.2 | Symbols: SVL2 | SHV3-like 2 | chr1:24992746-
24996005 REVERSE LENGTH=785
MNSRPSNPTKLVIRSSTLLFCGVVLIHLFAAQIDAQRSTSRWQTLNGESCSHFISFFCALFPRK
QENLCDAPLVIARGGFSGLYPDSSIAAYQLATLTSV
ADVVLWCDLQLTKDGLGICFPDLNLANASTIDRVYPNREKSYSVNGVTTKGWFPNDFSLTELQN
FLLIRGILSRTDRFDGNGYLISTIEDVVTTLNREGF
WLNVQHDAFYEQQNLSMSSFLLSVSRTVSIDFISSPEVNFFKKITGSFGRNGPTFVFQFLGKED
FEPTTNRTYGSILSNLTFVKTFASGILVPKSYILPL
DDEQYLVPHTSLVQDAHKAGLQVYVSGFANDVDIAYNYSSDPVSEYLSFVDNGDFSVDGVLSDF
PITASAAVDCFSHIGRNATKQVDFLVISKDGASGDY
PGCTDLAYEKAIKDGADVIDCSVQMSSDGVPFCLRSIDLRNSIAALQNTFSNRSTSVPEISSVP
GIFTFSLTWPEIQSLTPAISNPFRVYRIFRNPREKN
SGKLISLSQFLDLAKTYTSLSGVLISVENAAYLREKQGLDVVQAVLDTLTEAGYSNGTTTTKVM
IQSTNSSVLVDFKKQSKYETVYKIEETIGNIRDSAI
```

```
EDIKKFANAVVINKDSVFPNSDSFLTGQTNVVERLQKSQLPVYVELFRNEFVSQAYDFFSDATV
EINAYIYGAGINGTITEFPFTAARYKRNRCLGREEV
PPYMLPVNPGGLLNVMSPLSLPPAQAPNQDFIEADVTEPPLSPVIAKAPTSTPGTPSTIAQAPS
GQTRLKLSLLLSVFFLSLLLL
>AT1G77630.1 | Symbols:  | Peptidoglycan-binding LysM domain-
containing protein | chr1:29173726-29175387 FORWARD LENGTH=423
MKNPEKPLLLFLILASSLASMATAKSTIEPCSSKDTCNSLLGYTLYTDLKVTEVASLFQVDPVS
MLLSNSIDISYPDVENHVLPAKLFLKIPITCSCVDG
IRKSLSTHYKTRTSDTLGSIADSVYGGLVSPEQIQVANSETDLSVLDVGTKLVIPLPCACFNGT
DESLPALYLSYVVRGIDTMAGIAKRFSTSVTDLTNV
NAMGAPDINPGDILAVPLLACSSNFPKYATDYGLIIPNGSYALTAGHCVQCSCVLGSRSMYCEP
ASISVSCSSMRCRNSNFMLGNITSQQSSSGCKLTTC
SYNGFASGTILTTLSMSLQPRCPGPQQLAPLIAPPDNVPKELMYLPSPSPSPSPEFDDIAGGGS
SIAAVPAASPGGATVSSSNSIPGNPANGPGGSISIA
SCPLSYYSFIALLIPIGSCFFVF
>AT2G44790.1 | Symbols: UCC2 | uclacyanin 2 | chr2:18462182-
18463232 REVERSE LENGTH=202
MAMNGLSKMAVAAATALLLVLTIVPGAVAVTYTIEWTTGVDYSGWATGKTFRVGDILEFKYGSS
HTVDVVDKAGYDGCDASSSTENHSDGDTKIDLKTVG
INYFICSTPGHCRTNGGMKLAVNVVAGSAGPPATPTPPSSTPGTPTTPESPPSGGSPTPTTPTP
GAGSTSPPPPPKASGASKGVMSYVLVGVSMVLGYGL
WM
>AT3G07390.1 | Symbols: AIR12 | auxin-responsive family protein
| chr3:2365452-2366273 FORWARD LENGTH=273
MSLCLKIPLIKHQTTPEQNSAMASSSSSLLILAVACFVSLISPAISQQACKSQNLNSAGPFDSC
EDLPVLNSYLHYTYNSSNSSLSVAFVATPSQANGGW
VAWAINPTGTKMAGSQAFLAYRSGGGAAPVVKTYNISSYSSLVEGKLAFDFWNLRAESLSGGRI
AIFTTVKVPAGADSVNQVWQIGGNVTNGRPGVHPFG
PDNLGSHRVLSFTEDAAPGSAPSPGSAPAPGTSGSTTPGTAAGGPGNAGSLTRNVNFGVNLGIL
VLLGSIFIF
>AT3G16860.1 | Symbols: COBL8 | COBRA-like protein 8 precursor |
chr3:5759643-5762104 REVERSE LENGTH=653
MGLTRNFILWILLSSLFTAIQLTSSQRNTPPPRSKDADLCNGVFVSYTYLTGTKIKPNDTKNQP
YRFESEITVLNNGRDELKSWQVFVKFAHREILVSAT
NAVLSDGSSLPASVENGTIFAGFPSADLKTAIMTAGDVTQMEARVELVGTQFGVAPPSVPLPKN
ITLVNDGWSCPKPTQQGSNVLQVCCTPNPNITTSKI
GQKFLPRQEGDLTIMYDVTRAYQSSYSAQVTIENHNLLGRLDNWDLSFMWMKDEFLFSTKGAYP
SVVDSSDCITGPQAKYYKDLDFSNVMSCARRPHIID
LPLTKYNDTNVGRIPYCCRNGTILPRSMDPEKSKSVFQIEVYKMPPDLNISSITPPQSWQIKGN
LNPDYKCGPPLRVSSSQFPDTSGLPSNKSAFASWQV
VCNITQPTPPKCCVSFSSYFNDSVIPCKTCACGGCSSDRVARTCSTTSPALPLPYQALLIPFDN
RTKLTNAWAVLKNRKVPDPLPCGDNCGVSINWHLAT
DYRGGWTARVTLFNWGDTDFVDWFTAVELRNAAPGFQKAYSFNGSIIAVNGKNTTVLMEGLPGL
NYLLAEKDGKNPSEDFRIPGKQQSVISFTKKLTPGI
KVGSKDGFPTKVLFNGQECSLPSVLPTSNSHRKHVSTFLLILTPFLALLFLRI
>AT3G20520.1 | Symbols: SVL3 | SHV3-like 3 | chr3:7162845-
7165742 FORWARD LENGTH=729
```

```
MACPRVIFLILITFFILQTAFSSSWQTLSGKPPAVIARGGFSGMFPDSSIQAYQLVNITTSPDV
MLWCDLQLTKDGVGICFPNLKLDNGSNVIRIDPHYK
ERFSVDFTWKELSDVKLAQGVVSRPYIFDDVSSILAIEEVAKLTASGLWLNIQDSAFYAKHNLS
MRNSVVSLSRRLKVNFISSPGISFLKSMKNSVKPTV
TKLIFRFLKQEHIEPFTNQSYGSLAKNLSYIRTFSSGILVPKSYIWPVDSALYLQPHTSLVTDA
HKEGLQVFASEFANDFVIAYNYSYDPTAEYLSFIDN
GNFSVDGFLSDFPVTPYRAINCFSHVDPKRAKEQAKITIISKNGASGDFPGCTDLAYQRAASDG
ADILDCNVQMSKDKIPFCMSSFDLINSTNVIETSFR
NLSSVVSEINPRRSGIYTFSLTMSQIQTLKPTISNLEKDSGLFRNPRNNKAGKFLTLSEFLFLP
NRYSSLLGLLIEVENAAYLVEHQGISVVDAVLDELK
RATTQQNKTSARTILIQSTDKSVLMKFKEKNKMNHDELVYRVDDNIRDVADSAIKDIKNFAGSI
VISKKSVFPYKGFIILEKETNIASKLKSNGLRVYVE
RFSNECVTHAFDFYDDPTLEIDSFVRDVQIDGIITDFPATTARYRKNKCYGEFGLTTTGELITF
ANPMLLPPAEAPYPALLDSDVTEPPLPEARSQPPAS
SPSKAEEKAIEVPFAFIAMAILVCFFISV
>AT3G51330.1 | Symbols:  | Eukaryotic aspartyl protease family
protein | chr3:19053480-19056152 REVERSE LENGTH=529
MVVARQVFVLLSLLVVCWGLERCEASGKFSFEVHHMFSDRVKQSLGLDDLVPEKGSLEYFKVLA
QRDRLIRGRGLASNNEETPITFMRGNRTISIDLLGF
LHYANVSVGTPATWFLVALDTGSDLFWLPCNCGSTCIRDLKEVGLSQSRPLNLYSPNTSSTSSS
IRCSDDRCFGSSRCSSPASSCPYQIQYLSKDTFTTG
TLFEDVLHLVTEDEGLEPVKANITLGCGKNQTGFLQSSAAVNGLLGLGLKDYSVPSILAKAKIT
ANSFSMCFGNIIDVVGRISFGDKGYTDQMETPLLPT
EPSPTYAVSVTEVSVGGDAVGVQLLALFDTGTSFTHLLEPEYGLITKAFDDHVTDKRRPIDPEL
PFEFCYDLSPNKTTILFPRVAMTFEGGSQMFLRNPL
FIVWNEDNSAMYCLGILKSVDFKINIIGQNFMSGYRIVFDRERMILGWKRSDCFEDESLESTTP
PPPETEAPSPSASTPLPSLLPPPAAATPPQIDPRNS
TRNSGTGTAANLVPLASQLLLLLPLLAFL
>AT3G51350.1 | Symbols:  | Eukaryotic aspartyl protease family
protein | chr3:19060485-19063248 REVERSE LENGTH=528
MDVARQVFVLLSVLVVCWGFERCEATGKFGFEVHHIFSDSVKQSLGLGDLVPEQGSLEYFKVLA
HRDRLIRGRGLASNNDETPITFDGGNLTVSVKLLGS
LYYANVSVGTPPSSFLVALDTGSDLFWLPCNCGTTCIRDLEDIGVPQSVPLNLYTPNASTTSSS
IRCSDKRCFGSKKCSSPSSICPYQISYSNSTGTKGT
LLQDVLHLATEDENLTPVKANVTLGCGQKQTGLFQRNNSVNGVLGLGIKGYSVPSLLAKANITA
NSFSMCFGRVIGNVGRISFGDRGYTDQEETPFISVA
PSTAYGVNISGVSVAGDPVDIRLFAKFDTGSSFTHLREPAYGVLTKSFDELVEDRRRPVDPELP
FEFCYDLSPNATTIQFPLVEMTFIGGSKIILNNPFF
TARTQEGNVMYCLGVLKSVGLKINVIGQNFVAGYRIVFDRERMILGWKQSLCFEDESLESTTPP
PPEVEAPAPSVSAPPPRSLPPTVSATPPPINPRNST
GNPGTGGAANLIPLASQLLLLLPLLAFL
>AT4G16120.1 | Symbols: ATSEB1, COBL7, SEB1 | COBRA-like
protein-7 precursor | chr4:9116591-9119138 REVERSE LENGTH=661
MDSAPNFIPRLLLLSLLIVSIPLTSSQSDANTTNPSPSPPSDSDLCNGVFVSYTHTKGSKIPPN
DTANQPYRFESVITVLNHGRDELKSWRVFVKFAHRE
ILVSASNAVLSDGSSLPVSVENGTVFAGYPSSDLKSAIQTAGDVTQMQARVELVGTQFGVAPPN
VPLPKNITLATDGWKCPKATQKGTNVLQVCCIPDPD
```

```
YDNREIIDNEFLPRKDGDLTIMYDVVRSYSSNYMAQVTMENHNPLGRLDNWKLSFDWMRDEFIY
TMKGAYPSIVDSSDCVDGPQAKHYQDLDFSNVLSCA
RRPTVIDLPPTKYNDSTFGLIPFCCRNGTILPRSMDPSKSSSVFQMQVYKMPPDLNISALSPPQ
NWRINGTLNPDYKCGPPVRVSPSQFVDPSGLPSNRT
AFASWQVVCNITQPKDASPRCCVSFSAYFNDSIVPCKTCACGCSSNKAARACSATAPSLLLPQQ
ALLVPFENRTELTVAWAYLKQRPVPNPMPCGDNCGV
SINWHLATDYRGGWTARVTVFNWGETDFVDWFTAVQMKNAAPGFEKAYSFNASTIGINGKNNTI
FMEGLPGLNYLVAERDGENPLKNPRIPGKQQSVMSF
TKKLTPGINVPGGDGFPSKVFFNGEECSLPTILPMRSSQHRKHISVFLLALPVLALLILRA
>AT4G25240.1 | Symbols: SKS1 | SKU5 similar 1 | chr4:12930539-
12933563 FORWARD LENGTH=589
MAATCSLLASFLLCFALLSAVSFAADPFVSYDFRVSYLTASPLGVPQQVIAVNGQFPGPLLNAT
TNYNVVVNVFNHLDEPLLLTWPGIQMRRNSWQDGVL
GTNCPIPPRWNFTYQFQVKDQIGSFFYSPSLNFQRASGGFGPIVINNRDIIPIPFPQPDGELIF
IIGDWYTQDHKALRRALDSGKELGMPDGVLINGKGP
YKYNSSVPDGIDYLTFHVEPGKTYRIRVHNVGISTSLNFRIQNHSLLLVETEGHYTSQANFTDF
DVHVGQSYSFLVTMDQDATSDYYIVASARFVNETVW
QRVTGVAILHYSNSKGPVSGPLPVPKTDVSSPWSAMSQPKTIRQNTSASGARPNPQGSFHYGQI
NITNTYILRSLPPTIINGALRATLNGISFVNPSTPV
RLADRNKVKGAYKLDFPDRPFNRPLRLDRSMINATYKGFIQVVFQNNDTKIQSFHVDGYSFFVV
GMDFGIWSEDKKGSYNNWDAISRSTIEVYPGGWTAV
LISLDNVGVWNIRVENLDRWYLGEETYMRITNPEEDGKTEMDPPDNVLYCGALKNLQKEQHHSA
ATSILNGHLKLMLLMVLLASVFRFC
>AT4G26690.1 | Symbols: SHV3, MRH5, GPDL2 | PLC-like
phosphodiesterase family protein | chr4:13456793-13459890
REVERSE LENGTH=759
MRGLLRASSLLLCGVILIQLLAAQIHAQSKKPKSPWPTLTGDPPLVIARGGFSGLFPDSSYDAY
NFAILTSVPDAVLWCDVQLTKDALGICFPDLTMRNS
SSIEAVYPTRQKSYPVNGVPTSGWFTIDFSLKDLKDVNLIRGILSRSEKFDGNSNPIMTVQSVS
TQMKPSFFWLNVQHDAFYAQHNLSMSSFLVAASKTV
LIDFISSPEVNFFKKIAGRFGRNGPSLVFRFLGQDEFEPTTNRTYGSILSNLTFVKTFASGILV
PKSYILPLDDQQYLLPHTSLVQDAHKAGLEVFVSGF
ANDIDIAHDYSFDPVSEYLSFVDNGNFSVDGVLSDFPITASASLDCFSHVGRNATKQVDFLVIT
KDGASGDYPGCTDLAYKKAIKDGADVIDCSVQLSSD
GTPFCLSSIDLGNSTTVSLTAFRNRSTTVPELGSLGAIYTFSLTWAEIQTLTPAISNPYRVTSL
FRNPKQKNAGKLFSLSDFLSLAKNSTSLSGVLISVE
NAAYLREEQGLDVVKAVLDTLTQTGYSNSTATKVMIQSTNSSVLVDFKKQSQYETVYKVEENIR
DILDSAIEDIKKFADAVVIQKLSVFPVAQSFITTQT
NVVEKLQKSQLPVYVELFQNEFLSQPYDFFADATVEINSYITGAGINGTITEFPFTAARYKRNL
CLGRKETIPYMAPAQPGALLTLVSPTAFPPAEAPNP
VFTDADVTEPPLPPVTAKAPTSSPGTPSTNAQAPSGQTRITLSLLLSVFAMVLASLLLL
>AT4G28100.1 | Symbols:  | unknown protein; FUNCTIONS IN:
molecular_function unknown; INVOLVED IN: biological_process
unknown; LOCATED IN: anchored to plasma membrane, anchored to
membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING:
14 growth stages; BEST Arabidopsis thaliana protein match is:
unknown protein (TAIR:AT3G18050.1); Has 30201 Blast hits to
17322 proteins in 780 species: Archae - 12; Bacteria - 1396;
```

Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink). | chr4:13965300-13966697 REVERSE LENGTH=304
MKKSLTLLILLLCSLLFSTVLSNLLVEPVQPNTVPAFPVETQAQSCRLDLSNELFGGVNEACGR
NLDRSRCCPVLAAWLFAAHARSALQLPAPAPTPESS
DPDEPMKPDDSQKCVNTLQSALLTKQIKIPQPNSSCDAILCFCGIRLHQISSLSCPAAFNVSSG
FKNATPTAAVKNLEKECRNSSYSGCTRCLGALQKLK
VRGGNKKTTTERGTKMMSKDCQLMGLTWLLARNKTAYIPTVSAVLRAIMYSPHPPHLNKCSPDQ
ENMPLAVDSLQFQKSFSSSSHLFGVLPFLPLVLCIF
LFLL
>AT5G07190.1 | Symbols: ATS3 | seed gene 3 | chr5:2237610-2238488 FORWARD LENGTH=213
MTFPSLSVSFLFFAFIFVTHAFDLSIIQMQQGTCPYTVVVMTSCLSPESTRDQISIVFGDADGN
KVYAPKLGGLVRGPGGLGKCSTNTFQVRGQCLNDPI
CSLYINRNGPDGWVPESIEIYSEGSKSVKFDFSKSVPQLNTWYGHNNCNTTGRPSSPDLPPPHF
PPEFPPETPTTPPPPPPRPSAASRLGNGESVFLAFA
IATAIAAMVRWSY
>AT5G41280.1 | Symbols: | Receptor-like protein kinase-related family protein | chr5:16509532-16510729 FORWARD LENGTH=286
METTKKLSVLLCLFFTMNQAISESDSDEHMATFCNDSSGNFTRNTTYNTNLNTLLSTLSNQSSF
ANYYNLTTGLGSDTVHGMFLCIGDVNRTTCNACVKN
ATIEIAKNCTNHREAIIYYFSCMVRYSDKFFLSTLETKPNTYWSSDDPIPKSYDKFGQRLSDKM
GEVIIRSSLLSSSFTPYYLMDTTTFDNLYDLESVVQ
CSPHLDPKNCTTCLKLALQELTQCCGDQLWAFIFTPKCLVSFDTSNSSSLPPLPPPSRSGSFSI
RGNNKILVGMILAVSVFAFLGL
>AT5G41290.1 | Symbols: | Receptor-like protein kinase-related family protein | chr5:16512326-16513500 FORWARD LENGTH=287
METTKKLFALLCLFVTMNQAISVSDPDDMETFCMKSSRNTTSNTTYNKNLNTLLSTLSNQSSFA
NYYNLTTGLASDTVHGMFLCTGDVNRTTCNACVKNA
TIEIAKNCTNHREAIIYNVDCMVRYSDKFFLTTLETNPSYWWSSNDLIPKSFGKFGQRLSDKMG
EVIVRSSLLSSSFTPYYLMDTTRFDNLYDLESIVQC
TPDLDPRNCTTCLKLALQELTECCGNQVWAFIYTPNCMVSFDTYNSSLPPLPPPSRSGSFSHRG
NNKLLGGMVLAVSVSVFAFLSLV
>AT5G41300.1 | Symbols: | Receptor-like protein kinase-related family protein | chr5:16515004-16516102 FORWARD LENGTH=287
METTKKLSPIFCFSSLLCLFFTMNQAVSETDHMDTFCIDSSRNTTGNTTYNKNLNTMLSTFRNQ
SSIVNNYNLTTGLASDTVYGMFLCTGDVNITTCNNC
VKNATIEIVKNCTNHREAIIYYIDCMVRYSDKFFLSTFEKKPNSIWSGDDPIPKSLGPFKKRLY
KKMGEAIVRSSTLSSALTPYYYLDVTRFDGSYDLDS
LVQCSPHLNPENCTICLEYALQEIIDCCSDKFWAMIFTPNCFVNYMITTSPLPPLPSPYHHSGS
CSIRGNSEIFWGMIILAALVFTF
>AT5G42720.1 | Symbols: | Glycosyl hydrolase family 17 protein | chr5:17130535-17132763 FORWARD LENGTH=438
MRASVYSLILLFFSCLLHLSKSQPFLGVNYGLTADNLPPPSASAKLLQSTTFQKVRLYGSDPAV
IKALANTGIEIVIGASNGDVPGLASDPSFARSWVET
NVVPYYPASKIVLIAVGNEITSFGDNSLMSQLLPAMKNVQTALEAASLGGGKIKVSTVHIMSVL
AGSDPPSTAVFKPEHADILKGLLEFNSETGSPFAVN

```
PYPFFAYQDDRRPETLAYCLFQANPGRVDPNSNLKYMNMFDAQVDAVYSALNSMGFKDVEIMVA
ETGWPYKGDPEEAGATVENARAYNKNLIAHLKSGSG
TPLMPGRVIDTYLFALYDENLKPGKGSERAFGLFRPDLTMTYDIGLTKTTNYNQTSMAPLSPTR
PRLPPAAAPTRQTLPSPPQMILPSPVTPSDKNSGQT
DVHNSTPRSASLAHICRSLSISASMFFVSVLYALIILL
>AT5G49270.1 | Symbols: SHV2, DER9, MRH4, COBL9 | COBRA-like
extracellular glycosyl-phosphatidyl inositol-anchored protein
family | chr5:19972021-19974398 REVERSE LENGTH=663
MGVLLPIFFGVLLLFTVTPPSMSQLPPTIMVPAPAPAPISPSDLCNGIFLSYDFILGRKIPPND
TADQPYRFESVLTVLNNGREELKEWRVFVGFQHNEI
LISATDALIVNGTELPALVGNGTIFGGYPVSDLKTAIQTAGDLKQMTAEIELVGTQFMVAPPAV
PLPSNISLVNEGWLCPVPTLQSKRELTTCCIRDASI
IVNTTITTKFLPRQPGDLTIMYDVIRAYDQNYLTEVTMENHNPLGRLDHWELSFDWMRDEFIQK
MQGAYPTVVDATKCIFGPQSLIYTGLDFADVLTCER
RPIIIDLPPTKKDDSTLGNIPSCCRNGTILPRIMDPSKSVSVFTMQVAKMPPDFNRSALFPPQN
WRIKGTLNPDYSCGPPVRVTPTFYPDPSGMPTNKSS
FASWQIVCNITQAKTEIPKCCVSFSAFFNDSIIPCNTCACGCVSETRRTCSAETPSLLIPPDAL
LLPFENRTALTLAWNALKHKTLPNPMPCGDNCGVSI
NWHMASDYRGGWTVRITIFNWGEIDFPNWFLAVQMKKPALLGFEKAYSFNASLLSVDGGVNNTI
FMEGLPGLDYLVAEADEKDPKKKNIRIPGKQQSVIQ
FSKKLTPGINVAERDGFPAKVIFNGEECLLPDLLPMASGGRRNGAITVLSFITFYVAAFMVLL
>AT5G55480.1 | Symbols: SVL1 | SHV3-like 1 | chr5:22474277-
22477819 FORWARD LENGTH=766
MINMRDNPTMHVLQASKFLFLALILIQLLSTQLFAQRSKSPWQTLTGDAPLVIARGGFSGLLPD
SSLDAYSFVSQTSVPGAVLWCDVQLTKDAIGLCFPD
VKMMNASNIQDVYPKRKTSYLLNGVPTQDWFTIDFNFKDLTKVILKQGILSRSAAFDGNSYGIS
TVKDISTQLKPEGFWLNVQHDAFYAQHNLSMSSFLL
SISKTVIIDYLSSPEVNFFRNIGRRFGRNGPKFVFRFLEKDDVEVSTNQTYGSLAGNLTFLKTF
ASGVLVPKSYIWPIESQYLLPRTSFVQDAHKAGLEV
YASGFGNDFDLAYNYSFDPLAEYLSFMDNGDFSVDGLLSDFPLTASSAVDCFSHLGSNASSQVD
FLVISKNGASGDYPGCTDLAYTKAIKDGADVIDCSL
QMSSDGIPFCLSSINLGESTNVVQSPFRNRSTTVPEIGSLPGIYSFSLAWSEIQTLRPAIENPY
SREFTMFRNPRERSSGKFVSLSDFLNLAKNSSSLTG
VLISVENATYLREKQGLDAVKAVLDTLTEAGYSNKTTTTRVMIQSTNSSVLIDFKKQSRYETVY
KVEETIRDILDTAIEDIKKFADAVVISKKSVFPTSE
SFTTGQTKLVERLQKFQLPVYVEVFRNEFVSQPWDFFADATVEINSHVTGAGINGTITEFPLTA
ARYKRNSCLTRKDVPPYMIPVQPAGLLTIVSPASLP
PAEAPSPVFTDADVTEPPLPPVSARAPTTTPGPQSTGEKSPNGQTRVALSLLLSAFATVFASLL
LL
>AT5G58050.1 | Symbols: SVL4 | SHV3-like 4 | chr5:23494498-
23497386 REVERSE LENGTH=753
MLRFFILFSLFLHSSVAAPKTPAAAAAVPAKKWLTLNGQEPAVVARGGFSGLFPESSISANDLA
IGTSSPGFTMLCNLQMTKDGVGLCLSDIRLDNATTI
SSVFPKAQKTYKVNGQDLKGWFVIDYDADTIFNKVTLVQNIFSRPSIFDGQMSVSAVEDVLGTK
PPKFWLSVQYDAFYMEHKLSPAEYLRSLRFRGINVI
SSPEIGFLKSIGMDAGRAKTKLIFEFKDPEAVEPTTNKKYSEIQQNLAAIKAFASGVLVPKDYI
WPIDSAKYLKPATTFVADAHKAGLEVYASGFANDLR
```

```
TSFNYSYDPSAEYLQFVDNGQFSVDGVITDFPPTASQSITCFSHQNGNLPKAGHALVITHNGAS
GDYPGCTDLAYQKAIDDGADIIDCSVQMSKDGIAFC
HDAADLSASTTARTTFMSRATSVPEIQPTNGIFSFDLTWAEIQSVKPQIENPFTATGFQRNPAN
KNAGKFTTLADFLELGKAKAVTGVLINIQNAAYLAS
KKGLGVVDVVKSALTNSTLDKQSTQKVLIQSDDSSVLSSFEAVPPYTRVLSIDKEIGDAPKTSI
EEIKKHADAVNLLRTSLITVSQSFATGKTNVVEEMH
KANISVYVSVLRNEYIAIAFDYFSDPTIELATFIAGRGVDGVITEFPATATRYLRSPCSDLNKD
QPYAILPADAGALLTVADKEAQLPAIPPNPPLDAKD
VIDPPLPPVAKLASNGTEGGPPQTPPRSGTVAIAANLSLSLLAMMALGLLYTA
>AT5G58170.1 | Symbols: SVL5 | SHV3-like 5 | chr5:23540261-
23543092 REVERSE LENGTH=750
MLRFIIFFSLFIHLCVAAPQTPAAAAAVPAKKWLTLNGQEPAVVARGGFSGLFPESSASANDLA
IGTSSPGLTMLCNLQMTKDGVGLCLSDIILDNATTI
SSVFPKAQKTYKVNGQDLKGWFVLDYDADTIFNNVTLVQNIFSRPSIFDGQMSVSAVEDVLGTK
PPKFWLSVQYDAFYMEHKLSPAEYLRSLQFRGINVI
SSPEIGFLKSIGMDAGRAKTKLIFEFKDPEAVEPTTNKKYSEIQQNLAAIKAFASGVLVPKDYI
WPIDSAKYLKPATTFVADAHKAGLEVYASGFANDLR
TSFNYSYDPSAEYLQFVDNGQFSVDGVITDFPPTASQSITCFSHQNGNLPKAGHALVITHNGAS
GDYPGCTDLAYQKAVDDGADVIDCSVQMSKDGIAFC
HDAADLTASTTAMTIFMSRATSVPEIQPTNGIFSFDLTWAEIQSVKPQIENPFTATGFQRNPAN
KNAGKFITLADFLDFSKAKAVTGVMINIENAAYLAS
KKGLGVVDAVKSALAKSTLDKQSTQKVLIQSDDSSVLASFEAVPPYTRVLSIDKEIGGAPKPSV
DEIKKYAEAVNLLRTSLVTVSQSFTTGKTNVVEEMH
KGNISVYVSVLRNEYISVAFDYFSDPTIELATFISGSGVDGVITEFPATATRYLKSPCSDLNKE
QPYAILPAEAGGLVVVADKEAQPPASAPNPPLEAKD
VIDPPLPPVANLAASNATGGAQSHPPPASGTVANAANLGLSLLAMLALGV
>AT5G62210.1 | Symbols:  | Embryo-specific protein 3, (ATS3) |
chr5:24986521-24987477 REVERSE LENGTH=223
MECSLSSILLLLVLCFTHPIPSLTFEPLISSLDLHEEESCPYTVIVTTSCFSPDWSRDQVTIAL
GDADDNQVVAPRLDKPLSGGGGFEKCSSDTFQVKGK
CLNTICSVYIYRSGTDGWIPETVEIYKEGSKSVKFDFNKNVPENIWYGNNYCNNTDLPPPSPDF
PPFSPSIPPPSPPYFPPEPPSIPPPPPPSPPSAASG
RGSGQSLVVAFAAFVFAFAAVVV
```

Appendix II. *Arabidopsis thaliana* AGP protein sequences. Non-contiguous hydroxyproline motifs are bolded and highlighted in yellow, contiguous proline motifs are underlined, and transmembrane domains are highlighted in blue.

>AT1G03820.1 | Symbols: | unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; Has 1345 Blast hits to 1122 proteins in 102 species: Archae - 2; Bacteria - 28; Metazoa - 28; Fungi - 30; Plants - 109; Viruses - 0; Other Eukaryotes - 1148 (source: NCBI BLink). | chr1:960008-960676 REVERSE LENGTH=222
MASFALKPIFCFIAVFCFIVHNVEAREGKLFFSKFTHIDRPNNKDVAL**SPAPAP**GLAQANGRLGNGSFGPGSGMIPQTKESW**PS**SSTTTDEEFEKLMATFDEEKNTKLPEAFEEEEESEDSEDLNEPKDKYNNNNNNNGYTYTTNNYNDNGRGYGNEEEKQGMSDTRVMENGKYFYDTRGRNSEN**TPS**RGYENARGNDHTNEFETMEEYYKSLEGSQEEYEP
>AT1G03870.1 | Symbols: FLA9 | FASCICLIN-like arabinoogalactan 9 | chr1:982625-983368 REVERSE LENGTH=247
MATTRLTL**AP**LLLIAAVLLATKATAQ**PAAPAP**E**PA**GPINLTAILEKGGQFTTFIHLLNITQVGSQVNIQVNSSSEGMTVF**APT**DNAFQNLKPGTLNQL**SP**DDQVKLILYHV**SP**KYYSMDDLLSVSNPVRTQASGRDNGVYGLNFTGQTNQINVSTGYVETRISNSLRQQRPLAVYVVDMVLLPGEMFGEHKL**SP**I**APAP**KSKSGGVTDDSGSTKKAA**SPS**DKSGSGEKKVGLGFGLGLIVLCLKFLF
>AT1G15190.1 | Symbols: | Fasciclin-like arabinogalactan family protein | chr1:5227275-5228021 FORWARD LENGTH=248
MAKISSASCFRAIFLGALIIILCLPH**PS**TGVPLEELERAIAILRVRGRALFANAIITSDLLFDLLSDESLTLF**APT**DSMLFDLDMTHSLPFYVSTLRLHSVPLRLSLSGLRSLPNSSSL**PT**LL**PS**HRLLLTKHSSSNDSIFLDGVQLLIPGLFDGQHIAVHGLADLLPLT**APSSP**NRLVEDSTALAK**SP**WFLGSRF**SPAP**EPYFAFMDL**SPA**E**SPS**VEEV**SPSPS**WGEGEEDFIVGDEGGPLDGRNNGF
>AT1G28290.1 | Symbols: AGP31 | arabinogalactan protein 31 | chr1:9889331-9890843 REVERSE LENGTH=359
MGFIGKSVLVSLVALWCFTSSVFTEEVNHKTQ**TPS**L**APAPAP**YHHGHHHPHPPHHHHPHPHPHPHP**PA**K**SP**VKPPVK**AP**V**SPPA**KPPVKPPVYP**PT**K**AP**VKP**PT**KPPVKPPV**SPPA**KPPVKPPVYP**PT**K**AP**VKP**PT**KPPVKPPVYP**PT**K**AP**VKP**PT**KPPVKPPVYP**PT**K**AP**VKP**PT**KPPVKPPV**SPPA**KPPVKPPVYP**PT**K**AP**VKPPV**SPPT**KPPV**TP**PVYPPKFNRSLVAVRGTVYCKSCKYAAFNTLLGAKPIEGATVKLVCKSKKNITAETTTDKNGYFLLL**AP**KTVTNFGFRGCRVYLVKSKDYKCSKVSKLFGGDVGAELKPEKKLGKSTVVVNKLVYGLFNVGPFAFN**PS**CPK
>AT1G31250.1 | Symbols: | proline-rich family protein | chr1:11166658-11167500 REVERSE LENGTH=165
MANQYFRMAFLLCLFLSLSYQSIRIEAREGKACIGNCYGSS**AP**PVPPNASLSIPPNS**SPS**VKL**TP**PYA**SPS**VKL**TP**PYA**SPS**VR**PA**GT**TP**NA**SPS**VKL**TP**PYA**SPS**MR**PA**G**TP**NA**SPS**VKL**TP**PYA**SPS**VR**PT**GT**TP**NA**SPS**L**TP**PN**PSPS**EKFIPPNA**SP**FIHT
>AT1G35230.1 | Symbols: AGP5 | arabinogalactan protein 5 | chr1:12917184-12917585 FORWARD LENGTH=133
MASKSVVVFLFLALVASSVVAQ**AP**G**PAPT**I**SP**L**PATPTPS**Q**SP**RAT**APAPSPS**ANPP**PSAPT**T**AP**PVSQP**PT**E**SPPAPPT**ST**SPS**G**AP**GTNV**PS**GEAG**PA**Q**SP**LSG**SP**NAAAVSRVSLVGTFAGVAVIAALLL
>AT1G36150.1 | Symbols: | Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein |

chr1:13528257-13529470 FORWARD LENGTH=256
MK**PS**FVLLSIVLLLSSSLSDAADFG**SPS**QP**PS**M**APTP**Q**PS**NSTDCSSVIYSMVDCLSFLTVGST
D**PSPT**KTCCVGVKTVLNY**SP**KCLCSALESSREMGFVLDDTKALAMPKICNVPIDPNCDVS**TPA**A
STPV**SP**PVE**SPT**T**SPS**SAK**SPA**I**TPSSPA**VSH**SP**PPVRHS**SP**PVSHS**SP**PVSHS**SPPT**SRS**SPA**
VSHS**SP**VVAAS**SPV**KAVSSSTAS**SP**RAA**SPSPSPSPS**ISSSGILLVSKLFIAVVMVSSFLYILA
>AT1G51915.1 | Symbols: | cryptdin protein-related |
chr1:19293587-19293865 FORWARD LENGTH=67
MATERFSTMLISVLVLALVL**SP**ILPCQATRAHLDAETRMLRRVC**PS**CVCC**APAP**RGACCPCRCP
KNP
>AT1G68725.1 | Symbols: AGP19, ATAGP19 | arabinogalactan protein
19 | chr1:25809298-25810130 FORWARD LENGTH=248
MESNSIIWSLLLASALISSFSVNAQG**PA**A**SP**VTSTTT**AP**P**PT**TA**APPT**TA**AP**P**PT**TT**TP**PVSAA
QP**PASPV**T**P**PPAV**TPTSPPAP**KV**APV**I**SPATP**PPQPPQ**SPPA**S**APT**V**SP**PPV**SP**P**PAPTSP**P**PT**
PASPPPAPASPPPAPASPPPAPVSPPPVQ**APSP**ISLP**PAPAPAPT**KHKRKHKHKRHHH**APAPAP**
IP**PSPPSP**PVLTDPQDT**APAPSP**NTNGGNALNQLKGRAVMWLNTGLVILFLLAMTA
>AT2G04780.1 | Symbols: FLA7 | FASCICLIN-like arabinoogalactan 7
| chr2:1677488-1678252 FORWARD LENGTH=254
MAKMQLSIFIAVVALIVCSASAKTA**SPPAPV**LP**PTPAPAPAP**ENVNLTELLSVAGPFHTFLDYL
LSTGVIETFQNQANNTEEGITIFVPKDDAFKAQKNPPLSNLTKDQLKQLVLFHALPHYYSLSEF
KNLSQSGPVSTFAGGQYSLKFTDVSGTVRIDSLWTRTKVSSSVFSTDPVAVYQVNRVLLPEAIF
GTDVPPM**PAPAPAP**IVS**APS**D**SPS**VADSEGAS**SP**KSSHKNSGQKLLL**AP**ISMVISGLVALFL
>AT2G14890.1 | Symbols: AGP9 | arabinogalactan protein 9 |
chr2:6399679-6400755 FORWARD LENGTH=191
MARSFAIAVICIVLIAGVTGQ**APTSPPTATPAPPTPTTP**PPAA**TP**PPVS**AP**PPVTT**SP**PPVTT**A**
PPPANPPPPVS**SPPPASPPATP**PPVA**SP**PPPVA**SPPPATP**PPVA**TPPPAP**LA**SPP**AQ**VPAPAP**
TTKPD**SPSPSPS**S**SP**PL**PS**SD**AP**G**PS**TDSI**SPAPSPT**DVNDQNGASKMVSSLVFGSVLVWFMI
>AT2G20520.1 | Symbols: FLA6 | FASCICLIN-like arabinogalactan 6
| chr2:8840663-8841406 FORWARD LENGTH=247
MSSSLFSYVVLLIFLFTIPYIQSQ**PTAPAPT**TEK**SP**INLTAILEAGHQFTTLIQLLNTTQVGFQ
VSVQLNSSDQGMTIF**APT**DNAFNKLKPGTLNSLTYQQQIQLMLYHIIPKYYSLSDLLLASNPVR
TQATGQDGGVFGLNFTGQAQSNQVNVSTGVVETRINNALRQQFPLAVYVVDSVLLPEELFGTKT
TPTG**APAP**KSSTSSSDAD**SPA**ADDEHKSAGSSVKRTSLGIVVSFALFCCSVIYIA
>AT2G22470.1 | Symbols: AGP2, ATAGP2 | arabinogalactan protein 2
| chr2:9538400-9538795 REVERSE LENGTH=131
MNSKAMQALIFLGFLATSCLAQ**APAPAPT**TV**TP**P**PT**ALPPVTAE**TPSP**IA**SP**P**VPV**NE**PTPAPT**
T**SPT**T**SP**VA**SP**PQTD**APAP**G**PS**AGL**TPT**S**SPAP**GPDGAAD**APS**AAWANKAFLVGTAVAGALYAV
VLA
>AT2G23130.1 | Symbols: AGP17, ATAGP17 | arabinogalactan protein
17 | chr2:9844411-9845186 FORWARD LENGTH=185
MTRNILLTVTLICIVFITVGGQ**SPA**T**AP**IH**SPS**T**SP**HKPK**PTSPA**I**SPAAPTP**ESTE**APA**K**TP**V
E**AP**VE**APPSPTPA**S**TP**QI**SPPAPSP**EAD**TPSAP**EI**APS**AD**VPAPA**LTKHKKKTKKHKT**APAP**G**P**
ASELL**SPPAP**PGE**AP**GPG**PS**DAF**SPA**ADDQSGAQRISVVIQMVGAAAIAWSLLVLAF
>AT2G23990.2 | Symbols: ENODL11, AtENODL11 | early nodulin-like
protein 11 | chr2:10206835-10207911 REVERSE LENGTH=226
MVSLISIVSVVFLLFTTFYHFGEARIINVGGSLDAWKVPE**SP**NHSLNHWAESVRFQVGDALCSF
VMMVKIRMLVIVGYTFMFKYDSKIDSVLQVTKENYEKCNTQKPLEEHKDGYTTVKLDVSGPYYF
ISG**APS**GNCAKGEKVTVVVQ**SP**NHPKPG**PA**AV**TPT**LPPK**PS**T**TPAAPAPAPPTPSP**KSSTSTM**A**
PAPAPAKSSAVGLVAGNGIFWASTLVAVIGLAFA

>AT2G24450.1 | Symbols: FLA3 | FASCICLIN-like arabinogalactan protein 3 precursor | chr2:10393019-10393861 REVERSE LENGTH=280
MGLKVSSSLLCLTILLAVSSIVSAVNITRVLEKYPEFSTMTELLAKTEL**TP**IINKRQTITVLAL
NNDAIGSISGRPEEEVKNILMNHVVLDYFDELKLKALKEKSTLLTTLYQSTGLGQQQNGFLNCT
KSNGKIYFGSGVKG**AP**QTAEYITTVFRNPYNLSVVQISMPIV**AP**GLG**SP**VKVPPPPPMS**SPPAP**
SPKKGAA**TPAPAPA**DEGDYAD**AP**PGL**AP**ET**APA**S**APS**ESD**SPAPAP**DKSGKKKMAAADEAEP**PS**
SASNTGLSFGAVLVLGFVASFVGF
>AT2G25060.1 | Symbols: ENODL14, AtENODL14 | early nodulin-like protein 14 | chr2:10662308-10662930 FORWARD LENGTH=182
MFLSASMASSSLHVAIFSLIFLFSLAAANEVTVGGKSGDWKIP**PS**SSYSFTEWAQKARFKVGDF
IVFRYESGKDSVLEVTKEAYNSCNTTNPLANYTDGETKVKLDRSGPFYFISGANGHCEKGQKLS
LVVI**SP**RHSVI**SPAPSP**VEFEDG**PA**L**APAP**ISGSVRLGGCYVVLGLVLGLCAWF
>AT2G26720.1 | Symbols: | Cupredoxin superfamily protein | chr2:11384782-11385402 FORWARD LENGTH=206
MALIKNNIFFTSLLIFVTLFGVAVGGTVHKVGNTKGWTMIGGDYEAWASSRVFQVGDTLVFAYN
KDYHDVTEVTHNDFEMCESSKPLRRYKTGSDSISLTKPGLQHFICGVPGHCKKGQKLQIHVL**PA**
SLGHVAVPVPGPVRSQSSSS**SPSPSP**LVDPPVNN**AP**QYQMG**PTPA**SHSAASADFIFTFSFDLTL
IDLCTFFILFFILV
>AT2G28440.1 | Symbols: | proline-rich family protein | chr2:12161226-12162032 FORWARD LENGTH=268
MAKKLCFIVMLSICLLIFDFAGAQEE**SPSPA**AV**SP**GRE**PS**TD**SP**L**SPS**S**SP**EED**SP**L**SPS**S**SP**E
ED**SP**LP**PS**S**SP**EED**SP**L**APS**S**SP**EVD**SP**L**APS**S**SP**EVD**SP**QP**PS**S**SP**EAD**SP**LP**PS**S**SP**EAN**SP**
Q**SPA**S**SP**KPESLAD**SPSP**PPPPPQPE**SPSSPS**YPE**PAPVPAPS**DDDSDDDPEPETEYF**PSPAPS**
PELGMAQDIKASDAAGEELNDERGEDYGMSGLEKAGIAIGTILGVGAIVIGALVYKKRRDNMTR
ARYTYFTEGEFL
>AT2G31050.1 | Symbols: | Cupredoxin superfamily protein | chr2:13212150-13212752 FORWARD LENGTH=200
MALIKSNAFFTSLLILVALFGISVGGTVHKVGDSDGWTIMSVNYETWASTITFQVGDSLVFKYN
KDFHDVTEVTHNDYEMCE**PS**KPLARYETGSDIVILTKPGLQHFICGFPGHCDMGQKLQIHVL**PA**
SLGPVA**APVP**GPVRP**PS**SFS**SPS**Q**SP**LAE**SP**VNH**AP**VQYQMG**PSPAP**HSAASNSNVWIGLCFLP
LLSLLILV
>AT2G32300.1 | Symbols: UCC1 | uclacyanin 1 | chr2:13722510-13723464 FORWARD LENGTH=261
MASREMLIIISVLATTLIGLTVATDHTIGG**PS**GWTVGASLRTWAAGQTFAVGDNLVFSY**PA**AFH
DVVEVTKPEFDSCQAVKPLITFANGNSLVPLT**TP**GKRYFICGMPGHCSQGMKLEVNVV**PT**ATV**A**
PTAPLPNT**VPS**LN**APSPS**SVLPIQPLLPLNPVPVL**SPS**SS**TP**L**PS**SSLPLIPPL**SPA**L**SPA**TAA
GTSLPLFPG**SP**GSSSSTTSTKTVGTF**PS**STTGTTADLAGAD**SPPA**DSSSAAKTLVLGFGFMVAM
MLHLF
>AT2G33790.1 | Symbols: ATAGP30, AGP30 | arabinogalactan protein 30 | chr2:14293862-14295237 REVERSE LENGTH=239
MGIIGKSVSLTLFALLCFTSSVFTLGVNQPGSSDPFHSLPQHLPLPPIKL**PT**LP**PA**K**AP**IKL**PA**
YP**PA**K**AP**IKL**PT**LP**PA**K**AP**IKL**PT**LPPIKPPVLPPVYPPKYNKTLVAVRGVVYCKACKYAGVNN
VQGAKPVKDAVVRLVCKNKKNSISETKTDKNGYFMLL**AP**KTVTNYDIKGCRAFLVK**SP**DTKCSK
VSSLHDGGKGSVLKPVLKPGFSSTIMRWFKYSVYNVGPFAFE**PT**CPK
>AT2G35860.1 | Symbols: FLA16 | FASCICLIN-like arabinogalactan protein 16 precursor | chr2:15059859-15061810 FORWARD LENGTH=445
MDSSYGATKFLLLLFLTTSIATALPDNKPVPGQINSNSVLVALLDSHYTELAELVEKALLLQTL
EEAVGKHNITIF**AP**RNDALERNLDPLFKSFLLEPRNLKSLQSLLMFHILPKRIT**SP**QW**PS**LSHH

HRTLSNDHLHLTVDVNTLKVDSAEIIRPDDVIRPDGIIHGIERLLIPRSVQEDFNRRRSLRSIS
AVIPEG**AP**EVDPRTHRLKK**PSPA**V**PA**G**AP**PVLPIYDAM**SP**G**PS**L**APAPAP**GPGGPRGHFNGDAQ
VKDFIHTLLHYGGYNEMADILVNLTSLATEMGRLVSEGYVLTVL**AP**NDEAMAKLTTDQLSEPG**A**
PEQIMYYHIIPEYQTEESMYNAVRRFGKVKYDSLRFPHKVLAQEADGSVKFGHGDGSAYLFDPD
IYTDGRISVQGIDGVLFPKEE**TPA**TEIK**PAAP**VVKKVSKSRRGKLMEVACRMMGSRFIPCQ
>AT2G41905.1 | Symbols: | BEST Arabidopsis thaliana protein
match is: arabinogalactan protein 23 (TAIR:AT3G57690.1); Has
35333 Blast hits to 34131 proteins in 2444 species: Archae -
798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531;
Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink). |
chr2:17495766-17495951 FORWARD LENGTH=61
MEMKKIACGVVFAAASMTAVMAAEVG**APAP**G**PA**ASGASVAV**PA**LGSLVGASLVSLFAYYLN
>AT2G44790.1 | Symbols: UCC2 | uclacyanin 2 | chr2:18462182-
18463232 REVERSE LENGTH=202
MAMNGLSKMAVAAATALLLVLTIVPGAVAVTYTIEWTTGVDYSGWATGKTFRVGDILEFKYGSS
HTVDVVDKAGYDGCDASSSTENHSDGDTKIDLKTVGINYFICS**TP**GHCRTNGGMKLAVNVVAGS
AGP**PATPTPPS**S**TP**G**TPTTP**E**SPPS**GG**SPTPTTPTP**GAGST**SP**PPPPKASGASKGVMSYVLVGV
SMVLGYGLWM
>AT2G45470.1 | Symbols: FLA8, AGP8 | FASCICLIN-like
arabinogalactan protein 8 | chr2:18742797-18744059 REVERSE
LENGTH=420
MAASQTFSLLAFTFSLLAFASTVSSHNITQILAD**SP**DYSSFNSYLSQTKLADEINSRTTITVLV
LNNGAMSALAGKHPLSVIKSALSLLVLLDYYDPQKLHKISKGTTLSTTLYQTTGN**AP**GNLGFVN
ITDLKGGKVGFGSAASGSKLDSSYTKSVKQIPYNISILEID**AP**II**AP**GVLT**APAPS**ASLSNITG
LLEKAGCKTFANLLVSSGVLKTYESAVEKGLTVF**APS**DEAFKAEGVPDLTKLTQAEVVSLLEYH
ALAEYKPKGSLKTNKNNISTLATNGAGKFDLTTSTSGDEVILHTGV**APS**RLADTVLDA**TP**VVIF
TVDNVLL**PA**ELFGKSK**SPSPAPAP**EPVT**APTPSPA**D**APSPT**AA**SPPAPPT**DE**SP**ES**APS**D**SPT**G
SANSKSANAAVGVS**TPS**LFTALVTIAAIAVSVSLCS
>AT2G46330.1 | Symbols: AGP16, ATAGP16 | arabinogalactan protein
16 | chr2:19018730-19019108 REVERSE LENGTH=73
MASRNSVTGFALFSFVFAVILSLAGAQSL**APAPAPT**SDGTSIDQGIAYLLMVVALVLTYLIHPL
DASSSYSFF
>AT2G47930.1 | Symbols: AGP26, ATAGP26 | arabinogalactan protein
26 | chr2:19617219-19617629 REVERSE LENGTH=136
MSVSLFTAFTVLSLCLHTSTSEFQLSTISA**APS**FLPE**APS**SFSAS**TPA**M**SP**DT**SP**LF**PTP**GSSE
M**SPSPS**ESSIM**PT**I**PS**SL**SP**PNPDAV**TP**DPLLEV**SP**VG**SP**L**PA**SSSVCLVSSQLSSLLLVLLML
LLAFCSFF
>AT3G01700.1 | Symbols: AGP11, ATAGP11 | arabinogalactan protein
11 | chr3:258981-259391 FORWARD LENGTH=136
MARLFVVVALLALAVGTVFAAD**APS**A**APT**A**SPT**K**SPT**K**APA**A**AP**KSSAA**AP**KAS**SP**VAEE**PTP**E
DDYSAA**SPS**DSAE**APT**VS**SPPAPTP**EADG**PS**SDG**PS**SDG**PA**AAE**SP**KSGATTNVKLSIAGTVAA
AGFFIFSL
>AT3G01730.1 | Symbols: | unknown protein; Has 8 Blast hits to 8
proteins in 2 species: Archae - 0; Bacteria - 0; Metazoa - 0;
Fungi - 0; Plants - 8; Viruses - 0; Other Eukaryotes - 0
(source: NCBI BLink). | chr3:266628-266891 FORWARD LENGTH=87
MRAFSLMICVLVLATLIMVAESKSSYSGNDETL**TP**PPPM**APSPSPT**GTDGGHFGEPKTLLPPPN
AATFTTCPLLITSTLISALAFVF

>AT3G06360.1 | Symbols: AGP27, ATAGP27 | arabinogalactan protein 27 | chr3:1928976-1929353 FORWARD LENGTH=125
MASSILLTLITFIFLSSLSLS**SPT**TNTI**PS**SQTI**SPS**EEKI**SP**EI**AP**LL**PSPA**VSSTQTI**PS**SS
TLPEPENDDVSADPD**PA**F**APS**A**SPPA**SSLASLSSQ**AP**GVFIYFVFAAVYCFSLRLLAVSAI
>AT3G11700.1 | Symbols: FLA18 | FASCICLIN-like arabinogalactan protein 18 precursor | chr3:3698992-3700971 FORWARD LENGTH=462
MDRCIYGCSVITIFFSFFFLLNASALESGHHNITGSGQINSNSVLVALLDSRYTELAELVEKAL
LLQTLEDAVGRHNITIF**AP**RNEALERDLDPDFKRFLLQPGNLKSLQTLLLSHIIPKRVGSNQWP
EENSGRVKHVTLGHDQVLHLSKLKGTNGKRLVNSAVITRPDDLTRPDGLIHGIERLLIPRSVQE
DFNRRRNLRSISAVLPEG**AP**EIDPRTNRLKKSATAVSV**PA**G**SP**PVLPIESAM**AP**G**PS**L**APAPAP**
GPGGAHKHFNGDAQVKDFIHTLLHYGGYNEMADILVNLTSLATEMGRLVSEGYVLTVL**AP**NDEA
MGKLTTDQLSEPG**AP**EQIMYYHIIPEYQTEESMYNSVRRFGKVKYETLRFPHKVGAKEADGSVK
FGSGDRSAYLFDPDIYTDGRISVQGIDGVLFPEEKEEETVKK**PT**GPVKKVVQPRRGKLLEVACS
MLGAIGKDSYLSRC
>AT3G12660.1 | Symbols: FLA14 | FASCICLIN-like arabinogalactan protein 14 precursor | chr3:4019060-4019827 FORWARD LENGTH=255
MSSSLTIFFFFFASTFLYTSSNSFNITNILNEHDDFSNFNQLLSETQLASTINKRQTITVLVVS
NGALSSLSGQ**PT**SVIKKILSLHIVLDYYDQKKLKNLSKKTVLLTTLFQSSGLARGQQGFLNATV
MKNGDVAFGSAVPGSSLDAQLQDTVAALPFNISVLHISSAIMIDVKGDN**APT**A**SP**L**SP**VS**SP**PR
PAE**SP**NDDGQDFDEP**PS**S**AP**GAAADE**PS**ENAGSANGVSRNDSQ**PA**FAFTLLMSFIWWFMARLR
>AT3G13520.1 | Symbols: AGP12, ATAGP12 | arabinogalactan protein 12 | chr3:4409087-4409269 FORWARD LENGTH=60
MESMKMKLIVVLMVAIVAFSAVGNVAAQTE**APAPSPT**SDAAMFV**PA**LFASVAALASGFLF
>AT3G20570.1 | Symbols: ENODL9, AtENODL9 | early nodulin-like protein 9 | chr3:7186754-7187453 REVERSE LENGTH=203
MARNLKSMMLCGFGLLCFLMIVDRAYAREFTVGGATGWTV**PS**GSQVYSQWAEQSRFQIGDSLLF
VYQSNQDSVLQVTRDAYDSCNTD**SPT**AKFADGKTSVTLNHSGPYYFISGNKDNCKKNEKLVVIV
MADRSGNKNTAS**SPPSPAPAPS**GES**APSP**PVSGTFEM**TPAPTPT**TSED**TP**NSAASSLSFVAALL
GAALASTLFLH
>AT3G20865.1 | Symbols: AGP40 | arabinogalactan protein 40 | chr3:7308699-7308887 REVERSE LENGTH=62
MEMKNIFVALFISAVLVSSVSAATME**SPAPSP**GASSASTVAFPVVGSIVAASLSAFLALLLQ
>AT3G45230.1 | Symbols: | hydroxyproline-rich glycoprotein family protein | chr3:16569123-16569650 FORWARD LENGTH=175
MKLEFIIVAMMLSLVLVSGEILTKS**SPAPSP**DLAD**SP**LIHA**SPPS**KLGSHN**SPA**E**SP**IEYS**SP**P
EPETEH**SPSPSPA**N**SPS**V**SP**PLPNDSQ**SPS**SSA**SPSPSP**EASDVNHSDITGIEGEKL**PS**GSGGG
MSGGKKVGVAFGAIAAVCVVGVAGFVYKKRQENIRRSRYGYAAREIL
>AT3G46550.1 | Symbols: SOS5 | Fasciclin-like arabinogalactan family protein | chr3:17136612-17137874 REVERSE LENGTH=420
MANVISISHFTLLALPYLLLLLSSTAAAINVTAVLSSFPNLSSFSNLLVSSGIAAELSGRNSLT
LLAVPNSQFSSASLDLTRRLP**PS**ALADLLRFHVLLQFLSDSDLRRIP**PS**GSAVTTLYEASGRTF
FGSGSVNVTRD**PA**SGSVTIG**SPA**TKNVTVLKLLETKPPNITVLTVDSLIV**PT**GIDITASETL**TP**
P**PT**STSL**SP**P**PA**GINLTQILINGHNFNVALSLLVASGVITEFENDERGAGITVFV**PT**DSAFSDL
PSNVNLQSL**PA**EQKAFVLKFHVLHSYYTLGSLESITN**PV**Q**PT**LATEEMGAGSYTLNISRVNGSI
VTINSGVVLAVVTQTAFDQN**PV**SVFGVSKVLLPKELFPKSGQ**PV**AT**AP**PQEISL**SP**ESSSEQ**PS**
RLV**SP**PREIVSSGAVKRPLGFLVLWCWCIAFCYVLV
>AT3G52370.1 | Symbols: FLA15 | FASCICLIN-like arabinogalactan protein 15 precursor | chr3:19417549-19419549 FORWARD LENGTH=436

MDDLSKLLFFLLLTISITTALPDKPGSGQINSNSVLVALLDSHYTELAELVEKALLLQTLEEAV
GQHNITIF**AP**RNDALEKNLDPEFKSFLLQPKNLKSLQSLLMFHILPKRIT**SP**QFSSAVVSHRTL
SNDHLHFTNGKVNSAEITKPDDLTRPDGIIHGIERLLIPRSVQEDFNRRRSLRSIAAVLPEG**AP**
EVDPRTHRLKKK**PAP**I**PA**G**AP**PVLPVYDAM**SP**G**PS**L**APAPAP**GPGGPRHHFNGEAQVKDFIHTL
LHYGGYNEMADILVNLTSLATEMGRLVSEGYVLTVL**AP**NDEAMAKLTTDQLSEPG**AP**EQIMYYH
IIPEYQTEESMYNSVRRFGKIRYDSLRFPHKVEAQEADGSVKFGHGDGSAYLFDPDIYTDGRIS
VQGIDGVLFPEEK**TP**VEKKTGVPVVKK**AP**KPRRGKLMEVACTMLGSQF**PT**CQ
>AT3G57690.1 | Symbols: AGP23, ATAGP23 | arabinogalactan protein
23 | chr3:21384269-21384454 FORWARD LENGTH=61
MEMKKIACGVLFAAASMTAVMAAEEVG**APAP**G**PA**SAASAAL**PA**LGSLVGASLVSLFSYYLH
>AT3G60270.1 | Symbols: | Cupredoxin superfamily protein |
chr3:22278029-22278762 REVERSE LENGTH=187
MGSTAATALLLLLLLVAV**PA**VFAVTFQVGDNDGWTIGVEYTSWVSEKTFRVGDTLEFKYG**PS**HS
VAVVNKADYDGCETSR**PT**QSFSDGDTKIDLTKVGAIHFLCL**TP**GHCSLGMKLAVQVLAAVSLEP
P**PSPSAPSPSPSAPSPSPSAPSPSPSP**GNAENLKNAASKGIMSYGKIVVTMVLMYGVLN
>AT3G60280.1 | Symbols: UCC3 | uclacyanin 3 | chr3:22279867-
22280633 REVERSE LENGTH=222
MGSTVAAALLLFLAAV**PA**VFAATFKVGDISGWTSNLDYTVWLTGKTFRVGDTLEFVYGLSHSVS
VVDKAGYDNCDSSGATQNFADGDTKIDLTTVGTMHFLC**PT**FGHCKNGMKLAVPVLAA**APSPSTP**
SSPPSTPSTPSSPPSTPSTPSSPPSPPSPPSPSLP**PS**SLP**PS**A**SPPT**NG**TP**DSETL**TP**P**PAP**LP
PSL**SP**NAASKGVMSYGIIGVTMILMYAVMT
>AT3G60900.1 | Symbols: FLA10 | FASCICLIN-like arabinogalactan-
protein 10 | chr3:22499573-22500841 REVERSE LENGTH=422
MATSRAFTLFAFTLSLLTVASTVSGHNITQILSD**TP**EYSSFNNYLSQTKLADEINSRTTITVLV
LNNGAMSSLAGKHPLSVVKNALSLLVLLDYYDPLKLHQLSKGTTLTTTLYQTTGHALGNLGFVN
VTDLKGGKVGFGSA**AP**GSKLDSSYTKSVKQIPYNISVLEIN**AP**II**AP**GILT**APAPS**SAGVSNIT
GLLEKAGCKTFANLLVSSGVIKTFESTVEKGLTVF**APS**DEAFKARGVPDLTNLTQAEVVSLLEY
HALAEYKPKGSLKTNKDAISTLATNGAGKYDLTTSTSGDEVILHTGVG**PS**RLADTVVDE**TP**VVI
FTVDNVLL**PA**ELFGKSS**SPAPAP**E**PV**S**APTPTPA**K**SPSP**VE**APSPT**AA**SPPAP**PVDES**SP**EG**AP**
SD**SPT**SSENSNAKNAAFHVN**APA**LFTALVTIAATSLLL
>AT3G61640.1 | Symbols: AGP20, AtAGP20 | arabinogalactan protein
20 | chr3:22810283-22810629 REVERSE LENGTH=74
MASRNSVAVIALFAFVFAVI**SP**FAGAQSL**APAPSPT**SDGTSIDQGIAYLLMVVALVLTYLIHPL
DASSSSYTFF
>AT4G09030.1 | Symbols: AGP10, ATAGP10 | arabinogalactan protein
10 | chr4:5792249-5792632 FORWARD LENGTH=127
MASKSVVVLLFLALIASSAIAQ**AP**G**PAPT**R**SP**L**PSPA**QPPRTA**APTPS**I**TPTPTPTPS**A**TPT**A**A**
PVSPPAG**SP**L**PS**SA**SPPAPPT**SL**TP**DG**APV**AG**PT**GS**TP**VDNNNAATLAAGSLAGFVFVASLLL
>AT4G12730.1 | Symbols: FLA2 | FASCICLIN-like arabinogalactan 2
| chr4:7491598-7492809 REVERSE LENGTH=403
MAYLRRAATALVLIFQLHLFLSLSNAHNITRILAKDPDFSTFNHYLSATHLADEINRRQTITVL
AVDNSAMSSILSNGYSLYQIRNILSLHVLVDYFGTKKLHQITDGSTSTASMFQSTGSATGTSGY
INITDIKGGKVAFGVQDDDSKLTAHYVKSVFEKPYNISVLHISQVLT**SP**EAE**APT**A**SPS**DLILT
TILEKQGCKAFSDILKSTGADKTFQDTVDGGLTVFC**PS**DSAVGKFMPKFKSL**SPA**NKTALVLYH
GMPVYQSLQMLRSGNGAVNTLATEGNNKFDFTVQNDGEDVTLETDVVTAKVMGTLKDQEPLIVY
KIDKVLLPREIYKAVKTS**APAP**KSSKKKPKNAEADADG**PS**AD**APS**DDDVEVADDKNGAVSAMIT
RTSNVVTAIVGLCFGVWLM
>AT4G16980.1 | Symbols: | arabinogalactan-protein family |

chr4:9557087-9557581 FORWARD LENGTH=164
MASSFSSQAFFLLTLSMVLIPFSLAQ**AP**MM**APS**GSMSMPPMSSGGGSSVPPPVM**SP**MPMM**TP**PP
MPM**TP**PPMPM**TP**PPMPM**AP**PPMPMA**SP**PMMPM**TPS**T**SPSP**LTVPDM**PSP**PM**PS**GMES**SPSP**GPM
P**PA**MAA**SP**DSGAFNVRNNVVTLSCVVGVVAAHFLLV
>AT4G26320.1 | Symbols: AGP13 | arabinogalactan protein 13 |
chr4:13317235-13317414 REVERSE LENGTH=59
MEAMKMRLFVAVLVAAMAFSAVQQAAAVE**APAPSPT**SDASLAI**PA**FFASVATLAFGFLF
>AT4G27520.1 | Symbols: ENODL2, AtENODL2 | early nodulin-like
protein 2 | chr4:13750668-13751819 REVERSE LENGTH=349
MTFLKMKSLSFFFTILLSLSTLFTISNARKFNVGGSGAWVTNPPENYESWSGKNRFLVHDTLYF
SYAKGADSVLEVNKADYDACNTKNPIKRVDDGDSEISLDRYGPFYFISGNEDNCKKGQKLNVVV
ISARI**PS**TAQ**SP**HAA**AP**GSS**TP**GSM**TP**PGGAH**SP**KSS**SP**V**SPT**T**SP**PGST**TP**PGGAH**SP**KSSSA
V**SPA**T**SP**PGSM**AP**KSG**SP**V**SPT**T**SPPAP**PKST**SP**V**SPS**S**AP**MT**SPPAP**M**AP**KSSSTIP**PS**S**AP**M
T**SP**PGSM**AP**KSS**SP**VSN**SPT**V**SPS**L**AP**GGSTSS**SPS**D**SPS**GSAMG**PS**GDG**PS**AAGDIS**TPA**G**AP**
GQKKSSANGMTVMSITTVLSLVLTIFLSA
>AT4G28365.1 | Symbols: ENODL3, AtENODL3 | early nodulin-like
protein 3 | chr4:14033012-14033688 REVERSE LENGTH=199
MGLVMRFDLYLMFVMLMGLGFTISNGYKFYVGGKDGWV**PTPS**EDYSHWSHRNRFQVNDTLHFKY
AKGKDSVLEVTEQEYNTCNTTHPLTSLSDGDSLFLLSHSGSYFFISGNSQNCLKGQKLAVKVLS
TVHHSH**SP**RHT**SPSPSP**VHQELS**SP**G**PSP**GVE**PS**SDSNSRV**PAP**G**PA**T**AP**NSAGLVGPGMVVLV
IMISSLF
>AT4G30590.1 | Symbols: ENODL12, AtENODL12 | early nodulin-like
protein 12 | chr4:14935760-14936469 REVERSE LENGTH=190
MGIIVPVLTLVFLLFAKVSHGASNPRVILVGGSVGSWKVPD**SP**NNTLNHWAENNRFKVGDFIVW
KYDMKVDSVLQVTKEDYESCNTANPLKQYNDGNTKVALDKSGPYFFISG**AP**GNCAKGEKITLVV
LAERKSGGGSSSGD**AP**KV**SP**V**SPT**AQ**TPAPAP**G**PA**AAHNAAVGLKVASGWFLTAVVVGLAMA
>AT4G31370.1 | Symbols: FLA5 | FASCICLIN-like arabinogalactan
protein 5 precursor | chr4:15223838-15224674 REVERSE LENGTH=278
MGLKASLSLLSLTILLVFSKVVTANNITLAFQKYSKFSTMRDLFIKTKLIAAIDKYQTITVLAV
SNDAISSITNRSEVELRNILMTHVILDYYDELKLQGMREKSIMLTTLYQTTGLGEQMNGFLNVS
KSKGRVYFGSEVKN**SP**LNAEYVSTVYHNPYNLSIIQITMPIV**AP**GLSLAIFPPPPPYVHV**AP**Y**P**
TPMDASVV**PAP**G**PA**ADDN**SP**DSAVPK**TPPAPA**TD**TP**EAD**SPAPAPS**ADNEKIEAADKAK**PS**SSA
SKAGWSFDVILLLAFLASFAGF
>AT4G32490.1 | Symbols: ENODL4, AtENODL4 | early nodulin-like
protein 4 | chr4:15678811-15679556 REVERSE LENGTH=221
MVFVKMTDVYLMIVMLMGLGFSIELSNGHKFYVGGRDGWVL**TPS**EDYSHWSHRNRFQVNDTLYF
KYVKGKDSVLEVSEKEYNTCNTTHPLTSLSDGDSLFLLSRSDPFFFVSGNSGSCLKGQKLAVTV
MSTGHHSH**TP**RH**PSPSPSPS**A**SP**VRKALL**SPAP**IPVHKALS**SPAPTP**GVD**PS**HSEVL**APAP**G**PA**
AAVRNLAGSV**AP**GVISLGLVLVIMISSMV
>AT4G37450.1 | Symbols: AGP18, ATAGP18 | arabinogalactan protein
18 | chr4:17605926-17606734 REVERSE LENGTH=209
MDRNFLLTVTLICIVVAGVGGQ**SP**IS**SPT**K**SPTTPSAPT**T**SPT**K**SPA**VT**SPT**T**APA**K**TPT**ASAS
SPVE**SP**K**SPAP**VSES**SP**P**PTP**VPES**SP**PV**PAP**MVS**SP**VS**SP**PV**PAP**VAD**SPPAP**VA**AP**VADV**PA**
PAPSKHKKTTKKSKKHQA**APAPAP**ELLGP**PAPPT**E**SP**GPNSDAF**SP**G**PS**ADDQSGAASTRVLRN
VAVGAVATAWAVLVMAF
>AT4G40090.1 | Symbols: AGP3 | arabinogalactan protein 3 |
chr4:18581085-18581504 REVERSE LENGTH=139
MALKTLQALIFLGLFAASCLAQ**APAPAP**ITFLPPVE**SPSP**VV**TPT**AEP**PAP**VA**SP**PI**PA**NE**PTP**

V**PTTPPT**V**SPPT**T**SPT**T**SP**VA**SP**PKPYAL**AP**G**PS**G**PTPAPAPAP**RADGPVADSALTNKAFLVST
VIAGALYAVLA
>AT5G06390.1 | Symbols: FLA17 | FASCICLIN-like arabinogalactan
protein 17 precursor | chr5:1952939-1955047 FORWARD LENGTH=458
MDRRIYGGSAVIHLFLFFSVLIFSAASALSKNQ**SPS**SGSGQINSNSVLVALLDSRYTELAELVE
KALLLQTLEDAVGRHNITIF**AP**RNEALERDLDPEFKRFLLEPGNLKSLQTLLMFHIIPNRVGSN
QW**PS**EESGRVKHHTLGNDQVRLSNGQGKKMVDLAEIIRPDDLTRPDGLIHGIERLLIPRSVQED
FNRRRSLQSISAVLPEG**AP**EVDPRTNRLKK**PAAPVPA**G**SPPA**LPIQSAM**AP**G**PS**L**APAPAP**GPG
GKQHHFDGEAQVKDFIHTLLHYGGYNEMADILVNLTSLATEMGRLVSEGYVLTVL**AP**NDEAMAK
LTTDQLSEPG**AP**EQIVYYHIIPEYQTEESMYNSVRRFGKVKFDTLRFPHKVAAKEADGSVKFGD
GEKSAYLFDPDIYTDGRISVQGIDGVLFPQEEEVVESVKKPVKKIVQPRRGKLLEVACSMLGAF
GKDTYLSKCR
>AT5G06920.1 | Symbols: FLA21 | FASCICLIN-like arabinogalactan
protein 21 precursor | chr5:2142858-2143919 FORWARD LENGTH=353
MGCCSSDCFVYFILSIALAFMAISTTLR**SP**PDSE**PT**IPIAFSSS**SPS**LSLNASNTLRQSNFKAI
ATLLHI**SP**EIFLSS**SP**NTTLFAIEDASFFNTSSLHPLFLKQLLHYHTLPLMLSMDDLLKKPQGT
CL**PT**LLHHKSVQISTVNQESRTAEVNHVRITHPDMFLGDSLVIHGVIGPF**SP**LQPHSDHLIH**TP**
LCQSDTTNKTSNNEEVPVSIDWTRIVQLLSSNGFVPFAIGLHSVLNRIVNDHNHHKNLTGVTIL
A**TP**NLVSLSSA**SP**FLYEVVRHHILVQRLTYKDFASMSDKATVKTLDPYQDLTITRRNVNSSGGD
FMISGVEIVDPDMFSSSNFVIHGISHTLEIPHV
>AT5G10430.1 | Symbols: AGP4, ATAGP4 | arabinogalactan protein 4
| chr5:3277828-3278235 REVERSE LENGTH=135
MGSKIVQVFLMLALFATSALAQ**APAPTPT**A**TP**P**PATP**PPVA**TP**PPVA**TP**P**PA**A**TPAPATP**P**PA**A
TPAPAT**TPPS**V**APSPA**DV**PT**A**SPPAP**EG**PT**V**SPS**S**AP**G**PS**DA**SPAPS**AAFSNKAFFAGTAFAAI
MYAAVLA
>AT5G11740.1 | Symbols: AGP15, ATAGP15 | arabinogalactan protein
15 | chr5:3784318-3784503 FORWARD LENGTH=61
MAISKASIVVLMMVIISVVASAQSE**APAPSPT**SGSSAISASFVSAGVAAVAALVFGSALRI
>AT5G12880.1 | Symbols: | proline-rich family protein |
chr5:4068714-4068935 REVERSE LENGTH=73
MRRSW**PT**RLILLMVLTVVITMIAAAYGYSSVSSSKHKFPHYKYK**APS**P**PT**TY**SP**YRYF**SP**PPVT
DSDSAAYVR
>AT5G18690.1 | Symbols: AGP25, ATAGP25 | arabinogalactan protein
25 | chr5:6233776-6234126 REVERSE LENGTH=116
MAFSFLNKLLIIFIFIFISLSSS**SPT**ISLVQQL**SP**EI**AP**LL**PSP**GDAL**PS**DDGSGTI**PSSPSP**P
DPDTNDGSYPDPLAF**SP**FA**SP**PVS**SPS**PP**PS**L**PS**AGVLLISLIISSASFLAL
>AT5G24105.1 | Symbols: AGP41 | arabinogalactan protein 41 |
chr5:8152097-8152656 FORWARD LENGTH=63
MSGSRLFFGVSTIVSIIFAILLPMAHAQSA**APAPAPT**SDGTTIDQGIAYVLMLVALVLTYLIH
>AT5G26330.1 | Symbols: | Cupredoxin superfamily protein |
chr5:9241614-9242635 REVERSE LENGTH=187
MAAIIVAALACIVVMLRLSEAAVYKVGDSAGWTTIANVDYKLWASTKTFHIGDTVLFEYNPQFH
NVMRVTHPMYRSCNTSKPISTFTTGNDSITLTNHGHHFFFCGVPGHCLAGQKLDLHVLL**PA**SS**T**
PLSDP**PT**SSSS**SPPS**TTI**PA**AGVPG**PSPS**LAASL**PS**MVTAQIVAVVTLLVSLAFTNFAS
>AT5G40730.1 | Symbols: AGP24, ATAGP24 | arabinogalactan protein
24 | chr5:16301153-16301362 FORWARD LENGTH=69
MMMMTKMFVQIAVVCLLATMAVVSAHEGHHHH**APAPAP**G**PA**SSSTVVSATNMFTVLAIAAVALV
VGSNH

>AT5G44130.1 | Symbols: FLA13 | FASCICLIN-like arabinogalactan protein 13 precursor | chr5:17761128-17761871 FORWARD LENGTH=247
MAT**TP**LLLLLLTAVFLSTEITAQRA**APAP**G**PA**GPINITAILEKGGQFVTLIRLLNTTQIGNQIN
IQINSSSEGMTVL**APT**DNAFQNLKPGTLNKL**SP**DDQVKLILYHV**SP**KFYTLEDLLSVSNPVRTQ
ASGRDVGGVYGLNFTGQGNQVNVSTGVVETRLSTSLRQERPLAVYVVDMVLLPEEMFGERKI**SP**
M**AP**PPKSK**SP**DVSDDSESSKKAA**APS**ESEKSGSGEMNTGLGLGLGLVVLCLKFLL
>AT5G53250.1 | Symbols: AGP22, ATAGP22 | arabinogalactan protein 22 | chr5:21603715-21604007 FORWARD LENGTH=63
MASLKFPLEILAVFVIISVILLPIAQSHSS**SPAPAPT**SDGTSIDQGIAYVLMMVALALTYFIH
>AT5G53870.1 | Symbols: ENODL1, AtENODL1 | early nodulin-like protein 1 | chr5:21870033-21871228 REVERSE LENGTH=370
MSAIMKSLCFSFLILASFATFFSVADAWRFNVGGNGAWVTNPQENYNTWAERNRFQVNDSLYFK
YAKGSDSVQQVMKADFDGCNVRNPIKNFENGESVVTLDRSGAFYFISGNQDHCQKGQKLIVVVL
AVRNQ**PSAPA**H**SP**V**PS**V**SPT**QPPKSH**SP**V**SP**V**APA**S**APS**KSQPPRSSV**SPA**QPPKSS**SP**ISH**TP**
AL**SPS**HATSH**SPATPSPSP**K**SPSP**VSH**SPS**H**SPA**H**TPS**H**SPA**H**TPS**H**SPA**H**APS**H**SPA**H**APS**H**S**
PAH**APS**H**SPA**H**SPS**H**SPATP**K**SPSPS**S**SPA**Q**SPATPSP**M**TP**Q**SPSP**VS**SPSP**DQSA**APS**DQS**TP**
L**APSPS**ET**TPT**ADNIT**APAPSP**RTNSASGLAVTSVMSTLFSATFTFLMFA
>AT5G55730.1 | Symbols: FLA1 | FASCICLIN-like arabinogalactan 1 | chr5:22558375-22560392 REVERSE LENGTH=424
MAKKMSSLIIIFNILLLLTTQTHAHNVTRLLANH**PS**FSSFSHFLTQTHLADEINRRRTITVCAV
DNAAMSALTSKGYTLSTLKNILSLHVLLDYFGTKKLHQIRDGSALAATLFQATGA**AP**GTSGFVN
ITDLRGGKVGFGPDGGDLSSFFVKSIEEVPYNISIIQISRVL**PS**ETAA**APTPAPA**EMNLTGIMS
AHGCKVFAETLLTNPGASKTYQESLEGGMTVFCPGDDAMKGFLPKYKNLT**AP**KKEAFLDFLA**VP**
TYYSMAMLKSNNGPMNTLATDGANKFELTVQNDGEKVTLKTRINTVKIVDTLIDEQPLAIYATD
KVLLPKELFKASAVE**APAPAPAP**EDGDVAD**SP**KAAKGKAKGKKKKA**APSP**DNDPFGDSD**SPA**EG
PDGEADDATADDAGAVRIIGGAKAGLVVSLLCLFASSWLL
>AT5G56330.1 | Symbols: ATACA8, ACA8 | alpha carbonic anhydrase 8 | chr5:22813768-22816162 FORWARD LENGTH=350
MKISSLGWVLVLIFISITIVSS**APAP**KPPKPK**PAPAPTP**PKPK**PTPAPTP**PKPKPK**PAPTP**PKP
K**PAPAPTP**PKPK**PAPAPTP**PKPKPK**PAPTP**PNPK**PTPAPTP**PKPK**PAPAPAPTPAP**KPK**PAP**K**P**
APGGEVEDETEFSYETKGNKG**PA**KWGTLDAEWKMCGIGKMQ**SP**IDLRDKNVVVSNKFGLLRSQY
L**PS**NTTIKNRGHDIMLKFKGGNKGIGVTIRGTRYQLQQLHWH**SPS**EHTINGKRFALEEHLVHES
KDKRYAVVAFLYNLGASDPFLFSLEKQLKKITDTHASEEHIRTVSSKQVKLLRVAVHDASDSNA
RPLQAVNKRKVYLYKPKVKLMKKYCNISSY
>AT5G56540.1 | Symbols: AGP14, ATAGP14 | arabinogalactan protein 14 | chr5:22893243-22893425 FORWARD LENGTH=60
MEAMKMKLYVVVLVAVIAFSTVHQTVAAVD**APAPSPT**SDASSFI**PT**FFASVAVMAFGFFF
>AT5G60490.1 | Symbols: FLA12 | FASCICLIN-like arabinogalactan-protein 12 | chr5:24325916-24326665 REVERSE LENGTH=249
MEHSLIILLFTVLLLLTT**TP**GILSQ**PSPA**V**APAP**PG**PT**NVTKILEKAGQFTVFIRLLKSTGVAN
QLYGQLNNSDNGITIF**APS**DSSFTGLKAGTLNSLTDEQQVELIQFHVI**PS**YVSSSNFQTISNPL
RTQAGDSADGHFPLNVTTSGNTVNITSGVTNTTVSGNVYSDGQLAVYQVDKVLLPQQVFDPRP**P**
APAPAPSVSKSKKKKDDSDSSSDD**SPA**DASFALRNVGSVCDAVSFCVMSVMLAWFYL
>AT5G64310.1 | Symbols: AGP1, ATAGP1 | arabinogalactan protein 1 | chr5:25722018-25722413 FORWARD LENGTH=131
MAFSKSLVFVLLAALLISSAVAQ**SPAPAPS**NVGGRRI**SPAPSP**KKMT**APAPAP**EV**SPSPSPA**AA
L**TP**ESSA**SPPSP**PLAD**SPT**AD**SPA**L**SPS**AISD**SPT**E**AP**G**PA**QGGAVSNKFASFGSVAVMLTAAV
LVI

>AT5G65390.1 | Symbols: AGP7 | arabinogalactan protein 7 | chr5:26128883-26129275 REVERSE LENGTH=130
MNSKIIEAFFIVALFTTSCLAQ**APAPSPT**TTV**TP**PPVA**TP**P**PA**A**TPAPT**T**TP**P**PA**V**SPAPT**S**SP**
PSS**APSPS**SD**APT**A**SPPAP**EGPGV**SP**GEL**APTPS**DAS**AP**PPNAALTNKAFVVGSLVAAIIYAVV
LA
>AT2G13820.1 | Symbols: | Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein | chr2:5774295-5776279 REVERSE LENGTH=169
MAYATILMIFSVVALMSGERAHAAVDCSSLILNMADCLSFVTSGSTVVKPEGTCCSGLKTVVRT
GPECLCEAFKNSGSLGLTLDLSKAASL**PS**VCKVA**APPS**ARCGLSVSGDP**PA**T**AP**GL**SPT**AGAG**A**
PALSSGANAA**TP**VS**SP**RSSDASLLSVSFAFVIFMALISSFY
>AT2G27130.1 | Symbols: | Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein | chr2:11595379-11596395 FORWARD LENGTH=176
MLTTNTLAVLLLLFLSLCSGQ**SPPAP**EPIAADG**PSSP**VNCLVSMLNVSDCFSYVQVGSNEIKPE
AACCPELAGMVQS**SP**ECVCNLYGGGA**SP**RFGVKLDKQRAEQLSTICGVK**APSPS**LCSVLGF**PT**I
SPAGSEDSSSGSEGSDKDKKNGAMTTKYCGVALNSLALLLLFTFLSLS

>AT2G44290.1 | Symbols: | Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein | chr2:18305418-18306202 REVERSE LENGTH=205
MESRKIKVMATAIALIMVAMVVDAAGADKGKDKEECTAQLVGMATCLPYVQGKAK**SPTP**DCCSG
LKQVINSDMKCLCMIIQERNDPDLGLQVNVSLALAL**PS**VCHATADITKC**PA**LLHLDPN**SP**DAQV
FYQLAKGLNETVSAS**APT**GSASE**PT**SMSS**TP**GSSAGNNSGRTTSVPGTNHAQSFSKQWLGLEVV
AHFFVIFYIFILV
>AT2G44300.1 | Symbols: | Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein | chr2:18307468-18308286 REVERSE LENGTH=204
MESRKINLMATAIALIVVAMVVAAADDKTKDKEECTEQLVGMATCLPYVQGQAK**SPTP**DCCSGL
KQVLNSNKKCLCVIIQDRNDPDLGLQINVSLALAL**PS**VCHAAADVTKC**PA**LLHLDPN**SP**DAQVF
YQLAKGLNKTG**PA**S**APT**GS**SP**GPISI**SPT**SGSDDGNNSGRTTSVPGRNHAQSFYKQWLGLEVVF
HFFVIFYIFILV
>AT2G48130.1 | Symbols: | Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein | chr2:19685263-19685977 REVERSE LENGTH=183
MGYRRSYAITFVALVAALWSVTKAQ**PS**SSCVSTLTTL**SP**CLSYITGNST**TPS**QPCCSRLDSVIK
S**SP**QCICSAVN**SP**IPNIGLNINRTQALQLPNACNIQTPPLTQCNAATG**PT**AQP**PAPSPT**EK**TP**D
VTL**TPT**SLPGARSGVGGGSKT**VPS**VGTGSSSRNVDPLPLHFLMFAVLVVCTSSFL
>AT3G22600.1 | Symbols: | Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein | chr3:8006711-8007397 REVERSE LENGTH=170
MKMEMGLVFLTVFMAVMSSTMVSAQSSCTNALISM**SP**CLNYITGNST**SP**NQQCCNQLSRVVQS**S**
PDCLCQVLNGGGSQLGINVNQTQALGLPRACNVQTPPVSRCNTGGGGGGSTSD**SPA**E**SP**NSSGP
GNGSKTVPVGEGDGP**PS**SDGSSIKFSFPLIAFFSAVSYMAIF
>AT3G43720.1 | Symbols: | Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein | chr3:15615549-15617099 REVERSE LENGTH=193
MSNVVVIAVVLIVASLTGHVSAQMDM**SPS**SG**PS**G**AP**DCMANLMNMTGCLSYVTVGEGGGAAKPD

KTCC**PA**LAGLVES**SP**QCLCYLLSGDMAAQLGIKIDKAKALKLPGVCGVI**TP**D**PS**LCSLFGIPVG
APVAMGDEGA**SPA**Y**AP**GSMSGAE**SP**GGFGSG**PS**ASRGSD**APS**S**AP**YSLFLNLIIFPLAFAFYIF
C
>AT4G08670.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr4:5536765-5538210 REVERSE LENGTH=208
MKQSLLLSFVLLLLSSSSLV**TP**IHARNKSN**PA**K**SP**VG**APAP**G**PS**SSDCSTVIYSMMDCLGYLGV
GSNETKPEKSCCTGIETVLQYNPQCICAGLVSAGEMGIELNSTRALA**TP**KACKLSI**AP**PHCGII
TSGAT**TP**GA**SP**V**SPS**AG**APT**T**SPS**AAK**SP**ETSAT**SPS**SDE**TPS**MT**APSPS**SSGTNILSV**PA**LTI
VFVIVSSVAYISAFSN
>AT4G14815.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr4:8505122-8505760 FORWARD LENGTH=156
MKPRMCLILFIALMRVMSIVSAQSSCTNVLISM**AP**CLSFITQNTSL**PS**QQCCNQLAHVVRYSSE
CLCQVLDGGGSQLGINVNETQALALPKACHVETPPASRCHSGSSVNSHSEHGNGSKTVPREKSS
SDGSIKFSFPLLAILFTASYITLIYAKY
>AT5G09370.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr5:2909450-2910475 REVERSE LENGTH=158
MAYFSTATSLLLLVLSVS**SP**YVHGASDCDTLVITLFPCLPFISIGGTAD**TPT**ASCCSSLKNILD
TKPICLCEGLKK**AP**LGIKLNVTKSATLPVACKLNAPPVSACDSLPPASPP**T**ANGQ**APV**WGSGW**A**
PAPSPSKGNSLIPISGFSFVIVTALAMFRI
>AT5G64080.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr5:25645475-25646638 REVERSE LENGTH=182
MATHSSFTAT**TP**LFLIVLLSLSSVSVLGASHHHAT**APAPS**VDCSTLILNMADCLSFVSSGGTVA
KPEGTCCSGLKTVLKADSQCLCEAFKSSASLGVTLNITKASTL**PA**ACKLH**APS**IATCGLSV**APS**
T**AP**GL**AP**GVAAAGPETAGFL**AP**N**PS**SGNDGSSLI**PT**SFTTVLSAVLFVLFFSSA
>AT1G21090.1 | Symbols: | Cupredoxin superfamily protein |
chr1:7384854-7386199 FORWARD LENGTH=242
MGCSQKHLTSMLFFYFFCFLSLFSR**PS**LSATFLVDGVSVWK**SPT**VHTGDSVIFRHKYGYDLYIF
RNKDAFNVCNFTQATLLTKPNSTSFTWY**PS**RTGSYYFSFTNNTSLPKTCQLNQKLTVQVILAAA
SPPSQP**PA**T**APVPV**SEGGVIS**SPS**SYPWPLGPREGSAF**SP**G**PSPS**EITSVTVPGKDGVPFINSN
PAVPL**PT**GDVDSTSINPL**PT**STNSAHQVMMMTLTVKLGLCCVAMFLFLVL
>AT1G21880.2 | Symbols: LYM1 | lysm domain GPI-anchored protein
1 precursor | chr1:7680689-7682526 FORWARD LENGTH=416
MKIPEKPIFLIFVSLILASSLTFTATAKSTIEPCSSNDTCNALLGYTLYTDLKVSEVASLFQVD
PISILLANAIDISYPDVENHIL**PS**KLFLKIPITCSCVDGIRKSVSTHYKTR**PS**DNLGSIADSVY
GGLVSAEQIQEANSVND**PS**LLDVGTSLVIPLPCACFNGTDNSL**PA**VYLSYVVKEIDTLVGIARR
YSTTITDLMNVNAMG**AP**DVSSGDILAVPLSACASKFPRYASDFGLIVPNGSYALAAGHCVQCSC
ALGSRNLYCE**PA**SLAVSCSSMQCRNSNLMLGNITVQQTSAGCNVTTCDYNGIANGTILTMLTRS
LQPRCPGPQQF**AP**LL**AP**PDTVPRDVMY**APAPSP**DFDGPGSIAS**SP**RSSMLPGGGILPGN**PA**NG**P**
AGSISTASASSVSYFFITFLISIASFSLALSS
>AT1G61900.1 | Symbols: | unknown protein; FUNCTIONS IN:
molecular_function unknown; INVOLVED IN: biological_process
unknown; LOCATED IN: anchored to plasma membrane, plasma
membrane, anchored to membrane; EXPRESSED IN: 23 plant
structures; EXPRESSED DURING: 13 growth stages; BEST Arabidopsis

thaliana protein match is: unknown protein (TAIR:AT2G30700.1); Has 65 Blast hits to 65 proteins in 12 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 65; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink). | chr1:22882508-22884722 REVERSE LENGTH=433
MTRRAEFEMGLFVILQSMFLISLCSSQKPEEFLPEI**SP**DT**SP**QPFLPFI**APSP**MVPYINSTMPK
LSGLCSLNFSASESLIQTTSHNCWTVF**AP**LLANVMCCPQLDATLTIILGKASKETGLLALNRTQ
SKHCLSDLEQILVGKGASGQLNKICSIHSSNLTSSSCPVINVDEFESTVDTAKLLLACEKIDPV
KECCEEACQNAILDAATNISLKASETLTDNSDRINDCKNVVNRWLATKLD**PS**RVKETLRGLANC
KINRVCPLVFPHMKHIGGNCSNELSNQTGCCRAMESYVSHLQKQTLITNLQALDCATSLGTKLQ
KLNITKNIFSVCHISLKDFSLQVGNQESGCLL**PS**L**PS**DAIFDKDTGISFTCDLNDNI**PAP**W**PS**S
SLSSASTCKK**PV**RI**PA**L**PA**AASSQPRLHDEGVTRLVIFVLSMLLVMLLS

>AT1G63550.1 | Symbols: | Receptor-like protein kinase-related family protein | chr1:23569786-23570890 FORWARD LENGTH=324
MARIIITLTIPLFYFFFFSLLSHQTMSQPDHIFTVCN**PT**NNFTQTSSYETNRDTLLASLRESSS
LGHYSNATEGL**SP**DTVHGMFLCRGDITTASCVDCVQTATTEIASNCTLNKRAVIYYDECMVRYS
NVSFSSELEIV**PS**ITIYSLRS**AP**N**PT**RFNQTLTEKFSELIFNVSSSSLVPYFVEDQERVTQSEG
SYDLDTMVQC**SP**DLDIFNCTVCLRVAFFRISTCCGL**PS**YAKIF**TP**KCLLRFQTSVLL**SP**PP**SPS**
APPPR**SP**PPKS**SPPS**SLPQ**TPSP**PLVF**TP**PQNVPN**PS**GSFSFNVLKGNVIFGRIVVTMTALVFA
LVDL

>AT1G66970.2 | Symbols: SVL2 | SHV3-like 2 | chr1:24992746-24996005 REVERSE LENGTH=785
MNSR**PS**N**PT**KLVIRSSTLLFCGVVLIHLFAAQIDAQRSTSRWQTLNGESCSHFISFFCALFPRK
QENLCD**AP**LVIARGGFSGLYPDSSIAAYQLATLTSVADVVLWCDLQLTKDGLGICFPDLNLANA
STIDRVYPNREKSYSVNGVTTKGWFPNDFSLTELQNFLLIRGILSRTDRFDGNGYLISTIEDVV
TTLNREGFWLNVQHDAFYEQQNLSMSSFLLSVSRTVSIDFIS**SP**EVNFFKKITGSFGRNG**PT**FV
FQFLGKEDFE**PT**TNRTYGSILSNLTFVKTFASGILVPKSYILPLDDEQYLVPHTSLVQDAHKAG
LQVYVSGFANDVDIAYNYSSDPVSEYLSFVDNGDFSVDGVLSDFPITASAAVDCFSHIGRNATK
QVDFLVISKDGASGDYPGCTDLAYEKAIKDGADVIDCSVQMSSDGVPFCLRSIDLRNSIAALQN
TFSNRSTSVPEISSVPGIFTFSLTWPEIQSL**TPA**ISNPFRVYRIFRNPREKNSGKLISLSQFLD
LAKTYTSLSGVLISVENAAYLREKQGLDVVQAVLDTLTEAGYSNGTTTTKVMIQSTNSSVLVDF
KKQSKYETVYKIEETIGNIRDSAIEDIKKFANAVVINKDSVFPNSDSFLTGQTNVVERLQKSQL
PVYVELFRNEFVSQAYDFFSDATVEINAYIYGAGINGTITEFPFTAARYKRNRCLGREEVPPYM
LPVNPGGLLNVM**SP**LSLPPAQ**AP**NQDFIEADVTEPPL**SP**VIAK**APT**S**TP**G**TPS**TIAQ**APS**GQTR
LKLSLLLSVFFLSLLLL

>AT1G77630.1 | Symbols: | Peptidoglycan-binding LysM domain-containing protein | chr1:29173726-29175387 FORWARD LENGTH=423
MKNPEKPLLLFLILASSLASMATAKSTIEPCSSKDTCNSLLGYTLYTDLKVTEVASLFQVDPVS
MLLSNSIDISYPDVENHVL**PA**KLFLKIPITCSCVDGIRKSLSTHYKTRTSDTLGSIADSVYGGL
V**SP**EQIQVANSETDLSVLDVGTKLVIPLPCACFNGTDESL**PA**LYLSYVVRGIDTMAGIAKRFST
SVTDLTNVNAMG**AP**DINPGDILAVPLLACSSNFPKYATDYGLIIPNGSYALTAGHCVQCSCVLG
SRSMYCE**PA**SISVSCSSMRCRNSNFMLGNITSQQSSSGCKLTTCSYNGFASGTILTTLSMSLQP
RCPGPQQL**AP**LI**AP**PDNVPKELMYL**PSPSPSPSP**EFDDIAGGGSSIAAV**PA**A**SP**GGATVSSSNS
IPGN**PA**NGPGGSISIASCPLSYYSFIALLIPIGSCFFVF

>AT3G07390.1 | Symbols: AIR12 | auxin-responsive family protein | chr3:2365452-2366273 FORWARD LENGTH=273
MSLCLKIPLIKHQT**TP**EQNSAMASSSSSLLILAVACFVSLI**SPA**ISQQACKSQNLNSAGPFDSC
EDLPVLNSYLHYTYNSSNSSLSVAFVA**TPS**QANGGWVAWAIN**PT**GTKMAGSQAFLAYRSGGGA**A**

PV VKTYNISSYSSLVEGKLAFDFWNLRAESLSGGRIAIFTTVK**VPA**GADSVNQVWQIGGNVTNG
RPGVHPFGPDNLGSHRVLSFTEDA**AP**GS**APSP**GS**APAP**GTSGST**TP**GTAAGGPGNAGSLTRNVN
FGVNLGILVLLGSIFIF
>AT3G16860.1 | Symbols: COBL8 | COBRA-like protein 8 precursor |
chr3:5759643-5762104 REVERSE LENGTH=653
MGLTRNFILWILLSSLFTAIQLTSSQRN**TP**PPRSKDADLCNGVFVSYTYLTGTKIKPNDTKNQP
YRFESEITVLNNGRDELKSWQVFVKFAHREILVSATNAVLSDGSSL**PA**SVENGTIFAGF**PS**ADL
KTAIMTAGDVTQMEARVELVGTQFGV**APPS**VPLPKNITLVNDGWSCPK**PT**QQGSNVLQVCC**TP**N
PNITTSKIGQKFLPRQEGDLTIMYDVTRAYQSSYSAQVTIENHNLLGRLDNWDLSFMWMKDEFL
FSTKGAY**PS**VVDSSDCITGPQAKYYKDLDFSNVMSCARRPHIIDLPLTKYNDTNVGRIPYCCRN
GTILPRSMDPEKSKSVFQIEVYKMPPDLNISSI**TP**PQSWQIKGNLNPDYKCGPPLRVSSSQFPD
TSGL**PS**NKSAFASWQVVCNITQ**PTP**PKCCVSFSSYFNDSVIPCKTCACGGCSSDRVARTCSTT**S**
PALPLPYQALLIPFDNRTKLTNAWAVLKNRKVPDPLPCGDNCGVSINWHLATDYRGGWTARVTL
FNWGDTDFVDWFTAVELRNA**AP**GFQKAYSFNGSIIAVNGKNTTVLMEGLPGLNYLLAEKDGKN**P**
SEDFRIPGKQQSVISFTKKL**TP**GIKVGSKDGF**PT**KVLFNGQECSL**PS**VL**PT**SNSHRKHVSTFLL
IL**TP**FLALLFLRI
>AT3G20520.1 | Symbols: SVL3 | SHV3-like 3 | chr3:7162845-
7165742 FORWARD LENGTH=729
MACPRVIFLILITFFILQTAFSSSWQTLSGKPPAVIARGGFSGMFPDSSIQAYQLVNITTSPDV
MLWCDLQLTKDGVGICFPNLKLDNGSNVIRIDPHYKERFSVDFTWKELSDVKLAQGVVSRPYIF
DDVSSILAIEEVAKLTASGLWLNIQDSAFYAKHNLSMRNSVVSLSRRLKVNFISSPGISFLKSM
KNSVKPTVTKLIFRFLKQEHIEPFTNQSYGSLAKNLSYIRTFSSGILVPKSYIWPVDSALYLQP
HTSLVTDAHKEGLQVFASEFANDFVIAYNYSYDPTAEYLSFIDNGNFSVDGFLSDFPVTPYRAI
NCFSHVDPKRAKEQAKITIISKNGASGDFPGCTDLAYQRAASDGADILDCNVQMSKDKIPFCMS
SFDLINSTNVIETSFRNLSSVVSEINPRRSGIYTFSLTMSQIQTLKPTISNLEKDSGLFRNPRN
NKAGKFLTLSEFLFLPNRYSSLLGLLIEVENAAYLVEHQGISVVDAVLDELKRATTQQNKTSAR
TILIQSTDKSVLMKFKEKNKMNHDELVYRVDDNIRDVADSAIKDIKNFAGSIVISKKSVFPYKG
FIILEKETNIASKLKSNGLRVYVERFSNECVTHAFDFYDD**PT**LEIDSFVRDVQIDGIITDF**PA**T
TARYRKNKCYGEFGLTTTGELITFANPMLLPPAE**AP**Y**PA**LLDSDVTEPPLPEARSQPPAS**SPS**K
AEEKAIEVPFAFIAMAILVCFFISV
>AT3G51330.1 | Symbols: | Eukaryotic aspartyl protease family
protein | chr3:19053480-19056152 REVERSE LENGTH=529
MVVARQVFVLLSLLVVCWGLERCEASGKFSFEVHHMFSDRVKQSLGLDDLVPEKGSLEYFKVLA
QRDRLIRGRGLASNNEE**TP**ITFMRGNRTISIDLLGFLHYANVSVG**TPA**TWFLVALDTGSDLFWL
PCNCGSTCIRDLKEVGLSQSRPLNLY**SP**NTSSTSSSIRCSDDRCFGSSRCS**SPA**SSCPYQIQYL
SKDTFTTGTLFEDVLHLVTEDEGLEPVKANITLGCGKNQTGFLQSSAAVNGLLGLGLKDYSV**PS**
ILAKAKITANSFSMCFGNIIDVVGRISFGDKGYTDQME**TP**LL**PT**E**PSPT**YAVSVTEVSVGGDAV
GVQLLALFDTGTSFTHLLEPEYGLITKAFDDHVTDKRRPIDPELPFEFCYDL**SP**NKTTILFPRV
AMTFEGGSQMFLRNPLFIVWNEDNSAMYCLGILKSVDFKINIIGQNFMSGYRIVFDRERMILGW
KRSDCFEDESLESTTPPPPETE**APSPS**AS**TP**L**PS**LLPPPAAATPPQIDPRNSTRNSGTGTAANL
VPLASQLLLLLPLLAFL
>AT3G51350.1 | Symbols: | Eukaryotic aspartyl protease family
protein | chr3:19060485-19063248 REVERSE LENGTH=528
MDVARQVFVLLSVLVVCWGFERCEATGKFGFEVHHIFSDSVKQSLGLGDLVPEQGSLEYFKVLA
HRDRLIRGRGLASNNDE**TP**ITFDGGNLTVSVKLLGSLYYANVSVGTPPSSFLVALDTGSDLFWL
PCNCGTTCIRDLEDIGVPQSVPLNLY**TP**NASTTSSSIRCSDKRCFGSKKCS**SPS**SICPYQISYS
NSTGTKGTLLQDVLHLATEDENL**TP**VKANVTLGCGQKQTGLFQRNNSVNGVLGLGIKGYSV**PS**L
LAKANITANSFSMCFGRVIGNVGRISFGDRGYTDQEE**TP**FISV**APS**TAYGVNISGVSVAGDPVD

IRLFAKFDTGSSFTHLRE**PA**YGVLTKSFDELVEDRRRPVDPELPFEFCYDL**SP**NATTIQFPLVE
MTFIGGSKIILNNPFFTARTQEGNVMYCLGVLKSVGLKINVIGQNFVAGYRIVFDRERMILGWK
QSLCFEDESLESTTPPPPEVE**APAPS**VSAPPPRSLPPTVSATPPPINPRNSTGNPGTGGAANLI
PLASQLLLLLPLLAFL
>AT4G16120.1 | Symbols: ATSEB1, COBL7, SEB1 | COBRA-like
protein-7 precursor | chr4:9116591-9119138 REVERSE LENGTH=661
MDS**AP**NFIPRLLLLSLLIVSIPLTSSQSDANTTN**PSPS**PPSDSDLCNGVFVSYTHTKGSKIPPN
DTANQPYRFESVITVLNHGRDELKSWRVFVKFAHREILVSASNAVLSDGSSLPVSVENGTVFAG
Y**PS**SDLKSAIQTAGDVTQMQARVELVGTQFGVAPPN**VP**LPKNITLATDGWKCPKATQKGTNVLQ
VCCIPDPDYDNREIIDNEFLPRKDGDLTIMYDVVRSYSSNYMAQVTMENHNPLGRLDNWKLSFD
WMRDEFIYTMKGAY**PS**IVDSSDCVDGPQAKHYQDLDFSNVLSCARR**PT**VIDLPPTKYNDSTFGL
IPFCCRNGTILPRSMD**PS**KSSSVFQMQVYKMPPDLNISALSPPQNWRINGTLNPDYKCGPPVRV
SPSQFVD**PS**GL**PS**NRTAFASWQVVCNITQPKDA**SP**RCCVSFSAYFNDSIVPCKTCACGCSSNKA
ARACSAT**APS**LLLPQQALLVPFENRTELTVAWAYLKQR**PVP**NPMPCGDNCGVSINWHLATDYRG
GWTARVTVFNWGETDFVDWFTAVQMKNA**AP**GFEKAYSFNASTIGINGKNNTIFMEGLPGLNYLV
AERDGENPLKNPRIPGKQQSVMSFTKKL**TP**GINVPGGDGF**PS**KVFFNGEECSL**PT**ILPMRSSQH
RKHISVFLLAL**PV**LALLILRA
>AT4G25240.1 | Symbols: SKS1 | SKU5 similar 1 | chr4:12930539-
12933563 FORWARD LENGTH=589
MAATCSLLASFLLCFALLSAVSFAADPFVSYDFRVSYLTA**SP**LGVPQQVIAVNGQFPGPLLNAT
TNYNVVVNVFNHLDEPLLLTWPGIQMRRNSWQDGVLGTNCPIPPRWNFTYQFQVKDQIGSFFY**S**
PSLNFQRASGGFGPIVINNRDIIPIPFPQPDGELIFIIGDWYTQDHKALRRALDSGKELGMPDG
VLINGKGPYKYNSSVPDGIDYLTFHVEPGKTYRIRVHNVGISTSLNFRIQNHSLLLVETEGHYT
SQANFTDFDVHVGQSYSFLVTMDQDATSDYYIVASARFVNETVWQRVTGVAILHYSNSKG**PV**SG
PL**PVP**KTDVS**SP**WSAMSQPKTIRQNTSASGARPNPQGSFHYGQINITNTYILRSLPPTIINGAL
RATLNGISFVN**PSTPV**RLADRNKVKGAYKLDFPDRPFNRPLRLDRSMINATYKGFIQVVFQNND
TKIQSFHVDGYSFFVVGMDFGIWSEDKKGSYNNWDAISRSTIEVYPGGWTAVLISLDNVGVWNI
RVENLDRWYLGEETYMRITNPEEDGKTEMDPPDNVLYCGALKNLQKEQHHSAATSILNGHLKLM
LLMVLLASVFRFC
>AT4G26690.1 | Symbols: SHV3, MRH5, GPDL2 | PLC-like
phosphodiesterase family protein | chr4:13456793-13459890
REVERSE LENGTH=759
MRGLLRASSLLLCGVILIQLLAAQIHAQSKKPK**SP**W**PT**LTGDPPLVIARGGFSGLFPDSSYDAY
NFAILTSVPDAVLWCDVQLTKDALGICFPDLTMRNSSSIEAVY**PT**RQKSYPVNGV**PT**SGWFTID
FSLKDLKDVNLIRGILSRSEKFDGNSNPIMTVQSVSTQMK**PS**FFWLNVQHDAFYAQHNLSMSSF
LVAASKTVLIDFIS**SP**EVNFFKKIAGRFGRNG**PS**LVFRFLGQDEFE**PT**TNRTYGSILSNLTFVK
TFASGILVPKSYILPLDDQQYLLPHTSLVQDAHKAGLEVFVSGFANDIDIAHDYSFDPVSEYLS
FVDNGNFSVDGVLSDFPITASASLDCFSHVGRNATKQVDFLVITKDGASGDYPGCTDLAYKKAI
KDGADVIDCSVQLSSDG**TP**FCLSSIDLGNSTTVSLTAFRNRSTTVPELGSLGAIYTFSLTWAEI
QTL**TPA**ISNPYRVTSLFRNPKQKNAGKLFSLSDFLSLAKNSTSLSGVLISVENAAYLREEQGLD
VVKAVLDTLTQTGYSNSTATKVMIQSTNSSVLVDFKKQSQYETVYKVEENIRDILDSAIEDIKK
FADAVVIQKLSVFPVAQSFITTQTNVVEKLQKSQLPVYVELFQNEFLSQPYDFFADATVEINSY
ITGAGINGTITEFPFTAARYKRNLCLGRKETIPYM**APA**QPGALLTLV**SPT**AFPPAE**AP**NPVFTD
ADVTEPPLPPVTAK**APT**S**SP**G**TPS**TNAQ**APS**GQTRITLSLLLSVFAMVLASLLLL
>AT4G28100.1 | Symbols: | unknown protein; FUNCTIONS IN:
molecular_function unknown; INVOLVED IN: biological_process
unknown; LOCATED IN: anchored to plasma membrane, anchored to
membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING:

14 growth stages; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT3G18050.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink). | chr4:13965300-13966697 REVERSE LENGTH=304

MKKSLTLLILLLCSLLFSTVLSNLLVEPVQPNT**VPA**F**PV**ETQAQSCRLDLSNELFGGVNEACGRNLDRSRCCPVLAAWLFAAHARSALQL**PAPAPTP**ESSDPDEPMKPDDSQKCVNTLQSALLTKQIKIPQPNSSCDAILCFCGIRLHQISSLSC**PA**AFNVSSGFKNA**TPT**AAVKNLEKECRNSSYSGCTRCLGALQKLKVRGGNKKTTTERGTKMMSKDCQLMGLTWLLARNKTAYI**PT**VSAVLRAIMY**SP**HPPHLNKC**SP**DQENMPLAVDSLQFQKSFSSSSHLFGVLPFLPLVLCIFLFLL

>AT4G31840.1 | Symbols: ENODL15, AtENODL15 | early nodulin-like protein 15 | chr4:15401798-15402426 FORWARD LENGTH=177

MASSSLLVTIFLCISVFFFSSVNANEVTVGGKSGDWKIPPSSSFSFNEWAQKARFKVGDFIVFKYEAGKDSVLQVTREAYEKCNTT**SP**KASYTDGNTKVKLDQAGPVYFVSGTEGHCQKGQKLRLVVI**TP**RNSAF**SP**G**PSPS**EFDG**PA**V**APT**SGAAKLAGGFSVVFGLVLGLWAFFF

>AT5G07190.1 | Symbols: ATS3 | seed gene 3 | chr5:2237610-2238488 FORWARD LENGTH=213

MTF**PS**LSVSFLFFAFIFVTHAFDLSIIQMQQGTCPYTVVVMTSCL**SP**ESTRDQISIVFGDADGNKVY**AP**KLGGLVRGPGGLGKCSTNTFQVRGQCLNDPICSLYINRNGPDGWVPESIEIYSEGSKSVKFDFSKSVPQLNTWYGHNNCNTTGR**PSSP**DLPPPHFPPEFPPE**TPT**TPPPPPPR**PS**AASRLGNGESVFLAFAIATAIAAMVRWSY

>AT5G25090.1 | Symbols: ENODL13, AtENODL13 | early nodulin-like protein 13 | chr5:8647117-8647755 REVERSE LENGTH=186

MAQRTLVATFFLIFFLLTNLVCSKEIIVGGKTSSWKI**PSSPS**ESLNKWAESLRFRVGDTLVWKYDEEKDSVLQVTKDAYINCNTTN**PA**ANYSNGDTKVKLERSGPYFFISGSKSNCVEGEKLHIVVMSSRGGHTGGFFTGS**SPSPAPSPA**LLG**APT**V**APA**SGGSASSLTRQVGVLGFVGLLAIVLL

>AT4G12500.1 | Symbols: | Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein | chr4:7414369-7414902 REVERSE LENGTH=177

MASKNSTSLALFFALNILFFTLTTATDCRCNL**SP**KPRT**VPSP**K**VPSP**KY**PSPS**I**PSPSVPTPSVPTPS**V**PTPS**V**PSP**N**PTPVTP**PR**TP**GSSGNCPIDALRLGVCANVLSGLLNVQLGQ**PS**AQPCCSLIQGLVDLDAAICLCTALRANVLGINLNVPISLSVLLNVCNRRL**PS**DFQCA

>AT1G18280.1 | Symbols: | Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein | chr1:6294403-6295207 FORWARD LENGTH=180

MEAVRFAVAVVLVFCYVTSSNAQMT**SPP**SGGAGGDAHSLPCIQKLMPCQPYLHLA**TPPP**ATCCMPLNEIVAKDATCLCAVFNNVDMLKSLNLTKENALDLPKACGAKADVSLCKTSAGTNSSS**TPPATP**K**TPPA**SSTSTGTGSGSTGNA**APS**TAK**PT**SS**APA**INFGGLSFASAVVATLFF

>AT2G48140.1 | Symbols: EDA4 | Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein | chr2:19686442-19687444 FORWARD LENGTH=200

MEGLTLIVVMMSSFMLGGQGQQIS**TP**CTSSMISTF**TP**CLNFITGSSGGSV**TPT**AGCCDSLKTLTNTGMGCACLILTANVPL**PT**GFINRTLALALPRACKMGGVPIQCQAAG**TP**L**PAP**GQVPFLI**AP**PPQVSAF**SP**GASKAAGT**TPT**Q**APAP**D**TPA**DG**PT**G**PT**TKSGIRPVDQPMQ**PT**GLAQSST**SP**FLPLLFISLILLNL

>AT4G12490.1 | Symbols: | Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein |

chr4:7409830-7410378 REVERSE LENGTH=182
MASKNSASLALFFALNILFFTLTAGTNCRCN**PSP**KPRPLPNPKV**PSP**KV**PTPS**V**PSP**Y**VPTPSV**
PSPSVPTPSVPSPSVPSPN**PTP**VIPPR**TP**GSSGNCPIDALRLGVCANVLSGLLNVQLGQ**PSP**QP
CCSLIQGLVDLDAAVCLCTALRANVLGINLNVPISLSVLLNVCNRRL**PS**NFQCA
>AT4G15160.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr4:8646192-8647019 FORWARD LENGTH=275
MALLHKQNISFVILLLLGLLAVSYACDCSDPPK**PSP**HPVKPPKH**PA**KPPKP**PT**VKP**PT**H**TP**KP**P**
TVKPPPPYIPCPPPPY**TP**KP**PT**VKPPPPPYVKPPPP**PT**VKPPPPPYVKPPPP**PT**VKPPPP**PTP**Y
TPPP**PTP**Y**TP**PP**PT**VKPPPPPVV**TP**PP**PTPTP**E**AP**CPPPP**PTP**YPPPPKPETCPIDALKLGACV
DVLGGLIHIGLGKSYAKAKCCPLLDDLVGLDAAVCLCTTIRAKLLNIDLIIPIALEVLVDCGK**T**
PPPRGFKC**PTP**LKR**TP**LLG
>AT1G03103.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr1:747219-747955 FORWARD LENGTH=171
MILAILALVIATFLYGGATTVQAGCRDTLTSL**SP**CLYYLNGGSS**SPS**WSCCRQFSTVVQS**SP**EC
LCSVVNSNESSFYGFKFNRTLALNL**PT**ACNVQ**TPSPS**LCNTGGNV**PT**TL**PA**N**TP**VG**SP**RS**APSP**
SGTT**SPA**N**TPS**GSKKFPLSNESSSKSNVIILSFVSIALVLAII
>AT1G05450.2 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr1:1600004-1601086 FORWARD LENGTH=205
MNSNSFLISAALIFSLLSSN**SPT**SILAQIN**TP**C**SPS**MLSSVTGCTSFLTGGGSF**PT**SDCCGALK
SLTGTGMDCLCLIVTAGVPISIPINRTLAISLPRACGIPGVPVQCKASA**AP**L**PTP**G**PA**SFG**PT**T
SPTDSQTSDPEGSASFRP**PTSPT**TSQ**TP**NDKDLSGSGNGGDPMGF**AP**PP**PS**S**SPS**SSHSLKLSY
LLFAFAFTIIKFI
>AT4G22470.1 | Symbols: | protease inhibitor/seed storage/lipid
transfer protein (LTP) family protein | chr4:11840316-11841443
REVERSE LENGTH=375
MASTTIILFLSFSIIPLLTIVRADNHSVYCPPPPPCICICNPGPPPPQPDPQPPTPPTFQ**PA**PP
ANDQPPPPPQSTSPPPVATTPP**A**LPPKPLPPPLSPPQTTPPPPPAITPPPPPAITPPLSPPPPA
ITPPPPLATTPPALPPKPLPPPLSPPQTTPPPPPAITPPLSPPLVGICSKNDTELKICAGILAI
SDGLLTTGRAEPCCSIIRNVSDLDAVTCFCKSVG**AP**RFSL**SP**NFGIFFKVCGRRIPQGFSCPG**P**
SPTISPPPLPPQTLKPPPPQTTPPPPPAITPPLSPPLVGICSKNDTELKICAGILAISDGLLTT
GRAEPCCSIVRNVSDLDAVTCFCKSVGARRFSLSPNFGIFFKVCGRRIPQGFSCP
>AT2G45180.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr2:18626377-18626781 FORWARD LENGTH=134
MASKALAVTALLITLNLLFFTFVTSTKCP**PTTP**KPPK**TP**K**SP**KK**APA**VK**PT**C**PT**DTLKLGVCAD
LLGLVNVVVGSPPK**TP**CCTLLQGLANLEAAVCLCTALKANVLGINLNVPIDLTLLLNYCGKKVP
HGFQCS
>AT1G27950.1 | Symbols: LTPG1 | glycosylphosphatidylinositol-
anchored lipid protein transfer 1 | chr1:9740740-9741991 FORWARD
LENGTH=193
MKGLHLHLVLVTMTIVASIAAA**APAAP**GGALADECNQDFQKVTLCLDFATGKATI**PS**KKCCDAV
EDIKERDPKCLCFVIQQAKTGGQALKDLGVQEDKLIQL**PT**SCQLHNASITNCPKLLGI**SPSSP**D
AAVFTNNATT**TPVAPA**GK**SPATPA**TSTDKGGSASAKDGHAVVALAVALMAVSFVLTLPRHVTLG
M
>AT4G12480.1 | Symbols: pEARLI 1 | Bifunctional inhibitor/lipid-

transfer protein/seed storage 2S albumin superfamily protein |
chr4:7406371-7406877 REVERSE LENGTH=168
MASKNSASIALFFALNIIFFTLTAATDCGCN**PSP**KHK**PVPSP**KPK**PVPSP**KPK**PVPSPS**V**PSPS**
VPSPNPRPVTPPR**TP**GSSGNCPIDALRLGVCANVLSSLLNIQLGQ**PS**AQPCCSLIQGLVDLDAA
ICLCTALRANVLGINLNVPISLSVLLNVCNRKV**PS**GFQCA
>AT4G22460.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr4:11839160-11839561 REVERSE LENGTH=133
MALKDSLALLLLFNILFFTLTTATRSTNCPPPPGKHNKQK**PSPTPT**TGTCPKDALKVGVCVNAL
NLLNGL**TP**GTPPV**TP**CCSLIEGLVDLEAAICLCTALKASVLGINLTLPINLSLLLNICNREASR
DFQFP
>AT3G57310.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr3:21208200-21208511 REVERSE LENGTH=103
MKFTKLLFVASVMIVM**SP**L**APT**RATVVGGWGIEEKAACIVTNLMSCL**PA**ILKGSQPPAYCCEML
KEQQSCLCGYIK**SPT**FGHYVIPQNAHKLLAACGILYPKC
>AT3G18280.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr3:6267102-6267392 FORWARD LENGTH=96
MVMIKTTMVSLFALAAVLLMIL**APA**AEAVTC**SP**MQL**SP**CATAITSS**SPPS**ALCCAKLKEQRPCL
CGYMRN**PS**LRRFVS**TP**NARKVSKSCKLPIPRC
>AT1G32280.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr1:11646220-11646816 FORWARD LENGTH=112
MEKQIFCQFLVVMMLLSSSQIQGDLRKGCYDLGITVLMGCPDSIDKKL**PAPPTPS**EGCCTLVRT
IGMKCVCEIVNKKIEDTIDMQKLVNVAAACGRPL**AP**GSQCGSYRVPGA
>AT4G14805.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr4:8502374-8503199 REVERSE LENGTH=219
MATKITGVFILILTITFSSSSAVTATQQ**APS**S**SP**PVLTCTEELVMF**SP**CLPYVS**SP**PNNMSE**TP**
DPICCSVFTSSVHSSTGNCLCYLLRQPMILGFPLDRSRLISLSQICTDQNSEESFESLCSVSE**S**
PELPPLQSIQFTNPFVSGNNVSA**SP**QSVDL**AP**EV**SPS**SDLF**SP**ETATL**AP**PPPPPPLPVLQYFS
SDSLKIRNFWF**PS**TIIMTFATSILARI
>AT5G23820.1 | Symbols: | MD-2-related lipid recognition domain-
containing protein | chr5:8031386-8032809 FORWARD LENGTH=164
MAMSHVQPMLLLLVSLFFL**PA**LRGAIDFEYCAKNGNDYGTVTSIVV**SPS**VGPHEN**PT**ITINLFG
SASKNI**PA**GTLVYVAFRDGEFTGLLKTYNLCDVSACNNEAEIEAGTNFELTLSDVLYVGYDEEI
KYSVSLRRKTLEEEDPIIKMCVDFKV**PAPAPA**FVSI
>AT5G23840.1 | Symbols: | MD-2-related lipid recognition domain-
containing protein | chr5:8035874-8036777 FORWARD LENGTH=167
MAMSHIQPVLFFLASLFFL**PA**LCSAIDFEYCTKNGHDYGSISQIWV**SPS**DGPQEN**PT**ITIHLFG
SASKDISAGTLVYVTYRSGDFTGLLKTYDLCDVSACNPEYVIEAGTDFELTLSDVLYVGFDEEI
KYSVSLREKTSEEKNPIIKMCVDFKVPH**PA**FPP**PA**LVSM
>AT4G12360.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr4:7328164-7329104 REVERSE LENGTH=161

MAQTTTLILLLATLLVAATTVSGQGPHIPL**APSPS**VNEAMNCAAGLAVCL**PA**ITQRGP**PS**QECC
TAVETALTTQLSCLCGFIK**SP**MLLIPFNVTDFNALFSKTCGLTTDPNLCSETAAQ**AP**LPKTA**AP**
VPG**AP**KSDKDAASKLAGTGLVGIVVITIAAMFY
>AT1G62510.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr1:23136632-23137081 REVERSE LENGTH=149
MASRTTKSLALFLILNFLFFTTISACGNCGC**PSP**KPKHK**PSPSP**KPKPNPKPK**PTP**H**PSPSPA**I
AKCPRDALKLGVCANVLNGLLNVTLGKPPVEPCCTLIQGLADLEAAACLCTALKANILGINLNI
PLSLSLLLNVCSKKVPRGFQC
>AT3G22620.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr3:8008615-8009415 FORWARD LENGTH=203
MSKIISLVVAMIAVLALPIRGQQQPLSQC**TPS**MMTTV**SP**CMGFITNSSSNGT**SPS**SDCCNSLRS
LTTGGMGCLCLIVTGTVPFNIPINRTTAVSLPRACNMPRVPLQCQANI**APA**A**AP**G**PA**ATFG**PS**M
SPGPETDPIVPE**PTPA**AQ**TP**QSDTTRPF**TPS**VDGG**APT**SDDGGSTSR**PS**E**TPS**SAYAL**SPS**LLF
FSIALVALKFY
>AT5G56480.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr5:22870326-22870817 REVERSE LENGTH=113
MKEQIFCQFLVVMMLLSSSQIQGERCNDSGIEVLRGCPDSIDKEL**PTP**PR**PS**QGCCTLVRIIGM
ECVCEVINKEIEAAIDMQKLVNVAAACGRPL**AP**GSQCGSYLVPGGMIRH
>AT3G52130.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr3:19332081-19332458 REVERSE LENGTH=125
MMMKAMRVGLAMTLLMTITVLTIVAAQQEGLQQPPPPPMLPEEEVGGCSRTFFSALVQLIPCRA
AV**AP**F**SP**IP**PT**EICCSAVVTLGRPCLCLLANGPPLSGIDRSMALQLPQRCSANFPPCDIIN
>AT1G43665.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr1:16452981-16453804 REVERSE LENGTH=123
MKFITTLMVIAFVLSVL**VPT**QALRVLSEDKKVACIVTDLQVCLSALE**TP**IP**PS**AECCKNLKIQK
SCLCDYMEN**PS**IEKYLE**PA**RKVFAACGMPYPREDAIEVKATHHPVDVV**PT**HAHTSDVHA
>AT1G73560.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr1:27649768-27650456 REVERSE LENGTH=147
MEINKFLAVVVAVVVLYSVEATAQGGNPQLTACLQKLLPCQPYIHSLNPPPP**PS**CCGPMKEIVE
KD**AP**CLCIAFNNPEVLKALNLTKENALLLPKACGVNPDVSLCSKIA**TPSP**IA**SP**GSTNGTSSAS
TISFNRFSFLSAFVAMIFF
>AT2G37870.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr2:15859280-15859723 FORWARD LENGTH=115
MKCCKFVAVALMSLLISLASVEAAGECGRMPINQAAASL**SP**CL**PA**TKNPRGKVPPVCCAKVGAL
IRTNPRCLCAVML**SP**LAKKAGINPGIAIGVPKRCNIRNR**PA**GKRCGRYIVP
>AT5G46890.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr5:19036437-19036820 REVERSE LENGTH=127
MAYSKVALLLVFNVIFFTFVSSTSVPCPPPPPKSYHKK**PATPS**LK**PT**CKDALKLKVCANVLDVV
KVSLP**PT**SNCCALIKGLVDLEAAVCLCTALKANVLGINLNVPISLNVVLNHCGKK**VPS**GFKCA
>AT5G46900.1 | Symbols: | Bifunctional inhibitor/lipid-transfer

protein/seed storage 2S albumin superfamily protein |
chr5:19039954-19040337 REVERSE LENGTH=127
MAYSKIALLLIFNVIFFTLVSSTSVPCPPPPPKSHHKK**PATPSP**K**PT**CKDALKLKVCANVLDLV
KVSLP**PT**SNCCALIKGLVDLEAAVCLCTALKANVLGINLNVPISLNVVLNHCGKK**VPS**GFKCA
>AT1G73780.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr1:27743872-27744168 FORWARD LENGTH=98
MKGSCTKPVFFTCTILLILIVAQENRVAAVDPCN**PA**QL**SP**CLETIMKGSE**PS**DLCCSKVKEQQH
CICQYLKNPNFKSFLN**SP**NAKIIATDCHCPYPKC
>AT5G48605.1 | Symbols: | Putative membrane lipoprotein |
chr5:19710215-19710578 FORWARD LENGTH=91
MKTIFFFITFIVLVSSCTSNIMTKSISQVKSQFF**SPA**L**SP**NVD**PA**DEHIGH**SP**DDMKIIFCQQC
AFHCIEKKKNIGNCENSICRCTLEDIL
>AT4G12470.1 | Symbols: AZI1 | azelaic acid induced 1 |
chr4:7401371-7401856 REVERSE LENGTH=161
MASKNSASLALFFALNILFFTLTVATNCNCK**PSP**KPK**PVPSP**KPK**PV**QCPPPPR**PSVPSP**NPRP
V**TP**PR**TP**GSSGNSCPIDALKLGVCANVLSSLLNIQLGQ**PS**SQQCCSLIQGLVDVDAAICLCTAL
RANVLGINLNVPISLSVLLNVCNRKL**PS**GFQCA
>AT1G73890.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr1:27787903-27788658 REVERSE LENGTH=193
MASSTLLITLLISLSAFFLRMVLAQV**PA**TCASRLLSL**AP**CGPFVQGFAQL**PA**QPCCDSLNQIYS
QEATCLCLFLNNTSTL**SPA**FPINQTLALQLPPLCNI**PA**NSSTCSSSFPGE**APS**DSSSV**AP**P**PS**S
STGSQISQGAKNNSRVAA**TP**VAQM**AP**R**PT**SFMGLGYGLKSSGSKSEIQLTIFALAAIL**PA**ALLL
I
>AT5G48490.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr5:19647932-19648237 REVERSE LENGTH=101
MTSKKVAIMVIVMVMASLVVERSVAIDLCGMTQAELNECL**PA**VSKNN**PTSPS**LLCCNALKHADY
TCLCGYKN**SP**WLGSFGVDPKLASSLPKECDLTN**APT**C
>AT1G62790.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr1:23252369-23253481 FORWARD LENGTH=150
MTKTMMIFAAAMTVMALLL**VPT**IEAQTECVSKLVPCFNDLNTTT**TPV**KECCDSIKEAVEKELTC
LCTIYT**SP**GLLAQFNVTTEKALGLSRRCNVTTDLSACTAKG**APSP**KASLPP**PAPA**GNTKKDAGA
GNKLAGYGVTTVILSLISSIFF
>AT1G48750.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr1:18036019-18036303 FORWARD LENGTH=94
MVKVMWVSVLALAAAILLLTVPVAEGVTC**SP**MQLASCAAAMTSS**SPPS**EACCTKLREQQPCLCG
YMRN**PT**LRQYVS**SP**NARKVSNSCKI**PSPS**C
>AT5G48485.1 | Symbols: DIR1 | Bifunctional inhibitor/lipid-
transfer protein/seed storage 2S albumin superfamily protein |
chr5:19646317-19646625 REVERSE LENGTH=102
MASKKAAMVMMAMIVIMAMLVDTSVAIDLCGMSQDELNECK**PA**VSKEN**PTSPS**QPCCTALQHAD
FACLCGYKN**SP**WLGSFGVDPELASALPKQCGLAN**APT**C
>AT3G44100.1 | Symbols: | MD-2-related lipid recognition domain-
containing protein | chr3:15866162-15867273 REVERSE LENGTH=152

MAISQAQPLLLLLLSVFFL**PA**LHATSFTYCDKRLDPVKVTGVKI**SP**DPVVSGAAATFKIFGSTGEDISGGKVVIRVLYVGIPVHTETHDLCDETAC**PVAP**GSFVLSHSQTL**PS**I**TP**PGTYTLKMTINDKNGGRLTCISFKFKITVGSAVFAS

>AT3G59455.1 | Symbols: | Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein | chr3:21976840-21977130 FORWARD LENGTH=96

MKNISLMFIALVVLLTSF**PTPT**LSYCKESLHLCMQHLKLNDR**PT**WLKCCDRLIIPGPCMCKYIKDPVQWKEAYRLMASCGKTVPLNQSLKSYFKCG

>AT5G38160.1 | Symbols: | Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein | chr5:15225836-15226147 FORWARD LENGTH=103

MKFTGVVFILFVLGTML**SPVPV**KARVVKGSGEEVNVTCDATQLSSCVTAVSTG**APPS**TDCCGKLKEHETCLCTYIQNPLYSSYVT**SP**NARKTLAACDVAY**PT**C

>AT5G17340.1 | Symbols: | Putative membrane lipoprotein | chr5:5715736-5716218 REVERSE LENGTH=160

MKNLAILVVAMILFSSCVTSQVTAKNVDSFPLRTEELEWWHYNPFYPHFYPKPHWPF**PT**TGKALPPI**PA**GFHPIPFHPPPVVTKCLADCKDVKTCLADIAKAFFTRK**PA**IGLDCCASIQKMDEDCDKTVFGAYHNPFFDCVVKLHCSTKAKS**TPSAPTPA**

>AT4G22490.1 | Symbols: | Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein | chr4:11849933-11850295 REVERSE LENGTH=120

MASKSSTTISLIIILLISLAEANLLS**SPTPT**NNFGSCPRNPLQLGVCANVLGLANVTAGDPRARQCCTALNGLTNVQVTDCLCFIFRPIPLVFGIDVAVREIFFACNRVFPIGFQCPPPQ

>AT2G26370.1 | Symbols: | MD-2-related lipid recognition domain-containing protein | chr2:11224636-11226052 FORWARD LENGTH=173

MAISHAQPVLLLLSSLYFLSAFGAGAYNFENCKN**AP**IDYNYGITNVTRVEI**SP**YPVGPYDE**PT**ITISGFTSDDSYIIYRATIHVLYKYENVNSTIINYDLSDVMGEDPCSIEPGEKFVLTLSKVPGLQSLPHKDKSKIVISLVDEYGDDAEVPLLKMCVEFDN**PAPT**TTSVSA

>AT1G12100.1 | Symbols: | Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein | chr1:4095246-4095845 FORWARD LENGTH=132

MTFYKKNSEALFISFNLMFFAHVSGCNTCFP**PTP**IPNLNPISN**PTTPS**CSRDAIKLGVCAKILDVAVGTVIGN**PS**DTLCCSVLQGLVDLDAAVCLCTTIKANILGINIDLPISLSLLINTCGKKL**PS**DCICA

>AT5G38197.1 | Symbols: | Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein | chr5:15251024-15251314 FORWARD LENGTH=96

MKFVKLMVITFVIVVISSLILIKSEVAQSWI**PS**DLRLLCL**PA**MTTGGQ**PT**KDCCNTLIGQKESLCGYITNPLYYLFFT**SPA**ARKVLEVCNIPY**PT**C

>AT5G38195.1 | Symbols: | Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein | chr5:15246991-15247278 FORWARD LENGTH=95

MKFMKLMVITVVIVVMS**SP**ILIKSEVSSSCI**PT**ELMPCL**PA**MTTGGQ**PT**KDCCDKLIEQKECLCGYINNPLYSTFVS**SPV**ARKVLEVCNIPY**PS**C

>AT5G55450.1 | Symbols: | Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein | chr5:22467560-22467874 FORWARD LENGTH=104

MGKDNTRILMQFSALAMVLTAAIMVKEATSIPVCNIDTNDLAKCR**PA**VTGNNPPPPGPDCCAVA
RVANLQCLCPYKPYL**PT**VGID**PS**RVRPLLANCGVN**SPS**CF
>AT4G22666.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr4:11917161-11917733 FORWARD LENGTH=160
MAYTNQISAVVFLAVAI**AP**LLAEPQSTMFPEM**TP**ECATVMPDLLEKCFATGSV**TPT**EDCCTDLK
SATSTQVTCLCDNYIAN**PA**VSNITGPYSKAITTKCGVFDKYSCDGTSKGGEEKKGGSSSSNGKD
NGKSEGNGGRANSVAASMAMFGLLASLVFVMF
>AT3G51590.1 | Symbols: LTP12 | lipid transfer protein 12 |
chr3:19135828-19136654 REVERSE LENGTH=119
MAF**TP**KIITCLIVLTIYMA**SPT**ESTIQCGTVTSTLAQCLTYLTNSGPL**PS**QCCVGVKSLYQLAQ
T**TP**DRKQVCECLKLAGKEIKGLNTDLVAAL**PT**TCGVSIPYPISFSTNCDSISTAV
>AT5G38170.1 | Symbols: | Bifunctional inhibitor/lipid-transfer
protein/seed storage 2S albumin superfamily protein |
chr5:15227717-15228028 FORWARD LENGTH=103
MKFTALVFIVFVVGVMV**SP**VSIRATEVKLSGGEADVTCDAVQLSSCA**TP**MLTGVP**PS**TECCGKL
KEQQPCFCTYIKDPRYSQYVGSANAKKTLATCGVPY**PT**C
>AT1G24520.1 | Symbols: BCP1 | homolog of Brassica campestris
pollen protein 1 | chr1:8688699-8689058 FORWARD LENGTH=119
MGRQNIVVVVALVFMAIIGLAAAAS**SPSPS**A**SPS**K**APA**ASKTDHVE**AP**VTDDQIGTTDDDA**APT**
PGDGDVAVAGPLGSDSSYDN**APT**GSADSAKSGAAALGVSAVVVGVTSIAGSFLLL
>AT1G45063.1 | Symbols: | copper ion binding;electron carriers |
chr1:17033335-17034886 REVERSE LENGTH=369
MATARMKKIFSFVIVIFTLLFGCCSATVYKVGDSDGWTAKDHLYYHWTEDKEIHVGDSLIFEYD
HNLNDVTQVSGGLEYEFCDSSFPKAVYNTGHDVVTFTEPGSYYFITSNHTQCTSGQRLGVFVVH
D**PSSPSP**LPL**PS**KII**PS**RHVYKVGDSKSWGVYDSDFYYNWSKEKQFNVGDGLLFEYNNEVNGVY
EISGDLEFLNCD**PTSP**IAVHKTGHDIIKLTKPGIHYFISSEPGHCGAGLKLQVVVGTTLNVPKL
SPLERLTRNRLHICDRFISWGIQV**PS**LCLLCNALDETRQHVFFDCPFSHEVWSFFCSNARV**TP**P
RMFKDSARWLRHPCRDKKVAFILKLAYQASVYHIWRERNIRLNSNKSFP
>AT1G08500.1 | Symbols: ENODL18, AtENODL18 | early nodulin-like
protein 18 | chr1:2689110-2689881 FORWARD LENGTH=228
M**SPS**CSSCVNVLLIMCLMLLSLSADAYKNYTVGESTGWFDIQER**PS**ANYQKWADSKSFSLGDFL
IFNTDSNHSVVQTYDFKTYKDCDYDNNENNDTTEWSAAN**PS**AT**SPVPV**SIS**VP**LVKEGSNYFFS
GNYDGEQCKFGQHFMINVTHGQGLPDSS**SP**DDAA**AP**G**PS**ESSQSGDDEV**AP**DTIV**PA**NFDHPKD
IESADDDKEVHSKKSSSSTTKTSLFCFVFMGLFASF
>AT5G57920.1 | Symbols: ENODL10, AtENODL10 | early nodulin-like
protein 10 | chr5:23453634-23454256 FORWARD LENGTH=182
MSSVMMCCCLLLLFGLLSEGREILVGGKSNTWK**AP**ESRDETLNQWSGRTRFKIGDSLLWKYNAE
NDSVLQVRQTDYERCDRSEPIRGYKDGHTNIELKRSGPFYFISGEEGHCQRGEKLRVVVL**SP**NH
NRSVVD**APAP**VNIVL**SP**NYNRSVAA**AP**LNAHIMNKGSLNTAWSLLLLLPLGLLV
>AT1G22480.1 | Symbols: | Cupredoxin superfamily protein |
chr1:7934232-7935054 REVERSE LENGTH=174
MSTLLGCLVLIFSMVAQASSASLTVNWSLGTDY**TP**LTTGKTFSVGDTIVFNYGAGHTVDEVSEN
DYKSCTLGNSITSDSSGTTTIALTTTGPRYFICGIPGHCAAGMKLAVTVASNSSNGVAGGTT**TP**
TPFTGGGGGYN**PT**TTQAIPCAAWAVSCPLRALVATWAVVFYALALS
>AT5G07475.1 | Symbols: | Cupredoxin superfamily protein |
chr5:2364827-2365536 REVERSE LENGTH=192
MKKTSKIQFLFNLCIIFGVVVIRRCNATTYFVGDSSGWDISSDLESWTSGKRF**SP**GDVLMFQYS

STHSVYEVAKDNYQNCNTTDAIRTFTNGNTTVALSKPGNRFFVCGNRLHCFAGMRLLVNVEGNG
PSQ**AP**VG**SP**QAATSGILQ**PS**SKKNN**PA**TGVASSAARFVGDSGWRGTMGIFVYFMVFAFPFIWFC

>AT5G20230.1 | Symbols: ATBCB, BCB, SAG14 | blue-copper-binding protein | chr5:6826626-6827408 FORWARD LENGTH=196
MAGVFKTVTFLVLVFAAVVVFAEDYDVGDDTEWTRPMDPEFYTTWATGKTFRVGDELEFDFAAG
RHDVAVVSEAAFENCEKEKPISHMTVPPVKIMLNTTGPQYFICTVGDHCRFGQKLSITVVAAGA
TGGA**TP**GAGA**TPAP**GS**TPS**TGGT**TPPT**AGGTT**TPS**GSSGTT**TPA**GNAASSLGGATFLVAFVSAV
VALF

>AT1G64640.1 | Symbols: ENODL8, AtENODL8 | early nodulin-like protein 8 | chr1:24022482-24023151 REVERSE LENGTH=191
MGVMSLSKTMVVVVLQVMILLGQEIGKVSSTLYKVGDLDAWGIPIDAKVYSKWPKSHSFKIGDS
LLFLYP**PS**EDSLIQV**TPS**NFKSCNTKDPILYMNDGNSLFNLTQNGTLYFTSANPGHCTKYQKLL
VSVGTYSAEAEAL**SPS**SAAD**APS**YQNAFGSIPLSQKSSASSSLISAFSTVAASLACAVVGAIM

>AT1G79800.1 | Symbols: ENODL7, AtENODL7 | early nodulin-like protein 7 | chr1:30018549-30019217 FORWARD LENGTH=192
MMMMMMRSTCNLTLMLCICALVVASMAAEGPRDFKVGDEFGWRVPLQNDSAVYSHWASSNRFHI
GDSLSFVYDKDSVMEVDKWGFYHCNGSDPITAFDNGNSTFDLDRPGLFYFISGSNQHCTSGQRL
IVEVMHIHQHHDHDASMP**PS**M**SP**LSNSA**SP**YASASASSAASSL**PT**ACLLIPLFLTIASFRFISY

>AT1G23010.1 | Symbols: LPR1 | Cupredoxin superfamily protein | chr1:8147353-8149580 FORWARD LENGTH=581
MESLLCRRRIKRVMVLIIALTWLRSTCGELEDQLFEVGKLKMFVDDLPDMPRLYGFNSVHGIIK
PASLQIGMFSTKWKFHRDL**PATP**VFAYGTSRSKATVPG**PT**IETVYGVDTYVTWRNHLPKSHILP
WD**PT**I**SPATP**KHGGI**PT**VVHLHGGIHE**PT**SDGNADAWFTAGFRETGPKWTKTTLHYENKQQPGN
MWYHDHAMGLTRVNLLAGLVGAYILRHHAVE**SP**FQL**PT**GDEFDRPLIIFDRSFRKDGSIYMNAT
GNN**PS**IHPQWQPEYFGDVIIVNGKAWPRLNVRRRKYRFRIINASNARFFKFFFSNGLDFIVVGS
DSAYLSKPVMTKSILL**SPS**EIVDVVVDFYK**SPS**RTVVLAND**AP**YPY**PS**GDPVNEENGKVMKFII
NNESEDDTCTIPKKLINYPNADVSNAVLTRYISMYEYVSNSDE**PT**HLLVNGLPYE**AP**VTE**TP**KS
GTTEVWEVINLTEDNHPLHIHLGLFKVVEQTALLAAGLEEFKECMTKQNDAVKCQISKYARGKK
TAVTAHERGWKNVFKMMPGHVTRILVRFSYIHTNASYPFD**PT**QEPGYVYHCILDHEDNMMMRP
LKVII

>AT1G71040.1 | Symbols: LPR2 | Cupredoxin superfamily protein | chr1:26797201-26800224 REVERSE LENGTH=581
ME**PS**RRRMTRDMLLLIVTMAWLVTGDEGGIKQEERLFNLGKLEMFVDKLPHI**PT**LHGYHFVNGF
LKPKSLHIGMFFKKWKFHRDL**PATPV**FAYGTSKRSATVPG**PT**IEAVYGVDTYVTWRNHLPLHHI
LPWD**PT**I**SPA**IPKHGGI**PTV**VHLHGGIHE**PT**SDGNADSWFTAGFKETGSKWTKKTTHYVNKQQP
GNMWYHDHAAGLTRVNLLAGLLGSYILRHSSVE**SP**LRL**PT**GREFDRPLVIFDRSFRKDGSIYMN
ATGNN**PT**IHPQWQPEYFGDAIIVNGKAWPRLTVRRRKYRFRITNASNARFFRFFFSNGLDFIVV
GSDSAYLAKPVSTKSVLL**APS**EIVDVLVDFSKSTSKTAILANN**AP**YPY**PS**GDPVTEENSKVMKF
IINYKSEVDTSIIPKKLIEYP**PA**HVSTSTRTRYIAMFEYVSSIDE**PT**HLYINGLPYN**AP**VTE**TP**
KIGTSEVWEVINLTEDNHPLHIHLGLFKVLEQTALVKSEEFIECMTKRNDAVKCEISKYARGNK
TAVTVHERGWKNVFKMMPGHVTKILVRFSYIHSNESYSFDATQEPGYVYHCILDHEDNMMMRP
FAMVL

>AT2G43800.1 | Symbols: | Actin-binding FH2 (formin homology 2) family protein | chr2:18145721-18148721 FORWARD LENGTH=894
MTTIPFCFLFVAFFFSSSTADQRHHSRHLLHQPFF**PV**VTA**AP**PPYQPPVSSQP**PSPSP**HTHHHH
KKHLTTT**TP**PHEKHLFSSVANPPPPP**PSP**PHPNPFF**PS**SD**PT**STASHPP**PAP**PP**PA**SL**PT**F**PA**
NISSLLF**PT**HNKQSKP**PS**NGHIARLVTITASVISAAALLSLFAVFIIFIRRTRHRRRS**SPA**DDT
KSTRSDALQLFNA**SPS**DGSKKQKQHQQPPKYTSSHTSSEFLYLGTLVNSRSNGLEQQK**SP**ISLS

GGITGVLELPP**PA**SSSSSSSYSQYHKLG**SP**ELRPLPPLPKLQSF**TP**VYKSTEQLNPKRQDFDGD
DNENDEFF**SP**RGSSGRKQ**SPT**RVSDVDQIDNRSINGSGSNSC**SPT**NF**APS**LNA**SP**GTSLKPKSI
SPPVSLHSQISSNNGIPKRLC**PA**RPPPPPPPPPQVSEV**PA**TMSHSLPGDDSDPEKKVETMKPKL
KTLHWDKVRASSSRVMVWDQIKSNSFQVNEEMIETLFKVND**PT**SRTRDGVVQSVSQENRFLDPR
KSHNIAILLRALNVTADEVCEALIEGNSDTLGPELLECLLKM**APT**KEEEDKLKELKDDDDG**SPS**
KIG**PA**EKFLKALLNIPFAFKRIDAMLYIVKFESEIEYLNRSFDTLEAATGELKNTRMFLKLLEA
VLKTGNRMNIGTNRGDAHAFKLDTLLKLVDIKGADGKTTLLHFVVQEIIKFEGARVPF**TPS**QSH
IGDNMAEQSAFQDDLELKKLGLQVVSGLSSQLINVKKAAAMDSNSLINETAEIARGIAKVKEVI
TELKQETGVERFLESMNSFLNKGEKEITELQSHGDNVMKMVKEVTEYFHGNSETHPFRIFAVVR
DFLTILDQVCKEVGRVNERTVYGSMPLH**SPS**NQTA**TP**LFPVVINNNSRL**SPS**GSLDDDDGSF
>AT4G15200.1 | Symbols: AFH3, FH3 | formin 3 | chr4:8662993-
8665759 REVERSE LENGTH=764
MGRLRLAFLAISLVVFVCVSEEIFSRGGLNLLRFSVYGEDVAEQTWIHQNPRRKLISYPKKFSV
S**AP**NLAFG**PAPS**F**AP**GPG**PS**F**AP**G**PAP**NPRSYDWL**APA**S**SP**NEP**PA**E**TP**DES**SPSPS**EE**TPS**VV
APSQS**VP**GPPRPPPQREKKDDILMKLIIAVASTAVLTFVFVALMFLCCFKRNCNNAVGSRDGPR
DEGPLLRLSTGSTEN**SPT**VASTSRKMFSVASSKKRSFLSRVSLKRNGHEFSTAESSSAAGLPPL
KLPPGRS**AP**PPP**PA**A**AP**PPQPPPPPPPKPQPPPPPKIARPP**PAP**PKGA**AP**KRQGNTSSGDASDV
DSETG**AP**KTKLKPFFWDKMANPDQKMVWHEISAGSFQFNEEAMESLFGYNDGNKNKNGQKSTDS
SLRE**SP**LQYIQIIDTRKAQNLSILLRALNVTTEEVVDAIKEGNELPVELLQTLLKM**APT**SEEEL
KLRLYSGDLHLLG**PA**ERFLKILVDIPFAFKRIESLLFMISLQEEVSGLKEALGTLEVACKKLRN
SRLFLKLLEAVLKTGNRMNVGTFRGDAQAFKLDTLLKLSDVKGTDGKTTLLHFVVLEIIRSEGV
RALRLQSRSFSSVKTDDSNADSKLEDVKRAAIIDADGLAATLANISGSLTNAREFLKTMDEESD
FERALAGFIERADADFKWLKEEEERIMVLVKSSADYFHGKSAKNEGLRLFAIVRDFLIMLEKVC
REVKETTKTTNHSGKKESEMTTSDSNQ**PSP**DFRQRLF**PA**IAERRMDSSDDSDDEEDS**SPS**
>AT1G24150.1 | Symbols: ATFH4, FH4 | formin homologue 4 |
chr1:8549518-8551910 FORWARD LENGTH=725
MAAMLMQPWPPFLPHLTLVFLTLILFFPNQSFSQSD**SP**RNIETFFPNDTI**TP**PVQ**SP**VL**SP**PQN
PSSSSSDSDRGNILRAVLITAASTLLVAAVFFFLVHKCRRRRNRVGGVDNTLQPPVPPLAEAAL
AREGFTRFGGNVKGLILDENGLDVLYWRKLQQSQRDNKGGSFRKEIIHGDDEEKNVIYSKSKKK
SGPVTE**TP**LLRGRSSTSHSVIHNDNYRNATTTHPPHVKTDSFEFVKPD**PTP**PPPPPPPPIPVKQS
A**TP**PPPPPPPKLKNNG**PSP**PPPPPLKKTAALSSSASKKPP**PAP**RGSSSGESSNGQVKLKPLHWDK
VNPDSDHSMVWDKIDRGSFSFDGDLMEALFGYVAVGKK**SP**DDGGDKK**PS**SA**SPA**QIFILDPRKS
QNTAIVLKSLGMTRDELVESLMEGHDFHPDTLERLSRI**APT**KEEQSAILQFDGDTKMLADAESF
LFHLLKAVPCAFTRLNALLFRANYYPEISNHNKNLQTLDLACTELRSRGLFSVDGKTTLLNFVV
EEVVRSEGKRCVLNRRTNRSFSRSSSSSISEVISKEEQEKEYLRLGLPVVGGLSSEFTNVKKAA
AVDYDTVAATCLALTSRAKDARRVLAQSEGDNKEGVRFVKKMNEFLDSVEEEVKLAKEEEKKVL
ELVKRTTEYYQAGAVKGKNPLHLFVIVRDFLAMVDKVCVEIARNLQRRSSMGSTQQRNAVKFPV
LPPNFMSDRSRSDSGGSDSDM
>AT3G25500.1 | Symbols: AFH1, FH1, AHF1, ATFH1 | formin homology
1 | chr3:9251320-9254826 REVERSE LENGTH=1051
MLFFLFFFYLLLSSSSDLVFADRRVLHEPFFFPID**SP**P**PSPPSP**PPLPKLPFSST**TPPS**SSDPNA
SPFFPLY**PSSP**PPP**SPA**SFASF**PA**NISSLIVPHATK**SP**PNSKKLLIVAISAVSSAALVALLIAL
LYWRRSKRNQDLNFSDDSKTYTTDSSRRVYPPP**PA**T**APPT**RRNAEARSKQRTTTSSTNNNSSEF
LYLGTMVNQRGIDEQSLSNNGSSSRKLE**SP**DLQPLPPLMKRSFRLNPDVGSIGEEDEEDEFY**SP**
RGSQSGREPLNRVGLPGQNPRSVNNDTISCSSSSSG**SP**GRSTFISI**SPS**M**SP**KRSEPKPPVIS**T**
PE**PA**ELTDYRFVR**SPS**LSLASLSSGLKNSDEVGLNQIFR**SPT**VTSLTT**SP**ENNKKEN**SP**LSSTS
T**SP**ERRPND**TP**EAYLR**SPS**HSSAST**SP**YRCFQK**SP**EVL**PA**FMSNLRQGLQSQLLS**SPS**NSHGGQ
GFLKQLDALRSR**SPS**SSSSSVCS**SP**EKASHK**SP**VT**SP**KLSSRNSQSLSS**SP**DRDFSHSLDV**SP**R

ISNI**SP**QILQSRVPPPPPPPPPLPLWGRRSQVTTKADTISRP**PS**L**TPPS**HPFVI**PS**ENLPVTS**S**
PME**TP**ETVCASEAAEE**TP**KPKLKALHWDKVRASSDREMVWDHLRSSSFKLDEEMIETLFVAKSL
NNKPNQSQT**TP**RCVL**PSP**NQENRVLDPKKAQNIAILLRALNVTIEEVCEALLEGNADTLGTELL
ESLLKM**APT**KEEERKLKAYNDD**SP**VKLGHAEKFLKAMLDIPFAFKRVDAMLYVANFESEVEYLK
KSFETLEAACEELRNSRMFLKLLEAVLKTGNRMNVGTNRGDAHAFKLDTLLKLVDVKGADGKTT
LLHFVVQEIIRAEGTRLSGNNTQTDDIKCRKLGLQVVSSLCSELSNVKKAAAMDSEVLSSYVSK
LSQGIAKINEAIQVQSTITEESNSQRFSESMKTFLKRAEEEIIRVQAQESVALSLVKEITEYFH
GNSAKEEAHPFRIFLVVRDFLGVVDRVCKEVGMINERTMVSSAHKFPVPVNPMMPQPLPGLVGR
RQSSSSSSSSTSSSDEDEHNSISLVS
>AT3G05470.1 | Symbols: | Actin-binding FH2 (formin homology 2) family protein | chr3:1579667-1582547 REVERSE LENGTH=884
MVYFRQIFLMIIVVSLHCCKVRFFCIVANAKELDDWKVLTVENGERYRTHVGRYAGEEGGEKIK
LRVLEKFRALLDLIK**PS**TSRRRNLAESASF**SP**W**PAPSPSP**FPNGGPIE**SPA**YP**PAP**PRPIPPHL
RRPLPQRTHPLEQPEIQRRKHEKGGTFKKILVPVVASTASAIGFVVCVVGVFCLCARRKRKMNG
KTLSFKRKKGKSQSSTRKVSVN**PT**LDFLYLNSLGVDLERQNSVSVKEIRETEKDLNGINGGLLE
EEVKRSIETEISHDWDNASSYSTKEIVSVHENDEEQTVNSVSVPVVVINDSSDDDESFHSVGGG
SQYSNPRLSNASSASGSVNVGSSQRFSEHKLDIPECSRSDFGISVS**AP**PPPPPPPPPLPQFSNK
RIHTLS**SP**ETANLQTLSSQLCEKLCASSSKTSFPINVPNSQPRPPPPPPPPPQQLQVAGINK**TP**P
PPLSLDFSERRPLGKDG**AP**LPKLKPLHWDKVRA**TP**DRTMVWDKLRTSSFELDEEMIESLFGYTM
QSSTKNEEGKSK**TPSP**GKHLLEPKRLQNFTILLKALNATADQICSALGKGEGLCLQQLEALVKM
V**PT**KEEELKLRSYKGAVDELGSAEKFLRALVGVPFAFQRAEAMLYRETFEDEVVHLRNSFSMLE
EACKELKSSRLFLKLLEAVLKTGNRMNVGTIRGGAKAFKLDALLKLSDVKGTDGKTTLLHFVVQ
EISRSEGIRVSDSIMGRIMNQRSNKNR**TP**EEKEEDYRRMGLDLVSGLNTELRNVKKTATIDLEG
LVTSVSNLRDGLGQLSCLASEKLKGDEENRAFVSSMSSFLRYGEKSLEELREDEKRIMERVGEI
AEYFHGDVRGDEKNPLRIFVIVRDFLGMLDHVCRELRCVRVPN**SPSP**L**AP**FR
>AT5G54650.1 | Symbols: Fh5, ATFH5 | formin homology5 | chr5:22197856-22201649 REVERSE LENGTH=900
MVGMIRGGMGDQNWSRLVFWLILFSGLLVITLEENPEKDEIFLSQFM**APS**TGQVNEHMEETSWA
QRCWQDSDCVKEAVAEFNLCFPGSKDSRELFGLNHTNLKQTLLDCIQEKGKLNGHNPKYLELLS
SMLDIPRRNLATKPGS**SPSPSPS**RPPKRSRGPPRP**PT**RPK**SP**PPRKSSFP**PS**R**SP**PPP**PA**KKNA
SKNSTS**AP**V**SPA**KKKEDHEKTIIIAVVVTAVSTFLLAALFFLCCSRVCGNGSGGRKNDERPLLS
LSSSDYSVGSSINYGGSVKGDKQGHQSFNIYSNQGKMSSFDGSNSDTSDSLEERLSHEGLRNNS
ITNHGLPPLKPPPGRTASVLSGKSFSGKVEPLPPEPPKFLKVSSKKAS**AP**PPPV**PAP**QM**PS**SAG
PPRPPP**PAP**PPGSGGPKPPPPPGPKGPRPPPPMSLGPK**AP**RP**PS**G**PA**DALDDD**AP**KTKLKPFFW
DKVQANPEHSMVWNDIRSGSFQFNEEMIESLFGYAAADKNKNDKKGSSGQAALPQFVQILEPKK
GQNLSILLRALNATTEEVCDALREGNELPVEFIQTLLKM**APTP**EEELKLRLYCGEIAQLGSAER
FLKAVVDIPFAFKRLEALLFMCTLHEEMAFVKESFQKLEVACKELRGSRLFLKLLEAVLKTGNR
MNDGTFRGGAQAFKLDTLLKLADVKGTDGKTTLLHFVVQEIIRTEGVRAARTIRESQSFSSVKT
EDLLVEETSEESEENYRNLGLEKVSGLSSELEHVKKSANIDADGLTGTVLKMGHALSKARDFVN
SEMKSSGEESGFREALEDFIQNAEGSIMSILEEEKRIMALVKSTGDYFHGKAGKDEGLRLFVIV
RDFLIILDKSCKEVREARGRPVRMARKQGSTASASSE**TP**RQ**TPS**LDPRQKLF**PA**ITERRVDQSS
SDSD
>AT5G48360.1 | Symbols: | Actin-binding FH2 (formin homology 2) family protein | chr5:19595716-19598331 FORWARD LENGTH=782
MQNFWFAIFFFLLTC**APPSP**LSYASTVTLSRRLLYDYE**SP**LPLPL**SP**I**SP**PFFPLES**SPPSP**PP
PLP**PTPPT**TFAVF**PT**F**PA**NISALVLPRSSKPHHT**SPT**LLL**PA**LSAVLVIATVIGLALFLYGRHR
GQTRHLKNSHCSSSNTSSYGDEQSHITTNFNMAATT**SPS**EVFYLNTEESDHIRTGGTFFLKQD**S**
PEIRPLPPLPPRSFHHNNYETEVNEEDEEEEEDVFF**SP**MASLPGSANS**SPS**HSCSSSCSGWV**SP**

ARSFSITM**SP**PNPRYSDATNLQ**SPSP**ERLRVRKNYNGNGSSSLRMFSFWNQNMGFGFPRISSAS
T**SP**DRGFIR**TP**LSSLYSSVST**SP**DGLFRKFLDS**SP**PIWNDFSRNVKSVLLSHTASSRRDFVINI
GESSSQQSKV**PA**LPP**PT**RPPPLVP**PS**QPFVVQNDVKKQSFSDQPPKQLHWERLRSSSSKLSKEM
VETMFIANSSNPRDLPIQNQVLDPRKAQNIATLLQLLNLSTKDVCQALLDGDCDVLGAELLECL
SRL**APS**KEEERKLKSFSDGSEIG**PA**ERFLKELLHVPFVFKRVDALLFVANFHSEIKRLRKSFSV
VQVACEELRNSRMFSILLEAILKTGNMMSVRTNRCGDADAFKLDTLLKLVDVKGLDGRSSLLHF
VVQEMMKSEGSVRALEGIRNLNTELSNVKKSADIEYGVLRSNVSRICQGLKNIEALLLLSEESG
SYGDQWLKFKERMTRFLKTAAEEIVKIKIRESSTLSALEEVTEQFHGDASKEGHTMRIFMIVRD
FLSVLDQVCKEMGD
>AT1G70140.1 | Symbols: ATFH8, FH8 | formin 8 | chr1:26412688-26415048 REVERSE LENGTH=760
MAAMFNHPWPNLTLIYFFFIVVLPFQSLSQFD**SP**QNIETFFPISSL**SP**VPPPLLP**PS**SN**PSPPS**
NNSSSSDKKTITKAVLITAASTLLVAGVFFFCLQRCIIARRRRDRVGPVRVENTLPPYPPPPMT
SAAVTTTTLAREGFTRFGGVKGLILDENGLDVLYWRKLQSQRERSGSFRKQIVTGEEEDEKEVI
YYKNKKKTEPVTEIPLLRGRSSTSHSVIHNEDHQPPPQVKQSE**PTP**PPPP**PS**IAVKQS**APTPSP**
PPPIKKGS**SPSP**PPPPPVKKVGALSSSASKPP**PAP**VRGASGGETSKQVKLKPLHWDKVNPDSDH
SMVWDKIDRGSFSFDGDLMEALFGYVAVGKK**SP**EQGDEKNPKSTQIFILDPRKSQNTAIVLKSL
GMTREELVESLIEGNDFVPDTLERLARI**APT**KEEQSAILEFDGDTAKLADAETFLFHLLKSV**PT**
AFTRLNAFLFRANYYPEMAHHSKCLQTLDLACKELRSRGLFVKLLEAILKAGNRMNAGTARGNA
QAFNLTALLKLSDVKSVDGKTSLLNFVVEEVVRSEGKRCVMNRRSHSLTRSGSSNYNGGNSSLQ
VMSKEEQEKEYLKLGLPVVGGLSSEFSNVKKAACVDYETVVATCSALAVRAKDAKTVIGECEDG
EGGRFVKTMMTFLDSVEEEVKIAKGEERKVMELVKRTTDYYQAGAVTKGKNPLHLFVIVRDFLA
MVDKVCLDIMRNMQRRKVG**SP**I**SPS**SQRNAVKFPVLPPNFMSDRAWSDSGGSDSDM
>AT5G67470.1 | Symbols: ATFH6, FH6 | formin homolog 6 | chr5:26926835-26930212 FORWARD LENGTH=899
MKALQSRFFFFFFYIFFSVSVSSEAHRRILHQPLFPESS**TP**PPPDFQS**TPSP**PLPD**TP**DQPFF
PEN**PSTP**QQTLFPPPPPPVSADVNGGLPI**PT**ATTQSAKPGKKVAIVISVGIVTLGMLSALAFFL
YRHKAKHASDTQKLVTGGGDGGGSRRFQEDSGP**PT**TTSSTFLYMGTVE**PT**RVSASESNGGTNGP
VNS**SP**YRKLNSAKRSERYR**PSP**ELQPLPPLAKPPQ**PS**DN**SPS**AL**SPS**SSSSGEECRDTAFY**TP**H
GSAISSDDGYYTAFPRSANGSLPHSKRT**SP**RSKFGS**APT**TAASR**SP**EMKHVII**PS**IKQKLPPPV
QPPPLRGLESDEQELPYSQNKPKFSQPPPPPNRAAFQAITQEK**SP**VPPPRR**SP**PPLQ**TP**PPPPP
PPPL**AP**PPPPQKRPRDFQMLRKVTNSEATTNSTT**SPS**RKQAFK**TPSP**KTKAVEEVNSVSAGSLE
KSGDGDTD**PS**KPKLKPLHWDKVRASSDRATVWDQLKSSSFQLNEDRMEHLFGCNSGSS**AP**KEPV
RRSVIPLAENENRVLDPKKSQNIAILLRALNVTREEVSEALTDGNPESLGAELLETLVKM**APT**K
EEEIKLREYSGDVSKLGTAERFLKTILDIPFAFKRVEAMLYRANFDAEVKYLRNSFQTLEEASL
ELKASRLFLKLLEAVLMTGNRMNVGTNRGDAIAFKLDTLLKLVDIKGVDGKTTLLHFVVQEITR
SEGTTTTKDETILHGNNDGFRKQGLQVVAGLSRDLVNVKKSAGMDFDVLSSYVTKLEMGLDKLR
SFLKTETTQGRFFDSMKTFLKEAEEEIRKIKGGERKALSMVKEVTEYFHGNAAREEAHPLRIFM
VVRDFLGVLDNVCKEVKTMQEMSTSMGSASARSFRISATASLPVLHRYKARQDDTSSDSEHSSN
SST
>AT3G07540.1 | Symbols: | Actin-binding FH2 (formin homology 2) family protein | chr3:2404763-2407464 REVERSE LENGTH=841
MDGLCYVIFIIFSLLSCAF**SP**LSYA**SPA**TFSRRHLLQ**APV**TD**PS**TF**SP**PFFPLYSST**SP**PPP**PS**
PPQPLPP**APT**FATF**PA**NISALVLPR**SP**KPQ**TPS**RTLLIPAISAVLAAATLIALAFFFYGRWRG
QTSHFKDESKSLASDISQSQQQTLPCPPPRNNNTQNKLSV**APS**TSDVLYLGNVVTSSGSGFVKP
E**SP**DI**SP**LPPL**PA**RSFLLQHHSEANLDEEEEDDDFY**SP**LASIAGQESRDRRINPYSNCSCSISS
HSD**SPA**M**SPS**AAM**SP**PMNST**AP**HWSTNQNTH**SPSSP**ERTVRNNKRYGGQSLRMFSLWNQNLGFP
RISSAST**SP**ERGMIR**TP**DAYARSSMYSSVST**TP**DRFFRKVLDS**SP**PRWNDFSRNVKSLFLSSTS

A**SPA**RDFCINISESSRSLKSSWEKPELDTTQQRESAAAAVTLPPPQRPP**PA**MPEPPPLVP**PS**QS
FMVQKSGKKLSFSELPQSCGEGTTDRPKPKLKPLPWDKVR**PS**SRRTNTWDRLPYNSSNANSKQR
SLSCDLPMLNQESKVLDPRKSQNVAVLLTTLKLTTNDVCQALRDGHYDALGVELLESLARV**APS**
EEEEKKLISYSDDSVIKL**APS**ERFLKELLNVPFVFKRVDALLSVASFDSKVKHLKRSFSVIQAA
CEALRNSRMLLRLVGATLEAGMKSGNAHDFKLEALLGLVDIKSSDGRTSILDSVVQKITESEGI
KGLQVVRNLSSVLNDAKKSAELDYGVVRMNVSKLYEEVQKISEVLRLCEETGHSEEHQWWKFRE
SVTRFLETAAEEIKKIEREEGSTLFAVKKITEYFHVD**PA**KEEAQLLKVFVIVRDFLKILEGVCK
KMEVTSSLA
>AT1G20130.1 | Symbols: | GDSL-like Lipase/Acylhydrolase
superfamily protein | chr1:6977939-6980003 FORWARD LENGTH=534
MKRSSLVDSCSYSRIFRSIFCLLSFCIFFLTTTNAQVMHRRLWPWPLWPRPYPQPWPMNP**PTP**D
PSPK**PVAP**PG**PS**SK**PVAP**PG**PSP**C**PSP**PPKPQPKPP**PAPSPSP**C**PSP**PPKPQPKPVPP**PA**CP**PT**
PPKPQPK**PAP**PPEPK**PAP**P**PAP**K**PVP**C**PSP**PKP**PAPTP**K**PV**PPHGPPPK**PAPAPTPAPSP**K**PAP**
SPPKPENKTI**PA**VFFFGDSVFDTGNNNNLETKIKSNYRPYGMDFKFRVATGRFSNGMVASDYLA
KYMGVKEIV**PA**YLDPKIQPNDLLTGVSFASGGAGYN**PT**TSEAANAIPMLDQLTYFQDYIEKVNR
LVRQEKSQYKLAGLEKTNQLISKGVAIVVGGSNDLIITYFGSGAQRLKNDIDSYTTIIADSAAS
FVLQLYGYGARRIGVIG**TP**PLGCV**PS**QRLKKKKICNEELNYASQLFNSKLLLILGQLSKTLPNS
TFVYMDIYTIISQMLE**TPA**AYGFEETKKPCCKTGLLSAGALCKKSTSKICPNTSSYLFWDGVH**P**
TQRAYKTINKVLIKEYLHVLSK
>AT1G20120.1 | Symbols: | GDSL-like Lipase/Acylhydrolase
superfamily protein | chr1:6975504-6977123 FORWARD LENGTH=402
MLQDRVSGSLSSSKISRCVLFLSLFCFFLLTMHASANRLQRVPNPG**PSPAP**EPKPC**PSP**GPN**PA**
PATTKRTHNTTF**PA**IFAFGDSILDTGNNDYILTLIKANFLPYGMNFPDK**VPT**GRFCNGKI**PS**DF
IADYIGVK**PVVPA**YLRPGLTQEDLLTGVSFASGGSGYDPL**TP**IVVSAIPMSKQLTYFQEYIEKV
KGFVGKEKAEHIISKGLAIVVAGSDDLANTYYGEHLEEFLYDIDTYTSFMASSAASFAMQLYES
GAKKIGFIGV**SP**IGCIPIQRTTRGGLKRKCADELNFAAQLFNSKLSTSLNELAKTMKNTTLVYI
DIYSSFNDMIQNPKKYGFDEIDRGCCGTGLLELGPLCNKYTSLLCKNVSSFMFWDSYH**PT**ERAY
KILSQKFVENDMGPFYDN
>AT1G74460.1 | Symbols: | GDSL-like Lipase/Acylhydrolase
superfamily protein | chr1:27988150-27989765 REVERSE LENGTH=366
MKFCAIFVLFIVLAINGYDCKIVQFIFGDSLSDVGNNKNLPRSLATANLPFYGIDFGNGLPNGR
FTNGRTVSDIIGDKIGLPR**PV**AFLD**PS**MNEDVILENGVNYASGGGGILNETGGYFIQRFSLWKQ
IELFQGTQDVVVAKIGKKEADKFFQDARYVVALGSNDFINNYLMPVYSDSWKYNDQTFVDYLME
TLESQLKVLHSLGARKLMVFGLGPMGCIPLQRALSLDGNCQNKASNLAKRFNKAATTMLLDLET
KLPNASYRFGEAYDLVNDVITNPKKYGFDNSD**SP**CCSFYRIR**PA**LTCI**PA**STLCKDRSKYVFWD
EYH**PT**DKANELVANILIKRFDFMRADDGISH**APSPAP**DI**SPS**SDNN
>AT2G42990.1 | Symbols: | GDSL-like Lipase/Acylhydrolase
superfamily protein | chr2:17879056-17880200 FORWARD LENGTH=350
MATHYL**SPS**ILCIILTTLVSIAGAKI**PA**IIVFGDSSVDSGNNNFISTMARANFEPYGRDFPGGR
ATGRFCNGRLSSDFTSEAYGLK**PT**V**PA**YLD**PS**YNISDFATGVCFASAGTGYDNSTADVLGVIPL
WKEVEYFKEYQSNLSAYLGHRRAAKIIRESLYIVSIGTNDFLENYYTLPDRRSQFSISQYQDFL
VEIAEVFLKDIYRLGARKMSFTGI**SP**MGCLPLERVTNLDDPFSCARSYNDLAVDFNGRLRRLVT
KLNRELTGIKIYFANPYDIMWDIVTKPNLYGLEISSSACCGTGLFEMGFLCGQDNPLTCSDANK
FVFWDAFH**PT**ERTNQIVSDHFFKHLKNLFH
>AT1G20132.1 | Symbols: | GDSL-like Lipase/Acylhydrolase
superfamily protein | chr1:6981358-6983495 FORWARD LENGTH=383
MKSILIGFVFFLLSSFRSFFVTTTYSQVIHHRRLRPWPPPESGSG**PSP**G**PSPSP**HNKT**TPA**VFF
FGDSIIDTGNNNNLTTEMKCNF**SP**YGKDFPLGVATGRFSNGKVVSDYISEYLGVKPIV**PA**YFDP

NVQLEDLLTGVSFASGGSGYYHL**TP**KISRVKSMLEQLTYFQRHIARVKRLVGEEKTDQLLAKGL
SVVVAGSNDLAITYYGHGAQLLKDDIHYFTSKMANSAASFVMQLYEYGARQIAVLG**TP**PLGCVP
ILRTLKGGLRRECAQDINYASQLFNVKLSNILDQLAKNLPNSNLIYIDIYSAFSHILENSADYG
FEEIKRGCCGTGFVEAGPLCNRFTTFVCSNVSAYMFWDSLH**PT**QRFYKILTKILFEKYIHNLN
>AT1G71120.1 | Symbols: GLIP6 | GDSL-motif lipase/hydrolase 6 |
chr1:26821072-26822420 REVERSE LENGTH=362
MSSSSSMDLLMCLLLLI**SP**VVLAKSSSTV**PA**IFTFGDSIFDAGNNHYNKNCTAQADFPPYGSSF
FHR**PT**GRFTNGRTVADFISEFVGLPLQKPFLELQIQILNGTSNFSNGINFASAGSGLLLDTNKF
MGV**TP**IQTQLQQFQTLVEQNLIEKSIIQESLFLLETGSNDIFNYFLPFR**APT**L**SP**DAYVNAMLD
QVNKTIDQIYKLGARRIAFFSLG**PV**GCV**PA**RAMLPN**APT**NKCFGKMNVMAKMYNKRLEDIVNII
PTKYPGAIAVFGAVYGITHRFQTY**PA**RYGFSDVSNACCGNGTLGGLMQCGREGYKICNNPNEFL
FWDFYH**PT**EHTYRLMSKALWNGNKNHIRPFNLMALATNKITF
>AT3G48610.1 | Symbols: NPC6 | non-specific phospholipase C6 |
chr3:18011653-18013959 REVERSE LENGTH=520
MK**PS**SASRFSLTFSHFLTLYCLLTQTHVAQGSHQWQ**SP**IKTVVVLVLENRSFDHLLGWMKNSVN
PTINGVTGQECNPVPNSTQTICFTSDAEFVDPDPGHSFEAVEQQVFGSGPGQI**PS**MMGFVEQAL
SMPGNLSETVMKGFRPEAVPVYAELVKEFAVFDRWFSSIPG**PT**QPNRLFVYSATSHGSTSHVKK
QLAQGYPQKTIFDSLHSNDIDFGIYFQNI**PT**TLFYRNLRQLKYIFNLHQYDLKFKKDAAKGKL**P**
SLTVIEPRYFDLKGL**PA**NDDH**PS**HDVANGQKLVKEVYEALRS**SP**QWNETLLVITYDEHGGFYDH
VK**TP**YVGIPNPDGNTG**PAP**GFFKFDRLGVR**VPT**IMV**SP**WIQKGTVVSEAKG**PT**ESSEYEHSSI**P**
ATIKKLFNLSSNFLTHRDAWAATFEDVVSHLT**TP**RTDCPMTLPEV**AP**MRATEPKEDAALSEFQG
EVVQLAAVLNGDHFLSSFPEEIGKKMTVKQAHEYVKGATSRFIRASKEAMKLGADKSAIVDMRS
SLTTRPHN
>AT4G16230.1 | Symbols: | GDSL-like Lipase/Acylhydrolase
superfamily protein | chr4:9185298-9186447 FORWARD LENGTH=251
MSLLVFLCQIIVLSVLFFSEVCLAGKKI**PA**NFVFGDSLVDAGNNNYLATLSKANY**VP**NGIDFG**S**
PTGRFTNGRTIVDIVYQALGSDEL**TP**PYL**APT**TSGSLILNGVNYASGGSGILNSTGKLFGERIN
VDAQLDNFATTRQDIISWIGESEAAKLFRSAIFSVTTGSNDLINNYF**TPV**ISTLQRKVV**AP**EVF
VDTMISKFRLQLTRLYQLGARKIVVINIGPIGCIPFERESD**PA**AGNNCLAEPNEVGTNV
>AT5G63170.1 | Symbols: | GDSL-like Lipase/Acylhydrolase
superfamily protein | chr5:25338699-25340096 REVERSE LENGTH=338
MNSLVIQTTIVLVSVISVSIVHAGNI**PA**VIAFGDSILDTGNNNYLMTLTKVNFYPYGRDFVTRR
ATGRFGNGRI**PT**DLIAEGLGIKNI**VPA**YR**SP**FLEPNDILTGVSFASGGSGLDPMTARIQGVIWV
PDQLNDFKAYIAKLNSITGDEEKTRSIISNAVFVISAGNNDIAITYFTNPIRNTRYTIFSYTDL
MVSWTQSFIKELYNLGARKFAIMGTLPLGCLPGASNALGGLCLE**PA**NAVARLFNRKLADEVNNL
NSMLPGSRSIYVDMYNPLLELVKNPLRSGFI**SPT**RPCCC**APAAP**IPCLDASRYVFWDIAH**PS**EK
AYQTIIPPIIQQIQQSFA
>AT5G41890.1 | Symbols: | GDSL-like Lipase/Acylhydrolase
superfamily protein | chr5:16764292-16766680 REVERSE LENGTH=375
MDFTYRCSLKPFNCTFLLLWLSHFQAAQSFTNFIFGDSLVDVGNNNYIFTLSKADS**SP**YGIDF**A**
PSNGQ**PT**GRFTNGRTISDIVGEALGAK**SP**PPPYLEPNTEANTIRNGINYASGAAGILDDTGLLF
IGRVPLREQVSNFEKSREYMVRVIGENGTKEMLKNAMFTITIGSNDILNYIQ**PS**IPFFSQDKL**P**
TDVLQDSMVLHLTTHLKRLHQLGGRKFVVVGVGPLGCIPFARALNLI**PA**GKCSEQVNQVVRGYN
MKLIHSLKTLNNELRSEDYNTTFVYANSYDLFLKLVLNYQLFGLKNADKPCCGGYFPPFACFKG
PNQNSSQAACEDRSKFVFWDAYH**PT**EAANLIVAKALLDGDQTVA**TP**FNIRYLNDL
>AT4G28780.1 | Symbols: | GDSL-like Lipase/Acylhydrolase
superfamily protein | chr4:14215603-14217159 FORWARD LENGTH=367
MSTFLLTWIIMTVALSVTLFLMPQQTNAARAFFVFGDSLVDSGNNNYLVTTARAD**SP**PYGIDY**P**

TGR**PT**GRFSNGLNLPDIISEQIGSE**PT**LPIL**SP**ELTGEKLLIGANFASAGIGILNDTGVQFLNILRIGRQFELFQEYQERVSEIIGSDKTQQLVNGALVLMTLGGNDFVNNYFFPISTRRRQSSLGEFSQLLISEYKKILTSLYELGARRVMVTGTGPLGCV**PA**ELASSGSVNGEC**AP**EAQQAAAIFNPLLVQMLQGLNREIGSDVFIGANAFNTNADFINNPQRFGFVTSKVACCGQGAYNGQGVC**TP**LSTLCSDRNAYAFWDPFH**PT**EKATRLIVQQIMTGSVEYMNPMNLSTIMALDSRI

>AT3G26430.1 | Symbols: | GDSL-like Lipase/Acylhydrolase superfamily protein | chr3:9674419-9675889 FORWARD LENGTH=380

METNLLLVKCVLLASCLIHPRAC**SPS**CNF**PA**IFNFGDSNSDTGGLSASFGQ**AP**YPNGQTFFH**SPS**GRFSDGRLIIDFIAEELGLPYLNAFLDSIGSNFSHGANFATAGSTVRPPNATIAQSGV**SP**ISLDVQLVQFSDFITRSQLIRNRGGVFKKLLPKKEYFSQALYTFDIGQNDLTAGLKLNMTSDQIKAYIPDVHDQLSNVIRKVYSKGGRRFWIHNT**AP**LGCLPYVLDRF**PVPA**SQIDNHGCAIPRNEIARYYNSELKRRVIELRKELSEAAFTYVDIYSIKLTLITQAKKLGFRYPLVACCGHGGKYNFNKLIKCGAKVMIKGKEIVLAKSCNDVSFRVSWDGIHFTETTNSWIFQQINDGAFSDPPLPVKSACTR

>AT2G26870.1 | Symbols: NPC2 | non-specific phospholipase C2 | chr2:11457117-11459355 REVERSE LENGTH=514

MSIKAFALIQLLSVTILYNHVHAT**SP**IKTIVVVVMENRSFDHMLGWMKKLNPEINGVDGSESNPVSVSD**PS**SRKIKFGSGSHYVDPDPGHSFQAIREQVFGSNDTSMDPPPMNGFVQQAYSED**PS**GNMSASVMNGFEPDKVPVYKSLVSEFAVFDRWFASV**PS**STQPNRMFVHSGTSAGATSNNPISLAKGYPQRTIFDNLDDEEFSFGIYYQNI**PA**VLFYQSLRKLKYVFKFHSYGNSFKDHAKQGKL**PA**YTVIEQRYMDTLLE**PA**SDDH**PS**HDVYQGQKFIKEVYETLRA**SP**QWNETLLIITYDEHGGYFDH**VPTPV**RN**VPSP**DGIVGPDPFLFQFNRLGIR**VPT**IAV**SP**WIEKGTVVHGPNG**SP**F**PS**SEYEHSSI**PA**TVKKLFNLS**SP**FLTKRDEWAGTFENILQIRKEPRTDCPETLPEPVKIRMGEANEKALLTEFQQELVQLAAVLKGDNMLTTFPKEISKGMTVIEGKRYMEDAMKRFLEAGRMALSMGANKEELVHMKTSLTGRRP

>AT1G54790.2 | Symbols: | GDSL-like Lipase/Acylhydrolase superfamily protein | chr1:20441124-20443997 REVERSE LENGTH=408

MNITKMKLFYVILFFISSLQISNSIDFNY**PS**AFNFGDSNSDTGDLVAGLGIRLDLPNGQNSFKTSSQRFCDGRLVIDFLMDEMDLPFLNPYLDSLGLPNFKKGCNFAAAGSTIL**PA**N**PT**SV**SP**FSFDLQISQFIRFKSRAIELLSKTGRKYEKYLPPIDYYSKGLYMIDIGQNDIAGAFYSKTLDQVLASI**PS**ILETFEAGLKRLYEEGGRNIWIHNTGPLGCLAQNIAKFGTDSTKLDEFGCVSSHNQAAKLFNLQLHAMSNKFQAQYPDANVTYVDIFSIKSNLIANYSRFGKHFTKPLIDLNHLENVGYNKILNVLGFEKPLMACCGVGG**AP**LNYDSRITCGQTKVLDGISVTAKACNDSSEYINWDGIHYTEAANEFVSSQILTGKYSDPPFSDQMPFFLTLKF

>AT3G26820.1 | Symbols: | Esterase/lipase/thioesterase family protein | chr3:9881128-9885067 FORWARD LENGTH=634

MGVTLLRSIFGLCAVSSSSRVTDSYCSTKSYLRRRRTSASKQRLTEIKSVTS**TP**PP**PS**REARDFVGDGGGPPRWF**SP**LECRAQ**AP**N**SP**LLLFLPGIDGTGLGLIRHHKKLGEIFDIWCLHI**PV**SDR**TP**FKDLVKLIERTVKSENYRFPNRPIYLVGESIGACLALDVAARNPNVDLALILVN**PA**THVNNFMSKPLLGMLNVLPDGI**PT**LWEDVFGFKQG**AP**LTGILEAMSNEFSVQRMGGVGGGMLRDLFAVSANL**PT**LSRMFSKDTLLWKLEMLKSAIASVNSHIYSVKAETLIL**PS**GRDQWLLNEEDIVRYSRTLPNCIVRKLDDNGQFPLLEDSLDLATIIKLTCFYRRGKSHDYVSDYIK**PTP**FELQQLLDEHRLLMDAI**SPV**MLSTLEDGLLLKERNIHMRGLTHPMVFMYIQDSLVDPKMFDKYKLMGGVPVSNMNFYKLLREKAHVLLYPGGVREALHRKGEEYKLFWPEQSEFVRVASKFGAKIVPFGVVGEDDIFNIVLDSNDQRNIPILKDLMEKATKDAGNLRWKETKANWETKIAIIPGLVPKIPGRFYYYFGKPIDLAGKEKELKDKEKAQEVYLQAKSEVEQCIAYLKMKRECDPYRQLLPRMMYQASHGWSCEI**PT**FDL

>AT1G07230.1 | Symbols: NPC1 | non-specific phospholipase C1 | chr1:2220509-2222778 REVERSE LENGTH=533

MAFRRVLTTVILFCYLLISSQSIEFKNSQKPHKIQGPIKTIVVVVMENRSFDHILGWLKSTRPE
IDGLTGKESNPLNVSDPNSKKIFVSDDAVFVDMDPGHSFQAIREQIFGSNDTSGDPKMNGFAQQ
SESMEPGMAKNVMSGFKPEVLPVYTELANEFGVFDRWFAS**VPT**STQPNRFYVHSATSHGCSSNV
KKDLVKGFPQKTIFDSLDENGLSFGIYYQNI**PA**TFFFKSLRRLKHLVKFHSYALKFKLDAKLGK
LPNYSVVEQRYFDIDLF**PA**NDDH**PS**HDVAAGQRFVKEVYETLRS**SP**QWKEMALLITYDEHGGFY
DH**VPTPV**KG**VP**NPDGIIGPDPFYFGFDRLGVRV**PT**FLI**SP**WIEKGTVIHEPEG**PTP**HSQFEHSS
I**PA**TVKKLFNLKSHFLTKRDAWAGTFEKYFRIRD**SP**RQDCPEKLPEVKLSLRPWGAKEDSKLSE
FQVELIQLASQLVGDHLLNSYPDIGKNMTVSEGNKYAEDAVQKFLEAGMAALEAGADENTIVTM
R**PS**LTTRT**SPS**EGTNKYIGSY
>AT1G54010.1 | Symbols: | GDSL-like Lipase/Acylhydrolase
superfamily protein | chr1:20158854-20160747 REVERSE LENGTH=386
MMAKNCNLVSVLCVFLVLTLFNKPITVAGQNI**PA**VGLFTFGDSNFDAGNKQTLTKTLLPQTFWP
YGKSRDDPNGKFSDGLI**AP**DFLAKFMRIPIVIP**PA**LQPNVNVSRGASFAVADATLLG**AP**VESLT
LNQQVRKFNQMKAANWNDDFVKKSVFMIYIGANDYLNFTKNNPNADASTQQAFVTSVTNKLKND
ISLLYSSGASKFVIQTL**AP**LGCLPIVRQEFNTGMDQCYEKLNDLAKQHNEKIGPMLNELART**AP**
AS**AP**FQFTVFDFYNAILTRTQRNQNFRFFVTNASCCGVGTHDAYGCGFPNVHSRLCEYQRSYLF
FDGRHNTEKAQEMFGHLLFGADTNVIQPMNIRELVVY**PA**DEPMRESW**VPPT**SATVQLRESRGYE
YY
>AT1G75890.1 | Symbols: | GDSL-like Lipase/Acylhydrolase
superfamily protein | chr1:28493051-28495047 FORWARD LENGTH=379
MKRNSINIHHVTSFSS**SP**FWCVFFLVLLCKTSTNALVKQPPNET**TPA**IIVFGDSIVDAGNNDDI
MTTLARCNYPPYGIDFDGGI**PT**GRFCNGKVATDFIAGKFGIK**PS**I**PA**YRNPNLKPEDLLTGVTF
ASGGAGYVPFTTQLSTYLFIYKPLLFLKGGIALSQQLKLFEEYVEKMKKMVGEERTKLIIKNSL
FMVICGSNDITNTYFGL**PS**VQQQYDVASFTTLMADNARSFAQKLHEYGARRIQVFG**APPV**GC**VP**
SQRTLAGG**PT**RNCVVRFNDATKLYNVKLAANLGSLSRTLGDKTIIYVDIYDSLLDIILDPRQYG
FKVVDKGCCGTGLIEVALLCNNFAADVCPNRDEYVFWDSFH**PT**EKTYRIMATKYFERYV
>AT1G23500.1 | Symbols: | GDSL-like Lipase/Acylhydrolase
superfamily protein | chr1:8339668-8341072 FORWARD LENGTH=345
MNFSLLSTMLMALSSVCLFFVGYAQQFSGSVAVSALFAFGDSILDTGNNNNLNTLSKCNFFPYG
RNFIGGKATGRFGNGRVFSDMIAEGLNVKKLL**PA**YRDPNLSKNDL**PT**GVCFASGGSGLDERTAR
SQGVIWVPDQVKDFKEYIMKLNGVVRDKRKVNAIISNAVYLISAGNNDLAITY**PT**LMAQYTVST
YTDLLVTWTDNLLKSLYAMGARKFAVLGTLPLGCLPGARHTGGNFGNICLVPINQVAAIFNQKL
SAKLNNLHTILPGAKFVYVDMYNPLLNLINNPRASGFIDVADGCCCM**PTSPVP**CPDASQYVFWD
FAH**PS**EKSYMTI**AP**KIIEGIKKNLA
>AT2G42840.1 | Symbols: PDF1 | protodermal factor 1 |
chr2:17826327-17827426 REVERSE LENGTH=306
MRGMVSFAVWALFAALLSQQLFASVASVRFEDAKTYYL**SPPS**GSHG**TPPS**H**TPPS**SNCG**SP**PYD
PSPSTPSH**PSPPS**H**TPTPSTPS**H**TPTP**H**TPS**H**TPTP**H**TP**PCNCG**SPPS**H**PSTPS**H**PSTPS**H**PTP**
SHP**PS**GGYYS**SP**PPR**TP**VVV**TPPSP**IVDPG**TP**IIGG**SPPTP**IIDPG**TP**G**TP**FI**PAP**FPPITGTC
DYWRNH**PT**LIWGLLGWWGTVGGAFGTVSI**PS**SIPGFDPHMNLLQALSNTRSDPIGALYREGTAS
WLNSMVNHKFPFT**TP**QVRDHFVAGLSSNKAATKQAHTFKLANEGRLKPRV
>AT5G14920.1 | Symbols: | Gibberellin-regulated family protein |
chr5:4826598-4827761 FORWARD LENGTH=275
MALSLLSVFIFFHVFTNVVFAASNEESNALVSL**PTPT**L**PSPSPA**TKP**PSPA**LKP**PTPS**YKP**PT**L
PTTPIKP**PT**TKPPVKP**PT**IPV**TP**VKPPVS**TP**PIKLPPVQP**PT**YKP**PTPT**VKP**PS**VQP**PT**YKP**PT**
PTVKP**PT**T**SP**VKP**PTTP**PVQ**SP**PVQP**PT**YKP**PTSP**VKP**PT**T**TP**PVKP**PT**T**TP**PVQP**PT**YNP**PTT**
PVKP**PTAP**PVKP**PTP**PPVRTRIDCVPLCGTRCGQHSRKNVCMRACVTCCYRCKCVPPGTYGNKE
KCGSCYANMKTRGGKSKCP

>AT3G27416.1 | Symbols: | unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink). | chr3:10147371-10147886 REVERSE LENGTH=171
MTLKLLSLAVAALAITAVLAADP**PTPA**N**AP**KANGEKNST**SPSPA**AAA**SP**K**SPT**AS**APTSPPT**A**A**
PTMAKKNSTG**TPSPSPSSPSP**KSSSAK**TPA**S**SP**DSSSGDSSEG**PT**SSSD**APT**AS**SPPAPTP**EM**S**
PSSDDGTGASDGPEAS**APA**AGASSLVISVSGSVLAAVAWLFFI >AT1G76955.1 | Symbols: | Expressed protein | chr1:28919172-28919536 REVERSE LENGTH=78
MAK**TPS**SSSKLLVMMLFSFLALFIISHARVVFTD**TPS**NSY**AP**PIY**APVP**KECLKPPYCRGPPGE
SQFVYNERNYYQIN

>AT3G50580.1 | Symbols: | LOCATED IN: endomembrane system; EXPRESSED IN: leaf whorl, sepal, flower; EXPRESSED DURING: petal differentiation and expansion stage; BEST Arabidopsis thaliana protein match is: ECA1 gametogenesis related family protein (TAIR:AT1G44191.1); Has 132049 Blast hits to 55074 proteins in 2314 species: Archae - 393; Bacteria - 25609; Metazoa - 44661; Fungi - 17591; Plants - 17316; Viruses - 4153; Other Eukaryotes - 22326 (source: NCBI BLink). | chr3:18768767-18769519 REVERSE LENGTH=250
MKTSIVLVAAAFLCLVAF**PT**TTVGKYWPKIEGWPN**PS**EITRNELMLLNTGHSFGYGDSKVWKCT
YSNGS**APA**ISI**SPSTP**I**PSTPSTPSP**PP**PAP**KK**SP**PP**PTP**KK**SPSPPS**L**TP**FVPH**PTP**KK**SPSP**
P**PTPS**LPP**PAP**KK**SPSTPS**LPP**PTP**KK**SP**PPP**PS**HHSS**SPS**NPPHHQQNPWEHIERCMINMGPV
GMCRMQMEVSFYTRLFQVSDYCCNLVVNMKSECDDVAWGFFNDPFFVPLVRYTCHVTC

>AT5G11990.1 | Symbols: | proline-rich family protein | chr5:3872651-3873196 REVERSE LENGTH=181
MGLIVIIIALVMVMVVAS**SPS**DQTNVL**TP**LCISECSTC**PT**IC**SP**P**PS**K**PSPS**M**SP**P**PSPS**LPLS
S**SP**PPPPPHKH**SP**PPLSQSL**SP**PPLITVIHPPPPRFYYFES**TP**PPPPL**SP**DGKG**SPPS**V**PSSPP**
SPKGQSQGQQQPPYPFPYFYFYTASNATLLFSSSFLIALVVSSLSIFLNGSLV

>AT3G22120.1 | Symbols: CWLP | cell wall-plasma membrane linker protein | chr3:7795283-7796287 REVERSE LENGTH=334
MGSRSQNLSFLVLLLLGFVAVSYACDC**TP**PK**PSPAP**HKPPKHPVKPPKP**PA**VKPPKP**PA**VKP**PT**
PKP**PT**VKPHPKP**PT**VKPHPKP**PT**VKPHPKP**PT**VKPPHPKP**PT**KPHPHPKPPIVKP**PT**KPP**PSTP**
KP**PT**KPP**PSTP**KP**PT**TKPP**PSTP**KPPHHKPP**PTP**CPP**PTPTPTP**PVV**TPPTPTP**PVI**TPPTPTP**
PVV**TPPTPTP**PVI**TPPTPTP**PVI**TPPTPTP**PVV**TPPTPTP**PVV**TPPTPTPPTP**IPETCPIDTLK
LGACVDVLGGLIHIGLGKSHAKAECCPVLGGLLDLDAAVCLCTTIKLKLLNIDLVLPIALELLL
DCGK**TPPS**DFKC**PA**

>AT1G12665.1 | Symbols: | Encodes a Plant thionin family protein | chr1:4312398-4312667 FORWARD LENGTH=89
MEGKSLVAILVTMMVIGNLLPQTEAQKIPFMQCF**PA**CMIVCKSESKFPKFLTCPITCLKTCLHP
PSPP**PSP**PP**PSPSPSPSPSPSP**CENN

>AT1G12090.1 | Symbols: ELP | extensin-like protein | chr1:4090176-4090589 REVERSE LENGTH=137

MASSSIALFLALNLLFFTTISACGSC**TP**CGGGC**PSP**KPK**PTP**K**PTPSPS**SGSSKCPKDTLKLGV
CANVLNGLLDLTLGKPPVEPCCSLIQGLADVEAAVCLCTALKANILGINLNLPISLSLLLNVCS
KQLPPGFQC

>AT3G28790.1 | Symbols: | Protein of unknown function (DUF1216) | chr3:10813577-10815588 REVERSE LENGTH=608

MAKNLLAICLVFMVASSVVYEVQGTFLLKLYLRRKFPRRCIDF**AP**YAGKGMLMLVSNLEGGC**PA**
TREFKQFFSTFKSYMSFISSASISASKNIDVEMNGRCELLSKAMSALTGSKSSQSSELKMTMLS
MGKTLVEQKRQGSRMMSLKQKKELVVAMVKWTRMVITFVKSVAEKRGKSIDESSYGLDVDVNAS
IGSSSGSDGSSSSDNESSSNTKSQGTSSKSGSESTAGSIETNTGSKTEAGSKSSSSAKTKEVSG
GSSGNTYKDTTGSSSGA**SPS**G**SPTPTPSTPTPSTPTPSTPTPSTPTPSTPTPSTPAPSTPA**AGK
TSEKGSESASMKKESNSKSESESAASGSVSKTKETNKGSSGDTYKDTTGTSSG**SPS**G**SPS**G**SPT**
PSTSTDGKASSKGSASASAGASASASAGASASAEESAASQKKESNSKSSSSSSSTTSVKEVETQ
TSSEVNSFISNLEKKYTGNSELKVFFEKLKTSMSASAKLSTSNAKELVTGMRSAASKIAEAMMF
VSSRFSKSEETKTSMASCQQEVMQSLKELQDINSQIVSGKTVTSTQQTELKQTITKWEQVTTQF
VETAASSSSSSSSSSSSSSSSSQGSAKMAMKN

>AT3G52480.1 | Symbols: | unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 10 plant structures; EXPRESSED DURING: LP.04 four leaves visible, 4 anthesis, petal differentiation and expansion stage; Has 28 Blast hits to 28 proteins in 7 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 28; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink). | chr3:19452478-19453107 REVERSE LENGTH=209

MGL**TPT**ATLLVAVFAFCLVAVTAQFAYVLWWKRRFRRRSIAGSERDAFSSRGGDLTA**TP**P**PS**KE
LLYFFLFCLENKQFRIGSAT**AP**PL**PA**A**AP**PVNDVASKWSINGENLLCG**PS**ETLFTIAEDYTSES
DHRTGEIDPRGSIFTEDHVKDDEVEEEVVATDISDDEVDFSHYNQT**TP**FS**TP**CA**SP**PFY**TPSPS**
PIRDDLSKTVREESVYG

>AT5G27710.1 | Symbols: | unknown protein; Has 49 Blast hits to 49 proteins in 17 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 48; Viruses - 0; Other Eukaryotes - 1 (source: NCBI BLink). | chr5:9813069-9815018 FORWARD LENGTH=335

MVNSYVPLRFTIFISFSIAAASSFKLHSASH**SPS**SFPKATGDDLLSVLGP**PS**AASCLNPIVSRE
IKSCLKFLVPFKSDKPKPEFGRC**SP**RTGLCSGKIDAVERRSKFEEENSLIWWPPESVLELARLA
VDSGGDPGSIQRTLNPKMIPVPDVERSRKDKCQLTR**TP**YGRHFIAEEVNSYFEFLFHLIESRG**P**
SVGLNVSLSRYDLFHGHLFLASESGRLGILFHAKEY**PA**YDKKVFPYNMGYCQRGSDVKYNDSMN
LRNILWL**AP**L**PS**NS**SP**DWV**AP**GVLVVLDAHPDGIIYRDLIPDYVKFVRTIYEDDLGTTAVDVNY
LNVGAHEPDYQLFMC

>AT3G29300.1 | Symbols: | unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 12 plant structures; EXPRESSED DURING: 6 growth stages; Has 451 Blast hits to 349 proteins in 91 species: Archae - 0; Bacteria - 85; Metazoa - 81; Fungi - 88; Plants - 18; Viruses - 1; Other Eukaryotes - 178 (source: NCBI BLink). | chr3:11247011-11247652 FORWARD LENGTH=213

MAATLVMVVSLVMVLSLVLVLLAELYCSLLLRRRRHNSLNLPITTTVSTAARTTTTLNQAISTT
SNDNT**SPS**NINSS**SP**NPLYTGVIQ**TPT**KTHYNHEPYLQASLDLIQETIVNDSVDNFIYISNPMY

SNDATSK**PTTP**FE**TP**ES**SPS**RLETGESSSSSSGEEDNDHIIEVS**TPT**L**TP**MKDLPEKACSVSLK
NVETSASESNSSG**SP**YT**SPS**W
>AT2G30700.1 | Symbols: | unknown protein; BEST Arabidopsis
thaliana protein match is: unknown protein (TAIR:AT1G61900.1);
Has 68 Blast hits to 67 proteins in 13 species: Archae - 0;
Bacteria - 2; Metazoa - 0; Fungi - 0; Plants - 66; Viruses - 0;
Other Eukaryotes - 0 (source: NCBI BLink). | chr2:13082033-
13084384 REVERSE LENGTH=480
MPRGELAMGSLETVCWLKGCLVYRFLLFIIWLSSFQDVAAHDKLNEHSSRSTTSELANPPGIGV
SGPIQV**SPS**VIPKYA**SPA**LPW**TP**PMY**PT**FFPDTYEPKLTGK**CPT**DFQAISSVIDTAASDCSQPFA
ALVGNVICCPQFVSLLHIFQGQHNVKSNKLVLPDAVATDCFSDIVSILVSRRANMTI**PA**LCSVT
SSNLTGGSCPVTDVTTFEKVVNSSKLLDACRTVDPLKECCRPICQ**PA**IMEAALIISGHQMTVGD
KIPLAGSNNVNAINDCKNVVFSYLSRKL**PA**DKANAAFRILSSCKVNKACPLEFKE**PT**EVIKACR
NVA**APSPS**CCSSLNAYISGIQNQMLITNKQAIVCATVIGSMLRKGGVMTNIYELCDVDLKDFSV
QAYGMQQGCLLRSY**PA**DLIFDNTSGYSFTCDLTDNIA**AP**W**PS**SSSMSSLSLC**AP**EMSL**PA**L**PT**S
QTIKNHGFCNGGVGALRLIVLVFLLYVVVVRH
>AT3G11640.1 | Symbols: | unknown protein; FUNCTIONS IN:
molecular_function unknown; INVOLVED IN: biological_process
unknown; LOCATED IN: endomembrane system; BEST Arabidopsis
thaliana protein match is: unknown protein (TAIR:AT3G52480.1);
Has 36 Blast hits to 36 proteins in 7 species: Archae - 0;
Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 36; Viruses - 0;
Other Eukaryotes - 0 (source: NCBI BLink). | chr3:3674501-
3675061 FORWARD LENGTH=186
MHALTNSGAVLLAVLLFFLLILFAELSYIFCCRTGSL**PSPA**SKEVLFQFLMCCKKNHHHSRIEP
CTGIVSVQMEEDVVAAAVPEEEFAAADKWRVSRLLFTIEEEDLELEDDDDDVISAEVDNGFEVQ
VDIPVEYSGD**PTP**FL**TP**CN**SP**PYF**TPSPSP**GRDMDDVIDVYEVSSRNCCFNFQGSQMC
>AT2G39560.1 | Symbols: | Putative membrane lipoprotein |
chr2:16505124-16505825 REVERSE LENGTH=233
MRSLSSVGLALSIVFGCLLLALLAELYYLLWCKKRSTTRRPDFRNDYS**TP**GTRELLFIFCCSSS
TN**PS**SS**SPS**SSSFSNPKPIDTQQQCPLNNGFENVGGPGLVPRFLFTIMEETVEEMESEDVVSTK
GKSLNDLFLNMESGVI**TP**PYL**TP**RA**SPS**LF**TP**PL**TP**LLMESCNGRKEEISSLFESSSDAEFNRL
VRS**SP**LSSSH**SPS**S**SP**LSRFKFLRDAEEKLYKKKVMEIAEA
>AT5G07520.1 | Symbols: GRP18, ATGRP18, ATGRP-8 | glycine-rich
protein 18 | chr5:2380208-2381230 REVERSE LENGTH=228
MFSFLIFLLEAFQVVIATVVSIVFLVFAGLTLVGSATALTIT**TP**LFIIF**SP**IL**VPA**TIATAVIT
TGFTTGGALGAMAVALIRRRMGAK**PT**AEGTSSAQPLLKLPVYGGYGGFWGGKKFSGTFGNKPGG
GNPFGDISKWLGPGAAGGG**AP**GGLGGGGNPFGNISKWFGPGAAGGDASAAGG**APA**AEA**APA**AGA
APAAGA**APA**AGA**APA**AGA**APA**AGA**APA**AGGS**TPPT**W
>AT1G65720.1 | Symbols: | unknown protein; FUNCTIONS IN:
molecular_function unknown; INVOLVED IN: biological_process
unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 24 plant
structures; EXPRESSED DURING: 15 growth stages; Has 44 Blast
hits to 44 proteins in 14 species: Archae - 0; Bacteria - 0;
Metazoa - 0; Fungi - 0; Plants - 44; Viruses - 0; Other
Eukaryotes - 0 (source: NCBI BLink). | chr1:24440471-24441013
REVERSE LENGTH=180

MTSSSYLRFAIAVVAFLSITTITTARPCKTFLISSYSLSI**TP**ENPNLESDFTSTRFVTVFTIRR
LNPHHVVPFFVNRRHEKPQIQSDRSLPLISDNINSFRDRTRDILSVVVALLFGVGCGALTAATM
YLVWALVVNRQSYDFEEEEDDYENDESDAASLKKLGYVKI**PAPAPAPV**KEAA
>AT1G78460.1 | Symbols: | SOUL heme-binding family protein |
chr1:29518547-29519296 REVERSE LENGTH=219
METGLNILKLSLCVSLVVVGSYAQ**APAP**WN**PS**NGFRPGTCDHYEC**PT**YKLVEAGYGFEIRMYDA
ALWIST**SP**I**PS**LSMTQATKTGFRRLNRYIEGDNKSNVKMNMT**AP**VIAQA**TP**GRSVYTVSLYLPK
KNQQNPPQADDLHVRSTK**PT**YVAVRQIGGYVSNNVAKDEAAALMESLRDSNWILPIEKSKGKL**P**
AYFLAVYNP**PS**HTTARVINEIMVPFNM
>AT5G64720.1 | Symbols: | Protein of unknown function (DUF1278)
| chr5:25872346-25872813 REVERSE LENGTH=155
MATKSTSKPLLLSFLMMSYLISTFHVITVAEGRTLQFTKMATDHSGAGNLMDCWNAGLELKSCT
DEIVKFFLSQTGTSEPPVKGGIDKDCCGAIGLVVKDCWSVMFTSLGLTTMEGNNLREYCEFQAE
KSEL**SPSPAP**ETLAL**SP**VEITYPGLDY
>AT4G39340.1 | Symbols: | Protein of unknown function (DUF1278)
| chr4:18293129-18293512 REVERSE LENGTH=127
MASNTTFLFSTVTLLIIILLNTTVSGRDL**PA**ESSTNIAARLQSGGLMECWNALYELKSCTNEIVL
FFLNGETKLGVSCCESVDIITTNCW**PA**MLTSLGF**TP**EEANVLRGFCQNPNSGDS**SPAPSP**KIV
>AT5G62630.1 | Symbols: HIPL2 | hipl2 protein precursor |
chr5:25143719-25146390 REVERSE LENGTH=696
MAKTNQAITICSLLLLLLLSETTSHLLCSDSK**TP**VNNNETLQFCDSYKERSCCNSKDDLQLQNR
FNSMNISDSNCSSLLKSILCSKCDEFSGQLFGDDDSSLVPILCNSTSQDLCSKLWDSCQNISIV
S**SP**F**SPT**LLGGAT**SPS**TSSNSSTLTDLWKSQTEFCTAFGG**PS**QTNNNKTKCFNGEPVNRDTSDD
DEDDVK**TP**KGICLEKIGTGSYLNMVAHPDGSNRAFFSNQPGKIWLGTIPDQDSGKPMEIDES**TP**
FVDITDQVSFDTQFGMMGMAFHPKFAENGRFFASFNCDKVK**SP**GCSGRCACNSDVNCD**PS**KLPK
DDGT**TP**CRYQTVVSEYTANGTSS**SPS**TAKIGKASEVRRIFTMGLPYSSSHGGQILFGPDGYLYL
MTGDGGGVSDTHNFAQNKKSLLGKILRLDVDVM**PS**VSEISKLGLWGNYSIPKNNPFQGNENEQP
EIWALGLRNPWRCSFDSERPDYFLCADVGKDTYEEVDIITMGGNYGWRTYEGPYVF**SP**L**SP**FGE
NVSSDSNLTFPILGYNHSEVNKHEGSASIIGGYFYRSNTDPCSYGTYLYADLYANAMWAAIE**SP**
EDSGNFTDSLIPFSCSKD**SP**MKCTA**AP**GGASSG**PA**LGYIYSFGQDNNKDIHLLTSSGVYRIVR**P**
SRCNLACSKENTTASAGKQN**PA**GS**AP**PQPL**PS**SARKLCFSVFLLLSLLMMFLTLLD
>AT2G28405.1 | Symbols: LCR32 | low-molecular-weight cysteine-rich 32 | chr2:12153294-12154092 REVERSE LENGTH=83
MMGKHIQLSFAILIMFTIFVLGAVGDVDQGYKQQCYKTIDVNLCVTGECKKMCVRRFKQAAGMC
IKS**VPSAPAP**NRCRCIYHC
>AT1G30795.1 | Symbols: | Glycine-rich protein family |
chr1:10935831-10936160 FORWARD LENGTH=109
MASSTHSYFTTLALTLILIFRLIPETTASRHLNGKN**PA**VIGVTTTSEKYIV**PTP**LPPFLRPFFP
PLQFAA**AP**FGGNIPQPPL**PSP**P**PT**FLPCLPGFKFPPFQSRK**PTP**P
>AT4G38080.1 | Symbols: | hydroxyproline-rich glycoprotein
family protein | chr4:17883012-17883398 FORWARD LENGTH=128
M**AP**LKNSFVTSLVIALTFTSFFTSLSAHRHLLQS**TP**VTQP**PA**LTFPPLPKTTMPPV**PS**L**PTP**GQ
QTLPQPQ**PT**LPQ**PT**GLPPM**PS**TQI**PS**LPNQVQ**PT**IPNIPQINF**PS**NFPFNFPFNIPFL**TP**P**PS**K
>AT5G60650.1 | Symbols: | unknown protein; FUNCTIONS IN:
molecular_function unknown; INVOLVED IN: biological_process
unknown; LOCATED IN: endomembrane system; BEST Arabidopsis
thaliana protein match is: unknown protein (TAIR:AT2G28410.1);
Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0;

Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink). |
chr5:24374305-24374748 REVERSE LENGTH=147
MTYLTHSLLFFFSCLSMLLCVSIAGSR**PV**HG**PA**YTN**PS**AF**SP**QAYDFFHPKSSLSDHNNPPKS**S**
PSPLSL**SPSPS**KTSIVEPDSQGSKVSSDEHTTSESRREEEGRGETIGVVLGLSFIAFLSLGIYF
VIKKQRANIIRTIVTHSDA
>AT4G37160.1 | Symbols: sks15 | SKU5 similar 15 | chr4:17494820-17497124 REVERSE LENGTH=541
MKQTNLLVCKLFIGALFWLGSVLVNAEDPYMFYTWTVTYGTR**SP**LGVPQQVILINGQFPG**PA**IE
AVTNNNIVVNLINKLDEPFLITWNGVKQRRTSWQDGVLGTNCPIQPNSNWTYQFQLKDQIGTYT
YFASTSLHRASGAFGALNINQRSVIT**TP**Y**PTP**DGDFTLLVSDWFSNMTHKDLRKSLDAGSALPL
PDALLINGVSKGLIFTGQQGKTYKFRVSNVGIATSINFRIQNHTMSLIEVEGAHTLQESYESLD
VHVGQSMTVLVTLKASVRDYFIVASTRFTKPVLTTTASLRYQGSKNAAYGPLPIG**PT**YHIHWSM
KQARTIRMNLTANAARPNPQGSFHYGTIPINRTLVLANAATLIYGKLRYTVNRISYIN**PTTP**LK
LADWYNISGVFDFKTIIS**TPT**TG**PA**HIGTSVIDVELHEFVEIVFQNDERSIQSWHMDGTSAYAV
GYGSGTWNVTMRKRYNLVDAVPRHTFQVYPLSWTTILVSLDNKGMWNLRSQIWSRRYLGQELYV
RVWNDEKSLYTEAEPPLNVLYCGKAKRPL
>AT2G28410.1 | Symbols: | unknown protein; Has 18 Blast hits to 18 proteins in 5 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 18; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink). | chr2:12156202-12156549 FORWARD LENGTH=115
MTSKTCLVFFFSSLILTNFALAQDR**AP**HGLAYE**TPV**AF**SPS**AFDFFHTQPENPD**PT**FNPCSESG
C**SP**L**PV**AAKVQGASAKAQESDIVSISTGTRSGIEEHGVVGIIFGLAFAVMM
>AT2G47530.1 | Symbols: | Pollen Ole e 1 allergen and extensin family protein | chr2:19504220-19504852 FORWARD LENGTH=184
MASTGAATNLLLLAMVVVVATADYYAQPQPYVPK**PT**TTYT**SPV**K**TP**YLPKSNPDIAIEGFILCK
SGYKTYPIQGGKVKVVC**PV**VDSYGKLVAKVTISSY**PT**DLKGYFYFITYGLSHKVNNISSCKVKL
ES**SPV**FTCK**TPT**NVNKGVTG**AP**L**SP**DNSKFLSHDNLTLYTLEPFYFS**SPVAP**K**PV**Y
>AT2G34700.1 | Symbols: | Pollen Ole e 1 allergen and extensin family protein | chr2:14634298-14635597 REVERSE LENGTH=175
MGLVTKALKYLLLLSISLTFIYTNAVSSA**SP**M**TPPSSPA**KMSRRLVAVEGMVYCKSCKYSGVDT
LLEA**SP**LQGATVKLACNNTKRGVTMETKTDKNGYFFML**AP**KKLTTYAFHTCRAW**PT**NPG**PT**TAT
MTCT**VPS**KLNNGITGAMLK**PS**KTINIGEHDYVLFSVGPFAFE**PA**CAL
>AT3G16670.1 | Symbols: | Pollen Ole e 1 allergen and extensin family protein | chr3:5681500-5682111 FORWARD LENGTH=154
MAVIKKHITFSLVLLCLIVV**SP**MANAQSGLGGINVPIINGVLFCTING**AP**LNG**TPAPA**FANAVV
QLQCGNLNRVVAETITNIAGLFTFSTNGIQISL**PT**LLNDCRIV**VPTP**RSSCDATL**PS**TGQLISQ
LNLVGSIVSGLLNIVAIL**PT**GFI**PT**I
>AT5G15780.1 | Symbols: | Pollen Ole e 1 allergen and extensin family protein | chr5:5144898-5146297 REVERSE LENGTH=401
MERITTLWFWFSLMIFLGISINGGLSQGQQHVMKKTRSSAVVVGTVYCDTCFNGAFSK**SP**NHLI
SGALVAVECIDENSK**PS**FRQEVKTDKRGEFKVKLPFSVSKHVKKIKRCSVKLLSSSQPYCSIAS
SATSSSLKRLKSNHHGENTRVFSAGFFTFRPENQPEICSQKPINLRGSKPLLPD**PS**FPPPLQDP
PN**PSP**LPNLPIVPPLPNLPVPKLPVPDLPLPLVPPLLPPGPQKSASLHNKKSDSLKDKKTEALK
PNFFFPPNPLNP**PS**IIPPNPLI**PS**I**PTPT**LPPNPLI**PSPPS**LPPIPLI**PTPPT**L**PT**IPLL**PTPP**
TPTLPPI**PT**I**PT**LPPLPVLPPVPIVNP**PS**LPPPP**PS**F**PVP**LPPVPGLPGIPP**VP**LIPGIP**PAP**L
IPGIPPL**SPS**FSSHHQP

FORWARD LENGTH=359
MDPNHNNLLIFVAFFVLCLATNGVTGYATVTGSVFCDQCKDGERSLFDFPVSGIKISVTCADEN
GQVYMSREETTNWLGGYVMRFDG**TP**DLSNCYAQVSDNGVQQD**PS**SCSIASG**PA**QKLKLMFSFFG
IETFAADALLAQPVQ**PS**SFCPKP**PTAP**VMPPPQ**VPV**MPPPQ**VPV**KPHPK**VPV**I**SP**DP**PA**TLPPP
K**VPV**I**SP**DP**PT**TLPPPL**VPV**INLPPVT**SP**PQFKLPPLPQIPPMPFVE**PS**ACSHQLWMKPEYRCY
WRAIGPDTKVAVAFGLVAGRIYGTDMTVREALDGRGEAYKTLLREATTALLNSYNSLGFPYNSV
AVITYTNLALLGNSEHDVLMTAIRFIKANSGTCRFTVCN
>AT2G22510.1 | Symbols: | hydroxyproline-rich glycoprotein
family protein | chr2:9569327-9569701 REVERSE LENGTH=124
MAHWLSSLVIALTFTSFFTGLSASRHLLQS**TPA**I**TP**PVTTTFPPL**PT**TTMPPFP**PS**TSLPQ**PT**A
FPPL**PS**SQI**PS**LPN**PA**QPINIPNFPQINIPNFPISIPNNFPFNL**PT**SI**PT**IPFF**TP**P**PS**K
>AT3G16660.1 | Symbols: | Pollen Ole e 1 allergen and extensin
family protein | chr3:5676904-5677788 FORWARD LENGTH=180
MAMLKNKHMTVSLILVCLVVV**SP**MAEAQLGLGGSGGLGGLIGGLVGGLGGLVGGLVGGILNLVN
INGVVFCSLNG**APS**GTS**TPA**FANAGVELQCGRQNRVVSTATTNAAGLFSL**PT**DSIQMLLSTLLS
DCRVVVT**TP**LSTCNANL**PS**VGNLVSRLAMIGNSLTGLLNIISII**PA**GFGLLN
>AT5G13140.1 | Symbols: | Pollen Ole e 1 allergen and extensin
family protein | chr5:4170688-4171744 REVERSE LENGTH=267
MYLPLLYLLFLASCLTHQALGRGRGR**PSP**EGSLD**PS**SRITVVGVVYCDTCSINTFSRQSYFLQG
VEVHVTCRFKAS**SP**KTAEEVNISVNRTTNRSGVYKLEIPHVDGIDCVDGIAISSQCSAKILKTS
SDDNGGCSIPVFQTATNEVSIKSKQDRVCIYSLSALSYKPPHKNTSLCGNGGKKHHRKDEKVEK
KFRDSKFFWPYL**AP**YWFPWPYPDLPPL**PT**LPPF**PS**FPF**PS**LPFGNPNLAL**PA**FDWKNPVTWIPY
LPRFPPGDHNP
>AT5G20630.1 | Symbols: GLP3, GLP3A, GLP3B, ATGER3, GER3 |
germin 3 | chr5:6975315-6975950 REVERSE LENGTH=211
MKMIIQIFFIISLISTISFASVQDFCVADPKGPQ**SPS**GYSCKNPDQVTENDFAFTGLGTAGNTS
NIIKAAV**TPA**F**APA**YAGINGLGVSLARLDLAGGGVIPLHTHPGASEVLVVIQGTICAGFISSAN
KVYLKTLNRGDSMVFPQGLLHFQLNSGKG**PA**LAFVAFGSS**SP**GLQILPFALFANDL**PS**ELVEAT
TFLSDAEVKKLKGVLGGTN
>AT5G62630.1 | Symbols: HIPL2 | hipl2 protein precursor |
chr5:25143719-25146390 REVERSE LENGTH=696
MAKTNQAITICSLLLLLLLSETTSHLLCSDSK**TPV**NNNETLQFCDSYKERSCCNSKDDLQLQNR
FNSMNISDSNCSSLLKSILCSKCDEFSGQLFGDDDSSLVPILCNSTSQDLCSKLWDSCQNISIV
S**SP**F**SPT**LLGGAT**SPS**TSSNSSTLTDLWKSQTEFCTAFGG**PS**QTNNNKTKCFNGEPVNRDTSDD
DEDDVK**TP**KGICLEKIGTGSYLNMVAHPDGSNRAFFSNQPGKIWLGTIPDQDSGKPMEIDES**TP**
FVDITDQVSFDTQFGMMGMAFHPKFAENGRFFASFNCDKVK**SP**GCSGRCACNSDVNCD**PS**KLPK
DDGT**TP**CRYQTVVSEYTANGTSS**SPS**TAKIGKASEVRRIFTMGLPYSSSHGGQILFGPDGYLYL
MTGDGGGVSDTHNFAQNKKSLLGKILRLDVDVM**PS**VSEISKLGLWGNYSIPKNNPFQGNENEQP
EIWALGLRNPWRCSFDSERPDYFLCADVGKDTYEEVDIITMGGNYGWRTYEGPYVF**SP**L**SP**FGE
NVSSDSNLTFPILGYNHSEVNKHEGSASIIGGYFYRSNTDPCSYGTYLYADLYANAMWAAIE**SP**
EDSGNFTDSLIPFSCSKD**SP**MKCTA**AP**GGASSG**PA**LGYIYSFGQDNNKDIHLLTSSGVYRIVR**P**
SRCNLACSKENTTASAGKQN**PA**GS**AP**PQPL**PS**SARKLCFSVFLLLSLLMMFLTLLD
>AT3G45245.1 | Symbols: | ECA1 gametogenesis related family
protein | chr3:16578384-16578734 FORWARD LENGTH=116
MERMWATFMLALIILLTTSIQAKGNEKGNDIAG**APSPT**L**AP**QSENGLLPN**PA**SCLADVKTIPNC
VKAVKRFKLRNVTKKCCVILLYLPEDCFGYLFPVRWIYRILLKIACKILGHI
>AT1G63057.1 | Symbols: | unknown protein; FUNCTIONS IN:
molecular_function unknown; INVOLVED IN: biological_process

unknown; LOCATED IN: endomembrane system; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT1G63055.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink). | chr1:23382242-23382989 FORWARD LENGTH=116

MMNTKIVALLMVVMMVTMGMENILVQARHHHHHHHH**SPSPSPS**SDFDDSGSK**SPSPSP**KSKEKSSTYCMIGCSFEKCFHHGKSSTMTTGDVEKYDSCMKKCSKICNKKNENEDIYV

>AT2G34870.1 | Symbols: MEE26 | hydroxyproline-rich glycoprotein family protein | chr2:14710819-14711169 FORWARD LENGTH=116

MASSSLISVALTLILIFHLMPETTVARRLTDQK**PS**DEVVTTTTDEANNLPFPPGLPFGGVPPL**PS**LFPPF**VPSP**FPGNIPRLPFPFPF**PTSPPAPS**LPGFPGFTFPPLPFL**TP**PPL

>AT5G42560.1 | Symbols: | Abscisic acid-responsive (TB2/DP1, HVA22) family protein | chr5:17015573-17016969 FORWARD LENGTH=296

MIGSFLTRGLVMVLGYAY**PA**YECYKTVEKNRPEIEQLRFWCQYWILVACLTVFERVGDAFVSWVPMYSEAKLAFFIYLWYPKTRGTTYVYESFFRPYLSQHENDIDHSLLELRTRAGDMAVIYWQRVASYGQTRILEILQYVAAQS**TP**RPQPPQKRGGRANQ**APA**KPKK**APVP**QSEPEEVSLSSSSSSSSENEGNE**PT**RKVSG**PS**RPR**PT**VTS**VPA**ADPKNAGTTQIAQKSVA**SP**IVNPPQSTTQVEPMQIEEVEGEAESGNENPNPEGPKETVMEETIRMTRGRLRKTRSEESR

>AT5G66816.1 | Symbols: | unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT3G50610.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink). | chr5:26681495-26681800 FORWARD LENGTH=101

MKLSVYIILSILFISTVFYEIQFTEARQLRKTDDQDHDDHHFTVGYTDDFG**PTSP**GN**SP**GIGHKMKENEENAGGYKDDFE**PTTP**GH**SP**GVGHAVKNNEPNA

>AT1G63055.1 | Symbols: | unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT1G63057.2); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink). | chr1:23379739-23380445 FORWARD LENGTH=109

MVNTKIVALLMVVMMVTMGMENILVQANRRHHH**SPS**PF**PS**SDFDDSGSK**SPSPSP**KSKEKSSTYCMIGCSFEKCFHHGKSSSDVEKYDLCMKKCSKICNKKKESDDIYV

>AT5G26070.1 | Symbols: | hydroxyproline-rich glycoprotein family protein | chr5:9106341-9106649 FORWARD LENGTH=102

MVSLGLSFTLVFLLAILFTVAEANNNRKLLK**TPT**NYQPLY**SPSPSP**YR**SPV**TLPPPPPH**PA**YSRPVALP**PT**LPIPH**PSP**HAERFYYRQ**SP**PP**PS**GKPWWWLL

>AT3G18050.1 | Symbols: | unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT4G28100.1); Has 67 Blast hits to 66 proteins in 12 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 67; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink). | chr3:6180931-6182593 FORWARD LENGTH=335

MFQH**APS**SISLIIIFLFTIL**SPVSP**E**PV**TVQPFRVKS**SPPA**TI**PA**FPEQSDFSGCPLDLPEDLF
HGIKSACSGKKLHKGKCCPVLGAWLYSAYSTTALSRSISAAARNSSSSSAAVT**TP**EEDMPLLPD
DSETCVDGLGKSLRQRGIELTRPNETCDVVYCYCGIRLHPLSCSEAFRVNDEGRLVGDERVDRL
ETDCLSGSHNNADGF**SP**LLGCNKCLNSLYKLNPKKTSGTRN**PS**KEDRNRTAKMHNKDCVLMGLT
WLLAKNRTAYF**PT**VTSVLRAVMLNHDG**VP**RSCALGSDGMPLAVDSSEFSNG**SPT**SLQYPHHLVH
FLLYSVITLVLIRSW
>AT2G36695.1 | Symbols: | unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink). | chr2:15382702-15383298 FORWARD LENGTH=117
MTMACKNLVAILIIVSLFGLHQCDDDPHPKGIERCFTRFVGKPYSLEDLCCVE**AP**WICFAGYLT
DECIKRCPPLREGVP**PSP**PRRKGV**APSP**PRRKGV**APSPSP**MRKGV**APSPSP**MS
>AT1G24267.2 | Symbols: | Protein of unknown function (DUF1664) | chr1:8604451-8607241 REVERSE LENGTH=344
MAIPLGKLTILIGAGLVGSVLAKEGSLPDVSSFVSGALKMVFRQLKQEE**PA**KSASKPRNDTLMA
QVNSLRHELSLLSSNRPITIVTTAGSGGKKYGYIIIIGVIGYGYVWWKGWKLPDLMFATRRSLS
DACNSVGSQIDGFYTSLSGTKKELSSKIDGMGRSLDANTEIIQDTGREVMELQRGTENIKDDVK
FVFDAVENLVRKLIYRIEGNQDITLKGVGALHAQVRENKRIQESNKAL**PS**TSA**VPA**LEA**AP**M**TP**
SSRTLSLP**PASP**RESQ**SPS**TSNGAQQSRGPLQHTQSMSGLKEISENGTHSGETTGNTSSGLFSI
FSIPRIGRTRSVVNT**VPT**NSIGPQ
>AT5G48210.1 | Symbols: | Protein of unknown function (DUF1278) | chr5:19548459-19548770 FORWARD LENGTH=103
MWSNKFFFLL**PV**VCIVVFATSQPL**PSP**Q**PS**QLDCWSSLEVIPDCVPEIFRSITNGQFGNVG**PS**C
CHAFIGLDTECIPQMFIF**AP**LIP**PS**RRLRDHCSHKII**VP**
>AT1G61688.1 | Symbols: | Defensin-like (DEFL) family protein | chr1:22780793-22781353 REVERSE LENGTH=105
MAN**TP**KTLIAFVFSVIVIISYVHCHTTIAS**APS**SGE**PT**TYATG**PA**LSKHSHDNDGICFV**TPA**CF
APGQYEIGCIVYCHESHYKHYKCVNRSCCCYNTDKNASELK
>AT3G44430.1 | Symbols: | unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: mitochondrion; EXPRESSED IN: 20 plant structures; EXPRESSED DURING: 13 growth stages; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT5G41660.1); Has 6 Blast hits to 6 proteins in 1 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 6; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink). | chr3:16068301-16068921 FORWARD LENGTH=206
MFSDAGGGANRRRRDLALPINLLWLLLMLLFTTSVRYSKTLFQ**PS**RII**SPSPV**FNGGQKTSYGG
RRCG**PA**FALLFHPLLANS**SP**KSFFSSDAIPQSEAFII**SP**PFIYRSRSVST**TP**REYSHRNSLMAR
DVSWTVRCAISGGVVVGARRFLLMGDGSMALRQASKAAARFTDEGDNKSQNPNCKCGLLFEDGP
LFSWPNKISTSAQF
>AT5G56369.1 | Symbols: | Defensin-like (DEFL) family protein | chr5:22835446-22835823 FORWARD LENGTH=93
MTKIGFYLATYATIYIIL**SP**GLLATAARENLHHQCFCE**SPS**KCDCF**PTTPTPPT**SVNKSRKGGP
LCTTDGDCKHFCRPKKGVCNIDFETCICQ

>AT5G42785.1 | Symbols: | unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink). | chr5:17157069-17157693 REVERSE LENGTH=115

MKGRAYVMIFFFWALLTII**TP**MLVSWSQSLKNLPPLSNDQEQKIKDSGPRRMMGYSEEMHLAQEFVSHVEEEKLMME**PS**M**APTP**EDNHVM**PS**LSSLRIKQQEGLVAKRAHPMRLR

>AT4G20420.1 | Symbols: | Tapetum specific protein TAP35/TAP44 | chr4:11017041-11017469 FORWARD LENGTH=142

MSISRVYLCLIFLTFVS**SP**LVLCSR**SP**KLAAASAAIGKKHGKEHVH**SPA**MLFSEFPKVDSSSSMTHIDE**PA**TKSAIAGFFRYRLPFQGWPFHKY**AP**FPMG**TPT**N**PSVPV**TS**TPS**SGAAAAEEEETEK**VPSAPS**KGNRDGGNA

>AT3G26110.1 | Symbols: | Anther-specific protein agp1-like | chr3:9541573-9541959 FORWARD LENGTH=128

MARLHLALLLLFVAISAVIVSAAENEKP**PS**TTTAS**APT**AT**TPT**DGIGGDIAEG**PA**DDNAIGTTDNDDAAV**TP**DDANGDDEVAVAGPIGSDASYANYPPPQKEASGSDATVAFGFVSVVGAVAGSLFFF

>AT3G22053.1 | Symbols: | FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; BEST Arabidopsis thaliana protein match is: Receptor protein kinase-related (TAIR:AT3G22030.1); Has 46 Blast hits to 46 proteins in 2 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 46; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink). | chr3:7767433-7768299 FORWARD LENGTH=149

MHSSYSLSKCLVCFTILAIQTLIRRVSSLNRTNAYLNHKCLVIIKGNISGRASTRKTLTILLIISHLPKNFRMVSHTHLVARLQIL**SPS**YFNVGA**TPTVPS**VVLAMPP**PS**QGFVGDVKDTREQLYESFNKKTRDFLYNLMLKADKPKSR

>AT3G54040.1 | Symbols: | PAR1 protein | chr3:20014032-20015299 REVERSE LENGTH=183

MSFSSLKLPIFLILSSLLHAAIGENIVCENL**PT**NMCSFSISASGKRCILETANVAGEFTCRTSAVDVEGIVNHVETDECVSACGVDRKTVGISSDSLMEAGFAAKLCSSACLDYCPNILDLYFNLAAGEGAFLPDLCDAQRMNPQRSMMEFISSGA**AP**G**PV**SEI**AP**G**PT**SEEVS**SPA**L**APA**SM

>AT1G51913.1 | Symbols: | LOCATED IN: endomembrane system; BEST Arabidopsis thaliana protein match is: cryptdin protein-related (TAIR:AT1G51915.1); Has 22 Blast hits to 22 proteins in 3 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 22; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink). | chr1:19292930-19293249 FORWARD LENGTH=77

MANERSSSILFLSVLVLVLVF**SPT**LQCQAARVHLDVEGHMLLR**APV**PIRFC**PA**CACC**APAP**KGSCCPCRCTNNPQRM

>AT4G28160.1 | Symbols: | hydroxyproline-rich glycoprotein family protein | chr4:13980390-13980725 FORWARD LENGTH=111

MAYTLQNLTATIVLVSLILGCTEEVTGMRYIP**ISPTPS**ESKHSDFPVNM**VP**Q**PS**SLIPGFGRFLLP**PTP**KLPFLPYKDPLAV**APA**TSNG**TPS**NKAR**SPS**SGEDEVPPVPQP

>AT4G22010.1 | Symbols: sks4 | SKU5 similar 4 | chr4:11663429-11666463 FORWARD LENGTH=541

MRGSCKVSIVLLLVLINGVLGDNPYRFFTWKITYGDIYPLGVKQQGILINGQFPGPHIDAITNDNIIISVFNYLKEPFLISWNGVQQRKNSWQDGVVGTTCPIPPGKNFTYVIQVKDQIGSFYYF**PS**LAFHKAAGAFGAIRVWSRPRI**PVP**F**SP**PDGDFWLLAGDWYKTNHYVLRRLLEAGRNLPNPDGVLINGRGWGGNTFTVQPGKTYRFRISNVGVATSLNFRIQGHTMKLVEVEGSHTVQNIYTSLDIHLGQSYSVLVTANQ**AP**QDYYIVISSRFTRKVLTTTSILHYSNSRKGVSG**PVP**NG**PT**LDIASSLYQARTIRRNLTASGPRPNPQGSYHYGLIKPGRTIILANS**AP**WINGKQRYAVNGASFV**AP**D**TP**LKLADYFKIPGVFNLGSI**PTSPS**GGNGGYLQSSVMAANFREFIEVVFQNWENSVQSWHVSGYSFFVVGMDGGQW**TP**GSRAKYNLRDAVSRSTVQVYPRAWTAIYIALDNVGMWNIRSENWARQYLGQQFYLRVYTSSTSYRDEYPPPKNALMCGRAKGRHTRPF

>AT5G09480.1 | Symbols: | hydroxyproline-rich glycoprotein family protein | chr5:2951181-2951615 REVERSE LENGTH=144

MTNVRLPILVSSVLFLVFISSSLLF**PTPS**FAARVSHSLVQEEVKK**VP**EYTEPEEPE**VP**EEPEL**PSP**EEPEIPEEPEIPEEPE**VP**GEPE**VP**EEPEEPEEPEGSTFEF**PS**WF**PS**FPIPGVNGGLPKTEKTK**PT**STVEEVNVSNKKP

>AT1G12845.1 | Symbols: | unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; Has 8 Blast hits to 8 proteins in 4 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 8; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink). | chr1:4379023-4379382 REVERSE LENGTH=119

MALFSSVSVSAFLLIVVSVQWTLVCSESTILA**SPA**VLPYIN**AP**DMSSFF**PSPT**KDWPIETAT**SPVP**EPE**AP**G**PS**SGQLNGKISGRSMRLHPDISLVLAIVGICTFLCFNEPILVRFLHY

>AT5G10946.1 | Symbols: | unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink). | chr5:3456210-3457244 REVERSE LENGTH=55

MLLDSFFILGFSLFNHFSLSLYRERSKAADTCMLCMKERVNSSTHEL**APTPS**FKR

>AT2G17230.1 | Symbols: EXL5 | EXORDIUM like 5 | chr2:7494892-7495983 REVERSE LENGTH=363

MS**SPA**TTITFFFFFTLSSFFYITSSLQNNNNNKHTATVNSLNI**PS**AAAEITLVNPKLPPRSLSLTSSKKFEGSSNLVHLRYHMG**PV**LSS**SP**INIYVIWYGQWSRPHKSLIRDFLNSISDAK**APSPS**VSEWWRTASLYTDQTGSNVSRSVLIAGEYSDSKYSHGQHLTRLTIQEVIASAARSASF**PV**DHKNGMYLVLTSHDVTMQDFCRAVCGFHYFTF**PS**MVGYTMPYAWVGQSGKQCPEVCAYPFALPGYMGHGGPGELRPPNGETGVDGMVSVIGHELAEVVSNPLINAWYAGED**PTAPT**EIGDLCEGLYGSGGGGYIGQVMRDREGKTFNMNGKGGRKFLVQWIWNPNLKACSGPNSVD

>AT5G51105.1 | Symbols: | Protein of unknown function (DUF1278) | chr5:20777632-20778003 FORWARD LENGTH=123

MSIKSVVSLLVVVCIAASVNAQLPQFPPIMFPFPNPFK**PSP**GMAG**SP**DLSKCWSTVMDLPGCFEEIQQAVMSGKFEGVG**PA**CCKAFLDAEANCLPNL**PS**NPFFPPMMKHQCSKMA**SPPT**TTAY

>AT2G37478.1 | Symbols: CPuORF52 | conserved peptide upstream open reading frame 52 | chr2:15738139-15738270 FORWARD LENGTH=43

M**PSVPS**KLLPLLSLI**PS**L**TP**YHSGLSSNRLKPLEKCDWRGNLS

>AT1G41830.1 | Symbols: SKS6 | SKU5-similar 6 | chr1:15603892-15607802 REVERSE LENGTH=542

MMAVGRSGGTILLFCLSFFAAVTAE**SP**YRFFDWNVTYGDIYPLGVRQQGILINGQFPGPDIHSV
TNDNLIINVHNSLDEPFLISWNGVQNRRNSYVDGMYGTTCPIPPRSNYTYILQVKDQIGSFYYF
PSLAFHKAAGGFGGIRILSRPGIPVPFAD**PA**GDYTVLIGDWYKFNHTDLKSRLDRGRKL**PSP**DG
ILINGRSNGATLNVEQGKTYRLRISNVGLQDSLNFRIQNHRMKLVEVEGTHTLQTMFSSLDVHV
GQSYSVLITADQ**SP**RDYYVVVSSRFTDKIITTTGVLRYSGSS**TPA**SGPIPGG**PT**IQVDWSLNQA
RAIRTNLTASGPRPNPQGSYHYGLIPLIRTIVFGSSAGQINGKQRYGVNSVSF**VPA**D**TP**LKLAD
FFKISGVYKINSISDK**PT**YGGLYLDTSVLQVDYRTFIEIVFENQEDIVQSYHLNGYSFWVVGMD
GGQWKTGSRNGYNLRDAVSRSTVQVYPKSWTAIYIALDNVGMWNLRSEFWARQYLGQQLYLRVF
TSSTSLRDEYPIPKNSRLCGRARGRHTRPL
>AT5G67400.1 | Symbols: RHS19 | root hair specific 19 |
chr5:26894896-26896300 FORWARD LENGTH=329
MARFSLVVVVTLSLAISMFPDTTTAQLKTNFYGNSCPNVEQIVKKVVQEKIKQTFVT**IPA**TLRL
FFHDCFVNGCDASVMIQS**TPT**NKAEKDHPDNISLAGDGFDVVIKAKKALDAI**PS**CKNKVSCADI
LALATRDVVVAAKG**PS**YAVELGRFDGLVSTAASVNGNLPGPNNKVTELNKLFAKNKLTQEDMIA
LSAAHTLGFAHCGKVFNRIYNFNLTHAVD**PT**LNKAYAKELQLACPKTVDPRIAINMD**PTTP**RQF
DNIYFKNLQQGKGLFTSDQVLFTDGRSK**PT**VNDWAKNSVAFNKAFVTAMTKLGRVGVKTRRNGN
IRRDCGAFN
>AT5G52390.1 | Symbols: | PAR1 protein | chr5:21264281-21265173
REVERSE LENGTH=195
MALKTVFVAFMILLAIYSQTTFGDDVKCENLDENTCAFAVSSTGKRCVLEKSMKRSGIEVYTCR
SSEIEANKVTNIIESDECIKACGLDRKALGISSDALLESQFTHKLCSVKCLNQCPNVVDLYFNL
AAGEGVYLPKLCESQEGKSRRAMSEIRSSGIAMDTL**APV**G**PV**MLGEI**AP**E**PA**TSMDNMPY**VPAP**
SPY
>AT3G47295.1 | Symbols: | unknown protein; FUNCTIONS IN:
molecular_function unknown; INVOLVED IN: biological_process
unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 22 plant
structures; EXPRESSED DURING: 13 growth stages; Has 13 Blast
hits to 13 proteins in 2 species: Archae - 0; Bacteria - 0;
Metazoa - 0; Fungi - 0; Plants - 13; Viruses - 0; Other
Eukaryotes - 0 (source: NCBI BLink). | chr3:17428054-17428371
FORWARD LENGTH=71
MSFGTRLLLFLILTLPLVTSS**SP**NTLHVSGIVKTGTTSRFLMMTIEDYDD**PS**ANTRHD**PSVPT**N
AKADT**TP**
>AT3G47300.1 | Symbols: SELT | SELT-like protein precursor |
chr3:17428685-17429555 REVERSE LENGTH=209
MDKTQLILLGLPIFLLCSDLFNLF**TP**PPPKSQHQ**SPPS**ISETLDF**PA**QKSTGVGYGNTVEINFC
ISCSYKGTAVSMKKMLESVFPGLDVVLANY**PAPAP**KRILAKVVPVAQVGVIGLIMGGEQIFPMI
GIAQP**PA**WYHSLRANRFGSMASTWLLGNFLQSFLQSSGAFEVSCNGELVFSKLKEGRFPGEIEL
RDLISGTMTKPFVTGSY
>AT5G04030.1 | Symbols: | unknown protein; FUNCTIONS IN:
molecular_function unknown; INVOLVED IN: biological_process
unknown; LOCATED IN: endomembrane system; EXPRESSED IN: egg
cell; Has 1807 Blast hits to 1807 proteins in 277 species:
Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants -
385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink). |
chr5:1088462-1088692 REVERSE LENGTH=76
MALRMNHVRFLLILIVVAFVAGESFTVKQDSAKGDVIWEEVFHGDYGSW**SPTP**KIPRR**SPAP**IP
HDDF**AP**PRRLKA

>AT2G14900.1 | Symbols: | Gibberellin-regulated family protein | chr2:6404210-6405043 FORWARD LENGTH=108
MKIIVSILVLASLLLISSSLASATISDAFGSGAV**APAP**QSKDG**PA**LEKWCGQKCEGRCKEAGMK
DRCLKYCGICCKDCQC**VPS**GTYGNKHECACYRDKLSSKG**TP**KCP
>AT4G10457.1 | Symbols: SCRL1 | SCR-like 1 | chr4:6466900-6467270 REVERSE LENGTH=92
MKYGVLFMVSCGVMFLILSHVEEVEAMKKFGCNTTHPFPGKCGNNGKSSWVSDMKKL**PSAP**KNR
DIRCECSDR**PS**LARGMPGERVCRCDYDC
>AT4G13263.1 | Symbols: | Protein of unknown function (DUF784) | chr4:7711064-7711513 FORWARD LENGTH=149
MAKSLLMVMLVSIVMFYMACPIFSQEIHEDVAI**SPTP**FEEAN**SPA**MEYDMKLPHYSQKVFDFLE
TCAEK**PS**SICGGEIFQNVLDATTLVTDKCCRDILKIGKDCHLGLIKIFFSSYEYKNIASIPRSK
QTWNDCFRRVGSKIGVPVSIE
>AT2G21110.1 | Symbols: | Disease resistance-responsive (dirigent-like protein) family protein | chr2:9050290-9050850 REVERSE LENGTH=186
MGKNLGLVVSFYLCITFALGEYFSETRPI**TP**KQLVVTNLHFFFHDTLT**AP**N**PS**AILIAK**PT**HTR
GDNDS**SPSP**FGSLFALDDPLTVGPDPKSEKIGNARGMYVSSGKH**VPT**LTMYVDFGFTSGKFNGS
SIAVFSRNTITEKEREVAVVGGRGRFRMARGVAQLNTYYVNLTNGDAIVEYNVTLYHY
>AT1G76160.1 | Symbols: sks5 | SKU5 similar 5 | chr1:28578211-28581020 REVERSE LENGTH=541
MAGSASFAAALFIGLSLLFAVTAEDPYRFFEWNITYGDIYPLGVRQQGILINGAFPGPDIHSVT
NDNLIINVYNSLDEPFLLSWNGIQQRRNSFVDGVYGTTCPIPPGKNYTYILQMKDQIGSFYYF**P**
SLGFHKAAGGFGGIRILSRPRI**PVP**FPD**PA**GDTTVLIGDWYKANHTDLRAQLDNGKKLPLPDGI
LINGRSSGATLNVEQGKTYRFRISNVGLQDSLNFRIQDHKMKVVEVEGTHTLQTTFSSLDVHVG
QSYSVLVTADQ**TP**RDYYVVVSSRFTSNVLTTTGIFRYSNSAGGVSGPIPGG**PT**IQIDWSLNQAR
AIRTNLSASGPRPNPQGSYHYGMINTTRTIRLASSAGQVDGKQRYAVNSVSFK**PA**D**TP**LKIADY
FKIDGVYRSGSIQYQ**PT**GGGIYLDTSVMQVDYRTFVEIIFENSEDIVQSWHLDGYSFWVVGMDG
GQW**SP**DSRNEYNLRDAVARCTVQVY**PS**SWTAILIALDNVGMWNLRSEFWARQYLGQQLYLRVY**T**
PSTSLRDEYPIPKNALLCGRASGRSTRPL
>AT2G40530.1 | Symbols: | unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: leaf apex, inflorescence meristem, petal, flower; EXPRESSED DURING: 4 anthesis, petal differentiation and expansion stage; Has 5 Blast hits to 5 proteins in 2 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 5; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink). | chr2:16927600-16927917 FORWARD LENGTH=105
MDATKIKFDVILLSFLLIISGI**PS**NLGLSTSVRGTTRSEPEAFHGGKF**PA**MKMRKLM**AP**NMEVD
YSSDYYDGGSSSTT**SPSP**P**VP**DYDDIYRRQGD**VPSP**GIGH
>AT1G21850.1 | Symbols: sks8 | SKU5 similar 8 | chr1:7667803-7670530 REVERSE LENGTH=551
MEVKSVNTTAMILGLFFLISFVAAEDPYKFFEWHVTYGNI**SP**LKVAQQGILINGKFPGPDIAAV
TNDNLIINVFNHLDEPFLISWSGIRNWRNSYQDGVYGTTCPIPPGKNYTYALQVKDQIGSFYYF
PSLGFHKAAGGFGAIRISSRPRI**PVP**F**PAPA**GDYTVLIGDWYKTNHKDLRAQLDNGGKLPFPDG
ILINGRGSGATLNIEPGKTYRLRISNVGLQNSLNFRIQNHKMKLVEVEGTHTIQ**TP**FSSLDVHV
GQSYSVLITADQ**PA**KDYYIVVSSRFTSKILITAGVLHYSNSAG**PV**SGPIPE**AP**IQLRWSFDQAR

AIKTNLAASGPRPNPQGTYHYGKIKVTRTIKLASSAGNINGKQRYAVNSASFY**PT**D**TP**LKLADY
FKIAGVYNPGSIPDQ**PT**HGAIYPVTSVMQTDYKAFVEIVFENWEDIVQTWHLDGYSFFVVGMEL
GKWSAASRKVYNLNDAVSRCTVQVYPRSWTAIYVSLDNVGMWNLRSELWERQYLGQQFYMRVY**T**
PSTSLRDEYLIPKNALLCGRATGHHTT**TP**GPLSEGSERF
>AT4G26466.1 | Symbols: LRE | lorelei | chr4:13367645-13368629 REVERSE LENGTH=165
MELILLFFFLMALLVSLSSSSSISDGVFESQTSVSGRNLRHAKKKCEVNFEYMDYKVLTKRCKG
PAF**PA**KECCSAFKEFACPYVSQINDMNSDCAQTMFSYMNIYGNY**PT**GLFANECRERKDGLVCPL
PPLYSHNLNASTADS**TP**RFISLLISAATAVFALLVLT
>AT2G15340.1 | Symbols: | glycine-rich protein | chr2:6677339-6677698 FORWARD LENGTH=119
MALSGSQKKVFLLVIAIACLSSSLAEAWSWSSSDGNGWGWGSDGSSTSTSGPGS**TP**GDSNSGGS
NPGGPNSGGFS**SP**WGSWGWGWGSHGSG**SPAPS**NALFLLVIIKKKNLSDFLFYYIM
>AT2G28790.1 | Symbols: | Pathogenesis-related thaumatin superfamily protein | chr2:12354664-12355413 REVERSE LENGTH=249
MAKTSLPLAASFLLLISFSSAVDTSRLFLTVVNNCPFTVW**PA**IQPNAGHPVLEKGGFAL**PT**FTH
RSFN**VPT**THWSGRIWARTWCAHYNGKFSCLTGDCGNRLECNGLGG**APPA**SLAQFDLHHGGHHDF
SSYGVSLVDGYNVPMTV**TP**HEGHGVCPVVGCREDLIKTC**PA**HLQVRSHSGHVVACKSGCEAFHT
DELCCRGHYN**SP**NTCKASSHSLFFKHAC**PS**SFTFAHD**SPS**LMHDCA**SP**RELKVIFCH
>AT4G20990.1 | Symbols: ATACA4, ACA4 | alpha carbonic anhydrase 4 | chr4:11219772-11221126 FORWARD LENGTH=267
MDTNAKTIFFMAMCFIYLSFPNISHAHSEVDDE**TP**FTYEQKTEKGPEGWGKINPHWKVCNTGRY
Q**SPI**DLTNERVSLIHDQAWTRQYK**PAPA**VITNRGHDIMVSWKGDAGKMTIRKTDFNLVQCHWH**S**
PSEHTVNGTRYDLELHMVHTSARGRTAVIGVLYKLGEPNEFLTKLLNGIKAVGNKEINLGMIDP
REIRFQTRKFYRYIGSLTVPPCTEGVIWTVVKRVNTISMEQITALRQAVDDGFETNSR**PV**QDSK
GRSVWFYDPNV
>AT4G22900.1 | Symbols: | Protein of unknown function (DUF1191) | chr4:12010221-12011252 FORWARD LENGTH=343
MDSFSFCIIIMILLLSFHQSKSQLIQSTHLLDLMIRDYTIRNFKLNFNTGVTQKIYL**PS**NFSGI
DIDTVKLRCGSLRRYGAKIGEFHIGSGLTVEPCPERVMLIRQNFGSNWSSIYSTGYNLSGYNYK
LV**SPV**LGLLAYNANPDGVARNPYEVNVVGTDQNPILIDFLINKATNNT**SP**N**PT**KKNSSVLCACF
TSNSNTTFSEQV**SP**YVCKGTRQGHYALVMKTEAQKDDHEGGGSSGGVVASSTEVNGGNGGGKLS
RWKVAVGSVIGSGIGAILLGMLVVAMLVKGKKKAMREEMERRAYEEEALQVSMVGHVR**APTAP**G
TRTLPRISDDRYNKNTHLNSNYP
>AT1G21680.1 | Symbols: | DPP6 N-terminal domain-like protein | chr1:7613028-7615148 FORWARD LENGTH=706
MNISQIIFFASLLLRLHLSTAEHQNSNAGDGDTIIFTTLGRSHYEFDIFALSTTQP**PS**VSGELR
ITDGESVNFNGYF**PSPSPA**LLSLLPDETLIQMEDS**SP**LHLIYVTERNGTSSLYYDLVYGGNSDF
KTKRRSVLE**APS**RVQVPLLSRFDHLSGMTVNSFKDK**PS**LSGEFIVYVSTHESSGEPRASWTAVY
STELKTGLTRRL**TPS**GVADF**SPA**V**SPS**GNLTAVASYGERGWTGEVEELRTDIYVFLTRDGSHRV
KVVEHGGWPCWVDESTLYFHRRSEEDGWISVYRAILPENGPLTTESVTIQRV**TP**PGVHAF**TPA**T
SPNNHEFVAVATRRPGSDYRHVELFDLKRNEFIELTRLV**AP**KSHHLNPFL**SP**DSSRVGYHSCRG
DANGRR**SP**LLFLENIQTTTRDLSLFRIDGSF**PS**F**SP**GGDRIAYVKMPGVFVVKPDGSGQREVYK
GMAFSTAWDPVRPGIVYSSSG**PT**FATERTEVDVISIDVDAADKSSSVRRLTTNGKNNAFPW**PSP**
DGKRIVFRSGRTGHKNLYIMDAEKGESGGLWRLTEGAWTDTMCNW**SP**DGEWIAFASDRE**SP**GSG
SFELFLIHPNGTGLRKLIQSGTGGRTNHPIF**SP**DSKSLVFTSDYAGISAEPISNPHHYQPYGDI
FTVKLDGSNVRRLTHNSYEDG**TPA**W**AP**RFIHPNNVELQRRNDSRCSFEDCHWLNKY**PT**LKGRKI
SC

>AT3G18715.1 | Symbols: IDL4 | inflorescence deficient in
abscission (IDA)-like 4 | chr3:6441067-6441348 FORWARD LENGTH=93
MY**PT**RPHYWRRRLSINRPQAFLLLILCLFFIHHCDASRFSSSSVFYRNPNYDHSNNTVRRGHFL
GFLPRHL**PVPA**S**APS**RKHNDIGIQALL**SP**
>AT3G45252.1 | Symbols: | ECA1 gametogenesis related family
protein | chr3:16586074-16586364 FORWARD LENGTH=96
MQRKLAMNMIALIIMLSIFTQTTGNDL**APAP**H**PT**IPCLDNVKTIPNCVKAVFHFKFKEITETCC
TILLTLPDDCFGLLFPIPRVYHFLLSSACKNI
>AT3G27410.1 | Symbols: | unknown protein; Has 10 Blast hits to
10 proteins in 2 species: Archae - 0; Bacteria - 0; Metazoa - 0;
Fungi - 0; Plants - 10; Viruses - 0; Other Eukaryotes - 0
(source: NCBI BLink). | chr3:10145244-10145642 REVERSE
LENGTH=132
MASHFFLVFIFVILTVANGISAADSEKSQ**SPS**SSKTLADNK**PT**VDNSTKSLIDSLG**PS**QDYPDY
EIPLEL**AP**DGVVVVGDY**AP**I**SP**RG**TP**DTLAQSEADQDVKTSTSSSASRSSSTVLAIVVVMGGAS
LFFF
>AT4G26880.1 | Symbols: | Stigma-specific Stig1 family protein |
chr4:13508265-13508723 REVERSE LENGTH=152
MAFVKLLVSIAITTAITIAVITTITNNTTTIREYTSFD**APSTPT**IRPNRLLAQKEVGERNPNAA
DHCNRNPEIC**TP**YGGGGSNSTMTCCNNKCIDVSSDDNNCGACKNKCKFSQTCCRGQCVYVAYDK
RHCGQCNHPCELGEFCVYGLCNYA
>AT5G66920.1 | Symbols: sks17 | SKU5 similar 17 | chr5:26722963-
26725370 FORWARD LENGTH=546
MKMASRKTTSLLNHLLLLGALTLLSSLVIVKGE**SP**YKFYTWTVTYGII**SP**LGVPQQVILINGQF
PGPKLEVVTNDNIILNLINKLDQPFLLTWNGIKQRKNSWQDGVLGTNCPIQPNSNFTYKFQTKD
QIGTFNYF**PS**TAFHKAAGGFGAINVYARPGIPIPYPL**PT**ADFTLLVGDWFKTNHKTLQQRLDSG
GVLPFPDGMLINGQTQSTFSGDQGKTYMLRISNVGLSSTFNFRIQGHTMKVVEVEGSHVIQTDY
DSLDIHVGQSLAVLVTLNQ**SP**KDYYIVASTRFIRSKLSVMGLLRYSNSR**VPA**SGDP**PA**LPPGEL
VWSMRQARTFRWNLTANAARPNPQGSFHYGMI**SPT**KTFVFSNS**AP**LINGKQRYAVNGVSYVKSE
TPLKLADHFGISGVFSTNAIQS**VPS**N**SPPT**VATSVVQTSHHDFLEIVFQNNEKSMQSWHLDGYD
FWVVGFGSGQW**TPA**KRSLHNLVDALTRHTTQVYPESWTTILVSLDNQGMWNMRSAIWERQYSGQ
QFYLKVWNSVQSLANEYNPPDNLQLCGKAVGRHV
>AT1G21670.1 | Symbols: | LOCATED IN: cell wall, plant-type cell
wall; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 14
growth stages; CONTAINS InterPro DOMAIN/s: WD40-like Beta
Propeller (InterPro:IPR011659), Six-bladed beta-propeller, TolB-
like (InterPro:IPR011042); BEST Arabidopsis thaliana protein
match is: DPP6 N-terminal domain-like protein
(TAIR:AT1G21680.1); Has 8461 Blast hits to 5060 proteins in 1257
species: Archae - 79; Bacteria - 5567; Metazoa - 37; Fungi - 70;
Plants - 117; Viruses - 0; Other Eukaryotes - 2591 (source: NCBI
BLink). | chr1:7610409-7612520 FORWARD LENGTH=703
MKLTNALFFFFLICLLSLSSASSNKPQLSNGSTILFTTIGR**PT**FEFDIFTL**PT**SHRP**PSPA**DEH
RLTDGKSINFNGYFA**SPS**TALISLLPKRTQIQPQDVHLIYVTERAG**TPS**LNYDVVHSDNVGSRI
QVPLFSGEEQQSGMNVNSMKD**TP**VLTNGYLVHVSTHENPGKPMASWAAVYSTELRTKSTRRL**TP**
LGIADF**SPA**V**SPS**GKWTAVASFGEKGWTWSMVEKEISSDVYVFLTQDGTQRVKVVEQGGWPRWV
DDSTLYFHRKSDDGWISVYRAILPKTG**PV**TTKSVTIQRV**TP**PGLHAF**TPA**T**SP**NNNNFIAVATR
RPGSEIRHVELFDLKKNEFVELTRLV**SP**KSHHFNPFL**SP**DSSRVGYHSCRGDATGRK**TP**RNLLQ

SLKTTSNDLSLFRFDGAF**PS**I**SP**EGDRFAFVSFTGVFVVNPDGSGLRQLLPQMGFGTVWDPIRH
GIVYTSSG**PA**L**AP**GKSQIDILAINVD**APSPA**TAVKKLTTTGENNAFPW**PSP**DGKRIVFRSARSG
TKNLYIMDAEKGESGGLFRLTNGNWNDTIATW**SP**DGNWIVFASNREFPGTLLMNIYVVHPDGTG
LRKLAQNLTGLVSMHPMF**SP**DSKRIVFTTIYAGISAEQIGNPHFN**VPS**SEIFTVNLDGSGLTRL
THNSVEDGPPMWFPKIKATGDVAWPKRFG**PS**CSIQDFKTQNTTVKMKMNK**PA**TMSSMCV**VPS**Q
>AT5G04180.1 | Symbols: ATACA3, ACA3 | alpha carbonic anhydrase
3 | chr5:1147907-1149237 REVERSE LENGTH=277
MKTIILFVTFLALSSSSLADETETEFHYKPGEIAD**PS**KWSSIKAEWKICGTGKRQ**SP**INL**TP**KI
ARIVHNSTEILQTYYKPVEAILKNRGFDMKVKWEDDAGKIVINDTDYKLVQSHWH**APS**EHFLDG
QRLAMELHMVHKSVEGHLAVIGVLFREGEPNAFISRIMDKIHKIADVQDGEVSIGKIDPREFGW
DLTKFYEYRGSLT**TP**PCTEDVMWTIINKVGTVSREQIDVLTDARRGGYEKNAR**PA**QPLNGRLVY
LNEQS**SPSPTP**RLRIPRVGPV
>AT1G51915.1 | Symbols: | cryptdin protein-related |
chr1:19293587-19293865 FORWARD LENGTH=67
MATERFSTMLISVLVLALVL**SP**ILPCQATRAHLDAETRMLRRVC**PS**CVCC**APAP**RGACCPCRCP
KNP
>AT4G24350.1 | Symbols: | Phosphorylase superfamily protein |
chr4:12609637-12611328 FORWARD LENGTH=336
MDHTLSLLPFLFIIIISSFHVL**PVSPTPS**HKLKSATTIRKLNRGGPYIGLVTVIATEENAFLRS
VDFRPD**PT**HPFLDLSGRRFRIGKIHGKKVVYVRCGRGMVNGAAATQQMIDVFNVKGIVHFGIAG
NMNNSMSIGDVSIPKQITNAGLWDWLNPDKVKGVEDIAYLDVGNYNVPKGDGDNELGSIGYNYE
QLYSVTGHIN**AP**QNVFWINTTQEWLHLAADLEKMELSQCVNASLCLPKKPKLVVGLKAATAYIF
VDNAAYRNFLYDTFGVSSSDMESSAVAMTCASNGYPIIVIRGLSDLAGEGGDNTVHKFGSLAAT
NTAKAVLEFIRKLPPY
>AT3G47070.1 | Symbols: | LOCATED IN: thylakoid, chloroplast
thylakoid membrane, chloroplast, chloroplast envelope; EXPRESSED
IN: 22 plant structures; EXPRESSED DURING: 13 growth stages;
CONTAINS InterPro DOMAIN/s: Thylakoid soluble phosphoprotein
TSP9 (InterPro:IPR021584); Has 37 Blast hits to 37 proteins in
10 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0;
Plants - 37; Viruses - 0; Other Eukaryotes - 0 (source: NCBI
BLink). | chr3:17337205-17337507 REVERSE LENGTH=100
MVSSLLMSF**APA**TVRVYATSTKGGSGGPKEEKNPIDFVLGFMTKQDQFYETNPLLKKVDEKEGT
TTGGRGTVRGGKNS**APTPVP**KKSEGGFGGLGSLFKK
>AT2G04031.1 | Symbols: | Protein of unknown function (DUF784) |
chr2:1304324-1304779 REVERSE LENGTH=151
MAKSFSMVMLVSIIIFYMARPIYSRNKILEDIDI**SPTPV**EGTD**SPT**TEYEIELAHRLHEDYILA
CPKK**PS**SKCDDEIFNNMLDGT**TPV**TNECCHDILKTGKDCHLALVKIIFSTNDYKKIASKAIPKS
KQTWNDCVRRVGMEIG**APV**SFEP
>AT2G21740.1 | Symbols: | Protein of unknown function (DUF1278)
| chr2:9281986-9282363 REVERSE LENGTH=125
MASNTSFLFATIAILLVLNISGRTLPETEDSTNIAARLNGGGLMECWNALYELKSCTNEIVLFF
LNGETKLGVDCCQAVEVITTDCW**PA**MLTSLGFTSDETNVLRGFCQ**SP**NSGGS**SPAPS**SVKL
>AT5G34581.1 | Symbols: | hydroxyproline-rich glycoprotein
family protein | chr5:12920366-12921018 FORWARD LENGTH=161
MMISNLLACKFSEVSSMHRLLVLCCITDCFSE**PA**TRPHK**SP**PQTIQ**TPTPA**RSHSQSGSQQPPL
IRRKQQQPLACQHQIPYLQEQ**PS**QDPPETQFIPNHQNPNNHEEEDGEEDGDEFKEEDGEEVDEE
EQNPNVDYQELLDRLLALPGRQHLMILSQEPIP

>AT3G14860.2 | Symbols: | NHL domain-containing protein |
chr3:4998591-5000894 REVERSE LENGTH=493
MGKSQILLNAESFTCSSFNTFFTLWIFFTLHSFPFQAQA**APS**GSLIKHMSSVLKWTTGSSSKLS
QSDTNVLQFENGYLVETVVEGNDIGVVPYKIRVSDDGELYAVDELNSNIMKI**TP**PLSQYSRGRL
VAGSFQGKTGHADGK**PS**EARFNHPRGVTMDDKGNVYVADTLNLAIRKIGDSGVTTIAGGKSNIA
GYRDG**PS**EDAKFSNDFDVVYVR**PT**CSLLVIDRGNAALRQISLSEEDCDYQDDSSISLTDILLVI
GAVLIGYATCMLQQGFGNSFFSKTQVGSETSYEEEHPGKEKLSRPVHEKTITKEEPGW**PS**FGQL
LTDLCKLALEFITSHL**VPA**RFQTNPNLRPLKDRLIMPEDEQEPPRVQRHT**APAP**ISESRHAHLP
KADDSYPEHK**TP**KLRSSSVMKD**PT**LSSSKHHRTSSKRQDYAQFYASGEVAQPKIHKERSRRRHR
DKTTETEPK**PTPS**DTVKPVEYSNSSKFDHYNMRSSKYGPE**TP**FRF
>AT5G44570.2 | Symbols: | unknown protein; FUNCTIONS IN:
molecular_function unknown; INVOLVED IN: biological_process
unknown; LOCATED IN: endomembrane system; EXPRESSED IN: leaf
whorl, hypocotyl, sepal, flower, leaf; EXPRESSED DURING: petal
differentiation and expansion stage, LP.08 eight leaves visible;
Has 7 Blast hits to 7 proteins in 2 species: Archae - 0;
Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 7; Viruses - 0;
Other Eukaryotes - 0 (source: NCBI BLink). | chr5:17964439-
17966496 FORWARD LENGTH=140
MATKTSNFVSLRVSLFILLLFISSQVAIADAKHLQHSTTIKKSKRSSIQSTDAACAAATTT**APA**
PDAECSHT**SP**NAAVVTTAAENACIITAATNALSTSFVAVCFIADVYITKEKDDAGGSRITKKEA
DASCDDECKKIK
>AT5G10946.1 | Symbols: | unknown protein; FUNCTIONS IN:
molecular_function unknown; INVOLVED IN: biological_process
unknown; LOCATED IN: endomembrane system; Has 30201 Blast hits
to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396;
Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other
Eukaryotes - 2996 (source: NCBI BLink). | chr5:3456210-3457244
REVERSE LENGTH=55
MLLDSFFILGFSLFNHFSLSLYRERSKAADTCMLCMKERVNSSTHEL**APTPS**FKR
>AT3G29033.1 | Symbols: | glycine-rich protein | chr3:11024007-
11024889 FORWARD LENGTH=167
MVKVEISILSLYGFIAGDDDSQSS**SPSP**GHGQN**SPSP**RRYGQV**TP**NPRFHGYQAGGSSGRNYD**S**
PNHGRGTGQN**SP**NPGCYDQ**VTP**YQGSAQFYGYQADYADGYQGNLAGNPPGSGGYQGIGNVNSGG
YQVNQNSYENSCGYYQGDQVINGIQMVTVLIKELKDEED
>AT1G16950.1 | Symbols: | unknown protein; FUNCTIONS IN:
molecular_function unknown; INVOLVED IN: biological_process
unknown; LOCATED IN: endomembrane system; EXPRESSED IN: stamen;
EXPRESSED DURING: 4 anthesis; Has 17 Blast hits to 17 proteins
in 5 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0;
Plants - 17; Viruses - 0; Other Eukaryotes - 0 (source: NCBI
BLink). | chr1:5796083-5796364 FORWARD LENGTH=93
MARPRISISMICLLILIVGFVLQSSQARKVLVPYGTSKGLFLSALPKGNVP**PS**G**PS**DKGHT**SP**P
DDTDQRM**VP**EN**SP**EIYRRLES**VPSP**GVGH
>AT3G01960.1 | Symbols: | unknown protein; Has 13 Blast hits to
13 proteins in 5 species: Archae - 0; Bacteria - 0; Metazoa - 0;
Fungi - 0; Plants - 13; Viruses - 0; Other Eukaryotes - 0
(source: NCBI BLink). | chr3:324581-325087 REVERSE LENGTH=138

MEMNRRRSRGFMIAKLMPFCKAVK**SPTPS**QDLYNNVHYTSSTTTSYARQPDFNNYVTAT**TP**LPP
KVSFLLQ**PS**L**AP**EGKDMEKKLNMMAEKLIGRGINGVDDCVDARAASYISSVRERFKADQYCEKL
TTVISLDDEN
>AT5G35405.1 | Symbols: | Protein of unknown function (DUF784) |
chr5:13604277-13604726 REVERSE LENGTH=149
MAKSLLMVMLVSIVMFYMACPILSQEVHEDVVI**SPTP**FEKAI**SPA**MEYEMKLPHYSQKQYDFLE
ACVEK**PS**SICGGEIFQNVLDETNLVTDKCCRNILKIGKDCHLGLMKILFSSYEYKNIASIPRSK
QTWNDCLRRVGSKIG**VPV**SFE
>AT4G12420.1 | Symbols: SKU5 | Cupredoxin superfamily protein |
chr4:7349941-7352868 REVERSE LENGTH=587
MDLFKILLLVFFVNISFCFAADPYSFYNFEVSYITA**SP**LGVPQQVIAINGKFPG**PT**INVTTNEN
LVVNVRNKLDEGLLLHWNGIQQRRVSWQDGVLGTNCPIPPKWNWTYEFQVKDQIGSFFYF**PS**LH
FQRASGGFGSFVVNPRAIIPVPFS**TP**DGDITVTIGDWYIRNHTALRKALDDGKDLGMPDGVLIN
GKGPYRYNDTLVADGIDFETITVHPGKTYRLRVSNVGISTSLNFRIQGHNLVLAESEGSYTVQQ
NYTSLDIHVGQSYSFLVTMDQNASSDYYIVASARVVNETIWRRVTGVGILKYTNSKGKAKGQLP
PGPQDEFDKTFSMNQARSIRWNVSASGARPNPQGSFKYGSINVTDVYVLRNMPPVTISGKRRTT
LNGISFKN**PSTP**IRLADKLKVKDVYKLDFPKRPLTG**PA**KVATSIINGTYRGFMEVVLQNNDTKM
QSYHMSGYAFFVVGMDYGEWTENSRGTYNKWDGIARSTIQVYPGAWSAILISLDNPGAWNLRTE
NLDSWYLGQETYVRVVNPDENNKTEFGHPDNVLYCGALSKLQKPQKVSSSASKSIGFTSLSMVV
MALVMMMMLQH
>AT5G60630.1 | Symbols: | FUNCTIONS IN: molecular_function
unknown; INVOLVED IN: biological_process unknown; LOCATED IN:
endomembrane system; EXPRESSED IN: 17 plant structures;
EXPRESSED DURING: 6 growth stages; BEST Arabidopsis thaliana
protein match is: hydroxyproline-rich glycoprotein family
protein (TAIR:AT3G45230.1); Has 1807 Blast hits to 1807 proteins
in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi -
347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source:
NCBI BLink). | chr5:24369923-24370342 REVERSE LENGTH=139
MKFDFIIVALVMVSGVALLMVSGEISTEEI**SPA**IEHSSSLPQSETEM**SPSPT**MSNDYDY**PS**SSQ
LTESNDLNYTDSTRPGGEEASVGGENGGGGKKTGIAVVGSIAAASMVGFGGYVLKKRRENIRR
SRYGYASTEFF
>AT3G13275.1 | Symbols: | unknown protein; FUNCTIONS IN:
molecular_function unknown; INVOLVED IN: biological_process
unknown; LOCATED IN: endomembrane system; Has 29 Blast hits to
29 proteins in 10 species: Archae - 0; Bacteria - 0; Metazoa -
0; Fungi - 0; Plants - 29; Viruses - 0; Other Eukaryotes - 0
(source: NCBI BLink). | chr3:4289425-4289634 REVERSE LENGTH=69
MCCGRICMLCTCLLLVVIAIGFLFGFGVFKDGFHKIHESIHLECDPRFGCGGGLGRRGYGF**PAP**
AGGFN
>AT2G20870.1 | Symbols: | cell wall protein precursor, putative
| chr2:8982239-8982661 REVERSE LENGTH=140
MASSLITSAVIVVVLSLVLGSVEQVSGLRH**VP**K**SP**KTTDVKHPDFLVTIEPK**PT**ILIPGVGRFL
LPPKCKKPFYPYN**PV**TG**AP**LTGGGI**PS**YNGGQGAGPHTQLPGGDDTLVPNPGFE**APTPT**IGAGT
GSNGQVPPVPLP
>AT5G51480.1 | Symbols: SKS2 | SKU5 similar 2 | chr5:20910433-
20913153 FORWARD LENGTH=592

MAATDFFFAFVFSFALIFGFSFAGDPYVSYDFTLSYITA**SP**LG**VP**QQVIAVNGKFPGPVINATT
NYNVHVNVLNHLDEPLLLTWPGVQMRRNSWQDGVLGTNCPIPPNWNFTYDFQLKDQIGSYFY**SP**
SLNFQRASGGFGALIINNRDLVPIPFTEPDGEIIFIIGDWYTQNHTALRRILDSGKELGMPDGV
LINGKGPFKYNSSVPDGIEHETVNVDPGKTYRIRVHNVGISTSLNFRIQNHKLLLIETEGRYTS
QMNFTDFDVHVGQSYSFLVTMDQNATSDYYIVASARFVNETVWQRVTGVGILHYSNSKG**PA**SGP
L**PV**SATDVNHPWSAMNQPRAIKQNTSASGARPNPQGSFHYGQINITRTYILRSLP**PT**KINGKLR
ATLNGISFVN**PSTP**MRLADDHKVKGDYMLDFPDRPLDEKLPRLSSSIINATYKGFIQVIFQNND
TKIQSFHIDGYAFYVVAMDFGIWSEDRNSSYNNWDAVARSTVEVYPGAWTAVLISLDNVGVWNI
RVENLDRWYLGQETYMRIINPEENGSTEMDPPENVMYCGALQAMQKEQHHSSATKSMTNGQLIL
IFSMMMVLLSSFSSFC

>AT5G44580.1 | Symbols: | unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 18 plant structures; EXPRESSED DURING: 12 growth stages; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT5G44582.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink). | chr5:17979327-17980181 FORWARD LENGTH=98

MERKKFSSKFIHLLIVFLLLCTFLSRTESALPYHHELFLGRKRMNYYKPNSAIG**TPS**STSDH**AP**
GSNGRKLMSIYRPNGDIFTG**PS**GSGHGGGR**TPAP**

>AT1G34800.1 | Symbols: | Plant thionin family protein | chr1:12773164-12773442 REVERSE LENGTH=92

MGIQTCSVLIIVAILTMMFSAHIAQSNSITMCVKHCAQNECLKAAKK**PTP**EICDEACKKICNNQ
LFGDEKWF**VPAP**KGSSRICRW**AP**KYCQF

>AT1G34805.1 | Symbols: | Plant thionin family protein | chr1:12776578-12776856 REVERSE LENGTH=92

MGIQTCSVLIIVAILTMMFSAHIAQSNSITMCVKHCAQNECLKAAKK**PTP**EICDEACKKICNNQ
LFGDEKWF**VPAP**KGSSRICRW**AP**KYCQF

>AT1G34815.1 | Symbols: | Plant thionin family protein | chr1:12783406-12783684 REVERSE LENGTH=92

MGIQTCSVLIIVIILTMMFSAHIAQSNSITMCVKHCAQNECLKAAKKT**TP**EICDEACKKICNNQ
LFSDEKWF**VPSP**KGSSRICRW**AP**KYCQF

>AT1G34930.1 | Symbols: | Plant thionin family protein | chr1:12807121-12807399 REVERSE LENGTH=92

MGIQTNSVLIIVSILTMMFSAHIAQPNGITMCVKHCAQNECLKAAKK**PTP**EICDEACKKICNNQ
LFGDEKWF**VPAP**KGSSRICRW**AP**KYCQL

>AT1G34860.1 | Symbols: | Plant thionin family protein | chr1:12803708-12803986 REVERSE LENGTH=92

MGIQTRSVLIIVAILTMMFSAHIAQSNSITMCVKHCAQNECLKAAKK**PTP**KICDEACKKICNNQ
LFGDEKWF**VPAP**KGSSRICRW**AP**KYCQF

>AT1G77093.1 | Symbols: | Defensin-like (DEFL) family protein | chr1:28962508-28963149 FORWARD LENGTH=78

MASTKYLVLLFICLSVLL**TP**GLGTD**PVPTP**PGLHIPCGKGFTSKECNKYCTGVGYRRGYC**AP**DE
EYPQISSCYCKWRI

>AT1G51920.1 | Symbols: | unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: stem,

stamen; EXPRESSED DURING: 4 anthesis; Has 22 Blast hits to 22
proteins in 5 species: Archae - 0; Bacteria - 0; Metazoa - 0;
Fungi - 0; Plants - 22; Viruses - 0; Other Eukaryotes - 0
(source: NCBI BLink). | chr1:19294229-19294579 FORWARD LENGTH=78
MASFHSGKSIFLKLFVLFLLLVLPLSQSNATRIPR**AP**ISSRRPIC**PA**CVCCE**PAP**LGSCCRCCA
SPIVTQTHHHSQ**SP**
>AT3G30725.1 | Symbols: AtGDU6, GDU6 | glutamine dumper 6 |
chr3:12352509-12352844 FORWARD LENGTH=111
MR**PTP**KVEIWK**SPVP**YLFGGLFLLVLLIALALLSLVCTHQK**PS**SSSNNNHMDEEDDVGDKDAKP
ITREYLPKIVVILAGDNK**PT**CLA**VPV**VVPP**PT**SIFRCNCDNVTVIST
>AT1G34795.1 | Symbols: | Plant thionin family protein |
chr1:12769753-12770031 REVERSE LENGTH=92
MGIQTCSVLIIVIILTMMFSAHIAQSNSITMCVKHCAQNECLKAAKKT**TP**EICDEACKKICNNQ
LFSDEKWF**VPSP**KGSSRICRW**AP**KYCQF
>AT1G34820.1 | Symbols: | Plant thionin family protein |
chr1:12786817-12787095 REVERSE LENGTH=92
MGIQTCSVLIIVAILTMMFSAHIAQSNSITMCVKHCAQNECLKAAKK**PTP**EICDEACKKICNNQ
LFGDEKWF**VPAP**KGSSRICRW**AP**KYCQF
>AT1G35035.1 | Symbols: | Plant thionin family protein |
chr1:12811030-12811308 REVERSE LENGTH=92
MAIQTRSVLMIVAILTMMFSAHLAQSNSITMCVKSCAQNKCLKAAKK**PTP**KICDETCKKICNNQ
LFGDEKMF**VPAP**KGSSRICRWMPQYCQF
>AT2G28671.1 | Symbols: | unknown protein; FUNCTIONS IN:
molecular_function unknown; INVOLVED IN: biological_process
unknown; LOCATED IN: cellular_component unknown; Has 30201 Blast
hits to 17322 proteins in 780 species: Archae - 12; Bacteria -
1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0;
Other Eukaryotes - 2996 (source: NCBI BLink). | chr2:12300816-
12301715 FORWARD LENGTH=299
MLFPLLLFWRMLAVVP**PSPT**RKGTLLLLTIPLLLLFDG**TP**LLTGRF**AP**LGLANGSCPLKAGFAT
TPVTALAVGFEP**PS**ISCIKNTNVLSG**PA**L**AP**PPRAEPE**PATP**PPRAG**PAPAP**PPPRAG**PAPAP**P
PPRAE**PAPAP**PPRAG**PAPAP**PPRAG**PAPAP**PPRAE**PAP**GPPPRAE**PVPAPAPV**AEPVVGKEPEP
P**PA**KEPGTELAVGRRPGPEPLDGID**PVPAP**RPLPVEPEPDDPD**PAP**KPNPVEPEPLD**PVPAP**KP
I**PV**EPEPEPEPRVVVVAENAFSSSRSRAAAAKVTTRAITSKESIF
>AT1G21880.2 | Symbols: LYM1 | lysm domain GPI-anchored protein
1 precursor | chr1:7680689-7682526 FORWARD LENGTH=416
MKIPEKPIFLIFVSLILASSLTFTATAKSTIEPCSSNDTCNALLGYTLYTDLKVSEVASLFQVD
PISILLANAIDISYPDVENHIL**PS**KLFLKIPITCSCVDGIRKSVSTHYKTR**PS**DNLGSIADSVY
GGLVSAEQIQEANSVND**PS**LLDVGTSLVIPLPCACFNGTDNSL**PA**VYLSYVVKEIDTLVGIARR
YSTTITDLMNVNAMG**AP**DVSSGDILAVPLSACASKFPRYASDFGLIVPNGSYALAAGHCVQCSC
ALGSRNLYCE**PA**SLAVSCSSMQCRNSNLMLGNITVQQTSAGCNVTTCDYNGIANGTILTMLTRS
LQPRCPGPQQF**AP**LL**AP**PDTVPRDVMY**APAPSP**DFDGPGSIAS**SP**RSSMLPGGGILPGN**PA**NG**P**
AGSISTASASSVSYFFITFLISIASFSLALSS
>AT1G61900.1 | Symbols: | unknown protein; FUNCTIONS IN:
molecular_function unknown; INVOLVED IN: biological_process
unknown; LOCATED IN: anchored to plasma membrane, plasma
membrane, anchored to membrane; EXPRESSED IN: 23 plant
structures; EXPRESSED DURING: 13 growth stages; BEST Arabidopsis

thaliana protein match is: unknown protein (TAIR:AT2G30700.1); Has 65 Blast hits to 65 proteins in 12 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 65; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink). | chr1:22882508-22884722 REVERSE LENGTH=433
MTRRAEFEMGLFVILQSMFLISLCSSQKPEEFLPEI**SP**DT**SP**QPFLPFI**APSP**MVPYINSTMPK
LSGLCSLNFSASESLIQTTSHNCWTVF**AP**LLANVMCCPQLDATLTIILGKASKETGLLALNRTQ
SKHCLSDLEQILVGKGASGQLNKICSIHSSNLTSSSCPVINVDEFESTVDTAKLLLACEKIDPV
KECCEEACQNAILDAATNISLKASETLTDNSDRINDCKNVVNRWLATKLD**PS**RVKETLRGLANC
KINRVCPLVFPHMKHIGGNCSNELSNQTGCCRAMESYVSHLQKQTLITNLQALDCATSLGTKLQ
KLNITKNIFSVCHISLKDFSLQVGNQESGCLL**PS**L**PS**DAIFDKDTGISFTCDLNDNI**PAP**W**PS**S
SLSSASTCKKPVRI**PA**L**PA**AASSQPRLHDEGVTRLVIFVLSMLLVMLLS
>AT1G63550.1 | Symbols: | Receptor-like protein kinase-related family protein | chr1:23569786-23570890 FORWARD LENGTH=324
MARIIITLTIPLFYFFFFSLLSHQTMSQPDHIFTVCN**PT**NNFTQTSSYETNRDTLLASLRESSS
LGHYSNATEGL**SP**DTVHGMFLCRGDITTASCVDCVQTATTEIASNCTLNKRAVIYYDECMVRYS
NVSFSSELEI**VPS**ITIYSLRS**AP**N**PT**RFNQTLTEKFSELIFNVSSSSLVPYFVEDQERVTQSEG
SYDLDTMVQC**SP**DLDIFNCTVCLRVAFFRISTCCGL**PS**YAKIF**TP**KCLLRFQTSVLL**SP**PP**SPS**
APPPR**SP**PPKS**SPPS**SLPQ**TPSP**PLVF**TP**PQN**VP**N**PS**GSFSFNVLKGNVIFGRIVVTMTALVFA
LVDL
>AT1G77630.1 | Symbols: | Peptidoglycan-binding LysM domain-containing protein | chr1:29173726-29175387 FORWARD LENGTH=423
MKNPEKPLLLFLILASSLASMATAKSTIEPCSSKDTCNSLLGYTLYTDLKVTEVASLFQVDPVS
MLLSNSIDISYPDVENHVL**PA**KLFLKIPITCSCVDGIRKSLSTHYKTRTSDTLGSIADSVYGGL
V**SP**EQIQVANSETDLSVLDVGTKLVIPLPCACFNGTDESL**PA**LYLSYVVRGIDTMAGIAKRFST
SVTDLTNVNAMG**AP**DINPGDILAVPLLACSSNFPKYATDYGLIIPNGSYALTAGHCVQCSCVLG
SRSMYCE**PA**SISVSCSSMRCRNSNFMLGNITSQQSSSGCKLTTCSYNGFASGTILTTLSMSLQP
RCPGPQQL**AP**LI**AP**PDNVPKELMYL**PSPSPSPSP**EFDDIAGGGSSIAAV**PA**A**SP**GGATVSSSNS
IPGN**PA**NGPGGSISIASCPLSYYSFIALLIPIGSCFFVF
>AT3G07390.1 | Symbols: AIR12 | auxin-responsive family protein | chr3:2365452-2366273 FORWARD LENGTH=273
MSLCLKIPLIKHQT**TP**EQNSAMASSSSSLLILAVACFVSLI**SPA**ISQQACKSQNLNSAGPFDSC
EDLPVLNSYLHYTYNSSNSSLSVAFVA**TPS**QANGGWVAWAIN**PT**GTKMAGSQAFLAYRSGGGA**A**
PVVKTYNISSYSSLVEGKLAFDFWNLRAESLSGGRIAIFTTVK**VPA**GADSVNQVWQIGGNVTNG
RPGVHPFGPDNLGSHRVLSFTEDA**AP**GS**APSP**GS**APAP**GTSGST**TP**GTAAGGPGNAGSLTRNVN
FGVNLGILVLLGSIFIF
>AT3G16860.1 | Symbols: COBL8 | COBRA-like protein 8 precursor | chr3:5759643-5762104 REVERSE LENGTH=653
MGLTRNFILWILLSSLFTAIQLTSSQRN**TP**PPRSKDADLCNGVFVSYTYLTGTKIKPNDTKNQP
YRFESEITVLNNGRDELKSWQVFVKFAHREILVSATNAVLSDGSSL**PA**SVENGTIFAGF**PS**ADL
KTAIMTAGDVTQMEARVELVGTQFGV**APPS**VPLPKNITLVNDGWSCPK**PT**QQGSNVLQVCC**TP**N
PNITTSKIGQKFLPRQEGDLTIMYDVTRAYQSSYSAQVTIENHNLLGRLDNWDLSFMWMKDEFL
FSTKGAY**PS**VVDSSDCITGPQAKYYKDLDFSNVMSCARRPHIIDLPLTKYNDTNVGRIPYCCRN
GTILPRSMDPEKSKSVFQIEVYKMPPDLNISSI**TP**PQSWQIKGNLNPDYKCGPPLRVSSSQFPD
TSGL**PS**NKSAFASWQVVCNITQ**PT**PPKCCVSFSSYFNDSVIPCKTCACGGCSSDRVARTCSTT**S**
PALPLPYQALLIPFDNRTKLTNAWAVLKNRK**VP**DPLPCGDNCGVSINWHLATDYRGGWTARVTL
FNWGDTDFVDWFTAVELRNA**AP**GFQKAYSFNGSIIAVNGKNTTVLMEGLPGLNYLLAEKDGKN**P**
SEDFRIPGKQQSVISFTKKL**TP**GIKVGSKDGF**PT**KVLFNGQECSL**PS**VL**PT**SNSHRKHVSTFLL

IL**TP**FLALLFLRI
>AT3G20520.1 | Symbols: SVL3 | SHV3-like 3 | chr3:7162845-7165742 FORWARD LENGTH=729
MACPRVIFLILITFFILQTAFSSSWQTLSGKP**PA**VIARGGFSGMFPDSSIQAYQLVNITT**SP**DV
MLWCDLQLTKDGVGICFPNLKLDNGSNVIRIDPHYKERFSVDFTWKELSDVKLAQGVVSRPYIF
DDVSSILAIEEVAKLTASGLWLNIQDSAFYAKHNLSMRNSVVSLSRRLKVNFIS**SP**GISFLKSM
KNSVK**PT**VTKLIFRFLKQEHIEPFTNQSYGSLAKNLSYIRTFSSGIL**VP**KSYIW**PV**DSALYLQP
HTSLVTDAHKEGLQVFASEFANDFVIAYNYSYD**PT**AEYLSFIDNGNFSVDGFLSDF**PVTP**YRAI
NCFSHVDPKRAKEQAKITIISKNGASGDFPGCTDLAYQRAASDGADILDCNVQMSKDKIPFCMS
SFDLINSTNVIETSFRNLSSVVSEINPRRSGIYTFSLTMSQIQTLK**PT**ISNLEKDSGLFRNPRN
NKAGKFLTLSEFLFLPNRYSSLLGLLIEVENAAYLVEHQGISVVDAVLDELKRATTQQNKTSAR
TILIQSTDKSVLMKFKEKNKMNHDELVYRVDDNIRDVADSAIKDIKNFAGSIVISKKSVFPYKG
FIILEKETNIASKLKSNGLRVYVERFSNECVTHAFDFYDD**PT**LEIDSFVRDVQIDGIITDF**PA**T
TARYRKNKCYGEFGLTTTGELITFANPMLLP**PA**E**AP**Y**PA**LLDSDVTEPPLPEARSQP**PA**S**SPS**K
AEEKAIEVPFAFIAMAILVCFFISV
>AT3G51330.1 | Symbols: | Eukaryotic aspartyl protease family protein | chr3:19053480-19056152 REVERSE LENGTH=529
MVVARQVFVLLSLLVVCWGLERCEASGKFSFEVHHMFSDRVKQSLGLDDLVPEKGSLEYFKVLA
QRDRLIRGRGLASNNEE**TP**ITFMRGNRTISIDLLGFLHYANVSVG**TPA**TWFLVALDTGSDLFWL
PCNCGSTCIRDLKEVGLSQSRPLNLY**SP**NTSSTSSSIRCSDDRCFGSSRCS**SPA**SSCPYQIQYL
SKDTFTTGTLFEDVLHLVTEDEGLEPVKANITLGCGKNQTGFLQSSAAVNGLLGLGLKDYSV**PS**
ILAKAKITANSFSMCFGNIIDVVGRISFGDKGYTDQME**TP**LL**PT**E**PSPT**YAVSVTEVSVGGDAV
GVQLLALFDTGTSFTHLLEPEYGLITKAFDDHVTDKRRPIDPELPFEFCYDL**SP**NKTTILFPRV
AMTFEGGSQMFLRNPLFIVWNEDNSAMYCLGILKSVDFKINIIGQNFMSGYRIVFDRERMILGW
KRSDCFEDESLEST**TP**PPPETE**APSPS**AS**TP**L**PS**LLPP**PA**AA**TP**PQIDPRNSTRNSGTGTAANL
VPLASQLLLLLPLLAFL
>AT3G51350.1 | Symbols: | Eukaryotic aspartyl protease family protein | chr3:19060485-19063248 REVERSE LENGTH=528
MDVARQVFVLLSVLVVCWGFERCEATGKFGFEVHHIFSDSVKQSLGLGDLVPEQGSLEYFKVLA
HRDRLIRGRGLASNNDE**TP**ITFDGGNLTVSVKLLGSLYYANVSVG**TPPS**SFLVALDTGSDLFWL
PCNCGTTCIRDLEDIGVPQSVPLNLY**TP**NASTTSSSIRCSDKRCFGSKKCS**SPS**SICPYQISYS
NSTGTKGTLLQDVLHLATEDENL**TP**VKANVTLGCGQKQTGLFQRNNSVNGVLGLGIKGYS**VPS**L
LAKANITANSFSMCFGRVIGNVGRISFGDRGYTDQEE**TP**FISV**APS**TAYGVNISGVSVAGDPVD
IRLFAKFDTGSSFTHLRE**PA**YGVLTKSFDELVEDRRRPVDPELPFEFCYDL**SP**NATTIQFPLVE
MTFIGGSKIILNNPFFTARTQEGNVMYCLGVLKSVGLKINVIGQNFVAGYRIVFDRERMILGWK
QSLCFEDESLEST**TP**PPPEVE**APAPS**VS**AP**PPRSLP**PT**VSA**TP**PPINPRNSTGNPGTGGAANLI
PLASQLLLLLPLLAFL
>AT4G16120.1 | Symbols: ATSEB1, COBL7, SEB1 | COBRA-like protein-7 precursor | chr4:9116591-9119138 REVERSE LENGTH=661
MDS**AP**NFIPRLLLLSLLIVSIPLTSSQSDANTTN**PSPSPPS**DSDLCNGVFVSYTHTKGSKIPPN
DTANQPYRFESVITVLNHGRDELKSWRVFVKFAHREILVSASNAVLSDGSSLPVSVENGTVFAG
Y**PS**SDLKSAIQTAGDVTQMQARVELVGTQFGV**AP**PN**VP**LPKNITLATDGWKCPKATQKGTNVLQ
VCCIPDPDYDNREIIDNEFLPRKDGDLTIMYDVVRSYSSNYMAQVTMENHNPLGRLDNWKLSFD
WMRDEFIYTMKGAY**PS**IVDSSDCVDGPQAKHYQDLDFSNVLSCARR**PT**VIDLP**PT**KYNDSTFGL
IPFCCRNGTILPRSMD**PS**KSSSVFQMQVYKMPPDLNISAL**SP**PQNWRINGTLNPDYKCGPPVRV
SPSQFVD**PS**GL**PS**NRTAFASWQVVCNITQPKDA**SP**RCCVSFSAYFNDSIVPCKTCACGCSSNKA
ARACSAT**APS**LLLPQQALLVPFENRTELTVAWAYLKQRPVPNPMPCGDNCGVSINWHLATDYRG
GWTARVTVFNWGETDFVDWFTAVQMKNA**AP**GFEKAYSFNASTIGINGKNNTIFMEGLPGLNYLV

AERDGENPLKNPRIPGKQQSVMSFTKKL**TP**GINVPGGDGF**PS**KVFFNGEECSL**PT**ILPMRSSQH
RKHISVFLLALPVLALLILRA

>AT4G25240.1 | Symbols: SKS1 | SKU5 similar 1 | chr4:12930539-12933563 FORWARD LENGTH=589

MAATCSLLASFLLCFALLSAVSFAADPFVSYDFRVSYLTA**SP**LGVPQQVIAVNGQFPGPLLNAT
TNYNVVVNVFNHLDEPLLLTWPGIQMRRNSWQDGVLGTNCPIPPRWNFTYQFQVKDQIGSFFY**S**
PSLNFQRASGGFGPIVINNRDIIPIPFPQPDGELIFIIGDWYTQDHKALRRALDSGKELGMPDG
VLINGKGPYKYNSSVPDGIDYLTFHVEPGKTYRIRVHNVGISTSLNFRIQNHSLLLVETEGHYT
SQANFTDFDVHVGQSYSFLVTMDQDATSDYYIVASARFVNETVWQRVTGVAILHYSNSKGPVSG
PLPVPKTDVS**SP**WSAMSQPKTIRQNTSASGARPNPQGSFHYGQINITNTYILRSLP**PT**IINGAL
RATLNGISFVN**PSTPV**RLADRNKVKGAYKLDFPDRPFNRPLRLDRSMINATYKGFIQVVFQNND
TKIQSFHVDGYSFFVVGMDFGIWSEDKKGSYNNWDAISRSTIEVYPGGWTAVLISLDNVGVWNI
RVENLDRWYLGEETYMRITNPEEDGKTEMDPPDNVLYCGALKNLQKEQHHSAATSILNGHLKLM
LLMVLLASVFRFC

>AT4G26690.1 | Symbols: SHV3, MRH5, GPDL2 | PLC-like phosphodiesterase family protein | chr4:13456793-13459890 REVERSE LENGTH=759

MRGLLRASSLLLCGVILIQLLAAQIHAQSKKPK**SP**W**PT**LTGDPPLVIARGGFSGLFPDSSYDAY
NFAILTSVPDAVLWCDVQLTKDALGICFPDLTMRNSSSIEAVY**PT**RQKSY**PV**NGV**PT**SGWFTID
FSLKDLKDVNLIRGILSRSEKFDGNSNPIMTVQSVSTQMK**PS**FFWLNVQHDAFYAQHNLSMSSF
LVAASKTVLIDFIS**SP**EVNFFKKIAGRFGRNG**PS**LVFRFLGQDEFE**PT**TNRTYGSILSNLTFVK
TFASGILVPKSYILPLDDQQYLLPHTSLVQDAHKAGLEVFVSGFANDIDIAHDYSFDPVSEYLS
FVDNGNFSVDGVLSDFPITASASLDCFSHVGRNATKQVDFLVITKDGASGDYPGCTDLAYKKAI
KDGADVIDCSVQLSSDG**TP**FCLSSIDLGNSTTVSLTAFRNRSTTVPELGSLGAIYTFSLTWAEI
QTL**TPA**ISNPYRVTSLFRNPKQKNAGKLFSLSDFLSLAKNSTSLSGVLISVENAAYLREEQGLD
VVKAVLDTLTQTGYSNSTATKVMIQSTNSSVLVDFKKQSQYETVYKVEENIRDILDSAIEDIKK
FADAVVIQKLSVFPVAQSFITTQTNVVEKLQKSQLPVYVELFQNEFLSQPYDFFADATVEINSY
ITGAGINGTITEFPFTAARYKRNLCLGRKETIPYM**APA**QPGALLTLV**SPT**AFP**PA**E**AP**N**PV**FTD
ADVTEPPLPPVTAK**APT**S**SP**G**TPS**TNAQ**APS**GQTRITLSLLLSVFAMVLASLLLL

>AT4G28100.1 | Symbols: | unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: anchored to plasma membrane, anchored to membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 14 growth stages; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT3G18050.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink). | chr4:13965300-13966697 REVERSE LENGTH=304

MKKSLTLLILLLCSLLFSTVLSNLLVE**PV**QPNT**VPA**F**PV**ETQAQSCRLDLSNELFGGVNEACGR
NLDRSRCC**PV**LAAWLFAAHARSALQL**PAPAPTP**ESSDPDEPMKPDDSQKCVNTLQSALLTKQIK
IPQPNSSCDAILCFCGIRLHQISSLSC**PA**AFNVSSGFKNA**TPT**AAVKNLEKECRNSSYSGCTRC
LGALQKLKVRGGNKKTTTERGTKMMSKDCQLMGLTWLLARNKTAYI**PT**VSAVLRAIMY**SP**HPPH
LNKC**SP**DQENMPLAVDSLQFQKSFSSSSHLFGVLPFLPLVLCIFLFLL

>AT4G31840.1 | Symbols: ENODL15, AtENODL15 | early nodulin-like protein 15 | chr4:15401798-15402426 FORWARD LENGTH=177

MASSSLLVTIFLCISVFFFSSVNANEVTVGGKSGDWKIP**PS**SSFSFNEWAQKARFKVGDFIVFK
YEAGKDSVLQVTREAYEKCNTT**SP**KASYTDGNTKVKLDQAGPVYFVSGTEGHCQKGQKLRLVVI

TPRNSAF**SP**G**PSPS**EFDG**PA**V**APT**SGAAKLAGGFSVVFGLVLGLWAFFF
>AT5G25090.1 | Symbols: ENODL13, AtENODL13 | early nodulin-like protein 13 | chr5:8647117-8647755 REVERSE LENGTH=186
MAQRTLVATFFLIFFLLTNLVCSKEIIVGGKTSSWKI**PSSPS**ESLNKWAESLRFRVGDTLVWKY
DEEKDSVLQVTKDAYINCNTTN**PA**ANYSNGDTKVKLERSGPYFFISGSKSNCVEGEKLHIVVMS
SRGGHTGGFFTGS**SPSPAPSPA**LLG**APTVAPA**SGGSASSLTRQVGVLGFVGLLAIVLL
>AT5G42720.1 | Symbols: | Glycosyl hydrolase family 17 protein | chr5:17130535-17132763 FORWARD LENGTH=438
MRASVYSLILLFFSCLLHLSKSQPFLGVNYGLTADNLPP**PS**ASAKLLQSTTFQKVRLYGSD**PA**V
IKALANTGIEIVIGASNGDVPGLASD**PS**FARSWVETNVVPYY**PA**SKIVLIAVGNEITSFGDNSL
MSQLL**PA**MKNVQTALEAASLGGGKIKVSTVHIMSVLAGSDP**PS**TAVFKPEHADILKGLLEFNSE
TG**SP**FAVNPYPFFAYQDDRRPETLAYCLFQANPGRVDPNSNLKYMNMFDAQVDAVYSALNSMGF
KDVEIMVAETGWPYKGDPEEAGATVENARAYNKNLIAHLKSGSG**TP**LMPGRVIDTYLFALYDEN
LKPGKGSERAFGLFRPDLTMTYDIGLTKTTNYNQTSM**AP**L**SPT**RPRLP**PA**A**APT**RQTL**PSP**PQM
IL**PSPVTPS**DKNSGQTDVHNS**TP**RSASLAHICRSLSISASMFFVSVLYALIIILL
>AT5G49270.1 | Symbols: SHV2, DER9, MRH4, COBL9 | COBRA-like extracellular glycosyl-phosphatidyl inositol-anchored protein family | chr5:19972021-19974398 REVERSE LENGTH=663
MGVLLPIFFGVLLLFTV**TPPS**MSQLP**PT**IM**VPAPAPAP**I**SPS**DLCNGIFLSYDFILGRKIPPND
TADQPYRFESVLTVLNNGREELKEWRVFVGFQHNEILISATDALIVNGTEL**PA**LVGNGTIFGGY
PVSDLKTAIQTAGDLKQMTAEIELVGTQFMV**APPAVP**L**PS**NISLVNEGWLC**PVPT**LQSKRELTT
CCIRDASIIVNTTITTKFLPRQPGDLTIMYDVIRAYDQNYLTEVTMENHNPLGRLDHWELSFDW
MRDEFIQKMQGAY**PT**VVDATKCIFGPQSLIYTGLDFADVLTCERRPIIIDLP**PT**KKDDSTLGNI
PSCCRNGTILPRIMD**PS**KSVSVFTMQVAKMPPDFNRSALFPPQNWRIKGTLNPDYSCGPPVRV**T**
PTFYPD**PS**GM**PT**NKSSFASWQIVCNITQAKTEIPKCCVSFSAFFNDSIIPCNTCACGCVSETRR
TCSAE**TPS**LLIPPDALLLPFENRTALTLAWNALKHKTLPNPMPCGDNCGVSINWHMASDYRGGW
TVRITIFNWGEIDFPNWFLAVQMKK**PA**LLGFEKAYSFNASLLSVDGGVNNTIFMEGLPGLDYLV
AEADEKDPKKKNIRIPGKQQSVIQFSKKL**TP**GINVAERDGF**PA**KVIFNGEECLLPDLLPMASGG
RRNGAITVLSFITFYVAAFMVLL
>AT5G55480.1 | Symbols: SVL1 | SHV3-like 1 | chr5:22474277-22477819 FORWARD LENGTH=766
MINMRDN**PT**MHVLQASKFLFLALILIQLLSTQLFAQRSK**SP**WQTLTGD**AP**LVIARGGFSGLLPD
SSLDAYSFVSQTSVPGAVLWCDVQLTKDAIGLCFPDVKMMNASNIQDVYPKRKTSYLLNGV**PT**Q
DWFTIDFNFKDLTKVILKQGILSRSAAFDGNSYGISTVKDISTQLKPEGFWLNVQHDAFYAQHN
LSMSSFLLSISKTVIIDYLS**SP**EVNFFRNIGRRFGRNGPKFVFRFLEKDDVEVSTNQTYGSLAG
NLTFLKTFASGVLVPKSYIWPIESQYLLPRTSFVQDAHKAGLEVYASGFGNDFDLAYNYSFDPL
AEYLSFMDNGDFSVDGLLSDFPLTASSAVDCFSHLGSNASSQVDFLVISKNGASGDYPGCTDLA
YTKAIKDGADVIDCSLQMSSDGIPFCLSSINLGESTNVVQ**SP**FRNRSTTVPEIGSLPGIYSFSL
AWSEIQTLR**PA**IENPYSREFTMFRNPRERSSGKFVSLSDFLNLAKNSSSLTGVLISVENATYLR
EKQGLDAVKAVLDTLTEAGYSNKTTTTRVMIQSTNSSVLIDFKKQSRYETVYKVEETIRDILDT
AIEDIKKFADAVVISKKSVF**PT**SESFTTGQTKLVERLQKFQLPVYVEVFRNEFVSQPWDFFADA
TVEINSHVTGAGINGTITEFPLTAARYKRNSCLTRKDVPPYMI**PV**Q**PA**GLLTIV**SPA**SLP**PA**E**A**
PSPVFTDADVTEPPLPPVSAR**APT**T**TP**GPQSTGEK**SP**NGQTRVALSLLLSAFATVFASLLLL
>AT2G20300.1 | Symbols: ALE2 | Protein kinase superfamily protein | chr2:8756475-8759845 REVERSE LENGTH=744
MRNFAMLLLLILLLHSLASFPICFARLFPMSLPFTRSKAHQMHFFHPYLN**PS**V**APTPSPA**F**SP**N
PSRIPPLRHKGHHRHRRWHLRRNATAV**SPS**SHDCQQTCVEPLTS**TP**FG**SP**CGCVFPMKVQLLLS
V**AP**FSIF**PV**TNELEIEVAAGTYLEQSQVKIMGASADSENQGKTVVDINLVPLGEKFDNTTATLI

YQRFRHKKVPLNETVFGDYEVTHISYPGI**PS**S**SP**NGDVTGD**AP**GGLPIPINATTFANKSQGIGF
RTIAIIALSGFVLILVLVGAISIIVKWKKIGKSSNAVG**PA**L**APS**INKRPGAGSMFSSSARSSGS
DSLMSSMATCALSVKTFTLSELEKATDRFSAKRVLGEGGFGRVYQGSMEDGTEVAVKLLTRDNQ
NRDREFIAEVEMLSRLHHRNLVKLIGICIEGRTRCLIYELVHNGSVESHLHEGTLDWDARLKIA
LGAARGLAYLHEDSNPRVIHRDFKASNVLLEDDF**TP**KVSDFGLAREATEGSQHISTRVMGTFGY
V**AP**EYAMTGHLLVKSDVYSYGVVLLELLTGRRPVDMSQ**PS**GEENLVTWARPLLANREGLEQLVD
PALAGTYNFDDMAKVAAIASMCVHQEVSHRPFMGEVVQALKLIYNDADETCGDYCSQKDSSVPD
SADFKGDL**APS**DSSWWNL**TP**RLRYGQASSFITMDYSSGPLEDMENRPHSASSIPRVGGLILPNR
SGPLRPMRSRRNFFRLRGSMSEHGG**PS**SSRHLWSGNGDWL
>AT2G23300.1 | Symbols: | Leucine-rich repeat protein kinase family protein | chr2:9914608-9917130 FORWARD LENGTH=773
MKTQSA**SP**FFFVSFFFFFFFFSSLFLLSSALNSDGVLLLSFKYSVLLDPLSLLQSWNYDHDNPC
SWRGVLCNNDSRVVTLSLPNSNLVGSI**PS**DLGFLQNLQSLNLSNNSLNGSLPVEFFADKLRFL
DLSNNLISGEIPVSIGGLHNLQTLNLSDNIFTGKL**PA**NLASLGSLTEVSLKNNYFSGEFPGGGW
RSVQYLDISSNLINGSLPPDFSGDNLRYLNVSYNQISGEIPPNVGAGFPQNATVDFSFNNLTGS
IPD**SP**VYLNQKSISFSGNPGLCGG**PT**RNPCPI**PSSPA**TV**SPPT**S**TPA**LAAIPKSIGSNRETEPN
NNSNPRTGLRPGVIIGIIVGDIAGIGILALIFFYVYKYKNNKTVEKKNNHSLEAHEAKDTTSL**S**
PSSSTTTSSS**SP**EQSSRFAKWSCLRKNQETDETEEEDEENQRSGEIGENKKGTLVTIDGGEKEL
EVETLLKASAYILGATGSSIMYKTVLEDGTVLAVRRLGENGLSQQRRFKDFEAHIRAIGKLVHP
NLVRLRGFYWGTDEKLVIYDFVPNGSLVNARYRKGGS**SP**CHLPWETRLKIVKGLARGLAYLHDK
KHVHGNLK**PS**NILLGQDMEPKIGDFGLERLLAGDTSYNRASGSSRIFSSKRLTASSREFGTIG**P**
TPSPSPSSVGPI**SP**YC**AP**ESLRNLKPNPKWDVFGFGVILLELLTGKIVSIDEVGVGNGLTVEDG
NRALIMADVAIRSELEGKEDFLLGLFKLGYSCASQIPQKR**PT**MKEALVVFERYPISSSAK**SPS**Y
HYG
>AT5G56890.1 | Symbols: | Protein kinase superfamily protein | chr5:23010801-23015559 REVERSE LENGTH=1113
MENLTLLLRICLVSSVLVAASSSGSELL**SP**LS**SPPSP**LPETSKGFGQ**APS**N**SP**ESHKSDNVP**PS**
KASSQ**PS**LPPLADLA**AP**P**PS**DSVGGK**APA**G**VPV**VFVPN**APAPA**TI**PV**KDL**PV**A**SP**PVLQPI**TP**I
A**SP**PRFIPGD**AP**KEPPFSGRV**TPAP**VS**SP**VSDIPPI**PS**VALPP**PTPS**NVPPRNASNNHKPPPIE
KSI**AP**VA**SPPT**ISIDI**AP**PVH**PV**IPKL**TPS**S**SP**V**PT**S**TPT**KG**SP**RRNP**PT**TH**PV**FPIE**SPA**V**SP**
DHAAN**PV**KHPP**PS**DNGDDSK**SP**GA**APA**NETAKPL**PV**FPHKA**SPPS**I**APSAP**KFNRHSHHT**SPS**T
TPPPDS**TPS**NVHHH**PSSPSP**PPLSSHHQHHQERKKIAD**SPAPSP**LPPHLI**SP**KKSNRKGSM**TP**P
PQSHH**APS**PPIPDSLI**SPA**H**AP**VSFSMKRI**SPA**L**APSPT**QVFPLRSSSR**PS**KSRKFPLGPPLQA
FPPPPPNSDCSSTICLEPYTN**TP**PG**SP**CGCVWPIQVELRLSMALYDFFPMVSEFAREISAGVFM
KQSQVRIMGANAASEQPEKSIVLIDLVPLGDKFDNMTAMLTYQRFWSKKVYIDEPIFGGYDVIY
VRYPGL**PASPPT**SGMTIIDQGPYSGNNNGRAVKPLGVDVPRKPRKKELNGGSIAVIVLSAAAFI
GLCFVIVWFLVFRRQRDRRRLSKR**TP**LAR**PS**L**PS**LSK**PS**GSARSLTGSRFSSTSLSFESSI**AP**F
TLSAKTFTASEIMKATNNFDESRVLGEGGFGRVYEGVFDDGTKVAVKVLKRDDQQGSREFLAEV
EMLSRLHHRNLVNLIGICIEDRNRSLVYELIPNGSVESHLHGIDKAS**SP**LDWDARLKIALGAAR
GLAYLHEDS**SP**RVIHRDFKSSNILLENDF**TP**KVSDFGLARNALDDEDNRHISTRVMGTFGYV**AP**
EYAMTGHLLVKSDVYSYGVVLLELLTGRKPVDMSQPPGQENLVSWTRPFLTSAEGLAAIIDQSL
GPEISFDSIAKVAAIASMCVQPEVSHRPFMGEVVQALKLVSNECDEAKELNSLTSISKDDFRDD
TQAESSCGDSSARMARYPLLPNYDSEPDTERGLSTSEMYTGSGRFERQSNSGPLTSGRGKSFWQ
KMRRLSTGSLSEHG**TPT**VMLRSGSR
>AT5G60270.1 | Symbols: | Concanavalin A-like lectin protein kinase family protein | chr5:24257761-24259767 FORWARD LENGTH=668

MIRGLLLGIIWMIFCVCSSFQQE**TP**FVYNNFGHVDHLHLDGSARII**PS**GGILQLTNATNSQIGH
VFYEKPIEFKSSESVSFSTYFVCALL**PA**GD**PS**GHGMTFFVSHSTDFKGAEATRYFGIFNRNGST
STRVLAVELDTSLASDVKDISDNHVGIDVNSAESITSANASYFSDKEGKKIDIKLLSGDPIQVW
VDYEGTTLNVSL**AP**LRNKK**PS**RPLLSSTSINLTDILQGRRMFVGFSGSTGSSMSYQYILGWSFS
KSMASLPNIDISKLPKVPHSSTKKKST**SP**VLSVLLGLIAFIVLGILVVAYLYRRNLYSEVREEW
EKEYGPIRYSYKSLYKATKGFNRSEFLGRGGFGEVYKGTLPRSRELREVAVKRVSHDGEHGMKQ
FVAEIVSMRSLKHRSLVPLLGYCRRKHELLLVSEYMPNGSLDHYLFNHDRLSLPWWRRLAILRD
IASALSYLHTEADQVVIHRDIKAANVMLDAEFNGRLGDFGMSRLYDRGAD**PS**TTAAVGTVGYM**A**
PELTTMGASTGTDVYAFGVFLLEVTCGRR**PV**EPGLPEAKRFLIKWVSECWKRSSLIDARDPRLT
EFSSQEVEKVLKLGLLCANL**AP**DSR**PA**MEQVVQYLNGNLALPEFWPN**SP**GIGVL**SP**MAL**SPAP**L
VI**PS**LSFSSSSSNNSMFITHSVLYGSGR
>AT1G63600.1 | Symbols: | Receptor-like protein kinase-related
family protein | chr1:23581716-23582624 FORWARD LENGTH=302
MIQSLSSLFCFLSLLSLLFSTNLTISESDHIHMSTFCNQFSDNFTQTSTYETNRETVLSSLRLR
SSLGSYSNATAGI**SP**DTVRGMFLCRGDISETSCSDCVQTATLEISRNCTYQKEAFIFYEECMVR
YSDSSFFSLVDERPYIIRYSLSY**AP**NLDRFPQTLSDKMDELIINATS**SPS**LSS**TP**YFVEDQERV
KQFEGSFDIDSMAQC**SP**DLDPRNCTTCLKLAVQEMLECCNQSRWAQIF**TP**KCLLRYEATALS**SP**
PPPY**PSP**PP**PSSP**LF**SP**LL**PSP**PLFKRPQTASGFSGSSSINVIKGN
>AT1G10850.1 | Symbols: | Leucine-rich repeat protein kinase
family protein | chr1:3612228-3614343 FORWARD LENGTH=663
MASSSSSSSSSTVSFVVFAFTVFFCLV**TPA**RSSDVEALLSLKSSID**PS**NSISWRGTDLCNWQGV
RECMNGRVSKLVLEYLNLTGSLNEKSLNQLDQLRVLSFKANSLSGSIPNLSGLVNLKSVYLNDN
NFSGDFPESLTSLHRLKTIFLSGNRLSGRI**PS**SLLRLSRLYTLNVEDNLFTGSIPPLNQTSLRY
FNVSNNKLSGQIPLTRALKQFDESSFTGNVALCGDQIG**SP**CGI**SPAPS**AK**PTP**IPKSKKSKAKL
IGIIAGSVAGGVLVLILLLTLLIVCWRRKRRNQ**AP**REDRKGKGIAEAEGATTAETERDIERKDR
GFSWERGEEGAVGTLVFLGTSDSGETVVRYTMEDLLKASAETLGRGTLGSTYKAVMESGFIVTV
KRLKNARYPRMEEFKRHVEILGQLKHPNLVPLRAYFQAKEERLLVYDYFPNGSLFTLIHGTRAS
GSGKPLHWTSCLKIAEDLASALLYIHQNPGLTHGNLKSSNVLLGPDFESCLTDYGLSTLHDPDS
VEETSAVSLFYK**AP**ECRDPRKASTQ**PA**DVYSFGVLLLELLTGR**TP**FQDLVQEYGSDISRWVRAV
REEETESGEE**PT**SSGNEASEEKLQALLSIATVCVTIQPDNRPVMREVLKMVRDARAE**AP**FSSNS
SEH**SP**GRWSDTVQSLPRDDQVSI
>AT4G36180.1 | Symbols: | Leucine-rich receptor-like protein
kinase family protein | chr4:17120209-17123698 REVERSE
LENGTH=1136
MAMDISLFFIFLVIY**AP**LVSYADESQAEIDALTAFKLNLHDPLGALTSWD**PSTPAAP**CDWRGVG
CTNHRVTEIRLPRLQLSGRISDRISGLRMLRKLSLRSNSFNGTI**PT**SLAYCTRLLSVFLQYNSL
SGKLP**PA**MRNLTSLEVFNVAGNRLSGEI**PV**GL**PS**SLQFLDISSNTFSGQI**PS**GLANLTQLQLLN
LSYNQLTGEI**PA**SLGNLQSLQYLWLDFNLLQGTL**PS**AISNCSSLVHLSASENEIGGVI**PA**AYGA
LPKLEVLSLSNNNFSGT**VP**FSLFCNTSLTIVQLGFNAFSDIVRPETTANCRTGLQVLDLQENRI
SGRFPLWLTNILSLKNLDVSGNLFSGEIPPDIGNLKRLEELKLANNSLTGEIPVEIKQCGSLDV
LDFEGNSLKGQIPEFLGYMKALKVLSLGRNSFSGY**VPS**SMVNLQQLERLNLGENNLNGSFPVEL
MALTSLSELDLSGNRFSGAVPVSISNLSNLSFLNLSGNGFSGEI**PA**SVGNLFKLTALDLSKQNM
SGEVPVELSGLPNVQVIALQGNNFSGVVPEGFSSLVSLRYVNLSSNSFSGEIPQTFGFLRLLVS
LSLSDNHISGSIPPEIGNCSALEVLELRSNRLMGHI**PA**DLSRLPRLKVLDLGQNNLSGEIPPEI
SQSSSLNSLSLDHNHLSGVIPGSFSGLSNLTKMDLSVNNLTGEI**PA**SLALISSNLVYFNVSSNN
LKGEI**PA**SLGSRINNTSEFSGNTELCGKPLNRRCESSTAEGKKKKRKMILMIVMAAIGAFLLSL
FCCFYVYTLLKWRKKLKQQSTTGEKKR**SP**GRTSAGSRVRSSTSRSSTENGEPKLVMFNNKITLA
ETIEATRQFDEENVLSRTRYGLLFKANYNDGMVLSIRRLPNGSLLNENLFKKEAEVLGKVKHRN

ITVLRGYYAGPPDLRLLVYDYMPNGNLSTLLQEASHQDGHVLNWPMRHLIALGIARGLGFLHQS
NMVHGDIKPQNVLFDADFEAHISDFGLDRLTIR**SPS**RSAVTANTIGTLGYV**SP**EATLSGEITRE
SDIYSFGIVLLEILTGKRPVMFTQDEDIVKWVKKQLQRGQVTELLEPGLLELDPESSEWEEFLL
GIKVGLLCTATDPLDR**PT**MSDVVFMLEGCRVGPDV**PS**SAD**PT**SQ**PSPA**
>AT3G19300.1 | Symbols: | Protein kinase superfamily protein |
chr3:6690242-6693210 REVERSE LENGTH=663
MVLVYNQALLLAFFALLGFQYFPLLTEAGCPLDFTSSNFTLVASVCSNNTERAKCCRYMNAFVA
ISVARYANYTADLGVTSDLTEICITTISRTMELYGIPRNATIFCGLGTKILVNYDCEGLTT**VP**P
MLHSTSFGDVSRNCKLPLPPGDQCRNCLNSSISYLRSLVGADNSIKLSTCRDATYAVLASRVDN
SSALELASCFFN**VP**EL**SPTP**GS**SPS**SL**SP**EA**SP**PVLAD**SPS**SDNLVL**SP**LKSHHPYHLTM**VPT**I
GIAVTVFALVMVVVLIVLIQRKKRELDDSKSIACNLTRT**SPSP**RPRSMIHEGNSFGFRKFSYKE
IRKATEDFNAVIGRGGFGTVYKAEFSNGLVAAVKKMNKSSEQAEDEFCREIELLARLHHRHLVA
LKGFCNKKNERFLVYEYMENGSLKDHLHSTEK**SP**LSWESRMKIAIDVANALEYLHFYCDPPLCH
RDIKSSNILLDEHFVAKLADFGLAHASRDGSICFEPVNTDIRG**TP**GYVDPEYVVTHELTEKSDV
YSYGVVLLEIITGKRAVDEGRNLVELSQPLLVSESRRIDLVDPRIKDCIDGEQLETVVAVVRWC
TEKEGVAR**PS**IKQVLRLLYESCDPLHLGLAMAVEENKGRSLRGDSGFQSGDIRGLASSSSTTSR
SHCSRSFLLETG**SP**H**SP**PNGLSF
>AT5G03140.1 | Symbols: | Concanavalin A-like lectin protein
kinase family protein | chr5:737750-739885 REVERSE LENGTH=711
MLKLPPRFFSVYSTLIHILASFLCSSDVRGDF**PA**TRFDLGTLTLSSLKLLGDAHLNNGTIKLTR
ELS**VPT**STAGKALYGK**PV**KFRHPETK**SPA**SFTTYFSFSVTNLN**PS**SIGGGLAFVI**SP**DEDYLGS
TGGFLGLTEETGSGSGFVAVEFDTLMDVQFKDVNGNHVGLDLNAVVSAAVADLGNVDIDLKSGN
AVNSWITYDGSGRVLTVYVSYSNLKPK**SP**ILS**VP**LDLDRYVSDSMFVGFSGSTQGSTEIHSVDW
WSFSSSFEESSE**SP**PPMPN**SP**P**PSSPS**SSI**TPS**LSTVRRKTAD**PS**SSCRNKLCKK**SPA**AVAGVV
TAGAFFLALFAGVIIWVYSKKIKYTRKSESLASEIMK**SP**REFTYKELKLATDCFSSSRVIGNGA
FGTVYKGILQDSGEIIAIKRCSHISQGNTEFLSELSLIGTLRHRNLLRLQGYCREKGEILLIYD
LMPNGSLDKALYE**SPT**TLPWPHRRKILLGVASALAYLHQECENQIIHRDVKTSNIMLDANFNPK
LGDFGLARQTEHDK**SP**DATAAAGTMGYL**AP**EYLLTGRATEKTDVFSYGAVVLEVCTGRRPITRP
EPEPGLRPGLRSSLVDWVWGLYREGKLLTAVDERLSEFNPEEMSRVMMVGLACSQPDPVTR**PT**M
RSVVQILVGEADVPEVPIAK**PS**SSMSFSTSELLLTLQDSVSDCNEVL**AP**ISTTSCSSSEHDIFI
VGKDRSV
>AT5G58540.1 | Symbols: | Protein kinase superfamily protein |
chr5:23663400-23665182 FORWARD LENGTH=484
MSSKRRRRFNLRTVFSIIFLTFLPLNLNSQEIVEVFDSSQDHFLIQSRVYANHRSLID**TP**LPGK
D**PA**LDA**SPPSP**ESAILKDPLLPPPPPEGNE**TPSP**PRSG**VPT**Q**TP**E**TPPA**I**TP**LP**VP**L**APAPSPS**
PPV**SP**GTTKK**SP**KVYMIVGIVGGVFTVSVALIIIFLILTRKIPIKPWTNSGQLRDDLITDVPRL
QLSELQAACEDFSNVIGSFSDGTIYKGTLSTGAEIAVVSIVAGSRSDWSTTMDTQLLQKMHNLS
KVDHKNFLNVIGYCLEEEPFKRMLVFEY**AP**NGSLSEHLHSQYVEHLDW**PT**RLRIVMGIAYCLEH
MHNLNPPILLSNLDSSSVYLTEDNAAKVSDFSVINSIF**PS**KEGSSSKNLLE**PS**LLDPHTNVFNF
GAVLFEIISGKLPDPDSMLLEPK**PT**RDIVD**PT**LKTFQENVVERLLEVVRQCLNPYSDQR**PT**MRE
VVVKLREITGIEADAAMPRL**SP**RWWTELEIISTEGN
>AT5G11020.1 | Symbols: | Protein kinase superfamily protein |
chr5:3486439-3488983 REVERSE LENGTH=433
MKQIVITALVLLQAYVLHQSTCVMSLTTQE**SPSP**Q**PS**AF**TPA**L**SP**DYQQREKELHKQESNNMRL
VISLAATFSLVGIILLCSLLYWFCHRRRNLKSSGCGCSGITFLNRFSRSKTLDKRTTKQGTVSL
IDYNILEEGTSGFKESNILGQGGFGCVYSATLENNISAAVKKLDCANEDAAKEFKSEVEILSKL
QHPNIISLLGYSTNDTARFIVYELMPNVSLESHLHGSSQGSAITWPMRMKIALDVTRGLEYLHE
HCH**PA**IIHRDLKSSNILLDSNFNAKISDFGLAVVDGPKNKNHKLSGTVGYV**AP**EYLLNGQLTEK

SDVYAFGVVLLELLLGKKPVEKL**AP**GECQSIITWAMPYLTDRTKL**PS**VID**PA**IKDTMDLKHLYQ
VAAVAILCVQPE**PS**YRPLITDVLHSLIPLVPMELGGTLKTIKCASMDHC
>AT5G67280.1 | Symbols: RLK | receptor-like kinase |
chr5:26842430-26845126 REVERSE LENGTH=751
MMTTVAADLHRYLFLITVFLFFLCDKTSLALTTDGVLLLSFRYSIVDDPLYVFRSWRFDDE**TP**C
SWRGVTCDASSRHVTVLSL**PS**SNLTGTL**PS**NLGSLNSLQRLDLSNNSINGSF**PV**SLLNATELRF
LDLSDNHISGAL**PA**SFGALSNLQVLNLSDNSFVGELPNTLGWNRNLTEISLQKNYLSGGIPGGF
KSTEYLDLSSNLIKGSL**PS**HFRGNRLRYFNASYNRISGEI**PS**GFADEIPEDATVDLSFNQLTGQ
IPGFRVLDNQESNSFSGNPGLCGSDHAKHPCRDGEAT**SP**P**PSPTP**N**SPPA**LAAIPNTIGLTNHP
ISSKTGPKSKWDHK**PV**LIIGIVVGDLAGLAILGIVFFYIYQSRKRKTVTATSKWSTSSTDSKVS
KWYCLRKSVYVDGDCEEEEEESETSESESDEENPVGPNRRSGLDDQEKKGTLVNLDSEKELEIE
TLLKASAYILGATGSSIMYKAVLQDGTAVAVRRIAECGLDRFRDFEAQVRAVAKLIHPNLVRIR
GFYWGSDEKLVIYDFVPNGSLANARYRKVGS**SP**CHLPWDARLKIAKGIARGLTYVHDKKYVHGN
LK**PS**NILLGLDMEPKVADFGLEKLLIGDMSYRTGGS**AP**IFGSKRSTTSLEFG**PSPSPSPS**SVGL
PYN**AP**ESLRSIKPNSKWDVYSFGVILLELLTGKIVVVDELGQVNGLVIDDGERAIRMADSAIRA
ELEGKEEAVLACLKMGLACA**SP**IPQRRPNIKEALQVLERFPVHSSQQ
>AT2G20850.1 | Symbols: SRF1 | STRUBBELIG-receptor family 1 |
chr2:8975670-8979182 REVERSE LENGTH=775
MRSMRSGRDNNICFLGFLSFALISL**PS**LSLALTNPDDVAAINSLFLALE**SP**LLPGWVASGGDPC
GESWQGVLCNASQVETIILISANLGGELGVGLNMFTSLKAMDFSNNHIGGSI**PS**TL**PV**SLQNLF
LSGNNFTGTIPESLSSLKSLSVMSLNNNLLSGKIPDVFQDLGLMINIDLSSNNLSGPLP**PS**MQN
LSTLTSLLLQNNHLSGELDVLQDLPLKDLNVENNLFNGPIPEKLLSIPNFIKGGNLFNVTI**APS**
PSPE**TPPSPTSP**KRPFFGP**PSP**NASAGHGQAHVR**SPPS**DHH**PS**R**PTP**QGKEDSFTSKRIIWISI
LGAFSFVVLALVCLLCGRKCLRKREDSEQLSKPHLTSEYGRAREGSRSNASMLP**PS**NTFNKDKE
ARPKERVGGASKLHGGAERSVGSESKQESHEIDMNGNAMDLMH**PS**SIPPIKRVIAKATE**PA**EAS
LKRTTSKSHGPLTAVKHFTVASLQQHTNSFSHENLIGTGMLGSVYRAELPGGKLFAVRKLDKK**S**
PNHEEEGKFLELVNNIDRIRHANIVQLVGFCSEHSQRLLIHEYCRNGTLHDLLHIDDRLKIELS
WNVRVRIALEAAKALEYLHEICDP**PS**IHRNFKSANILLDDDIRVHVSDCGL**AP**LISSGAVSQLS
GQLLAAYGYG**AP**EFEYGIYTMKCDVYSFGVVMLELLTGRKSYDKKRDRGEQFLVRWAIPQLHDI
DALAKMVD**PS**LKGDY**PA**KSLSHFADVISRCVQSEPEYRPLMSEVVQDLSDMIQREHRRNDSNGD
NQYTGRR
>AT4G03390.1 | Symbols: SRF3 | STRUBBELIG-receptor family 3 |
chr4:1490912-1494553 REVERSE LENGTH=776
MAAKRSIYCLLLLPLLLSLLIWI**PS**ISLAATNPDDVAAINGLFAALG**AP**VLPGWIASGGDPCGE
AWQGIICNVSDIISITVNAANLQGELGDNLAKFTSIRGIDFSNNRIGGSI**PS**TLPVTLQHFFLS
ANQFTGSIPESLGTLSFLNDMSLNDNLLSGELPDVFQNLVGLINLDISSNNISGTLP**PS**MENLL
TLTTLRVQNNQLSGTLDVLQGLPLQDLNIENNLFSGPIPDKLLSIPKFLHEGNPFNATMINSTS
T**APS**L**SPS**L**SPT**K**PAPT**RPFSGVPPPPNERNRGKVADG**PS**DSEGSSSENSKGKNSSHTKKIILI
AFAGVLVFIILVLAILLLLPKCARRREHANRVFKPHQVGADRGSRENALENG**TP**VLPPPGRSEK
VQREPFKKAGEEPKVLHDLERLRR**PAP**ISRQESQDIDFSMLMPPPPPPPPPPPPPPLDEKVTVM
PII**SP**ERPVKKT**SP**KRLPLTSVKHYSIASLQQYTESFAQENLIGSGMLGSVYRARLPNGKLFAV
KKLDKRASEQQQDHEFIELVNNIDMIRHSNIVELVGYCAEHDQRLLVYEYCSNGTLQDGLHSDD
EFKKKLSWNTRVSMALGAARALEYLHEVCEPPIIHRNFKSANVLLDDDLSVLVSDCGL**AP**LISS
GSVSQLSGQLLAAYGYG**AP**EFDSGIYTWQSDVYSFGVVMLELLTGRMSYDRDRSRGEQFLVRWA
IPQLHDIDALGKMVD**PS**LNGQY**PA**KSLSHFADIISRCVQSEPEFRPLMSEVVQDLLDMIRRERH
GSGDSTAD
>AT4G18760.1 | Symbols: AtRLP51, RLP51 | receptor like protein
51 | chr4:10308163-10309458 REVERSE LENGTH=431

MKP**PS**LPLLLLLLHLSATISA**APS**L**SPTPSPT**T**SP**IPPHK**PS**SSS**SP**LDPKQLKALESLNI**PT**V
KDPCNHR**PT**TKSTSSSVVTCDTS**SP**FRLVTSISFTNCSTDLSISTTALRAL**SPS**LTSLSFLNC**P**
SL**SP**PPRLPDSLHSFTAVSSFLRRRNGLSGVFLARLVNLTDLTVSS**VPV**STSGLFVILGNMHEI
VSLTISHANLSGNIPKSFHSNLTFIDLSDNLLKGSI**PT**SITLLSNLKSLNLSKNTISGDIPDSI
GDLISLKNLSLSSNKLSGPIPDSISSIPELTHLDLSGNQLNGTIPRFISKMKYLTHLNLANNAF
HGVLPFNASFIKNLEVFKIGGNSDLCYNHSVLSSKMKLGIAQCDKHGLPL**SP**PPQKEDSNSDYD
YGNEDDTSEKKKEEHHGPNKVVLGVAIGLSSLVFLIIFMILLAKWCG
>AT2G25440.1 | Symbols: AtRLP20, RLP20 | receptor like protein
20 | chr2:10826735-10829402 FORWARD LENGTH=671
MKTMSKSSLRLHFLSLLLLCCV**SPS**SFVIIRFITHNHFDGLVRCHPHKFQALTQFKNEFDTRRC
NHSNYFNGIWCDNSKVRSQSYDYGTVSVELSNQTVASSSFIIFATLISLTTT**SP**PL**PS**L**PS**LF**P**
TLRNLTKLTVLDLSHNHFSGTLKPNNSLFELHHLRYLNLEVNNFSSSL**PS**EFGYLNNLQHCGLK
EFPNIFKTLKKMEAIDVSNNRINGKIPEWLWSLPLLHLVNILNNSFDGFEGSTEVLVNSSVRIL
LLESNNFEGAL**PS**LPHSINAFSAGHNNFTGEIPLSICTRTSLGVLDLNYNNLIGPVSQCLSNVT
FVNLRKNNLEGTIPETFIVGSSIRTLDVGYNRLTGKLPRSLLNCSSLEFLSVDNNRIKDTFPFW
LKALPKLQVLTLSSNKFYGPI**SP**PHQGPLGFPELRILEISDNKFTGSLSSRYFENWKASSAMMN
EYVGLYMVYEKNPYGVVVYTFLDRIDLKYKGLNMEQARVLTSYSAIDFSRNLLEGNIPESIGLL
KALIALNLSNNAFTGHIPQSLANLKELQSLDMSRNQLSGTIPNGLKQLSFLAYISVSHNQLKGE
IPQGTQITGQLKSSFEGNVGLCGLPLEERCFDNSA**SPT**QHHKQDEEEEEEQVLHWKAVAMGYGP
GLLVGFAIAYVIASYKPEWLTKIIGPNKRRN
>AT3G19020.1 | Symbols: | Leucine-rich repeat (LRR) family
protein | chr3:6559174-6562044 REVERSE LENGTH=956
MTRRTMEKPFGCFLLLFCFTISIFFYSAAALTDEEASFLTRRQLLALSENGDLPDDIEYEVDLD
LKFANNRLKRAYIALQAWKKAFYSDPFNTAANWVGPDVCSYKGVFC**APA**LDD**PS**VLVVAGIDLN
HADIAGYLPPELGLLTDVALFHVNSNRFCGVIPKSLSKLTLMYEFDVSNNRFVGPF**PT**VALSW**P**
SLKFLDIRYNDFEGKLPPEIFDKDLDAIFLNNNRFESTIPETIGKSTASVVTFAHNKFSGCIPK
TIGQMKNLNEIVFIGNNLSGCLPNEIGSLNNVTVFDASSNGFVGSL**PS**TLSGLANVEQMDFSYN
KFTGFVTDNICKLPKLSNFTFSYNFFNGEAQSC**VP**GSSQEKQFDDTSNCLQNRPNQKSAKECLP
VVSRPVDCSKDKCAGGGGGGSN**PSP**K**PTPTP**K**AP**EPKKEINPPNLEE**PS**KPKPEE**SP**KPQQ**PSP**
KPE**TPS**HE**PS**NPKEPKPE**SP**KQE**SP**KTEQPKPKPE**SP**KQE**SP**KQE**AP**KPEQPKPKPE**SP**KQESS
KQEPPKPEE**SP**KPEPPKPEE**SP**KPQPPKQE**TP**KPEE**SP**KPQPPKQE**TP**KPEE**SP**KPQPPKQE**TP**
KPEE**SP**KPQPPKQEQPPKTE**AP**KMG**SP**PLE**SPVP**NDPYDA**SP**IKKRRPQP**PSPS**TEETKTT**SP**Q
SPPVH**SP**PPPPPVH**SP**PPPVF**SP**PPPMH**SP**PPPVY**SP**PPPVH**SP**PPPPVH**SP**PPPVH**SP**PPPVH
SPPPPVH**SP**PPPVH**SP**PPPVH**SP**PPPVQ**SP**PPPPVF**SP**PP**PAP**IY**SP**PPPPVH**SP**PPPVH**SP**PP
PPVH**SP**PPPVH**SP**PPPVH**SP**PPPVH**SP**PPPVH**SP**PP**PSP**IY**SP**PPPVF**SP**PPK**PVTP**LP**PATSP**
MAN**APTPS**SSESGEIS**TP**VQ**APTP**DSEDIE**APS**DSNH**SPV**FKS**SPAPSP**DSEPEVE**APVPS**SEP
EVE**AP**KQSEA**TPS**S**SPPS**SN**PSP**DVT**APPS**EDNDDGDNFILPPNIGHQYA**SP**PPPMFPGY
>AT1G12040.1 | Symbols: LRX1 | leucine-rich repeat/extensin 1 |
chr1:4070160-4072394 FORWARD LENGTH=744
MLFPPLRSLFLFTLLLSSVCFLQIKADHDDESDLGSDIKVDKRLKFENPKLRQAYIALQSWKKA
IFSDPFNFTANWNGSDVCSYNGIYC**APSPS**YPKTRVVAGIDLNHADMAGYLASELGLLSDLALF
HINSNRFCGE**VP**LTFNRMKLLYELDLSNNRFVGKFPKVVLSL**PS**LKFLDLRYNEFEGKI**PS**KLF
DRELDAIFLNHNRFRFGIPKNMGN**SPV**SALVLADNNLGGCIPGSIGQMGKTLNELILSNDNLTG
CLPPQIGNLKKVTVFDITSNRLQGPL**PS**SVGNMKSLEELHVANNAFTGVIP**PS**ICQLSNLENFT
YSSNYFSGRPPICAASLLADIVVNGTMNCITGLARQRSDKQCSSLLARPVDCSKFGCYNIF**SP**P
P**PT**FKM**SP**EVRTLPPPIYVYS**SP**PPP**PS**SKM**SPT**VRAY**SP**PPP**PS**SKM**SPS**VRAY**SP**PPPPYSK
M**SPS**VRAYPPPPP**PSPSP**PPPYVYS**SP**PPPYVYS**SP**PPPPYVYS**SP**PPPPYVYS**SP**PPPYVYS**S**
PPPPYVYS**SP**PPPP**PSP**PPPCPES**SP**PPPVVYY**APV**TQ**SP**PPP**SPV**YYPPVTQ**SP**PP**PSP**VYYP

PVTN**SP**PP**PSP**VYYPPVTY**SP**PP**PSPV**YYPQV**TPSP**PP**PSP**LYYPPV**TPSP**PP**PSPV**YYPPV**TP**
SPPP**PSPV**YYPPV**TPSP**PP**PSP**VYY**PS**ETQ**SP**PP**PT**EYYY**SPS**Q**SP**P**PT**KACKEGHPPQA**TPS**Y
EPPPEYSYSS**SP**PP**PSPT**SYFPPM**PS**VSYDA**SP**PPP**PS**YY
>AT4G28380.1 | Symbols: | Leucine-rich repeat (LRR) family
protein | chr4:14039756-14040931 REVERSE LENGTH=391
MALSITFPVLISVLLYHFSLVSTAKHSLHHHTSHSTNPKQHLQQAYRALKAWKKVIYSDPKNLT
ADWVG**PS**VCSYTGIFC**APSPS**NPNTLVVAGIDLNHGDIAGFLPEAIGLLSDLALIHLNSNRFCG
ILPRSFANLSLLYELDLSNNRFVGPFPDVVLAL**PS**LKYLDLRYNEFEGPLPPKLFSNPLDAIFV
NNNRLTSLIPRDFTGTTASVVVFANNDFSGCLP**PT**IARFADTLEELLLINSSLSGCLPPEVGYL
YKLRVLDMSYNSLVG**PVP**YSLAGLGHLEQLNLEHNMFTGT**VP**LGVCVL**PS**LLNVTVSYNYFSEE
EGICRNLTSRGIAIDDRYNCLPDKPLQRPQKVCDAVLEHPIDCYDHEC**SP**M**AP**LV**APS**TAG**PS**I
APG**PA**DT
>AT5G25550.1 | Symbols: | Leucine-rich repeat (LRR) family
protein | chr5:8894179-8895480 FORWARD LENGTH=433
MRIYQ**PT**LLIFTTVVLLSISAV**AP**GGSRQLLYTRDDPITIPPYLIFENVRLERAYVALQAWKRA
MISDPWNLTTNWFGSRVCDYNGVVCSESLDDPLVKTVSGVDLNQGDIAGHLPEELGLLTDIALF
HVNSNRFCGTL**PV**GFSQLSLLFELDLSNNRFAGKFPEVVIGLPKLKYLDLRYNEFEGELPESLF
DKDLDALFLNSNRFRSKI**PV**NMGN**SPV**SVLVLASNRFEGCIP**PS**FGKMGKTLNEIILMDNGLQS
CIPNDMGLLQNVTVLDISYNWLVGELPKSMGQMENLEVLNVERNMLSGLIPDELCSLEKLRDFR
YGSNYFTGE**PA**TCRYLENYNYTMNCFKDVRDQRSMMECKMFLSK**PV**DCDSFKC**SP**GSSCF**SP**P**P**
SQI**SPS**SQPL**APAPSPTSP**PLS**TP**P**PA**RPCPPVY**SP**PPPPPLSL**APS**MN
>AT4G13340.1 | Symbols: | Leucine-rich repeat (LRR) family
protein | chr4:7758610-7760892 FORWARD LENGTH=760
MKKTIQILLFFFFLINLTNALSISSDGGVLSDNEVRHIQRRQLLEFAERSVKITVD**PS**LNFENP
RLRNAYIALQAWKQAILSDPNNFTSNWIGSNVCNYTGVFC**SPA**LDNRKIRTVAGIDLNHADIAG
YLPEELGLLSDLALFHVNSNRFCGTVPHRFNRLKLLFELDLSNNRFAGKF**PT**VVLQL**PS**LKFLD
LRFNEFEGTVPKELFSKDLDAIFINHNRFRFELPENFGD**SPV**SVIVLANNRFHGCV**PS**SLVEMK
NLNEIIFMNNGLNSCL**PS**DIGRLKNVTVFDVSFNELVGPLPESVGEMVSVEQLNVAHNMLSGKI
PASICQLPKLENFTYSYNFFTGE**APV**CLRLPEFDDRRNCLPGR**PA**QR**SP**GQCKAFLSRPPVNCG
SFSCGRSV**SP**RPPVV**TP**LPP**PS**L**PSP**PP**PAP**IFS**TPPT**LT**SP**PP**PSP**PPPVY**SP**PPPPPPPPPV
Y**SP**PPPPPPPPPPPVY**SP**PPPPPPPPPPPPVY**SP**PP**PSP**PPPPPPVY**SP**PPPPPPPPPPPVY**SP**
PPPPVYS**SP**PPP**PSPAPTPV**YCTRPPPPPPH**SP**PPPQF**SP**PPPEPYYYS**SP**PPPHS**SP**PPH**SP**P
PPH**SP**PPPIYPYL**SP**PPP**PTPV**S**SP**P**PTPV**Y**SP**PPPPPCIEPPPPPPCIEY**SP**PPPPPVVHYSS**S**
PPPPPVYYS**SP**PPPPVYYS**SP**PPPPPVHYS**SP**PPPEVHYH**SP**P**PSPV**HYS**SP**PPP**PSAP**CEE**SP**
P**PAPV**VHH**SP**PPPMVHH**SP**PPPVIHQ**SP**PP**PSP**EYEGPLPPVIGVSYA**SP**PPPPFY
>AT3G24480.1 | Symbols: | Leucine-rich repeat (LRR) family
protein | chr3:8901154-8902638 REVERSE LENGTH=494
MKNNTTQSLLLLLLFFFFFFEISHSLSISSN**AP**LSDTEVRFIQRRQLLYYRDEFGDRGENVTVD
PSLIFENPRLRSAYIALQAWKQAILSDPNNITVNWIGSNVCNYTGVFCSKALDNRKIRTVAGID
LNHADIAGYLPEELGLLTDLALFHVNSNRFCGTVPHKFKQLKLLFELDLSNNRFAGKF**PT**VVLH
L**PS**LKFLDLRFNEFEGTVPKELFSKNLDAIFINHNRFRFELPENFGD**SPV**SVIVLANNHFHGCI
PTSLVEMKNLNEIIFMNNGLNSCL**PA**DIGRLKNVTVFDVSFNELVGPLPESVGGMVEVEQLNVA
HNLLSGK**IPA**SICQLPKLENFTYSYNFFTGE**APV**CLRLSEFDDRRNCLPGR**PA**QRSSRQCSAFL
SR**PS**VDCGSFGCGRSVVK**PSP**PIVALPPPPP**PSP**PLPPPVY**SP**P**PSP**PVF**SP**P**PSP**PVY**SP**PPP
PSIHYS**SP**PPPPVHHS**SP**PP**PSP**EFEGPLPPVIGVSYA**SP**PPPPFY
>AT1G49490.1 | Symbols: | Leucine-rich repeat (LRR) family
protein | chr1:18317563-18320106 REVERSE LENGTH=847

MERPFGCFFILLLISYTVVATFDDE**PS**FPENADLTKDLEQKCFSINKVDPNLKFENDRLKRAYI
ALQAWKKAIYSDPFKTTANWVGSDVCSYNGVYC**APA**LDDDSLTVVAGVDLNHADIAGHLPPELG
LMTDLALFHINSNRFCGIIPKSLSKLALMYEFDVSNNRFVGQFPEVSLSW**PS**LKFLDLRYNEFE
GSL**PS**EIFDKDLDAIFLNNNRFESVIPGTIGKSKASVVTFANNKFSGCIPKSIGNMKNLNEIVF
TGNNLTGCFPNEIGLLNNVTVFDASKNGFVGSL**PS**TLSGLASVEQLDLSHNKLTGFVVDKFCKL
PNLDSFKFSYNFFNGEAESC**VP**GRNNGKQFDDTNNCLQNR**PS**QK**PA**KQCLPVVSRPVDCSKDKC
SGGSNGGS**SPSP**NPPRTSEPK**PS**KPEPVMPK**PS**DSSKPE**TP**K**TP**EQ**PSP**KPQPPKHE**SP**KPEEP
ENKHELPKQKE**SP**KPQ**PS**KPED**SP**KPEQPKPEE**SP**KPEQPQIPE**PT**K**PVSP**PNEAQG**PTP**DDPY
DA**SPV**KNRR**SP**PPPKVEDTRVPPPQPPM**PSPSPPSP**IY**SP**PPPVH**SP**PPPVYS**SP**PPPHVY**SP**P
PPVA**SP**PP**PSP**PPPVH**SP**PPPPVF**SP**PPPVF**SP**PP**PSP**VY**SP**PP**PS**H**SP**PPPVY**SP**PP**PT**F**SP**P
PTHNTNQPPMG**APTPT**Q**APTPS**SETTQ**VPTPS**SESDQSQIL**SP**VQ**APTPV**QSS**TPS**SE**PT**QV**PT**
PSSSESYQ**AP**NL**SP**VQ**APTPV**Q**APT**TSSETSQ**VPTPS**SESNQ**SPS**Q**APTP**ILE**PV**H**APTP**NSK**P**
VQ**SPTPS**SEPVS**SP**EQSEEVE**AP**E**PTPV**N**PS**SV**PS**S**SPS**TDTSIPPPENNDDDDDGDFVLPPHI
GFQYA**SP**PPPMFQGY
>AT2G43290.1 | Symbols: MSS3 | Calcium-binding EF-hand family
protein | chr2:17991308-17991955 REVERSE LENGTH=215
MVRIFLLYNILNSFLLSLVPKKLRTLFPLSWFDKTLHKN**SPPSPS**TML**PSPS**SSS**APT**KRID**PS**
ELKRVFQMFDKNGDGRITKEELNDSLENLGIYIPDKDLTQMIHKIDANGDGCVDIDEFESLYSS
IVDEHHNDGETEEEDMKDAFNVFDQDGDGFITVEELKSVMASLGLKQGKTLDGCKKMIMQVDAD
GDGRVNYKEFLQMMKGGGFSSSN
>AT1G51940.1 | Symbols: | protein kinase family protein /
peptidoglycan-binding LysM domain-containing protein |
chr1:19296092-19298941 REVERSE LENGTH=651
MNLTFYIFFLSLL**PS**FSSSKPMNCSDTTRLCSSFLAFKPNQNQSFSVIQSMFDVLPQDITADIS
GGYFFIKKNCSCLTTTHQYTTNTTFTIRQNVGYVYNVTVSAYSGLAFPPNTTRAARAGAVVSVQ
LLCGCSSGLWNYLMSYVAMAGDSVQSLSSRFGVSMDRIEDVNGILNLDNITAGDLLYIPLDSVP
GEPYETSKINP**PAPSPAPA**SSLANGNISDDQVNHTAKSGSHVPYIWIVGGLGVVLALLVLCILV
CICLRSSSCSSSEEDGNGHNFQILRKSGFFCGSGRYNCCRSGDFRQTNGETQVVAIPKALGDGM
FEIEKPMVFTYEEIRAATDEFSDSNLLGHGNYGSVYFGLLREQEVAVKRMTATKTKEFAAEMKV
LCKVHHSNLVELIGYAATVDELFVVYEYVRKGMLKSHLHDPQSKGN**TP**LSWIMRNQIALDAARG
LEYIHEHTKTHYVHRDIKTSNILLDEAFRAKISDFGLAKLVEKTGEGEISVTKVVGTYGYL**AP**E
YLSDGLATSKSDIYAFGVVLFEIISGREAVIRTEAIGTKNPERRPLASIMLAVLKN**SP**DSMNMS
SLKEFVDPNMMDLYPHDCLFKIATLAKQCVDDDPILRPNMKQVVISLSQILLSSIEWEATLAGN
SQVFSGLVQGR>AT2G13800.1 | Symbols: ATSERK5, SERK5, BAK8 |
somatic embryogenesis receptor-like kinase 5 | chr2:5753276-
5757065 FORWARD LENGTH=601
MEHGSSRGFIWLILFLDFVSRVTGKTQVDALIALRSSLSSGDHTNNILQSWNATHV**TP**CSWFHV
TCNTENSVTRLDLGSANLSGELVPQLAQLPNLQYLELFNNNITGEIPEELGDLMELVSLDLFAN
NISGPI**PS**SLGKLGKLRFLRLYNNSLSGEIPRSLTALPLDVLDISNNRLSGDIPVNGSFSQFTS
MSFANNKLRPR**PASPSPSPS**GTSAAIVVGVAAGAALLFALAWWLRRKLQGHFLDV**PA**EEDPEVY
LGQFKRFSLRELLVATEKFSKRNVLGKGRFGILYKGRLADDTLVAVKRLNEERTKGGELQFQTE
VEMISMAVHRNLLRLRGFCM**TPT**ERLLVYPYMANGSVASCLRERPEGN**PA**LDWPKRKHIALGSA
RGLAYLHDHCDQKIIHLDVKAANILLDEEFEAVVGDFGLAKLMNYNDSHVTTAVRGTIGHI**AP**E
YLSTGKSSEKTDVFGYGVMLLELITGQKAFDLARLANDDDIMLLDWVKEVLKEKKLESLVDAEL
EGKYVETEVEQLIQMALLCTQSSAMERPKMSEVVRMLEGDGLAERWEEWQKEEMPIHDFNYQAY
PHAGTDWLIPYSNSLIENDY**PS**GPR
>AT1G63570.1 | Symbols: | Receptor-like protein kinase-related
family protein | chr1:23575450-23576304 FORWARD LENGTH=284

MARIILT**AP**LFYFFFSLLSHQTMSQPQHMHTFCSVDSFTQTSSYETNRNILLTTLSLTSSLVHY
LNATIGL**SP**DTVYGMFLCRGDINTTSCSDCVQTAAIEIATNCTLNKRAFIYYDECMVRYSNVSF
FSEFESK**PV**IVRYSLRS**AP**NSNRFNQTLSNKLDQLIPNV**SPS**TLIPYFVEDQERVTQLEGSYDL
VSMIQC**SP**DLD**PS**NCTICLRFAYATVSTCCG**VPS**SALIF**TP**KCILRYRTFVL**PSPAPSPS**SLPP
I**SPT**S**SP**PLSLPPQLPPPLSQPPPPLST
>AT4G18640.1 | Symbols: MRH1 | Leucine-rich repeat protein
kinase family protein | chr4:10260481-10263577 FORWARD
LENGTH=678
MMGCGFHFPWFFFLIIGLQ**AP**LSLSLTSQGSALLKFRARVNSDPHGTLANWNVSGINDLCYWSG
VTCVDGKVQILDLSGYSLEGTL**AP**ELSQLSDLRSLILSRNHFSGGIPKEYGSFENLEVLDLREN
DLSGQIPPELSNGLSLKHLLLSGNKFSDDMRIKIVRLQSSYEVRLKK**SP**KL**SP**LAVLGCINRKL
GHCVSRNRIIQVKKVEAIVFRIKATSRRFLKAF**PSF**LEETDIYKRRELLEETSNLAAE**PAPSAP**
SPSPGIITEA**SP**RSSGSF**PA**VTNAKKRRPPLVPPV**PSP**DKGST**SP**DISKNQPQDNKQSKGSKHV
WLYVVIAVASFVGLLIIVAVIFFCRKRAVKSIGPWKTGLSGQLQKAFVTGVPKLNRSELETACE
DFSNIIETFDGYTVYKGTLSSGVEIAVASTAIAESKEWTRAMEMAYRRKIDTLSRINHKNFVNL
IGYCEEDDPFNRMMVFEY**AP**NGTLFEHLHDKETEHLDWSARMRIIMGTAYCLQHMHGMNPPMAH
TDFNSSEIYLTDDYAAKVSEIPFNLEARLNPKKHVSGDLEQTSLLLPPEPEANVHSFGVLMLEI
ISGKLSFSDEYGSIEQWASKYLEKDDLGEMID**PS**LKTFKEEELEVICDVIRECLKTEQRQR**PS**M
KDVAEQLKQVINI**TP**EKA**TP**RS**SP**LWWAELEILSSEAT
>AT3G19320.1 | Symbols: | Leucine-rich repeat (LRR) family
protein | chr3:6696395-6698073 REVERSE LENGTH=493
M**AP**NYY**PS**IFLILSVHFLSILATGNNHNTYRKSLEIIIGGGSDSTNYN**SPAPSP**EPEDYLPLPP
PPQ**TP**PPPPPPQSLPP**PSPSP**EPEHYPPPPYHHYI**TPSP**PPPRPLPPPPPPPLHFS**SP**LIKKVY
PVIKNFQTLIEDDPKSILKTWVGTDICAQDKYIGLECAKLPGTNDLALASIQFNNFNLGGKKLR
LDNFLNKLEEVTIFHANSNNFVGSVPNFSKLKYLFELDLSNNKLSGEF**PS**SVLKATNLTFLDLR
FNSFSGSVPPQVFNLDLDVLFINNNNLVQRLPENLGSITALYLTFANNRFTGPIPGSIGDIKSL
QEVLFLNNKLTGCLPYQIGNLNRATVFDVELNQLTGPIPYSFGCLKKMEQLNLARNNFYGTIPE
IVCELSALKNLSLSYNYFTQVGPKCRTLIKRKILDVGMNCILDLTNQR**TP**WECAKFFLRKQSCP
NFKSFFYLPCGKDPHRIKPDQEGSDGQA**SP**PVSYNALNPDRIRNL
>AT3G04690.1 | Symbols: ANX1 | Malectin/receptor-like protein
kinase family protein | chr3:1273386-1275938 REVERSE LENGTH=850
MSGKTRILFFLTCLSFLLVF**PT**RSNGQDLALSCGTSEASADQDKKKWEPDTKFLKTGNSIHATA
TYQD**PS**LLST**VP**YMTARIFT**APA**TYEIPIKGDKRHLLRLYFY**PS**TYTGLNISNSYFTVEANDVT
LLSNFSAAITCQALTQAYLVKEYSL**APT**DKDVLSIKF**TPS**DKYRDAFAFINGIEVIQMPELFDT
AALVGFTDQTMDAKTANLQSMFRLNVGGQDIPGSQDSGGLTRTWYND**AP**YIFSAGLGVTLQASN
NFRINYQNMPVSI**APA**DIYKTARSQGPNGDINLKSNLTWMFQIDKNFTYILRLHFCEFQLSKIN
QKVFNIYINNRTAQADT**TPA**DIIGWTGEKGIPMYKDYAIYVDANNGGEEITLQM**TPS**TFGQPEY
YDSSLNGLEIFKMDTMKNLAGPNPE**PSP**MQAEEEVKKEFKNEKRHAFIIGSAGGVLAVLIGALC
FTAYKKKQGYQGGDSHTSSWLPIYGNSTTSGTKSTISGKSNNGSHLSNLAAGLCRRFSLPEIKH
GTQNFDDSNVIGVGGFGKVYKGVIDGTTKVAVKKSNPNSEQGLNEFETEIELLSRLRHKHLVSL
IGYCDEGGEMCLVYDYMAFGTLREHLYNTKKPQLTWKRRLEIAIGAARGLHYLHTGAKYTIIHR
DVKTTNILVDENWVAKVSDFGLSKTGPNMNGGHVTTVVKGSFGYLDPEYFRRQQLTEKSDVYSF
GVVLFEILCAR**PA**LN**PS**LPKEQVSLGDWAMNCKRKGNLEDIIDPNLKGKINAECLKKFADTAEK
CLNDSGLER**PT**MGDVLWNLEFALQLQETADGTRHR**TP**NNGGSSEDLGRGGMAVNVAGRDDVSDL
SSEDNTEIFSQIVNPKGR
>AT4G39110.1 | Symbols: | Malectin/receptor-like protein kinase
family protein | chr4:18222483-18225119 REVERSE LENGTH=878

MEIRKKPNIFTVLVIDFSSK**PS**MALLLAILLFLSG**PS**ASAVAAAAVG**PA**TGFK**PA**DDILIDCGS
KSSSK**TP**DGRVFKSDQETIQYIEAKEDIQVS**APPS**DKVA**SP**IYLTARIFREEATYKFHLTRPGW
HWVRLHFLAFPNDKFDLQQATFSVLTEKYVLLHNFKISNNNNDSQAAVQKEYLVNMTDAQFALR
FRPMKSSAAFINAIEVVS**AP**DELISDSGTALF**PV**IGFSGLSDYAYQSVYRVNVGGPLIMPQNDT
LGRTWIPDKEFLKDENLAKDVKT**TPS**AIKYPPEV**TP**LI**AP**QTVYATAVEMANSLTIDPNFNVSW
NF**PS**N**PS**FNYLIRLHFCDIVSKSLNDLYFNVYINGKTAISGLDLSTVAGNLA**AP**YYKDIVVNAT
LMGPELQVQIGPMGEDTGTKNAILNGVEVLKMSNSVNSLDGEFGVDGRTTGMGKHGMVATAGFV
MMFGAFIGLGAMVYKWKKRPQDWQKRNSFSSWLLPIHAGDSTFMTSKGGSQKSNFYNSTLGLGR
YFSLSELQEATKNFEASQIIGVGGFGNVYIGTLDDGTKVAVKRGNPQSEQGITEFQTEIQMLSK
LRHRHLVSLIGYCDENSEMILVYEFMSNGPFRDHLYGKNL**AP**LTWKQRLEICIGSARGLHYLHT
GTAQGIIHRDVKSTNILLDEALVAKVADFGLSKDVAFGQNHVSTAVKGSFGYLDPEYFRRQQLT
DKSDVYSFGVVLLEALCAR**PA**INPQLPREQVNLAEWAMQWKRKGLLEKIIDPHLAGTINPESMK
KFAEAAEKCLEDYGVDR**PT**MGDVLWNLEYALQLQEAFTQGKAEETENAKPDVV**TP**GSVPVSD**PS**
PITPSVTTNEAATVPV**PA**KVEENSGTAVDEHSGTAMFTQFANLNGR

>AT3G56100.1 | Symbols: MRLK, IMK3 | meristematic receptor-like kinase | chr3:20817074-20819517 REVERSE LENGTH=719

MEFITQNQAITSLSMINTDIDQPKASLRSRFLLHLIICLLFFVPPCSSQAWDGVVITQADYQGL
QAVKQELIDPRGFLRSWNGSGFSACSGGWAGIKCAQGQVIVIQLPWKSLGGRISEKIGQLQALR
KLSLHDNNLGGSIPMSLGLIPNLRGVQLFNNRLTGSI**PA**SLGVSHFLQTLDLSNNLLSEIIPPN
LADSSKLLRLNLSFNSLSGQIPVSLSRSSSLQFLALDHNNLSGPILDTWGSKIRGTL**PS**ELSKL
TKLRKMDISGNSVSGHIPETLGNISSLIHLDLSQNKLTGEIPISISDLESLNFFNVSYNNLSG**P**
VPTLLSQKFNSSSFVGNSLLCGYSVS**TP**C**PT**L**PSPSP**EKERK**PS**HRNLSTKDIILIASGALLIV
MLILVCVLCCLLRKKANETKAKGGEAGPGAVAAKTEKGGEAEAGGETGGKLVHFDGPMAFTADD
LLCATAEIMGKSTYGTVYKATLEDGSQVAVKRLRER**SP**KVKKREKLVVFDYMSRGSLATFLHAR
GPDVHINW**PT**RMSLIKGMARGLFYLHTHANIIHGNLTSSNVLLDENITAKISDYGLSRLMTAAA
GSSVIATAGALGYR**AP**ELSKLKKANTKTDVYSLGVIILELLTGK**SPS**EALNGVDLPQWVATAVK
EEWTNEVFDLELLNDVNTMGDEILNTLKLALHCVDA**TPS**TRPEAQQVMTQLGEIRPEETTATTS
EPLIDVPEASASTSQ

>AT5G39020.1 | Symbols: | Malectin/receptor-like protein kinase family protein | chr5:15616917-15619358 FORWARD LENGTH=813

MNCNVLFLLSVLVSVTAGVTAAYH**PT**DVFLFNCGDTSNNVDNSGRNWTVESRQILSSNLVNASF
TSEASYQKAGVSRIPYMKARIFRSEFTYSF**PVTP**GSIFLRLYFY**PT**QYKSGFDAVNSFFSVKVN
GFTLLRNFNADSTVQASIPLSNSLIKEFII**PV**HQTLNLTF**TPS**KNLLAFVNGIEIVSMPDRFYS
KGGFDNVLRNVSSDVDFQIDNSTAFESVHRLNVGGQIVNEVDDSGMFRRWLSDDSFGNSGSIVN
VPGVKINYTEK**TPA**YV**AP**YDVYATSRLMGNSSNLMFNLTGMFLTVDAGYNYLVRLHFCETLPQV
TKAGQRVFSIFVEDKMAKKETDVIRLSGGPRIPMYLDFSVYVGFESGMIQPELRLDL**VP**LKDTN
QTYYDAILSGVEILKLNDSDGNLARPNPELLVSTDS**TP**DDSNV**TP**PIKGKPHVLVIILIVVGSV
IGLATFIVIIMLLIRQMKRKKNKKENSVIMFKLLLKQYIYAELKKITKSFSHTVGKGGFGTVYR
GNLSNGRTVAVKVLKDLKGNGDDFINEVTSMSQTSHVNIVSLLGFCYEGSKRAIISEFLEHGSL
DQFISRNKSL**TP**NVTTLYGIALGIARGLEYLHYGCKTRIVHFDIKPQNILLDDNFCPKVADFGL
AKLCEKRESILSLIDTRGTIGYI**AP**EVVSRMYGGISHKSDVYSYGMLVLDMIGARNKVETTTCN
GSTAYFPDWIYKDLENGDQTWIIGDEINEEDNKIVKKMILVSLWCIRPC**PS**DRPPMNKVVEMIE
GSLDALELPPK**PS**RHISTELVLESSSLSDGQEAEKQTQTLDSTII

>AT5G28680.1 | Symbols: ANX2 | Malectin/receptor-like protein kinase family protein | chr5:10719437-10722013 REVERSE LENGTH=858

MNEKLRILFSFLCFFYVLLV**SPS**QSNGQDISLSCGASE**PA**VDQDKKKWEPDTKFLK**TP**NTVH**AP**
ATYQD**PS**LLSTVPYMTSRIFT**APA**TYEI**PV**KGDKRHMLRLHFY**PS**TYTGLNILDSYFSVAANDL

TLLSNFSAAITCQALTQAYLVREYSL**APS**EKDVLSIIF**TPS**DKHPKAFAFINGIEVIPMPELFD
TASLVGFSDQTSDTKTANLQTMFRLNVGGQDIPGSQDSGGLTRTWYND**AP**YIFSAGLGVTLQAS
NNFRIDYQKM**PV**ST**APA**DVYKTARSQGPNGDINMKSNLTWMFQVDTNFTYIMRLHFCEFQLAKI
NQKVFNIFINNRTAQGDTN**PA**DILGWTGGKGI**PT**YKDYAIYVDANTGGGGEEISLQM**TPS**TFGQ
PEYYDSQLNGLEIFKIDTMKNLAGPNPK**PSP**MQANEDVKKDFQGDKRITAFVIGSAGGVAAVLF
CALCFTMYQRKRKFSGSDSHTSSWLPIYGNSHTSATKSTISGKSNNGSHLSNLAAGLCRRFSLS
EIKHGTHNFDESNVIGVGGFGKVYKGVIDGGTKVAIKKSNPNSEQGLNEFETEIELLSRLRHKH
LVSLIGYCDEGGEMCLIYDYMSLGTLREHLYNTKRPQLTWKRRLEIAIGAARGLHYLHTGAKYT
IIHRDVKTTNILLDENWVAKVSDFGLSKTGPNMNGGHVTTVVKGSFGYLDPEYFRRQQLTEKSD
VYSFGVVLFEVLCAR**PA**LN**PS**LSKEQVSLGDWAMNCKRKGTLEDIIDPNLKGKINPECLKKFAD
TAEKCLSDSGLDR**PT**MGDVLWNLEFALQLQETADGSRHR**TPS**NGGGSVDLGGGGGGVTVNISAG
ESDLGDDLSSEENSGIFSQIVNPKGR
>AT5G01550.1 | Symbols: LECRKA4.2 | lectin receptor kinase a4.1
| chr5:214517-216583 REVERSE LENGTH=688
MLVLFLLLTI**PT**RAQRTTTE**TP**KTEFIFRGFSGNQSNIVTTGAATIKLDGLLRLTDRNSNVTGT
SFYHK**PV**RLLETNTSSTNSTIRSFSTSFVFVII**PT**SSSNGGFGFTFTL**SPTP**DRTGAESAQYLG
LLNKANDGNSTNHVFAVEFDTVQGFKDGADRTGNHIGLNFNSLTSDVQE**PV**VYYDNEDPNRKED
FPLQSGDPIRAILDYDG**PT**QTLNLTVY**PA**NLKSR**PV**RPLISR**PVP**KLSQIVQEEMYVGFTAATG
RDQSSAHYVMGWSFSSGGDLLTEDTLDLLELPRPPPNTAKKRGYNSQVLALIVALSGVTVILLA
LLFFFVMYKKRLQQGEVLEDWEINHPHRLRYKDLYAATDGFKENRIVGTGGFGTVFRGNLS**SPS**
SDQIAVKKI**TP**NSMQGVREFIAEIESLGRLRHKNLVNLQGWCKQKNDLLLIYDYIPNGSLDSLL
YSRPRQSGVVLSWNARFKIAKGIASGLLYLHEEWEKVVIHRDIK**PS**NVLIEDDMNPRLGDFGLA
RLYERGSQSNTTVVVGTIGYM**AP**ELARNGKSSSASDVFAFGVLLLEIVSGRR**PT**DSGTFFLADW
VMELHARGEILHAVDPRLGFGYDGVEARLALVVGLLCCHQR**PT**SR**PS**MRTVLRYLNGDDDVPEI
DNDWGYSDSSRSDLGSNFEGYVSSDRASSSV**PS**FSVTRVSSSSVISGR
>AT1G75640.1 | Symbols: | Leucine-rich receptor-like protein
kinase family protein | chr1:28403600-28407022 REVERSE
LENGTH=1140
MAATVIFFLHFAAIFFSRFHHTSAISSETQALTSFKLSLHDPLGALESWNQS**SPSAP**CDWHGVS
CFSGRVRELRLPRLHLTGHL**SP**RLGELTQLRKLSLHTNDINGA**VPS**SLSRCVFLRALYLHYNSF
SGDFPPEILNLRNLQVLNAAHNSLTGNLSDVTVSKSLRYVDLSSNAISGKI**PA**NFSADSSLQLI
NLSFNHFSGEI**PA**TLGQLQDLEYLWLDSNQLQGTI**PS**ALANCSSLIHFSVTGNHLTGLIPVTLG
TIRSLQVISLSENSFTGT**VP**VSLLCGYSGYNSSMRIIQLGVNNFTGIAK**PS**NAACVNPNLEILD
IHENRINGDF**PA**WLTDLTSLVVLDISGNGFSGGVTAKVGNLMALQELRVANNSLVGEI**PT**SIRN
CKSLRVVDFEGNKFSGQIPGFLSQLRSLTTISLGRNGFSGRI**PS**DLLSLYGLETLNLNENHLTG
AI**PS**EITKLANLTILNLSFNRFSGE**VPS**NVGDLKSLSVLNISGCGLTGRI**PV**SISGLMKLQVLD
ISKQRISGQL**PV**ELFGLPDLQVVALGNNLLGGV**VP**EGFSSLVSLKYLNLSSNLFSGHIPKNYGF
LKSLQVLSLSHNRISGTIPPEIGNCSSLEVLELGSNSLKGHI**PV**YVSKLSLLKKLDLSHNSLTG
SIPDQISKDSSLESLLLNSNSLSGRIPESLSRLTNLTALDLSSNRLNSTI**PS**SLSRLRFLNYFN
LSRNSLEGEIPEALAARFTN**PT**VFVKNPGLCGKPLGIECPNVRRRRRRKLILLVTLAVAGALLL
LLCCCGYVFSLWKWRNKLRLGLSRDKKG**TPS**RTSRASSGGTRGEDNNGGPKLVMFNNKITLAET
LEATRQFDEENVLSRGRYGLVFKATFRDGMVLSVRRLMDGASITDATFRNQAEALGRVKHKNIT
VLRGYYCGPPDLRLLVYDYMPNGNLATLLQEASHQDGHVLNWPMRHLIALGIARGLSFLHSLSI
IHGDLKPQNVLFDADFEAHLSEFGLDRLTAL**TPA**EE**PS**TSS**TP**VGSLGYI**AP**EAGLTGETSKES
DVYSFGIVLLEILTGKKAVMFTEDEDIVKWVKRQLQKGQIVELLEPGLLELDPESSEWEEFLLG
IKVGLLCTGGDVVDR**PS**MADVVFMLEGCRVG**PA**ISLSAD**PTSPTSPA**ATAVS
>AT5G38990.1 | Symbols: | Malectin/receptor-like protein kinase
family protein | chr5:15608824-15611466 FORWARD LENGTH=880

MICHVLVIFTILVSAVVDATASYE**PT**DVFLINCGDTSNNMDYSGRNWTTENPKFMSSNAVDDAS
FTSSASYQESGIPQ**VP**YLKARIFRYDFTYSF**PVSP**GWKFLRLYFY**PT**RYGSDFDAVKSFFSVNV
NRFTLLHNFSVKASIPESSSLIKEFI**VPV**NQTLDLTF**TPSP**NSLAFVNGIEIISMPDRFYSKGG
FDDVVRNVGRDVDFEIDNSTAFETVYRVNVGGKVVGDVGDSGMFRRWLSDEGFLLGINSGAIPN
ITGVKINYTDK**TPA**YV**AP**EDVYTTCRLMGNKD**SP**ELNLNFNLTWLFEVDAGFAYIVRLHFCETQ
PEVNKTGDRVFSIFFGYQLAMREMDVFRLSGGFRLPMYLDFKVLVDADGTSQR**PS**LRVDL**TP**YK
EDY**PT**YYDAILSGVEILKLSNSDGNLAGLNPIPQL**SP**PPQSI**TP**LKGKGKSSHVLPIIIAVVGS
AVALAFFVLVVVLVVMKRKKKSNESSVDTTNK**PS**TNSSWGPLLHGTGSTNTKSASSL**PS**DLCRR
FSIYEIKSATNDFEEKLIIGVGGFGSVYKGRIDGGATLVAVKRLEITSNQGAKEFDTELEMLSK
LRHVHLVSLIGYCDDDNEMVLVYEYMPHGTLKDHLFRRDKASDPPLSWKRRLEICIGAARGLQY
LHTGAKYTIIHRDIKTTNILLDENFVAKVSDFGLSRVG**PT**SASQTHVSTVVKGTFGYLDPEYYR
RQILTEKSDVYSFGVVLLEVLCCRPIRMQSVPPEQADLIRWVKSNFNKRTVDQIIDSDLTADIT
STSMEKFCEIAIRCVQDRGMERPPMNDVVWALEFALQLHETAKKKNDNVESLDLM**PS**GEVGTTT
DGEDDLFSRTTGHVGKSTTTDDSVLVVGDERSGSSWGVFSEINEPKAR
>AT3G22030.1 | Symbols: | Receptor protein kinase-related |
chr3:7759317-7760738 FORWARD LENGTH=253
MFSSYSLCKCLVSFHILAIQVLISCASSLNLTNEYLNHKCRVIKENTNQEVSTRKNSTLSAVEL
LPLTLQMVSHTGPIVEIPNL**SPS**YSNVAVTLTG**PS**A**TPA**MLP**PSP**GCPRNKGGIIWYDQCFLDV
SMINDN**SP**RRMNYDNIFSMHNPNNVRGNVNSFNKKTTEFLYKLIGKADRLDVDGINFLYYAAGE
MRLGKQTLFAMVQCAKDILSCKDCLEWSIKELSKCCDGKQGARVVGTICNLRYELYPFLRT
>AT2G21480.1 | Symbols: | Malectin/receptor-like protein kinase
family protein | chr2:9202753-9205368 REVERSE LENGTH=871
MEIRKKPNIPMCLVLDSSSRPFMTLLFTILLFLTGLASAVGAVGG**SPT**AGFK**PA**DDILIDCGSK
SSTK**TP**EGRVFKSDSETVQYIEAKDDIQVS**APPS**DKL**PSP**IYLTAKIFREEAIYKFHLTRPGWH
WVRLHFFAFPNDKFDLQQATFSVLTEKYVLLHNFKLSNDNNDSQATVQKEYLLNMTDAQFALRF
KPMKGSAAFINGIELVS**AP**DELISDAGTSLFPVNGFSGLSDYAYQSVYRVNVGGPLI**TP**QNDTL
GRTW**TP**DKEYLKDENLAKDVKTN**PT**AIIYPPGV**TP**LI**AP**QTVYATGAEMADSQTIDPNFNVTWN
F**PS**N**PS**FHYFIRLHFCDIISKSLNDLYFNVYINGKTAISGLDLSTVAGDLS**AP**YYKDIVVNSTL
MTSELQVQIGPMGEDTGKKNAILNGVEVLKMSNSVNSLDGEFGVDGQRASMGKQGMVATAGFVM
MFGAFVGLGAMVYKWKKRPQDWQKRNSFSSWLLPIHAGDSTFMTSKTGSHKSNLYNSALGLGRY
FSLSELQEVTKNFDASEIIGVGGFGNVYIGTIDDGTQVAIKRGNPQSEQGITEFHTEIQMLSKL
RHRHLVSLIGYCDENAEMILVYEYMSNGPFRDHLYGKNL**SP**LTWKQRLEICIGAARGLHYLHTG
TAQGIIHRDVKSTNILLDEALVAKVADFGLSKDVAFGQNHVSTAVKGSFGYLDPEYFRRQQLTD
KSDVYSFGVVLLEALCAR**PA**INPQLPREQVNLAEWAMLWKQKGLLEKIIDPHLVGAVNPESMKK
FAEAAEKCLADYGVDR**PT**MGDVLWNLEYALQLQEAFSQGKAEAEEVE**TP**KPVAV**PAAAPTSPA**A
TTAAASERPVSQTEEKDDSTVDQHSGTTMFTQFASLNGR
>AT4G38830.1 | Symbols: CRK26 | cysteine-rich RLK (RECEPTOR-like
protein kinase) 26 | chr4:18122339-18124943 FORWARD LENGTH=665
MLSLLLPLISLLFQIQCFTVKSQPVPLNQICSNVTGNFTVN**TP**YAVNLDRLISSLSSLRRNVNG
FYNISVGDSDEKVNSISQCRGDVKLEVCINCIAMAGKRLVTLCPVQKEAIIWYDKCTFRYSNRT
IFNRLEI**SP**HTSITGTRNFTGDRDSWEKSLRGLLEGLKNRASVIGRSKKNFVVGETSG**PS**FQTL
FGLVQC**TP**DISEEDCSYCLSQGIAKI**PS**CCDMKMGSYVM**SPS**CMLAY**AP**WRFYDPVDTDD**PS**S**V**
PATPSRPPKNETRSVTQGDKNRGVPKALIFASASVAIVVLFIVLLVVFLKLRRKENIRNSENKH
ENENISTDSMKFDFSVLQDATSHFSLENKLGEGGFGAVYKGVLSDGQKIAVKRLSKNAQQGETE
FKNEFLLVAKLQHRNLVKLLGYSIEGTERLLVYEFLPHTSLDKFIFDPIQGNELEWEIRYKIIG
GVARGLLYLHQDSRLRIIHRDLKASNILLDEEM**TP**KIADFGMARLFDIDHTTQRYTNRIVGTFG
YM**AP**EYVMHGQFSFKTDVYSFGVLVLEIISGKKNSGFSSEDSMGDLISFAWRNWKEGVALNLVD

KILMTMSSYSSNMIMRCINIGLLCVQEKVAER**PS**MASVVLMLDGHTIALSE**PS**K**PA**FFSHSNAV
SDSSSSLGHNAKTSNYNSNTELYPR
>AT1G28340.1 | Symbols: AtRLP4, RLP4 | receptor like protein 4 |
chr1:9940175-9943252 FORWARD LENGTH=626
MMLRFILASLLLSSFSLYSSLAR**PAP**YALRISCGARKNVR**TPPT**YALWFKDIAYTGGV**PA**NAT**T**
PTYI**TP**PLKTLRYFPISEGPNNCYNIVR**VP**KGHYSVRIFFGLVDQ**PS**FDKEPLFDISIEGTQIS
SLKSGWSSQDDQVFAEALIFLLGGTATICFHSTGHGD**PA**ILSIEILQVDDKAYSFGEGWGQGVI
LRTATRLTCGTGKSRFDEDYRGDHWGGDRFWNRMRSFGKSAD**SP**RSTEETIKKASV**SP**NFYPEG
LYQSALVSTDDQPDLTYSLDVEPNRNYSVWLHFAEIDNTITAEGKRVFDVVINGDTFFEDVDII
KMSGGRYAALVLNATVTVSGRTLTVVLQPKAGGHAIINAIEVFEIITAEFKTLRDEVSALQKMK
KALGL**PS**RFGWNGDPCVPPQHPWSGANCQLDKNTSRWFIDGLDLDNQGLKGFLPNDISKLKHLQ
SINLSENNIRGGI**PA**SLGSVTSLEVLDLSYNSFNGSIPETLGELTSLRILNLNGNSLSGKV**PA**A
VGGRLLHRASFNFTDNAGLCGIPGL**PA**CGPHLSSGAKIGIAFGVSLAFLLIVACAMIWWKRRQN
ILRAQQIAARG**AP**YAKKRTHVSHDIQMSRHGHNNHGQARTAVENG**PS**LLS
>AT2G40270.1 | Symbols: | Protein kinase family protein |
chr2:16822136-16824327 REVERSE LENGTH=489
MLFKMRSFVAFVLLLSWFGSCCSLKDQAVDFLKSEDSLKKDLSSDEDSTYLKAFGFHRKTLVRN
PYKDL**PS**RKDRKNRVVAATT**TPS**S**SP**E**PAP**KHVSTKASTVSEPQKRSSTQDV**SPSPSAP**LAN**SP**
IPRNSHSSVPLVVGCVGGAFFLLLVATGLYFFTSKAGKTVNPWRTGLSGQLQKVFVTGIPVLKR
SEIEAACEDFSNVIGSCPIGKLFKGTLSSGVEIAVASFATTTAKDWKDSTEIHFRKKIEMLSKI
NHKNFANLLGYCEEKEPFTRILIFEY**AP**NGSLFEHLHYKESEHLDWGMRLRIAMGLAYCLDHMH
QLNPPIAHTNLVSSSLQLTEDYAVKVSDFSFGSSETETNINNNTVIDTHISALNPEDNIYSFGL
LLFEMITGKLIESVNKPDSVDSSLVDFLRGETLAKMVD**PT**LESYDAKIENIGEVIKSCLRTDPK
ER**PT**MQEVTGWLREITGL**SP**NDA**TP**KL**SP**LWWAELEVLSTA
>AT1G70690.1 | Symbols: HWI1, PDLP5 | Receptor-like protein
kinase-related family protein | chr1:26652099-26653381 FORWARD
LENGTH=299
MIKTKTTSLLCFLLTAVILMN**PS**SS**SPT**DNYIYAVC**SPA**KF**SPS**SGYETNLNSLLSSFVTSTAQ
TRYANFT**VPT**GKPE**PT**VTVYGIYQCRGDLD**PT**ACSTCVSSAVAQVGALCSNSYSGFLQMENCLI
RYDNKSFLGVQDKTLILNKCGQPMEFNDQDALTKASDVIGSLGTGDGSYRTGGNGNVQGVAQCS
GDLSTSQCQDCLSDAIGRLKSDCGMAQGGYVYLSKCYARFSVGGSHARQ**TP**GPNFGHEGEKGNK
DDNGVGKTLAIIIGIVTLIILLVVFLAFVGKCCRKLQDEKWCK
>AT2G31880.1 | Symbols: SOBIR1, EVR | Leucine-rich repeat
protein kinase family protein | chr2:13554920-13556845 FORWARD
LENGTH=641
MAV**PT**GSANLFLRPLILAVLSFLLLSSFVSSVEWLDIDSSDLKALQVIETELGVNSQRSSASDV
NPCGRRGVFCERRHSATTGEYVLRVTRLVYRSRSLTGTI**SP**VIGMLSELKELTLSNNQLVNAVP
VDILSCKQLEVLDLRKNRFSGQIPGNFSSLSRLRILDLSSNKLSGNLNFLKNLRNLENLSVANN
LFSGKIPEQIVSFHNLRFFDFSGNRYLEG**PAP**VMSSIKLQT**SP**HQTRHILAE**TPT**S**SPT**NKPNN
STTSK**AP**KG**AP**KPGKLKKKKKKSKKKKVAAWILGFVVGAIGGTISGFVFSVLFKLIIQAIRGSE
KPPG**PS**IF**SP**LIKKAEDLAFLENEEALASLEIIGRGGCGEVFKAELPGSNGKIIAVKKVIQPPK
DADELTDEDSKFLNKKMRQIRSEINTVGHIRHRNLLPLLAHVSRPECHYLVYEYMEKGSLQDIL
TDVQAGNQELMW**PA**RHKIALGIAAGLEYLHMDHNPRIIHRDLK**PA**NVLLDDDMEARISDFGLAK
AMPDAVTHITTSHVAGTVGYI**AP**EFYQTHKFTDKCDIYSFGVILGILVIGKL**PS**DEFFQHTDEM
SLIKWMRNIITSEN**PS**LAIDPKLMDQGFDEQMLLVLKIACYCTLDDPKQRPNSKDVRTMLSQIK
H
>AT1G34290.1 | Symbols: AtRLP5, RLP5 | receptor like protein 5 |
chr1:12498000-12498800 FORWARD LENGTH=266

MINYRHIVFCLCVMVVVDSRL**TP**YLAAIEQVDPIVKIVLPIVGRFDPEEFVTSWQGNNPCEWFG
TNCLEGIIIGISFISLNLIGTI**SP**HFADLTSLRVIDLSHNRLKCTIPFEITKLKNLTIVDVSYN
QLHGEVPRVRGIVILTERNPNIESTCLL**VPSPT**RNKNK**PT**VLVLLLGILVGLVVAGGASFGFYL
YRIRKQPKRLQEPNEAVTLTQQQSSDESLVSDESYVISLQLQYRVLRRFSWVSKGPLLLTRQLK
TNQNPHLPYM
>AT5G54380.1 | Symbols: THE1 | protein kinase family protein |
chr5:22077313-22079880 REVERSE LENGTH=855
MVFTKSLLVLLWFLSCYTTTTSSALFNPPDNYLISCGSSQNITFQNRIFVPDSLHSSLVLKIGN
SSVATSTTSNNSTNSIYQTARVFSSLASYRFKITSLGRHWIRLHF**SP**INNSTWNLTSASITVVT
EDFVLLNNFSFNNFNGSYIFKEYTVNVTSEFLTLSFI**PS**NNSVVFVNAIEVVS**VP**DNLIPDQAL
ALN**PSTP**FSGLSLLAFETVYRLNMGGPLLTSQNDTLGRQWDNDAEYLHVNSSVLVVTAN**PS**SIK
Y**SPS**VTQET**AP**NMVYATADTMGDANVA**SPS**FNVTWVL**PV**DPDFRYFVRVHFCDIVSQALNTLVF
NLYVNDDLALGSLDLSTLTNGLKVPYFKDFISNGSVESSGVLTVSVGPDSQADITNATMNGLEV
LKISNEAKSLSGVSSVKSLLPGGSGSKSKKKAVIIGSLVGAVTLILLIAVCCYCCLVASRKQRS
T**SP**QEGGNGHPWLPLPLYGLSQTLTKSTASHKSATASCISLASTHLGRCFMFQEIMDATNKFDE
SSLLGVGGFGRVYKGTLEDGTKVAVKRGNPRSEQGMAEFRTEIEMLSKLRHRHLVSLIGYCDER
SEMILVYEYMANGPLRSHLYGADLPPLSWKQRLEICIGAARGLHYLHTGASQSIIHRDVKTTNI
LLDENLVAKVADFGLSKTG**PS**LDQTHVSTAVKGSFGYLDPEYFRRQQLTEKSDVYSFGVVLMEV
LCCR**PA**LNPVLPREQVNIAEWAMAWQKKGLLDQIMDSNLTGKVN**PA**SLKKFGETAEKCLAEYGV
DR**PS**MGDVLWNLEYALQLEETSSALMEPDDNSTNHIPGIPM**AP**MEPFDNSMSIIDRGGVNSGTG
TDDDAEDATTSAVFSQLVHPRGR
>AT3G08680.1 | Symbols: | Leucine-rich repeat protein kinase
family protein | chr3:2638591-2640590 FORWARD LENGTH=640
awkward newlines to ensure no whitespace in output
MMKIIAAFLFLLVTTFVSRCLSADIESDKQALLEFASLVPHSRKLNWNSTIPICASWTGITCSK
NNARVTALRLPGSGLYGPLPEKTFEKLDALRIISLRSNHLQGNI**PS**VILSLPFIRSLYFHENNF
SGTIPPVLSHRLVNLDLSANSLSGNI**PT**SLQNLTQLTDLSLQNNSLSGPIPNLPPRLKYLNLSF
NNLNGSV**PS**SVKSF**PA**SSFQGNSLLCG**AP**L**TP**CPENTT**APSPSPTTPT**EGPGTTNIGRGTAKKV
LSTGAIVGIAVGGSVLLFIILAIITLCCAKKRDGGQDSTAVPKAKPGRSDNKAEEFGSGVQEAE
KNKLVFFEGSSYNFDLEDLLRASAEVLGKGSYGTTYKAILEEGTTVVVKRLKEVAAGKREFEQQ
MEAVGRI**SP**HVNV**AP**LRAYYFSKDEKLLVYDYYQGGNFSMLLHGNNEGGRAALDWETRLRICLE
AARGISHIHSASGAKLLHGNIK**SP**NVLLTQELHVCVSDFGI**AP**LMSHHTLI**PS**RSLGYR**AP**EAI
ETRKHTQKSDVYSFGVLLLEMLTGKAAGKTTGHEEVVDLPKWVQSVVREEWTGEVFDVELIKQQ
HNVEEEMVQMLQIAMACVSKHPDSR**PS**MEEVVNMMEEIR**PS**GSGPGSGNRAS**SP**EMIRSSD**SP**V
>AT1G34300.1 | Symbols: | lectin protein kinase family protein |
chr1:12503450-12505939 FORWARD LENGTH=829
MAVK**TP**FLKLLPLLLLLLHFPFSFSTIPLGSVIYASGSNQNW**PSP**NSTFSVSF**VPSPSP**NSFLA
AVSFAGSVPIWSAGTVDSRGSLRLHTSGSLRLTNGSGTTVWDSKTDRLGVTSGSIEDTGEFILL
NNRSVPVWSSFDN**PT**DTIVQSQNFTAGKILRSGLYSFQLERSGNLTLRWNTSAIYWNHGLNSSF
SSNLS**SP**RLSLQTNGVVSIFESNLLGGAEIVYSGDYGDSNTFRFLKLDDDGNLRIYSSASRNSG
PVNAHWSAVDQCLVYGYCGNFGICSYNDTNPICSC**PS**RNFDFVDVNDRRKGCKRKVELSDCSGN
TTMLDLVHTRLFTYEDDPNSESFFAGS**SP**CRANCLSSVLCLASVSMSDGSGNCWQKHPGSFFTG
YQW**PS**V**PS**TSYVKVCGPVVANTLERATKGDDNNSKVHLWIVAVAVIAGLLGLVAVEIGLWWCCC
RKNPRFGTLSSHYTLLEYASG**AP**VQFTYKELQRCTKSFKEKLGAGGFGTVYRGVLTNRTVVAVK
QLEGIEQGEKQFRMEVATISSTHHLNLVRLIGFCSQGRHRLLVYEFMRNGSLDNFLFTTDSAKF
LTWEYRFNIALGTAKGITYLHEECRDCIVHCDIKPENILVDDNFAAKVSDFGLAKLLNPKDNRY
NMSSVRGTRGYL**AP**EWLANLPITSKSDVYSYGMVLLELVSGKRNFDVSEKTNHKKFSIWAYEEF
EKGNTKAILDTRLSEDQTVDMEQVMRMVKTSFWCIQEQPLQR**PT**MGKVVQMLEGITEIKNPLCP

KTISEVSFSGNSMSTSHASMFVASG**PT**RSSSFSATRSFQTMGITSSG**PA**STRISEGSMLGS
>AT2G42800.1 | Symbols: AtRLP29, RLP29 | receptor like protein
29 | chr2:17808157-17809545 REVERSE LENGTH=462
MTMKRAL**PSPS**SLLFFFLLI**TP**LFLCQENRVSASMP**PS**ESETLFKIMESMSSDQQWRQSHPNPC
APGSSWPGIECKTGPDHLSHVSRLDFGS**AP**N**PS**CKSSASF**PS**SIFTLPFLQSVFFFNCFTHF**PT**
TIMFPIKLIPNSSLQQLSLRSN**PS**LSGQIPPRISSLKSLQILTLSQNRLTGDIP**PA**IFSLKSLV
HLDLSYNKLTGKIPLQLGNLNNLVGLDLSYNSLTGTIP**PT**ISQLGMLQKLDLSSNSLFGRIPEG
VEKLRSLSFMALSNNKLKGAFPKGISNLQSLQYFIMDNNPMFVALPVELGFLPKLQELQLENSG
YSGVIPESYTKLTNLSSLSLANNRLTGEI**PS**GFESLPHVFHLNLSRNLLIGVVPFDSSFLRRLG
KNLDLSGNRGLCLNPEDEFSVVKTGVDVCGKNVSSGGGLSVHSSKKKSQASRYYRSCFFANALF
PFALFLGLHQRWVL
>AT3G36659.1 | Symbols: | Plant invertase/pectin methylesterase
inhibitor superfamily protein | chr3:2100189-2100983 FORWARD
LENGTH=264
MEFNKTYRIIVIFTVSILVTISTHVESRSD**SP**FS**TPAPAP**EVMKPIA**SPA**N**SPA**EIYTD**SPV**A**S**
PE**APA**DSF**SP**KNPPVEIDIY**SPA**A**SP**E**APVASP**E**APA**ESSSA**APS**GIKVSS**SP**KSLLNPPLL**SP**
EIKTICGKTDNPPLCESSV**SP**LL**TP**QLKPNTSSVLILAIQASITATKAAMAIVEKVDASDCQEL
YDDAVVNLEDAVNAVKSSDIATVNTNLSAAMTDYSTCNDGFEESGEPNPLAYVADKLTKMVSNC
LAISTLIK
>AT3G26610.1 | Symbols: | Pectin lyase-like superfamily protein
| chr3:9777025-9780267 FORWARD LENGTH=470
MKTVKSLPILAMLLGVIIVAAAISTVSVEGRKHHVKKIKPKHRRHSKN**TPT**G**SPAPAP**Y**PS**TNE
GVFNIFSYGAKGDGVSDDSKALVGAWKAACKVVGGKVEI**PA**GTQFLVKAVTLQGPCKEETVVQI
EGILV**AP**EKIGSWPNSSLFQWLNFKWVSHLTIQGSGTLNGRGYNWWNLDTYQTQTRNKYIPPMK
PTALRFYSSNNVTVRDISIVN**SP**LCHLKFDDSDGVKVNNITIS**SP**EN**SP**NTDGIHLQNTRNVEI
QHSNIACGDDCVSIQTGSSNVHIHHINCGPGHGISIGGLGKDKSVACVSDIIVEDISIQNTLAG
VRIKTWQGGLGVVKNLTFSNIQVKDVKVPIVIDQYYCDKSKCKNQTRAVSISGVKYNNIVGSFT
VQPVRIACSNNVPCMDVDLMDIRLR**PS**GGIRGLQTHQQQQALCWNSYGKTQGPLV**PS**SIGYCLR
KSNIGGYYSQKVSRSYDKIC**PS**
>AT1G02460.1 | Symbols: | Pectin lyase-like superfamily protein
| chr1:504897-507099 REVERSE LENGTH=491
MLTSTYNRNQVLGFMTLILTLMSLSEARNHHHKEKHKHNNHNHHSSKPKP**PS**SSISQP**PTP**PPG
PPD**SPAPS**LP**PSPS**DD**PA**DDNNGIYNVRKYGAVGDGETDDTEAFKTAWDSSCNNENNTDSVLLV
PYGYTFMIQSTIFTGPCRSYQFFQVDGTIV**TP**DGPESW**PS**NISKRQWLVFYRVNGMALKGEGVI
DGRGQKWWDLPCKPHRSVNKSAIVTGPCD**SP**IALRFFMSSNLRVEGLQIKN**SP**QFHFRFDGCQG
VHVESLHIT**AP**PL**SP**NTDGIHIENSNSVTIYNSIISNGDDCVSIGSGSYDVDIRNLTCGPGGHG
ISIGSLGNHNSRACVSNITVRDSVIKYSDNGVRIKTWQGGSGSVSGVTFNNIHVDSVRNPIIID
QYYCMTKDCANKTSAVFVSDIAYQGIKGTYDIR**SP**PMHFGCSDAVPCTNLTLSDIELL**PA**KGEI
VLDPFCWNAYGIAEELSIPPVWCLMSDPPKGLQGSLVDKCGSS
>AT4G20040.1 | Symbols: | Pectin lyase-like superfamily protein
| chr4:10847348-10848963 REVERSE LENGTH=483
MSLYSFISLLLLFLITISMEFKETSSLRRDMTKLLKLQEKIQERLAV**TPT**L**SPV**S**SPS**SH**SP**KM
VGKVIYPIGYGAD**PT**GGQDSSDAILEALTDAFQLQTGLEMLPRVADLGGLVIDLQGGSYMIGKP
LRF**PS**SGGGNLVVKGGTFRASELFPGDRHLVELVASNAKKPMKM**SP**EESFSDQKDQSSGIFYED
VTFQDVLFDSRFRGGGILVIDSARIRITNCYFLHFTTQGIKVQGGHETYISNSFLGQHSTVGGD
REERGFTGTGIDISSNDNAITDVVIFSAGIGISLNGGANMVTGVHCYNKATWFGGIGILVKSHL
TRIDNCYLDYTGIVIEDPVHVHVTNALFLGDANIVLRSVHGKISGVNIVNNMFSGTAKNNFPIV

KLEGEFHDINQVVIDQNNAEGMMLKSTTGKAMVSANGTRWIADF**SPV**LVFPNRINHYQHSFFAQ
SGQI**PA**NAVTNVSNNMVVVETDRAVTGTVSIIAYQ
>AT3G44990.1 | Symbols: XTR8, ATXTR8, XTH31 | xyloglucan endo-transglycosylase-related 8 | chr3:16447280-16448678 REVERSE LENGTH=293
MALSLIFLALLVLC**PS**SGHSQR**SPSP**GYY**PS**SRV**PTSP**FDREFRTLWGSQHQRREQDVVTLWLD
KSTGSGFKSLRPYRSGYFGASIKLQPGFTAGVDTSLYLSNNQEHPGDHDEVDIEFLGT**TP**GKPY
SLQTNVFVRGSGDRNVIGREMKFTLWFD**PT**QDFHHYAILWNPNQIVFFVDDVPIRTYNRKNEAI
F**PT**RPMWVYGSIWDASDWATENGRIKADYRYQPFVAKYKNFKLAGCTADSSSSCRP**PSPAP**MRN
RGLSRQQMAALTWAQRNFLVYNYCHDPKRDHTQ**TP**EC
>AT1G69295.1 | Symbols: PDCB4 | plasmodesmata callose-binding protein 4 | chr1:26050492-26051843 REVERSE LENGTH=222
MSVLLPLCLIISMFTYSNAAYCLCKEGNEQVLQKAIDYACGNGADCTQIQ**PT**GACYQPNTVKNH
CDVAVNSYYQKKASSGATCDFNGAA**SPS**T**TPPS**TASNCLTGSSSSG**TPT**TG**TPT**TG**TPT**SG**TPT**
TG**TPT**TG**TPT**TG**TPT**SG**TPT**SGFPNTG**TP**NTGTNTGMPNSNGM**PT**SSSSSVFPGTTLG**PT**GSGG
LGDPNAGEKLSVRTNTVVFLLTGVAAMLVI
>AT1G09460.1 | Symbols: | Carbohydrate-binding X8 domain superfamily protein | chr1:3053901-3055090 FORWARD LENGTH=330
MAKQSYFFVFTFLFLSLLSYCSSTTTHHDVLNP**PT**VF**PT**N**PT**T**TPT**ATFPPVTI**TPT**N**PA**TTVP
IVPPVTTIPP**PT**L**TP**PPVITIPP**PT**L**TP**PVTNPVTNPVTQYP**PT**Q**PS**GT**VPVPVPV**V**AP**PVVSN
SPSVSGQSWCVAKPGASQVSLQQALDYACGIADCSQLQQGGNCY**SP**ISLQSHASFAFNSYYQKN
PSPQSCDFGGAASLVNTN**PS**TGSCIYQTGSSTS**TP**MTAGTT**TPTPS**TQTVNQPPVTS**TP**II**PT**G
GGIIGVG**TPPA**IFN**PA**N**PT**SNTLNN**PS**SGGLAGGYGFDG**SP**NENN**PT**SSDSTHLKIHFGHAMVA
TLILHAVLFH
>AT1G13830.1 | Symbols: | Carbohydrate-binding X8 domain superfamily protein | chr1:4739999-4740926 REVERSE LENGTH=197
MKVLLGLLLLLSLTNSSSAIYCLCKDGIGDTELQTSIDYACGTLADCNPIHDKGTCYQPDTIKS
HCDWAVNSYFQNAAQVPGSCNFSGTATTNPNP**PS**NLANGCIY**PSSPS**STR**SPPS**T**TPPT**GT**TPT**
NGT**TP**FPG**TP**FPG**TP**FPG**TP**PVFG**PT**GVFN**PS**NPGSGASSLGTSSVFTLCFSLLAFLWGSDVRF
GFSHV
>AT1G18650.1 | Symbols: PDCB3 | plasmodesmata callose-binding protein 3 | chr1:6419036-6420413 REVERSE LENGTH=184
MAVFVLVMILLAMAGHSSGTWCVCKEGLSEAMLQKTLDYACGAGADCGPIHQTGPCFNPNTVKS
HCSYAVNSFFQKKGQSLGTCDFAGTATFSASD**PS**YTTCPF**PA**SASGSGTT**TP**VTT**TPS**TR**VPT**T
TNTRPYTI**TPS**TGGGLGI**PS**GINPDYTD**PS**FGFKLQ**SP**RFGFIVLFTLFLPFYLFS
>AT3G47400.1 | Symbols: | Plant invertase/pectin methylesterase inhibitor superfamily | chr3:17465629-17467888 FORWARD LENGTH=594
MLRGIFHICLLASFLLLPFSSAVHDSGFTGGTD**AP**PPWDHNV**SP**PPET**APSPTPT**S**SPS**TT**SPP**
SPGPVA**APSP**INNGSVSGDMTWWCNK**TP**HAETCNYYFRKSSQNNINLRPPRFRSEFLRMLVKVA
LDQAVITHSQTVKFG**PS**CTNNQRKAAWSDCVNLFQNTVAQLNRTLKGLN**PA**ASSDVKCTDFDAQ
TWLSTAQTNIETCRSGSEDLNVSDFVMPVISNKNLSDLIGNCLAVNGVLMKQHDHTTTANHKEY
F**PS**WVSRHERRLLVSASLAKS**SP**HLVVAQDRSGHFRSIQAAINFAARRRFKSRFVIYVKKGVYR
ENIDVGNDNHNIMLVGDGERKTIITSGRSVQHGYTTYNSATGGFGGQRFVAKDMTFINTAGPLR
GQAVAVRSSSDLSVFYRVGIHGFQDTLYIHSQRQFFRECYISGTIDFIFGNAAVVFQNCMILVR
RPLHGQANIITAQGRGDPFQNTGITIHSSRIIAASDLKPVIRAYKTYLGRPWQAYSRVTIMKTY
IDNSI**SP**LGW**SP**WLRGSNFALNTVFYGEYKNFGPGSSTRWRVRWKGFHAITSTAVASRFTVGSL
IAGGSWL**PA**TGVPFKSGL

>AT5G21105.1 | Symbols: | Plant L-ascorbate oxidase | chr5:7172727-7177409 FORWARD LENGTH=588
MSYDEHTSSSFTYISQMGVWWIVLVVAVLTHTASAAVREYHWEVEYKYW**SP**DCKEGAVMTVNGE
FPG**PT**IKAFAGDTIVVNLTNKLTTEGLVIHWHGIRQFG**SP**WADGAAGVTQCAINPGETFTYNFT
VEKPGTHFYHGHYGMQRSAGLYGSLIVDVAKGKSERLRYDGEFNLLLSDWWHEAI**PS**QELGLSS
KPMRWIGEAQSILINGRGQFNCSLAAQFSNNTSLPMCTFKEGDQC**AP**QILHVEPNKTYRIRLSS
TTALASLNLAVQGHKLVVVEADGNYI**TP**FTTDDIDIYSGESYSVLLTTDQD**PS**QNYYISVGVRG
RKPNTTQALTILNYVT**APA**SKL**PSSP**PPV**TP**RWDDFERSKNFSKKIFSAMG**SPSP**PKKYRKRLI
LLNTQNLIDGYTKWAINNVSLV**TPATP**YLGSVKYNLKLGFNRK**SP**PRSYRMDYDIMNPPPFPNT
TTGNGIYVFPFNVTVDVIIQNANVLKGIVSEIHPWHLHGHDFWVLGYGDGKFKPGIDEKTYNLK
NPPLRNTAILYPYGWTAIRFVTDNPGVWFFHCHIEPHLHMGMGVVFAEGLNRIGKVPDEALGCG
LTKQFLMNRNRN
>AT1G02550.1 | Symbols: | Plant invertase/pectin methylesterase inhibitor superfamily protein | chr1:536483-537211 FORWARD LENGTH=242
MKKS**SP**LLCFSLALFSLLS**SPS**SSTRIISSI**VPS**A**APSPA**V**APT**TDGDVDENDFSAFTQWNILN
LTDLKSTFKNLPDFSKLNISSLHV**SPA**VGSVCSNTDYAAECIVSILPLLRDFRKFEPKPIDVLR
MEMSALYEKANATLDLAKRLIVDKS**TP**RDVADVLDLCVDNYESLLDDLKDASVAVDDGDFERLE
SVVSAAIADVVTCSDAFAESSELE**SP**MANVDDFLKKLCSNVLAISQMIHM
>AT3G55260.1 | Symbols: HEXO1, ATHEX2 | beta-hexosaminidase 1 | chr3:20489317-20492858 FORWARD LENGTH=541
MSTNLLRLILLFITLSITSSLS**TPSPA**D**SP**PYLWPL**PA**EFSFGNETLSVD**PT**VTLIVAGNGGGS
LIIRAAFDRYMGIIFKHASGRGSLLSRIRFLKMVEYDITSLKIVVHSDSEELQLGVDESYTLMV
SKKNEQSIVGAATIEANTVYGALRGLETFSQLCAFDYITKSVQIYK**AP**WYIQDKPRFGYRGLLI
DTSRHYLPIDVIKQIIESMSFAKLNVLHWHIVDEQSFPLE**TPT**YPNLWKGAYSRWERYTVEDAS
EIVRFAKMRGINVMAEVDVPGHAESWGTGYPDLW**PS**LSCREPLDVTKNFTFDVISGILADMRKI
FPFELFHLGGDEVNTDCWKNTTHVKEWLQGRNFTTKDAYKYFVLRAQQIAISKNW**TP**VNWEETF
SSFGKDLDPRTVIQNWLVSDICQKAVAKGFRCIFSNQGYWYLDHLDVPWEEVYNTEPLNGIED**P**
SLQKLVIGGEVCMWGETADTSVVLQTIWPRAAAAAERMWSTREAVSKGNITLTALPRLHYFRCL
LNNRGV**PAAP**VDNFYARRPPLGPGSCYAQ
>AT1G66250.1 | Symbols: | O-Glycosyl hydrolases family 17 protein | chr1:24693063-24695416 FORWARD LENGTH=505
MASLLHLLLLSLSLLVLASA**SPSPPA**DEGSYIGVNIGTDLSDMPH**PT**QVVALLKAQEIRHIRLY
NADPGLLIALANTGIKVIISIPNDQLLGIGQSNSTAANWVKRNVIAHY**PA**TMITAVSVGSEVLT
SLSNA**AP**VLVSAIKNVHAALLSANLDKLIKVS**TP**LSTSLILDPFP**PS**QAFFNRSLNAVIVPLLS
FLQSTNSYLMVNVYPYIDYMQSNGVIPLDYALFKPIPPNKEAVDANTLVRYSNAFDAMVDATYF
AMAFLNFTNIPVLVTESGW**PS**KGETNEPDATLDNANTYNSNLIRHVLNKTG**TP**KRPGIAVSTYI
YELYNEDTKAGLSEKNWGLFNANGEPVYVLRLTNSGSVLANDTTNQTYCTAREGADTKMLQAAL
DWACGPGKIDC**SP**IKQGETCYEPDNVVAHANYAFDTYYHQTGNNPDACNFNGVASITTTD**PS**HG
TCVFAGSRGNGRNGTSVNIT**APS**ANSTTSSGIRSDLYYSRGIWSILTVMILNVANIL
>AT1G02360.1 | Symbols: | Chitinase family protein | chr1:472138-473116 REVERSE LENGTH=272
MAQQHSFLLLCFFLSISYLLSSAQTEATSIERLVPRDLYNKIFIHKDNTAC**PA**NGFYTYESFVQ
ATRRFPRFGSVG**SP**VTQRLEVAAFLAQISHETTGGWAT**AP**DGPYAWGLCFKEEV**SP**QSTYCDSS
DTQWPCFPNKTYQGRGPIQLSWNYNYG**PA**GRALGFDGLRNPETVSNNSVIAFQTALWFWM**TP**Q**S**
PK**PS**CHDVMIGKYR**PT**AADLAANRTGGFGLTTNIINGGLECGIPGDGRVNDRIGFFQRYTGLFK
VATGPNLDCENQRPYA
>AT4G26010.1 | Symbols: | Peroxidase superfamily protein |

chr4:13200653-13201688 FORWARD LENGTH=310
MRSITALFFLFCFL**APS**ALAQLRTGFYSRSCPRAESIVASVVANRFRSDKSITAAFLRMQFHDC
FVRGCDASLLIDPRPGR**PS**EKSTGPNASVRGYEIIDEAKRQLEAACPRTVSCADIVTLATRDSV
ALAGGPRFS**VPT**GRRDGLRSNPNDVNLPG**PT**IPVSASIQLFAAQGMNTNDMVTLIGGGHSVGVA
HCSLFQDRLSDRAME**PS**LKSSLRRKCS**SP**ND**PT**TFLDQKTSFTVDNAIYGEIRRQRGILRIDQN
LGLDRSTSGIVSGYASSNTLFRKRFAEALVKMGTIKVLTGRSGEIRRNCRVFNN
>AT1G56710.1 | Symbols: | Pectin lyase-like superfamily protein
| chr1:21258206-21259922 REVERSE LENGTH=434
MSLPNSLFYTFFLILFSTIRIAQSIYL**SPSPAP**N**PA**YNDNDNI**APT**VFDVTSFGAIGDCSTDDT
SAFKMAWDAACMSTGPKSALLLVPYTFCFLVK**PT**TFNGPCRTNLVLQIDGFIV**SP**DGPRSW**PS**N
YQRQWMMFYRVNGLSIQGSGVINGRGQKWWNLPCKPHKGLNGTTQTGPCD**SP**VAIRLFQSSKVR
IQGINFMNSAQFHVRFDNCSDVVVDSVIIK**APA**S**SP**NTDGIHIENTHNVQIRNSMISNGDDCIS
IGAGCFNVDIKNVTCG**PS**HGISIGSLGVHNSQAYVSNITVTNSTIWNSDNGVRIKTWQGGSGSV
SRIVFSNILMVNVRNPIMIDQYYCQTNNCANQTSAVIISDVLYANIKGTYDLR**SP**PIHFGCSDS
VPCTNLTLTEVDLF**PS**KGQHLENPFCWNAYGSMKITVPPVYCLLD**AP**PDF
>AT4G29360.1 | Symbols: | O-Glycosyl hydrolases family 17
protein | chr4:14451562-14453687 REVERSE LENGTH=534
MGQRLNLVFWIFVSILAFLNFGMASKIGICYGRNADNL**PSP**NRVSELIQHLNIKFVRIYDANID
VLKAFANTGIELMIGVPNADLLAFAQFQSNVDTWLSNNILPYY**PS**TKITSISVGLEVTE**AP**DNA
TGLVL**PA**MRNIHTALKKSGLDKKIKISSSHSLAILSRSFP**PS**SASFSKKHSAFLKPMLEFLVEN
E**SP**FMIDLYPYYAYRDSTEKVPLEYALFESSSQVVD**PA**TGLLYSNMFDAQLDAIYFALTAMSFK
TVKVMVTESGW**PS**KG**SP**KETAA**TP**ENALAYNTNLIRHVIGDPG**TPA**KPGEEIDVYLFSLFNENR
KPGIESERNWGMFYANGTNVYALDFTGENT**TP**V**SPT**NSTTGT**SPSPS**S**SP**IINGNSTVTIGGGG
GGGTKKWCIASSQASVTELQTALDWACGPGNVDCSAVQPDQPCFEPDTVLSHASYAFNTYYQQS
GASSIDCSFNGASVEVDKD**PS**YGNCLYMI**APA**TDGFNRTMAGNITGNITAID**SP**LA**SPS**STNEA
FRQMVVAVSVLLPCFVVCSSIW
>AT2G03505.1 | Symbols: | Carbohydrate-binding X8 domain
superfamily protein | chr2:1064317-1065094 FORWARD LENGTH=168
MRLLLSLLFLLALTTYSSATYCLCRDGVGEKDLQTSIDYACGVLKDCNPIHEKGPCYQPNTIKS
HCDWAVNTYFQRFGQISGSCNFSGTATTSQNL**PS**TVVTGCLY**PSSP**GSAGT**TPT**TG**TPS**GTQTF
PGP**PA**FG**PA**GDFD**PS**GNNG**APS**LFISIALSLGFSVVIAFL
>AT2G30933.1 | Symbols: | Carbohydrate-binding X8 domain
superfamily protein | chr2:13164179-13166517 REVERSE LENGTH=227
MDKGFRLVTSLLLLSFFFSGTKAITEPIEEEKDIT**TP**LATN**PT**T**TPT**TVVPNSDSDASAVAT**TP**
LTI**PSSP**HGVAYPGDQSWCVARENVAKMALQAALDYACGIGGADCSEIQEGGNCYNPNSLRAHA
SFAFNSYYQKNPI**PS**SCNFDGTAITISAD**PS**LGSCHF**PS**TSTSESILNVTSEDGLGLFGRI**PS**H
PTPKPEASTSSSRTLISFLYFLCFCVLYPLVITLT
>AT1G79480.1 | Symbols: | Carbohydrate-binding X8 domain
superfamily protein | chr1:29897905-29899267 REVERSE LENGTH=397
MESYRVFTTICLLLCLFLSANFFTHYVVDARKSVGFEREPKKVMMIKALKHTSLLQKMMTQLNL
AQPLDYSSSSNTQPYGVSTTLTLPPYVSLPPLSVPGN**AP**PFCINPPN**TPPS**SSYPGL**SP**PPGPI
TLPNPPDSSSNPNSNPNPPESSSNPNPPDSSSNPNSNPNPPVTVPNPPESSSNPNPPDSSSNPN
SNPNPPESSSNPNPPVTVPNPPESSSNPNPPESSSNPNPPITIPYPPESS**SP**NPPEIV**PSP**PES
GY**TP**GPVLGPPYSEPG**PSTPT**GSI**PSPS**SGFLPPIVYPPPM**APPSPS**V**TPT**SAYWCVAK**PS**VPD
PIIQEAMNFACGSGADCHSIQPNGPCFKPNTLWAHASFAYNSYWQRTKSTGGSCTFGGTGMLVT
VD**PS**FNGCHFDFF
>AT1G26450.1 | Symbols: | Carbohydrate-binding X8 domain
superfamily protein | chr1:9149003-9150063 REVERSE LENGTH=197

MAVLLPLFLLSFMFTYSNAAVCVCKDANELDLQKVIDFACGGGADCAQIQTTGACYQPNTLKNH
CDVAVNSYYQKKASTGATCDFNGAAVIST**SPPS**TTSSCLSSSSSNG**TPT**AGY**PS**TGNSTTA**SP**G
TTN**PS**TGNSTNSTL**PT**NDK**PT**SSTITFPDSTTMG**PS**SSTSGDPNGGEELSVRTTTIILLTTIAA
VALRV
>AT1G29380.1 | Symbols: | Carbohydrate-binding X8 domain
superfamily protein | chr1:10284209-10285671 FORWARD LENGTH=315
MTHRI**AP**RCFSTIMLLCLLYT**SP**LDVIAQGQGGQGDIPVVN**PTAP**GGSTT**TPT**ITQ**PSPPS**LTF
PG**PTTPT**GGYPPLDGT**TPT**GGYPPLYGT**TP**PGGGDVGGGGGGYGGG**TP**GGGGGGGGDTGAGAGG
GGYGGGGDTGAGGGVGSGQWCIAKANA**SPT**SLQVALDYACGYGGADCGQIQQGAACYEPNTIRD
HASFAFNSYYQKHPGSDSCNFGGAAQLTSTD**PS**KGSCHFSSSSGTVST**SPPS**QM**SP**PDFS**SPPS**
STYPPPIT**TPT**TGITGSGPPFGVAE**PT**GLPNSATSVSHSLLSIFTAVGILMPLLRENYL
>AT1G78820.1 | Symbols: | D-mannose binding lectin protein with
Apple-like carbohydrate-binding domain | chr1:29634401-29635768
REVERSE LENGTH=455
MLRFDYLLITALAISTVSVVMAQVPPEKQFRVLNEPGY**AP**YITEYDASYRFLN**SP**NQNFFTIPF
QLMFYNT**TPS**AYVLALRVGTRRDMSFTRWIWDANRNNPVGDNSTLSFGRNGNLVLAELNGQVKW
QTNTANKGVTGFQILPNGNMVLHDKHGKFVWQSFDH**PT**DTLLVGQSLKVNGVNKLVSRTSDMNG
SDGPYSMVLDNKGLTMYVNKTG**TP**LVYGGWTDHDFRGTVTFAVTREFDNLTE**PS**AYELLLE**PAP**
Q**PA**TNPGNNRRLLQVRPIGSGGGTLNLNKINYNGTISYLRLGSDGSLKAFSYF**PA**ATYLEWEET
FAFFSNYFVRQCGL**PT**FCGDYGYCDRGMCVGC**PTP**KGLLAWSDKC**AP**PKTTQFCSGGKGKAVNY
YKIVGVEHFTGPYVNDGQG**PT**SVNDCKAKCDRDCKCLGYFYKEKDKKCLL**AP**LLGTLIKDANTS
SVAYIKY
>AT5G61130.1 | Symbols: PDCB1 | plasmodesmata callose-binding
protein 1 | chr5:24587438-24589369 REVERSE LENGTH=201
MAALVLSLLLLSLAGHSSASWCVCKTGLSDTVLQATLDYACGNGADCN**PT**KPKQSCFNPDNVRS
HCNYAVNSFFQKKGQ**SP**GSCNFDGTA**TPT**NSD**PS**YTGCAF**PT**SASGSSGSTTV**TP**GTTNPKG**SP**
TTTTLPGSGTN**SP**YSGN**PT**NGVFGGNSTGGTTGTGINPDYTTDSSAFALKNSSKLFICLLLIAS
SGFCSFLML
>AT4G19410.1 | Symbols: | Pectinacetylesterase family protein |
chr4:10582188-10584766 REVERSE LENGTH=391
MGRLKQCWSSLLVLAVLVIGTGAVPITYLQSAVAKGAVCLDGS**APA**YHFDKGFGSGVNNWIVHM
EGGGWCTDVASCNERKGTMKGSSKFMNKDFGFSGILGGKQSTNPDFYNWNRIKVRYCDGSSFTG
NVEAVN**PA**NKLFFRGARVWRAVVDDLMAKGMKNAQNAILSGCSAGALAAILHCDTFRAILPRTA
SVKCVSDAGYFIHGKDITGGSYIQSYYSKVVALHGSAKSLPVSCTSKMKPELCFFPQYVV**PS**MR
TPLFVINAAFDSWQIKNVL**APT**AVDKGKEWKNCKLDLKKCSAAQLKTVQGFRDQMMRAL**SPV**HS
TPSRGLFLDSCHAHCQGGSAASWSGDKGPQVANTRIAKAVGNWFYGRSAFQKIDC**PSPT**CN**PT**C
PAISTED
>AT4G01890.1 | Symbols: | Pectin lyase-like superfamily protein
| chr4:816210-818428 FORWARD LENGTH=468
MLKLSRDPILCITTLILIITFSLLSYGTEARLHHQASQP**PSPSP**NPND**PS**K**SPS**RSQDLDHEVV
YDVRKYGAVGNGVADDTVSFKTAWDSACSNNKNNTASVLHVPYGFTFMIRSTIFTGPCRSYQYF
QVDGTIVPRDGPKSW**PS**SLNKRQWLAFYRINGMALQGAGVIDGRGQNWWDLPCKPHQQNVNKTK
LAGPCE**SPA**ALRFFMSSNVIVKGLSIKN**SP**QVHLKLDGCHVVHINSLRII**SPPASP**NTDGIHIE
NSNSVEIYNSVISNGDDCVSIGPGAYDIDIRNITCGPGGHGISIGSLGEKNSHACVSNVTVRDS
FIKFSENGVRIKTWQGGSGSVSGVTFDNIHVDTVRNPIIIDQYYCTTKSCANKTSAVFVNDIVY
QSIKGTYDIR**SP**PMHFGCSNNVPCTNLTLSNIELL**PS**KEDIVVGPFCWNAYGITDEFSVPLISC
LKSN**PS**TSLLSGLSGRCG**SP**
>AT3G54920.1 | Symbols: PMR6 | Pectin lyase-like superfamily

protein | chr3:20345311-20348477 FORWARD LENGTH=501
MLLQNFSNTIFLLCLFFTLLSATKPLNLTLPHQH**PSP**DSVALHVIRSVNESLARRQLS**SPS**SSS
SSSSSSSSSSCRTGNPIDDCWRCSDADWSTNRQRLADCSIGFGHGTLGGKNGKIYVVTDSSDNN
PTN**PTP**GTLRYGVIQEEPLWIVFSSNMLIRLKQELIINSYKTLDGRGSAVHITGNGCLTLQYVQ
HIIIHNLHIYDCK**PS**AGFEKRGRSDGDGISIFGSQKIWVDHCSMSHCTDGLIDAVMGSTAITIS
NNYFTHHDEVMLLGHDDNY**AP**DTGMQVTIAFNHFGQGLVQRMPRCRRGYIHVVNNDFTEWKMYA
IGGSGN**PT**INSQGNRYS**APS**D**PS**AKEVTKRVDSKDDGEWSNWNWRTEGDLMENGAFFVASGEGM
SSMYSKASSVDPKAASLVDQLTRNAGVFGGPRDDQGQSGNSY**SP**YGGDGGGGGSSGGSSGGGMD
VMGGTTRGSSSSSGDDSNVFQMIFGSD**APS**RPRLTLLFSLLMISVLSLSTLLL
>AT5G48140.1 | Symbols: | Pectin lyase-like superfamily protein
| chr5:19518907-19520412 REVERSE LENGTH=395
MGRVHFGVSAFFVFCLLGLSANAKIFNIN**SP**PGSDITNALLKAFNEACQF**PT**KSTVMIPKGEYK
LGEIVMMGPCK**AP**IRIALLGTVKADGNANGKEKWVAFRNINGFKLNGGGVFDGEGNAAWRVNNC
HKTFNCKKLPISIRFDFVTDAKIRGITSLDAKHFHINVIGAKNVTFEDVKII**APAESP**NTDGIH
VGRSDGIKIINSFISTGDDCVSVGDGMKNLLVERVTCGPGHGISIGSLGRYSHEENVSGIKIIN
CTLQETDNGLRIKTW**PS**AACTTTASDIHFENILLKNVSNPILIDQEYCPWNQCNKQK**PS**TIKLA
NISFKKIRGTSGNKDAVKLLCSKGYPCQNVEVGDVNIQYTGADG**PA**TFQCSNV**SP**KLVGTQIPK
ACS**SP**VTKPPK
>AT5G01050.1 | Symbols: | Laccase/Diphenol oxidase family
protein | chr5:18209-20812 REVERSE LENGTH=586
MPRVHHSLSNQAFLVLLLFSSIASAAIVEHVLHVKDVVV**TP**LCKEQMIPIVNGSLPG**PT**INVRE
GDTLVVHVINKSTYNVTIHWHGVFQLKSVWMDGANMITQCPIQ**PS**NNFTYQFDITGQEGTLLWH
AHVVNLRATIHGALIIRPRSGRPYPFPKPYKEVPLIFQQWWDTDVRLLELR**PAPV**SDAYLINGL
AGDSYPCSKNRMFNLKVVQGKTYLLRIINAALNTHLFFKIANHNVTVVAVDAVYT**TP**YLTDVMI
L**TP**GQTIDAILTADQPIGTYYMAIIPYFSAIGV**PASP**DTK**PT**RGLIVYEGATSSS**SPT**KPWMP**P**
ANDI**PT**AHRFSSNITSLVGGPHW**TPVP**RHVDEKMFITMGLGLDPC**PS**NAKCVGPLDQRLAGSLN
NRTFMIPERISMQEAYFYNITGVYTDDFPDQPPLKFDFTKFEQH**PT**NSDMEMMFPERKTSVKTI
RFNSTVEIVLQNTGIL**TP**ESHPMHLHGFNFYVLGYGFGNYDPIRDARKLNLFNPQMHNTVGVPP
GGWVVLRFIANNPGIWLFHCHMDAHLPLGIMMAFIVQNG**PT**RETSL**PSPPS**NLPQCTRD**PT**IYD
SRTTNVDMSY
>AT3G07830.1 | Symbols: | Pectin lyase-like superfamily protein
| chr3:2499454-2500966 REVERSE LENGTH=397
MGSYFGISTIFVICLLGFSANAHEVLRLTS**SP**GSDITQALLRAFTTACQ**SPTP**RKVVIPKGQFK
LGEIMMSGPCK**SPV**EITLLGTVLADGNSIHGKEKWVVFQRMDGFRLNGGGTFDGEGNAAWRVNN
CHKTFECKKLPISIRFDFVTNAEIRDISSIDAKNFHINVIGAKNMTFDNVKIM**APAESP**NTDGI
HLGRSVGVSIINSRISTGDDCVSVGDGMVNLLVKNVVCGPGHGISVGSLGRYGHEQDVSGIRVI
NCTLQETDNGLRIKTW**PS**AACSTTASNIHFENIILRNVSNPILIDQEYCPWNQCNKQKSSSIKL
ANISFRRIRGTSGNKDAVKLLCSKGYPCENVQVGDINIQYTGADG**PA**TFMCSNVRPKLVGTQFP
KACN**TPPV**LTQPK
>AT2G16230.1 | Symbols: | O-Glycosyl hydrolases family 17
protein | chr2:7035463-7038326 REVERSE LENGTH=503
MALSILFLLLFILFSI**SPS**NAQSFIGVNYGLLSDNLPP**PS**QTAKLLQSTSIQKVRLYNADSSII
TSLVGTGIGIVIGVANGDL**PS**IASDLNIASQWINSNVLPFY**PA**SNIILINVGNEVLLSNDLNLV
NQLL**PA**MQNVQKALEAVSLGGKIKVSTVHAMTVLGNSEP**PS**AGSF**APS**YQAGLKGILQFLSDTG
SPFAINPYPFFAYQSDPRPETLAFCLFQPNPGRVDSNTGIKYMNMFDAQVDAVHSALKSIGFEK
VEVLVAETGW**PS**TGDSNEVG**PS**VENAKAYNGNLIAHLRSMVG**TP**LMPGKSIDTYIFALFDENLK
PG**PS**FEQSFGLFKPDLSMAYDIGLTKTTSSQTSQ**SP**QLGKVTSMGWCVPKEDATQEQLQDSLDW

VCGQGIDCGPIMPGGVCFEPNNVASHTAYAMNLYFQK**SP**EN**PT**DCDFSKTARITSEN**PS**YSSCV
YPRAGDGSITGEVTKYVTSDKATEKNGSECFSSLYLARFIISIYFFCLF**PS**LRIM
>AT3G07820.1 | Symbols: | Pectin lyase-like superfamily protein | chr3:2496597-2498028 REVERSE LENGTH=391
MGGYFGVSTIFIICLLGISANAEVFTIGSSSGSDITQALLKAFTSACQSS**SPS**KVVIPKGEFKL
GEIEMRGPCK**AP**IEVTLQGTVKADGNAIQGKEKWVVFGNIDGFKLNGGGAFDGEGNAAWRVNNC
HKTFECKKLPISIRFDFILNSEIRDISSIDAKNFHINVLGAKNMTMNNIKIV**AP**ED**SP**NTDGIH
LGRSDGVKILNSFISTGDDCISVGDGMKNLHVEKVTCGPGHGISVGSLGRYGHEQDVSGIKVIN
CTLQETDNGLRIKTW**PS**AACSTTASDIHFEDIILKDVSNPILIDQEYCPWNQCNKQKASTIKLV
NISFKNIRGTSGNKDAVKLLCSKGYPCQNVEIGDIDIKYNGADG**PA**TFHCSNV**SP**KILGSQ**SP**K
ACS**APA**A
>AT4G32380.1 | Symbols: | Pectin lyase-like superfamily protein | chr4:15633641-15635420 FORWARD LENGTH=354
MANDWGRNKLDCWLCFEKVTGLVLTGSGVLNTHGESWWSSVALQSRPVAVRFFGCQNILYNGLT
QIN**SP**RNHITILDSNNATLSNLHLI**APA**S**SP**NTDGIDISHSQNINIMSSTIKTGDDCVAIKRNS
YNINVTYVTCGPGHGISIGSLGEGGASEVVQNVNVRHCTFTGTQNGARIKTWPGGQGFVKNILY
EDITLINANFPIIIDQQYRDNAGQYKQSAGATAVKVSDVTFRSFTGTCA**AP**IAIKLDCDPNTGC
DNIVMEQINIASS**SP**K**TP**LTSYCKFAHVVSRFVSIPITCSFHTEDSQ**PA**SLNPQ**PSAP**YAI**SPT**
TPHTQPH**APT**TQPPLFFRFYTNFKAFLGRLGRNC
>AT1G09790.1 | Symbols: COBL6 | COBRA-like protein 6 precursor | chr1:3168568-3170819 REVERSE LENGTH=454
MGAMLNLLLVVTVILCSIL**SPT**RFMIMIDKMVADGYDPLDPFGKIIIKWDLLLS**SP**GQHHVQVT
LENMQEYRHVEKPGWKLSWHWLNQEVIWDMKGAETTEQGNCSAFASSGNLPHCCLER**PT**IVDLL
PGASLNVQVANCCRGGVLTSMSQDHANHVSAFHMTVGS**SP**DGPEEFNM**PS**NFDIGVPGYSCDNA
TSV**SPT**KFSTDKGRRKTQALATWEAVCVYSQFRS**SPSP**KCCVSLSAFYYQNIVPC**PT**CSCGCSS
SHCVKDGELPPYLEQKHDPDEEV**SPV**VKCSDHMCPIRIHWHVKVNYREYWRVKITATNFNTMKN
YTNWNLVVLHPNLKSVQQVFSFNYKSL**TP**YQNSINDTGMFWGVQFYNDVLLQEGKIGNVQTELL
LKKDMGNFTFREGWAFPRRILFNGDECVM**PSP**DDFPRLPKSAHSSSSSSAVISSVSVVFCFLLH
HLLLLV
>AT1G69940.1 | Symbols: ATPPME1, PPME1 | Pectin lyase-like superfamily protein | chr1:26343549-26344971 REVERSE LENGTH=361
MGYTNVSILLGLLMVFV**TP**MVFADDV**TP**IPEGKPQVAQWFNANVGPLAQRKGLD**PA**LVAAEA**AP**
RIINVNPKGGEFKTLTDAIKS**VPA**GNTKRVIIKM**AP**GEYKEKVTIDRNKPFITLMGQPNAMPVI
TYDGTAAKYGTVDSASLIILSDYFMAVNIVVKNT**APAP**DGKTKGAQALSMRISGNFAAFYNCKF
YGFQDTICDDTGNHFFKDCYVEGTFDFIFGSGTSMYLGTQLHVVGDGIRVIAAHAGKSAEEKSG
YSFVHCKVTGTGGGIYLGRAWMSHPKVVYAYTEMTSVVN**PT**GWQENK**TPA**HDKTVFYGEYKCSG
PGSHKAKRVPFTQDIDDKEANRFLSLGYIQGSKWLLPP**PA**L
>AT5G07410.1 | Symbols: | Pectin lyase-like superfamily protein | chr5:2345852-2347276 FORWARD LENGTH=361
MRYTNVSILLGMLVIFV**SP**MVFADDL**TP**IPEGKPQVVQWFNTHVGPLAQRKGLD**PA**LVAAEA**AP**
RIINVNPKGGEFKTLTDAIKS**VPA**GNTKRVIIKMAHGEYREKVTIDRNKPFITLMGQPNAMPVI
TYDGTAAKYGTVDSASLIILSDYFMAVNIVVKNT**APAP**DGKTKGAQALSMRISGNFAAFYNCKF
YGFQDTICDDTGNHFFKDCYVEGTFDFIFGSGTSMYLGTQLHVVGDGIRVIAAHAGKSAEEKSG
YSFVHCKVTGTGGGIYLGRAWMSHPKVVYAYTEMTSVVN**PT**GWQENK**TPA**HDKTVFYGEYKCSG
PGSHKAKRVPFTQDIDDKEANCFLSLGYIQGSKWLLPP**PA**L
>AT5G46940.1 | Symbols: | Plant invertase/pectin methylesterase inhibitor superfamily protein | chr5:19058230-19058760 REVERSE LENGTH=176

MKFLLYLVTFFVLSNGLANGQTLIRNSCKKATAT**SP**KFKYNLCVTSLETNPQAKTAKDLAGLVM
ASTKNAVTKATTLKGTVDKIIKGKKVNKMTAMPLRDCLQLYTDAIGSLNEALAGVKSRNY**PT**VK
TVLSAAMD**TPS**TCETGFKERK**APSPV**TKENDNLYQMILIPLAFTNMLK
>AT4G19820.1 | Symbols: | Glycosyl hydrolase family protein with chitinase insertion domain | chr4:10767436-10768614 REVERSE LENGTH=366
MSSTKLISITFFLSLLLRFSSAQTVVKATYWFAESE**SP**LAQIDSSLFTHLFCAFADINTLTYQV
IVSSRNKPKFSTFTQTVRRRN**PT**VKTLLSIGGDFTYNFAFASMASN**PT**SRKLFISSSIKLARSC
GFHGLDLNWKY**PS**ITTEMDNFGKLLREWRLAVEAEARSSGKPRLLLTAAVFYSYSYYSVLHPVN
AVADSLDWVNLVAYDFYESGSSRVTC**SPAP**LYDPITTG**PS**GDAGVRAWTQAGL**PA**KKAVLGFPL
YGYAWCLTDAKNHNYYANSSG**PA**I**SP**DGSIGYDQIRRFIVDNKATMVYNSNLVQNYCYAKKTWI
GYDDNQSIVMKVKYAKQRGLLGYFSWHIGADDNSRLSRAGSVLFQF
>AT3G06260.1 | Symbols: GATL4 | galacturonosyltransferase-like 4 | chr3:1893804-1894859 REVERSE LENGTH=351
MASRSLSYTQLLGLLSFILLLVTTTTMAVRVGVILHK**PSAPT**L**PV**FRE**APA**FRNGDQCGTREAD
QIHIAMTLDTNYLRGTMAAVLSLLQHSTCPENLSFHFLSLPHFENDLFTSIKSTFPYLNFKIYQ
FDPNLVRSKISKSIRQALDQPLNYARIYLADII**PS**SVDRIIYLDSDLVVVDDIEKLWHVEMEGK
VVA**AP**EYCHANFTHYFTRTFWSDPVLVKVLEGKRPCYFNTGVMVVDVNKWRKGMYTQKVEEWMT
IQKQKRIYHLGSLPPFLLIFAGDIKAVNHRWNQHGLGGDNFEGRCRTLHPGPISLLHWSGKGKP
WLRLDSRKPCIVDHLW**AP**YDLYRSSRHSLEE
>AT2G41400.1 | Symbols: | Pollen Ole e 1 allergen and extensin family protein | chr2:17259509-17259961 FORWARD LENGTH=150
MEKSNVFLLSILLMASLFYISHAFKLWSYTIDRVSIRGVVYCSLDGD**PSAP**PVSNATVYIECPG
SNSTLAQAVTNQVGVFTLVFN**PA**NTSFDN**PS**KCDIKVNL**PT**NSCFIYPPGGVLIASVVNEIYGG
SISLENFIVIASYAATTFL**SP**G
AT5G60920.1 | Symbols: COB | COBRA-like extracellular glycosyl-phosphatidyl inositol-anchored protein family | chr5:24511466-24513932 REVERSE LENGTH=456
MESFFSRSTSIVSKLSFLALWIVFLISSSSFTSTEAYDALDPEGNITMKWDVMSW**TP**DGYVAVV
TMFNFQKYRHIQ**SP**GWTLGWKWAKKEVIWSMVGAQTTEQGDCSKYKGNIPHCCKKD**PT**VVDLLP
G**TP**YNQQIANCCKGGVMNSWVQD**PA**TAASSFQISVGAAGTTNKTVRVPRNFTLMGPGPGYTCG**P**
AKIVR**PT**KFVTTDTRRTTQAMMTWNITCTYSQFLAQR**TPT**CCVSLSSFYNETIVGC**PT**CACGCQ
NNRTESGACLDPD**TP**HLASVV**SPPT**KKGTVLPPLVQCTRHMCPIRVHWHVKQNYKEYWRVKITI
TNFNYRLNYTQWNLVAQHPNLDNITQIFSFNYKSL**TP**YAGLNDTAMLWGVKFYNDFLSEAGPLG
NVQSEILFRKDQSTFTFEKGWAFPRRIYFNGDNCVMPPPDSYPFLPNGGSRSQFSFVAAVLLPL
LVFFFFSA
>AT5G67460.1 | Symbols: | O-Glycosyl hydrolases family 17 protein | chr5:26921984-26923274 REVERSE LENGTH=380
MRHLLYIFLFIAFLSFAGAGQES**TP**IEALNLNNNLLQLPQTTDLDLAVSVSGNKTITEISSSII
KAETWLKTHILSRY**PS**TKITTIVIFFPDSCQTTQHSSDLVLSSLKNIYHSLTRWGLENNIKVSS
GFSYQCLNNPKSSEMFKPVLIFLKTINSTFTINPPQNFLT**SP**HNHRDLLHSVEKLGSLSFNKVN
FLNPEPEQESTTTMTRRNLRSLVNLSSKFTTSF**PTLPSPSP**ET**SP**IHSSIG**SPSPPT**IPYFPE**P**
SQ**SP**MESNQGISLPPCLPYH**PAPSP**QPVKKKNVEGLWCVAK**PS**VAAETLQQSLDFACGQGGANC
DEIKPHGICYYPDTVMAHASYAFNSYWQKTKRNGGTCSFGGTAMLITTD**PS**YQHCRFVLS
>AT3G55430.1 | Symbols: | O-Glycosyl hydrolases family 17 protein | chr3:20549806-20552004 REVERSE LENGTH=449
MAK**APPS**ISLLLLLCAAVFLTI**PA**VISAIGVNYGTLGNLPP**PT**QVANFIKTQTSIDSVKIFDVN
PDILRAFAGTGISVVVTVPNGDI**PA**LANGRQAARWVSVNILPFHPQTKIKYISVGNEILLTGDN

NMINNLL**PA**MRNLNNALVRAGVRDVKVTTAHSLNIIAYDLTG**APS**SGRFRPGWDKGIL**AP**ILAY
HRRTK**SP**FMVNPYPYFGFDPKNVNFAIFR**TP**YKAVRDPFTRHVYTNMFDALMDSTYSAMKALGY
GDVNIVVGETGW**PS**ACD**AP**WC**SPA**NAAWFNLNIIKRAQGQG**TP**LMPNRRFETYIFGLFNEEGKP
G**PT**AERNWGLFRADF**SP**VYDVGLLRNGQGGGGR**PA**L**PAPS**TAGGKWCVARSGATNTQLQDSINW
VCGQGVDCKPIQAGGSCFN**PS**SLRTHASFVMNAYFQSHGRTDGACNFSGTGMIVGNN**PS**NGACK
Y
>AT3G23770.1 | Symbols: | O-Glycosyl hydrolases family 17
protein | chr3:8565556-8567196 FORWARD LENGTH=476
M**TP**FALFLFTLLALSSSCCSAIGPQNNRTVLALASRIGINYGKLGNNLPFPYQSINLIKTIKAG
HVKLYDADPETLKLLSTTNLYVTIMVPNNQIISIGADQAAADNWVATNVLPFHPQTRIRFVLVG
NEVLSYSSDQDKQIWANL**VPA**MRKVVNSLRARGIHNIKVG**TP**LAMDALRSSFP**PS**SGTFREDIA
VPVMLPLLKFLNGTNSFFFLDVYPYFPWSTDPVNNHLDFALFESNSTYTDPQTGLVYTNLLDQM
LDSVIFAMTKLGYPNISLAISETGWPNDGDIHETGANIVNAATYNRNLIKKMTANPPLG**TPA**RR
G**AP**I**PT**FLFSLFNENQKPGSGTERHWGILNPDG**TP**IYDIDFSGRRSFSGFDSLPK**PS**NNVPFKG
NVWCVAVDGADEAELGQALNFACGRSNATCAAL**AP**GGECY**AP**VTVTWHASYAFSSYWAQFRNQS
SQCYFNGLARETTTNPGNERCKF**PS**VTL
>AT2G19440.1 | Symbols: | O-Glycosyl hydrolases family 17
protein | chr2:8418164-8419806 REVERSE LENGTH=478
MNLLAFVVGFGIMGIVMVDGLGVNWGTMATHKLPPKKVVQMLKDNNINKVKLFDADETTMSALS
GSGLEVMVAIPNDQLKVMGSYDRAKDWVHKNVTRYNFNGGVNITFVAVGNEPFLKSYNGSFINL
TF**PA**LQNIQNALNEAGLGSSVKATVPLNADVYD**SPS**SN**PVPS**AGRFRPDIIGQMTQIVDFLGNN
S**AP**ITINIYPFLSLYGNDDFPLNYAFFDGAKPVDDNGIAYTNVFDANFDTLVSALKAVGHGDMP
IIVGEVGW**PT**EGDKHANSGSAYRFYNGLLPRLGENRG**TP**LR**PT**YIEVYLFGLLDEDAKSI**AP**GE
FERHWGIFKFDGQPKFPIDLSGQGQNKLLIGAENVTYQPKKWCMFNTEAKDLTKLAANIDYACT
FSDCTALGYGSSCNTLDANGNASYAFNMYFQVKNQDEDACIFQGLATITTKNISQGQCNFPIQI
VASTASSFSSSLVLLIAGVWFLLSGVMFEV
>AT4G14080.1 | Symbols: MEE48 | O-Glycosyl hydrolases family 17
protein | chr4:8118697-8120292 REVERSE LENGTH=478
MSLLAFFLFTILVFSSSCCSATRFQGHRYMQRKTMLDLASKIGINYGRRGNNL**PSP**YQSINFIK
SIKAGHVKLYDADPESLTLLSQTNLYVTITVPNHQITALSSNQTIADEWVRTNILPYYPQTQIR
FVLVGNEILSYNSGNVSVNL**VPA**MRKIVNSLRLHGIHNIKVG**TP**LAMDSLRSSFP**PS**NGTFREE
ITGPVMLPLLKFLNGTNSYFFLNVHPYFRWSRNPMNTSLDFALFQGHSTYTDPQTGLVYRNLLD
QMLDSVLFAMTKLGYPHMRLAISETGWPNFGDIDETGANILNAATYNRNLIKKMSA**SP**PIG**TPS**
RPGLPI**PT**FVFSLFNENQKSGSGTQRHWGILHPDG**SP**IYDVDFTGQ**TP**LTGFNPLPK**PT**NNVPY
KGQVWCVPVEGANETELEETLRMACAQSNTTCAAL**AP**GRECYEPVSIYWHASYALNSYWAQFRN
QSIQCFFNGLAHETTTNPGNDRCKF**PS**VTL
>AT4G26830.1 | Symbols: | O-Glycosyl hydrolases family 17
protein | chr4:13494726-13496487 REVERSE LENGTH=455
MAVSFLPYFLILSFLSAIDAHSGMVGVNYGRIANNL**PSP**EKVVNLLKSQGINRIKIFDTDKNVL
TALANSKIKVIVALPNELLSSAASHQSFADNWIKTHIMPYF**PA**TEIEAIAVGNEVFVD**PT**I**TP**Y
LVNAMKNIHTSLVKYKLDKAIKIS**SP**IALSALANSYP**PS**SGSFKPELIEPVVKPMLALLQQTSS
YLMVNAYPFFAYAANADKISLDYALFKENAGNIDSGTGLKYNSLFDAQIDAVYAALSAVGFKGV
KVMVTETGW**PS**VGDENEIGASESNAAAYNAGLVKRVLTGKG**TP**LR**PT**EPLNVYLFALFNENQKP
G**PT**SERNYGLFYPNEGKVYNVPFTKKST**TP**VNGNRGKVPVTHEGHTWCVSNGEVAKEKLQEALD
YACGEGGADCRPIQPGATCYHPESLEAHASYAFNSYYQKNSRRVGTCFFGGAAHVVTQPPRYGK
CEF**PT**GH
>AT1G11820.2 | Symbols: | O-Glycosyl hydrolases family 17
protein | chr1:3991144-3993327 REVERSE LENGTH=511

MAFTSMVSTVPVLFFFFTLLLISANSSSLSHNIKVQEQDKDPFVGFNIGTDVSNLL**SPT**ELVKF
LQAQKVNHVRLYDADPELLKALAKTKVRVIISVPNNQLLAIGSSNSTAASWIGRNVVAYYPETL
ITAISVGDEVLTTV**PS**S**AP**LLL**PA**IESLYNALVASNLHTQIKVS**TP**HAASIMLDTFP**PS**QAYFN
QTWHSIMVPLLQFLSKTG**SP**LMMNLYPYYVYMQNKGVVPLDNCLFEPL**TPS**KEMVDPNTLLHYT
NVLDAMVDAAYVSMKNLNVSDVAVLVTESGW**PS**KGDSKEPYATIDNADTYNSNLIKHVFDRTG**T**
PLHPEMTSSVYIYELFNEDLR**AP**PVSEASWGLFYGNS**TP**VYLLHVSGSGTFLANDTTNQTYCIA
MDGVDAKTLQAALDWACGPGRSNCSEIQPGESCYQPNNVKGHASFAFNSYYQKEGRASGSCDFK
GVAMITTTD**PS**HGSCIFPGSKKVGNRTQTVVNSTEVAAGEATSRSLSRGFCVTIMILVTFSIL
>AT2G01630.1 | Symbols: | O-Glycosyl hydrolases family 17 protein | chr2:279541-281955 REVERSE LENGTH=501
MAALLLLFLFLFASSALSQDSLIGVNIGTEVTNM**PSPT**QVVALLKSQNINRVRLYDADRSMLLA
FAHTGVQVIISVPNDQLLGISQSNATAANWVTRNVAAYY**PA**TNITTIAVGSEVLTSLTNAASVL
VSALKYIQAALVTANLDRQIKVS**TP**HSSTIILDSFP**PS**QAFFNKTWDPVIVPLLKFLQSTG**SP**L
LLNVYPYFDYVQSNGVIPLDYALFQPLQANKEAVDANTLLHYTNVFDAIVDAAYFAMSYLNFTN
IPIVVTESGW**PS**KGG**PS**EHDATVENANTYNSNLIQHVINKTG**TP**KHPGTAVTTYIYELYNEDTR
PGPVSEKNWGLFYTNG**TP**VYTLRLAGAGAILANDTTNQTFCIAKEKVDRKMLQAALDWACGPGK
VDCSALMQGESCYEPDDVVAHSTYAFNAYYQKMGKASGSCDFKGVATVTTTD**PS**RGTCVFPGSA
KSNQTLGNNTSAL**APS**ANSTTSGCIPKYYHHPHASFGDLTLLSLLLIIALVFL
>AT5G23870.3 | Symbols: | Pectinacetylesterase family protein | chr5:8046060-8049999 REVERSE LENGTH=451
MKTTTRLLDLTAAMVLVVYVSF**SP**PLVSGEPGRRVSMTLVRDAAALGAFCLDGSL**PA**YHLDRGF
GAGSNNWILQFEGGGWCNDIASCVERAKTRRGSTRYMSKTVVFTGVLSNNASQNPDFYNWNKVR
LRYCDGASFAGDSQFGNGTSLLYFRGQRIWNAIILDLLPKGLAKAHKALLTGCSAGGLSTFLHC
DNFTSYLPKNASVKCMSDAGFFLDAIDVAANRTMRSFYSQLVSLQGIQKNLD**PS**CTHAFFPE**PS**
LCFFPQYVLRFIK**TP**FFILNSAYDVFQFHHGLVP**PS**ADQTGRWNRCKLNVTACNPHQLDALQGF
RKDMLGALMNFFRNSTRGGMFINSCFDHCQSALEETWL**SPTSP**RINNKTIAETVGDWYFGRGEE
AKEIGCPYPCDKTCHNLI**PA**STSDFLASNASGSGHNSRGTHLTFLFLLLNFFFFFVISKFSKKD
YVT
>AT2G30210.1 | Symbols: LAC3 | laccase 3 | chr2:12887735-12889827 REVERSE LENGTH=570
MESFRRFSLLSFIALLAYFAFLASAEHHVHQFVI**TPTPV**KRLCRTHQSITVNGQYPG**PT**LVVRN
GDSLAITVINRARYNISIHWHGIRQLRNPWADGPEYITQCPIRPGQTYTYRFKIEDQEGTLWWH
AHSRWLRATVYGALIIYPRLG**SP**YPFSMPKRDIPILLGEWWDRNPMDVLKQAQFTGAAANVSDA
YTINGQPGDLYRCSRAGTIRFPIFPGETVQLRVINAGMNQELFFSVANHQFTVVETDSAYTKPF
TTNVIMIGPGQTTNVLLTANQRPGRYYMAARAYNSAN**AP**FDNTTTTAILQYVN**APT**RRGRGRGQ
I**APV**F**PV**LPGFNDTATATAFTNRLRYWKR**AP**VPQQVDENLFFTVGLGLINCANPN**SP**RCQGPNG
TRFAASMNNMSFVLPRSNSVMQAYYQG**TP**GIFTTDFPPVPPVQFDYTGNVSRGLWQPIKGTKAY
KLKYKSNVQIVLQDTSIV**TP**ENHPMHLHGYQFYVVGSGFGNFNPRTD**PA**RFNLFDPPERNTIG**T**
PPGGWVAIRFVADNPGAWFMHCHIDSHLGWGLAMVFLVENGRGQLQSVQ**AP**PLDLPRC
>AT5G04310.1 | Symbols: | Pectin lyase-like superfamily protein | chr5:1203356-1207352 REVERSE LENGTH=518
MVAHERRIHNLQK**PT**CICIIWFCLLVSLSHHGRASSTSASIFNLSLPHQHPFPEHVVLNVQRKL
NDSLSRRQLLTYQQDDGTTAS**SP**I**PS**CITGNPIDDCWRCDPNWSANRQRLADCSIGFGQGTLGG
KGGQFYLVTDSSDNDAANPIPGTLRHAVIQPEPLWIIFSSDMGIKLKHELIIGSYKTIDGRGTN
IQITGHGCLTIQQVSHVIIHNVHIHHCK**PS**GNTLVAS**SPT**HVGFRGVSDGDGISVSASHHIWVD
HCSLGYCADGLIDVILASTAVTISNNYFSHHDEVMLLGHDDRYTADKGMQVTIAFNHFGEGLVQ
RMPRCRHGYIHVVNNDFTAWEMYAIGGSA**SPT**INSQGNRYT**AP**IDPNAKEVTKRVDSNEKHWSG
WNWRTEGDVMVNGAFFV**PS**GDGV**SPA**YARATSVQPKAAAIIDQLTVNAGVFGD**PS**GRNGQGGSF

PGITNGGGTITRGYSKSG**PA**GGGSGSDSDDGLFTLIFGNNSGAVALRPGQVWSILLIIILYWYI
PHHTRS
>AT2G39640.1 | Symbols: | glycosyl hydrolase family 17 protein |
chr2:16525171-16527012 REVERSE LENGTH=549
MAKTIRSFILPFLLIVAGVIFQLSAVTSAIGINYGTLGNLQPPQQVVDFIKTKTTFDSVKIYDA
NPDILRAFAGSEINITIMVPNGNI**PA**MVNVANARQWVAANVLPFQQQIKFKYVCVGNEILASND
NNLISNL**VPA**MQSLNEALKASNLTYIKVT**TP**HAFTISYNRN**TPS**ESRFTNDQKDIFTKILEFHR
QAK**SP**FMINAYTFFTMDTNNVNYAIFG**PS**NAITDTNTQQTYTNMFDAVMDATYSAMKALGYGDV
DIAVGETGW**PT**ACDASWC**SP**QNAENYNLNIIKRAQVIG**TP**LMPNRHIDIFIFALFNEDGKPG**PT**
RERNWGIFKPDF**SP**MYDVGVLKGGG**SP**LPFPPINNNGKWCVGKPEATLMQLQANIDWVCSHGID
C**TP**I**SP**GGICFDNNNMTTRSSFIMNAYYQSKGCVDVVCDFSGTGIVTSTN**PS**TSTCPIPIGEGG
GGNGAKSKSANWCMAKQEATETQLQANIDWVCSQGIDCKPI**SP**GGICFDNNNMKTRSTFIMNAY
YESKGYSKDACDFRGSGIVTTTN**PS**TSTCV**VPA**SVTL
>AT4G17180.1 | Symbols: | O-Glycosyl hydrolases family 17
protein | chr4:9646541-9648045 FORWARD LENGTH=475
MGSGVGVALFALSLLLVSHEVESAIGVNWGTLSFHKMR**PS**TVVDLLKANKITKVKLFDANPDAL
RALMGTGIQVMIGIPNEMLSTFNSDLFVQQNLSRFIGKNGADIRYVAVGNEPFLTGYGGQFQNY
VV**PT**MVNLQQSLVRANLASYVKLVVPCNADAYQSN**VPS**QGMFRPELTQIMTQLVSFLNSNG**SP**F
VVNIYPFLSLYGNSDFPQDYAFFEGSSHPVPDGPNTYYNAFDGNFDTLVAALTKLGYGQMPIVI
GEIGW**PT**DGAVGANLTAARVFNQGLISHVLSNKG**TP**LRPG**SPPA**DVYLFGLLDEGAKSTLPGNF
ERHWGIFSFDGQAKYRLNLGLGNRGLKNAKNVQYL**PS**RWCVAH**PS**RDMTQVGDHLRLACSEADC
TTLNDGGSCSQLGEKDNISYAFNSYYQMQMQHEKSCDFDGLGMVTFLD**PS**VGDCRFLVGVTDIG
LSSSTEPMARWSIYHICIGLLIWTLTM
>AT3G16850.1 | Symbols: | Pectin lyase-like superfamily protein
| chr3:5748692-5750981 FORWARD LENGTH=455
MKMPVALVWLLAFTILLISGEGNNAICKENFKLDPRPHSVSILEFGAVGDGKTLNTIAFQNAVF
YLKSFADKGGAQLYVPPGKWLTGSFNLTSHLTLFLEKGATILA**SP**D**PS**HWDVV**SP**L**PS**YGRGIE
LPGKRYRSLINGDNLIDVVITGENGTFDGQGAAWWEWLESGSLNYSRPHIIEFVSSKHILISNL
TFLN**APA**INIHPVYCSQIHIRKVLIETSVD**SP**HVLGV**AP**DSSDNVCIEDSTINVGHDAVSLKSG
WDQYGIHYGR**PT**TAVHIRNLRLK**SPT**GAGISFGSEMSGGVSDVTVERLNIHSSLIGVAFRTTRG
RGGYIRNITISDVDLTSVDTAIVANGHTGSHPDDKFDRDALPVVTHIVMRNFTGVDIGVAGNLT
GIGE**SP**FTSICLADIHLQTRSEESWICSNVSGFSDDV**SP**EPCQELMS**SPS**SCFAGGSIYEGDAT
AQSYYSW
>AT3G17070.1 | Symbols: | Peroxidase family protein |
chr3:5821048-5823165 FORWARD LENGTH=339
MKPKSKVAESTAASCFLVMSLLCSCIIGDQMETNNEGLSYSYYEKTCPKVEEIVRSSLSSMFIL
D**PTSPA**ALLRLMFHDCQVQGCDASILLEPIRDQQFTELDSAKNFGIRKRDLVGSIKTSLELECP
KQVSCSDVIILAARDAVALTGGPLISVPLGRKDSLS**TPS**KHVADSELP**PS**TADVDTTLSLFANK
GMTIEESVAIMGAHTIGVTHCNNVLSRFDNANATSENMDPRFQTFLRVACPEF**SPT**SQAAEATF
VPNDQTSVIFDTAYYDDAIAGRGNLRIDSEIGADPRTRPFVEAFAADQDRFFNAFSSAFVKLSS
YKVLTGNEGVIRSVCDKVD
>AT2G18980.1 | Symbols: | Peroxidase superfamily protein |
chr2:8233419-8235294 REVERSE LENGTH=323
MKNQSSFSIVALLLIFFSSSVFAQLQTNFYRKSCPNVETIVRNAVRQKFQQTFVT**APA**TLRLFF
HDCFVRGCDASILLA**SPS**EKDHPDDKSLAGDGFDTVAKAKQALDRDPNCRNKVSCADILALATR
DVVVLTGGPNYPVELGRRDGRLSTVASVQHSLPQ**PS**FKLDQLNTMFARHGLSQTDMIALSGAHT
IGFAHCGKFSKRIYNF**SP**KRPID**PT**LNIRYALQLRQMCPIRVDLRIAINMD**PTSP**NTFDNAYFK

NLQKGMGLFTSDQVLFSDERSRSTVNSFASSEATFRQAFISAITKLGRVGVKTGNAGEIRRDCS
RVN
>AT2G36700.1 | Symbols: | Pectin lyase-like superfamily protein
| chr2:15384706-15386421 REVERSE LENGTH=333
MLFFILFLSII**SP**IESVDQRIHH**PS**KCDHLSKF**PT**KGFTMVLKVSLNGCGRFKRVQDAIDASIG
SSQSKTLILIDFGIYRERFIVHENKNNLVVQGMGYSRTSIEWNNTTASSNGTFSSFSVAVFGEK
FTAYNISFKNT**APAP**NPGAVDAQAVALKVVGDKAAFYGCGFYGNQDTLLDQEGRHFFKGCFIEG
SIDFIFGNGRSLYEDCTLHSIAKENTIGCITANGKDTLKDRTGFVFVNCKITGSARVWLGRAWR
PYARVIFSKTYMSRVVSLDGWNDMGDPKTQRTVYYGEHRCYGPGANHSKRVTYAKLLSDVEA**AP**
FTNISFIDGEEWL
>AT3G13560.1 | Symbols: | O-Glycosyl hydrolases family 17
protein | chr3:4425484-4427284 REVERSE LENGTH=505
MLLPRWFAEALLLLLSILACSNAAFIGVNIGTDLTNMPP**PS**DIVTLLKSQQITHVRLYDANSHM
LKAFANTSIEVMVGVTNEEILKIGRF**PS**AAAAWVNKNVAAYI**PS**TNITAIAVGSEVLTTIPHV**A**
PILASALNNIHKALVASNLNFKVKVS**SP**MSMDIMPKPFP**PS**TSTF**SPS**WNTTVYQLLQFLKNTG
SFFMLNAYPYYGYTTANGIFPLDYALFKQL**SPV**KQIVDPNTLLHYNSMFDAMVDAAYYSMEALN
FSKIPVVVTETGW**PS**SGGSDEAAATVANAETFNTNLIKRVLNNSGP**PS**QPDIPINTYIYELYNE
DKRSGPVSERNWGILFPNGTSVYPLSLSGGSSSAALNGSSMFCVAKADADDDKLVDGLNWACGQ
GRANCAAIQPGQPCYLPNDVKSHASFAFNDYYQKMKSAGGTCDFDGTAITTTRD**PS**YRTCAYTG
SLNANATNGNFPPDALG**PASP**LGGNANARIIFSYHLPIL**AP**LALTLLQLLLQHDRLL
>AT5G01930.1 | Symbols: MAN6, AtMAN6 | Glycosyl hydrolase
superfamily protein | chr5:361189-362867 REVERSE LENGTH=448
MKDQLGFRIVLCSAVFIILTQNRALADLDSESHEVNSESVGEEQWEMVQRKGMQFTLNGQPFYV
NGFNTYWMMTLAADNSTRGKVTEVFQQASAVGMTVGRTWAFNDGQWRALQK**SPS**VYDEEVFKAL
DFVLSEARKYKIRLILSLVNNWDAYGGKAQYVKWGNASGLNLTSDDDFFTN**PT**LRNFYQSHVRT
VLNRVNTFTNITYKND**PT**IFAWELMNEPRC**PS**D**PS**GDKLQSWIQEMAVFVKSLDAKHLVEIGLE
GFYG**PSAPA**RTRFNPNPYAAQVGTDFIRNNQVLGIDFASVHVYPDSWI**SPA**VSNSFLEFTSSWM
QAHVEDAEMYLGMPVLFTEFGVSAHDPGFNTSFRDMMLNTVYKMTLNSTRKGGAGAGSLVWQVF
PQGAEFMDDGYAVYLTRAHTASKIISLQSKRLAIFNSLCSWRCRWGCKKKNQTALDALLSHDEL
>AT5G47500.1 | Symbols: | Pectin lyase-like superfamily protein
| chr5:19271262-19272845 REVERSE LENGTH=362
MAQLTNSLNYLFSVSLLLFVSFHCLCFRFSLVAACSNSTDDQQIQHHHHRKWVG**PS**GHKVITVS
LNGHAQFRSVQDAVDSIPKNNNKSITIKI**AP**GFYREKVV**VPA**TKPYITFKGAGRDVTAIEWHDR
ASDLGANGQQLRTYQTASVTVYANYFTARNISFTNT**APAP**LPGMQGWQAVAFRISGDKAFFSGC
GFYGAQDTLCDDAGRHYFKECYIEGSIDFIFGNGRSMYKDCELHSIASRFGSIAAHGRTCPEEK
TGFAFVGCRVTGTGPLYVGRAMGQYSRIVYAYTYFDALVAHGGWDDWDHKSNKSKTAFFGVYNC
YGPGAAATRGVSWARALDYESAHPFIAKSFVNGRHWI**AP**RDA
>AT5G60950.1 | Symbols: COBL5 | COBRA-like protein 5 precursor |
chr5:24527157-24528005 REVERSE LENGTH=204
MESLFSTMIVLLLVSFSCLISTEALTSNYGNITVKWDLLNW**TP**DGYVAVVTAYNYQKQRSIPGW
KMSWRGTKKEVIWNMLGAKTTGQGGCSMFKGNIPQSCVRK**PT**VVDLLPG**TP**FNQQIANCCKSGV
LKPGSESAFQLSVGSAGNSVKTARM**PA**NFMFT**AP**KQQYICG**PS**KNVR**PT**RFTTADKRRITAALM
TWNITCVFHKAT
>AT3G29810.1 | Symbols: COBL2 | COBRA-like protein 2 precursor |
chr3:11728212-11730158 FORWARD LENGTH=441
MNILFSRFSFLLLFLCSWTSFTFTTTEAYDALDPYGNITIKWDIMSWTGDGYVAVVTIFNFQQY
RHIE**AP**GWQLGWSWMKKEVIWSMVGGQATEQGDCSKFKGNIPHCCKK**TPA**IVDLLPG**TP**YNQQI
SNCCRGGVISAWAQD**PA**TAISSFQISVGQSGTTNTTVR**AP**RNITLK**AP**GPGYTCG**PA**KLVK**PS**R

FISADKRRKTQSLLTWNITCTYSQFLARK**TPT**CCVSLSAFYNETIVPC**PT**CSCGCQNSSQAGTC
VDPKIASVV**PA**LGKNNLEPLLQCTQHMCPIRVHWHVKTSYKEYWRVKVAITNFNYNMNYSQWNL
VVQHPNFDNLTKLFSFNYKPLNPYLNINDTAMLWGIKFYNDFLSQAGPVGNVQSELLFQKNPLE
FTFEKGWAFPRRIYFNGDNCVMPPPDSYPWLPNA**SP**NIAT**SP**FVILLITFLSVLILM
>AT3G07320.1 | Symbols: | O-Glycosyl hydrolases family 17
protein | chr3:2332324-2333925 REVERSE LENGTH=460
MSLLLHLFALSLLISVSGAKFSGRPGINYGQLGNNL**PSPS**DSVNLIKSLNAKRVKLYDANPKIL
AALNGTDITVSVMVPNELLVNISKSASLSDDWIRSNILPFY**PT**TKIRYLLVGNEILSLPDSELK
SSLV**PA**MRKIQRSLKSLGVKKVKVGTTLATDVLQSSFP**PS**SGEFREDISGLIMKPMLQFLNRTK
SFLFVDVYPYFAWAQD**PT**HVDLDYAIFESTNVTVTDPVSNLTYHNLFDQMIDAFVFAMKRVGYP
DIRIWVAETGWPNNGDYDQIGANIYNAATYNRNVVKKLAADPPVG**TPA**RPGKVL**PA**FVFALYNE
NQKTGPGTERHFGLLHPNGTQVYGIDLSGKTEYKESL**PAP**ENNDLYKGKIWCVVAKGANWTQLG
DALSYACSQGNNTCDPIQRGGPCQKPDLTVLHASYAFSSYWAQFRKIGGTCSFNGLATQTIKD**P**
SYGRCEF**PS**VTL
>AT4G11050.1 | Symbols: AtGH9C3, GH9C3 | glycosyl hydrolase 9C3
| chr4:6747498-6751311 REVERSE LENGTH=626
MGSRTTISILVVLLLGLVQLAISGHDYKQALSKSILFFEAQRSGHLPPNQRVSWRSHSGLYDGK
SSGVDLVGGYYDAGDNVKFGLPMAFTVTTMCWSIIEYGGQLESNGELGHAIDAVKWGTDYFIKA
HPEPNVLYGEVGDGKSDHYCWQRPEEMTTDRRAYKIDRNNPGSDLAGETAAAMAAASIVFRRSD
PSYSAELLRHAHQLFEFADKYRGKYDSSITVAQKYYRSVSGYNDELLWAAAWLYQATNDKYYLD
YLGKNGDSMGGTGWSMTEFGWDVKYAGVQTLVAKVLMQGKGGEHTAVFERYQQKAEQFMCSLLG
KSTKNIKK**TP**GGLIFRQSWNNMQFVTSASFLATVYSDYLSYSKRDLLCSQGNI**SPS**QLLEFSKS
QVDYILGDNPRATSYMVGYGENYPRQVHHRGSSIVSFNVDQKFVTCRGGYATWFSRKGSDPNVL
TGALVGGPDAYDNFADQRDNYEQTE**PA**TYNN**AP**LLGVLARLISGSTGFDQLLPGV**SPTPSP**VII
K**PAP**VPQRK**PT**KP**PA**AS**SPSP**ITISQKMTNSWKNEGKVYYRYSTILTNRSTKTLKILKISITKL
YGPIWGVTKTGNSFSF**PS**WMQSL**PS**GKSMEFVYIHSA**SPA**DVLVSNYSLE
>AT1G71980.1 | Symbols: | Protease-associated (PA) RING/U-box
zinc finger family protein | chr1:27098250-27099881 FORWARD
LENGTH=448
MNRALVLLLYVCTVSCLASSKVILMRNNITLSFDDIEANF**APS**VKGTGEIGVVYVAEPLDACQN
LMNKPEQSSNET**SP**FVLIVRGGCSFEEKVRKAQRAGFKAAIIYDNEDRGTLIAMAGNSGGIRIH
AVFVTKETGEVLKEYAGFPDTKVWLI**PS**FENSAWSIMAVSFISLLAMSAVLATCFFVRRHRIRR
RTSRSSRVREFHGMSRRLVKAM**PS**LIFSSFHEDNTTAFTCAICLEDYTVGDKLRLLPCCHKFHA
ACVDSWLTSWRTFCPVCKRDARTSTGEP**PA**SES**TP**LLSSAASSFTSSSLHSSVRSSALLIG**PS**L
GSL**PT**SISF**SPA**YASSSYIRQSFQSSSNRR**SP**PISVSRSSVDLRQQAA**SPSPSPS**QRSYISHMA
SPQSLGY**PT**I**SP**FNTRYM**SP**YR**PSPS**NA**SPA**MAGSSNYPLNPLRYSESAGTF**SP**YASANSLPDC
>AT4G34980.1 | Symbols: SLP2 | subtilisin-like serine protease 2
| chr4:16656929-16659223 REVERSE LENGTH=764
MASSTIVLLLFLSFPFISFAASQAAKTFIFRIDGGSM**PS**IF**PT**HYHWYSTEFAEESRIVHVYHT
VFHGFSAVV**TP**DEADNLRNH**PA**VLAVFEDRRRELHTTR**SP**QFLGLQNQKGLWSESDYGSDVIIG
VFDTGIWPERRSFSDLNLGPIPKRWRGVCESGARF**SP**RNCNRKIIGARFFAKGQQAAVIGGINK
TVEFL**SP**RDADGHGTHTSSTAAGRHAFKASMSGYASGVAKGV**AP**KARIAAYKVCWKDSGCLDSD
ILAAFDAAVRDGVDVISISIGGGDGIT**SP**YYLDPIAIGSYGAASKGIFVSSSAGNEGPNGMSVT
NL**AP**WVTTVGASTIDRNF**PA**DAILGDGHRLRGVSLYAGVPLNGRMFPVVYPGKSGMSSASLCME
NTLDPKQVRGKIVICDRGS**SP**RVAKGLVVKKAGGVGMILANGASNGEGLVGDAHLI**PA**CAVGSN
EGDRIKAYASSHPNPIASIDFRGTIVGIK**PAP**VIASFSGRGPNGL**SP**EILKPDLI**AP**GVNILAA
WTDAVG**PT**GL**PS**DPRKTEFNILSGTSMACPHVSGAAALLKSAHPDW**SPA**VIRSAMMTTTNLVDN
SNRSLIDESTGKSA**TP**YDYGSGHLNLGRAMNPGLVYDITNDDYITFLCSIGYGPKTIQVITR**TP**

VRC**PT**TRK**PSP**GNLNY**PS**ITAVF**PT**NRRGLVSKTVIRTATNVGQAEAVYRARIE**SP**RGVTVTVK
PPRLVFTSAVKRRSYAVTVTVNTRNVVLGETGAVFGSVTWFDGGKHVVR**SP**IVVTQMDTL

>AT1G35630.1 | Symbols: | Protease-associated (PA) RING/U-box zinc finger family protein | chr1:13163041-13164484 REVERSE LENGTH=318
MNYSWITIMSLLVICKLASAKVVLIGKNTILSFDDVEATF**TP**IVRNSGECGILYVAEPLEACSD
ITNMAEKRSKYRSSYVLIVLGGCSFEEKVRKAQKAGYKAAIVYNDGYDELLVPMAGNSSGVDIH
GLLVTRASGEVLKGYADQDEMKLWLIPGFGISSWSIMGITFISLLAMSAILATCFVVRRHQIRQ
SVRDLPHGGQGLSCMPRDLLQSM**PT**EVYSGVLEESSTSVTCAICIDDYCVGEKLRILPCKHKYH
AVCIDSWLGRCRSFCPVCKQNPRTGND**VPPA**SET**TP**LI**SPSP**NSITSLQSFYDLPIVVRVYL

>AT4G38220.2 | Symbols: | Peptidase M20/M25/M40 family protein | chr4:17925251-17926919 FORWARD LENGTH=433
MSLLRLLLVVVVLHLSAVAGDDAIVSRFQEYLRINTVQPNPEYYKAVDFIISQAKPLSLESQTI
EFVKGKPLLLLKWVGSD**PT**L**PA**FLLNSHTDVVPFEDSKWTHHPLQAHMDHHGDIYARGSQDMKC
VGMQYLEAIRKLQASGFKPLRSVYLSFVPDEEIGGHDGAEKFAESQLFKSLNIAIVLDEGL**PSP**
TESYRVFYGER**SP**WWLVIKAKGPPGHGAKLYDNSAMENLLKSIESIRRFRASQFDLLKAGGIAE
GDVVSVNMAFLKAG**TPSPT**GFVMNLQ**PS**EAEAGFDIR**VPPS**VDAEALERRLVEEW**APA**ARNMSF
ELGQFKQKLTGKQFLTAADDSNPWWGLLENAVKEAGGRTSKPEIF**PA**STDARYFRKAG**VPA**FGF
SPISN**TPS**LLHDHNEYLGKAEYLKGIEVYVSIIKAYASYESKSGSRDEL

>AT1G09750.1 | Symbols: | Eukaryotic aspartyl protease family protein | chr1:3157541-3158960 FORWARD LENGTH=449
MASSSLHFFFFLTLLLPFTFTTATRDTCATA**AP**DGSDDLSIIPINAKC**SP**F**APT**HVSASVIDTV
LHMASSDSHRLTYLSSLVAGKPK**PT**SVPVASGNQLHIGNYVVRAKLG**TP**PQLMFMVLDTSNDAV
WLPCSGCSGCSNASTSFNTNSSSTYSTVSCSTAQCTQARGLTC**PS**S**SP**Q**PS**VCSFNQSYGGDSS
FSASLVQDTLTL**AP**DVIPNFSFGCINSASGNSLPPQGLMGLGRGPMSLVSQTTSLYSGVFSYCL
PSFRSFYFSGSLKLGLLGQPKSIRY**TP**LLRNPRR**PS**LYYVNLTGVSVGSVQVPVDPVYLTFDAN
SGAGTIIDSGTVITRFAQPVYEAIRDEFRKQVNVSSFSTLGAFDTCFSADNENV**AP**KITLHMTS
LDLKLPMENTLIHSSAGTLTCLSMAGIRQNANAVLNVIANLQQQNLRILFDVPNSRIGI**AP**EPC
N

>AT5G07030.1 | Symbols: | Eukaryotic aspartyl protease family protein | chr5:2183600-2185717 REVERSE LENGTH=455
M**PTP**FNN**PS**NPKAYNTMSTLVLFLQLFSILPLALGLNHPNCDLTKTQDQGSTLRIFHID**SP**C**SP**
FKSS**SP**LSWEARVLQTLAQDQARLQYLSSLVAGRSVVPIASGRQMLQSTTYIVKALIG**TPA**QPL
LLAMDTSSDVAWIPCSGCVGC**PS**NTAF**SPA**KSTSFKNVSCS**AP**QCKQVPN**PT**CGARACSFNLTY
GSSSIAANLSQDTIRLAADPIKAFTFGCVNKVAGGGTIPPPQGLLGLGRGPLSLMSQAQSIYKS
TFSYCL**PS**FRSLTFSGSLRLG**PT**SQPQRVKYTQLLRNPRRSSLYYVNLVAIRVGRKVVDLP**PA**A
IAFN**PS**TGAGTIFDSGTVYTRLAKPVYEAVRNEFRKRVK**PT**TAVVTSLGGFDTCYSGQVKV**PT**I
TFMFKGVNMTM**PA**DNLMLHSTAGSTSCLAMAA**AP**ENVNSVVNVIASMQQQNHRVLIDVPNGRLG
LARERCS

>AT4G21326.1 | Symbols: ATSBT3.12, SBT3.12 | subtilase 3.12 | chr4:11346685-11349653 FORWARD LENGTH=754
MGIVKGRSRAGLFIGFLFIVNVGFCVFAQESSNEERKIYVVHLGVRRHDDSELVSESHQRMLES
VFESAEAARESIVYNYHHGFSGFAARLTDSQAKQLSDRPDVFSV**AP**NRKVELQSTRIYDYLGL**S**
PSF**PS**GVLHESNMGSDLVIGFLDSGVWPE**SPA**YNDEGLEPIPKHWGKGKCVAGEDFD**PA**KHCNKK
LVGAKYFTDGFDENNSGISEEDFM**SP**RGYRGHGTMVSSIAASSFVPNVSYGGL**AP**GVMRGA**AP**K
ARIAMYKIVWDRALLMSSTATMVKAFDEAINDGVDVLSISLASA**AP**FRPIDSITGDLELGSFHA
VMKGIPVIAGASNTGPEAYTVANVFPWMLTVAATNIDRTFYADMTFGNNITIIGQAQYTGKEVS
AGLVYIEHYKTDTSGMLGKVVLTFVKEDWEMASALATTTINKAAGLIVARSGDYQSDIVYNQPF

IYVDYEVGAKILRYIRSSS**SPT**IKISTGKTLVGRPIATQVCGFSSRGPNGL**SPA**ILKPDIA**AP**G
VTILGATSQAYPDSFGGYFLGTGTSYA**TP**VVAGLVVLLKALHPDW**SPA**ALKSAIMTTAWKTD**PS**
GEPIFAEGEPRKLADPFDYGAGLVNAERAKDPGLVYDMNIDDYIHYFCATGYNDTSITIITGK**P**
TKCS**SP**L**PS**ILDLNY**PA**ITIPDLEEEVTVTRTVTNVGPVDSVYRAVVEPPRGVEIVVEPETLVF
CSNTKKLGFKVRVSSSHKSNTGFFFGSFTWTDGTRNVTIPLSVRIRVLNP
>AT5G43100.1 | Symbols: | Eukaryotic aspartyl protease family protein | chr5:17299264-17302718 FORWARD LENGTH=631
MDRSSLLLLLFTTTTISIFFFDLTTADE**SP**MIFPLSYSSLPPRPRVEDFRRRRLHQSQLPNAHM
KLYDDLLSNGYYTTRLWIG**TP**PQEFALIVDTGSTVTYVPCSTCKQCGKHQDPKFQPELSTSYQA
LKCNPDCNCDDEGKLCVYERRYAEMSSSSGVLSEDLISFGNESQL**SP**QRAVFGCENEETGDLFS
QRADGIMGLGRGKLSVVDQLVDKGVIEDVFSLCYGGMEVGGGAMVLGKI**SP**PPGMVFSHSDPFR
SPYYNIDLKQMHVAGKSLKLNPKVFNGKHGTVLDSGTTYAYFPKEAFIAIKDAVIKEI**PS**LKRI
HGPDPNYDDVCFSGAGRDVAEIHNFFPEIAMEFGNGQKLIL**SP**ENYLFRHTKVRGAYCLGIFPD
RDSTTLLGGIVVRNTLVTYDRENDKLGFLKTNCSDIWRRLA**AP**E**SPAPTSP**ISQNKSSNI**SPSP**
ATSE**SPT**SHLPGVFRVGVITFEVSISVNNSSLKPKFSEIADFIAHELDIQSAQVRLLNFSSSGN
EYRLKWGVFPPQSSEYISNTTALNIMLLLKENRLRLPGQFGSYKLLEWKAEQKKKQSWWEKHLL
GVVGGAMISLLVTSVMIKLALVWRRRKQEEATYEPVNAAIKEQELQPLSSETSNA

Appendix III. *Medicago truncatula* AGP protein sequences.

>Medtr1g016850.1 _ transmembrane protein, putative _ HC _
chr1_4528072_4527427 _ 20130731 ; MatureChain: 25-61
AE**APAPSPTSPA**TAV**SPS**FIAGVVAAVVALAGSTLRI
>Medtr1g052120.1 _ transmembrane protein, putative _ HC _
chr1_21007175_21006594 _ 20130731 ; MatureChain: 23-61
ATHEV**PSPAPAP**G**PA**SGASTTVVGSLVGASLLSFFALFQ
>Medtr3g028270.1 _ transmembrane protein, putative _ HC _
chr3_8965132_8965667 _ 20130731 ; MatureChain: 26-67
QDL**SPS**L**APAPSP**DAGAAGSVTSSMAMIGASIVLSMLAIFKN
>Medtr5g008670.1 _ transmembrane protein, putative _ LC _
chr5_1865715_1865512 _ 20130731 ; MatureChain: 22-67
RDHYHHHRHVHG**PAPAP**G**PS**SDAA**SP**GSILGVSLFSFVAYYLCNHA
>Medtr5g035640.1 _ transmembrane protein, putative _ HC _
chr5_15508476_15508987 _ 20130731 ; MatureChain: 24-76
QNNGEDGINIKVSNDMPGMVM**APAPTP**KSSASL**PT**LTYSAAILIFLPFMLSFF
>Medtr6g029260.1 _ transmembrane protein, putative _ HC _
chr6_10069246_10068704 _ 20130731 ; MatureChain: 27-59
AD**APAPSPT**SDATTLFV**PT**IIASFVALVFGLLF
>Medtr6g029330.1 _ transmembrane protein, putative _ HC _
chr6_10106936_10106129 _ 20130731 ; MatureChain: 27-59
AD**APAPSPT**SDATTLFV**PT**IIASFVALLFGLLF
>Medtr7g085780.1 _ transmembrane protein, putative _ HC _
chr7_33252752_33253550 _ 20130731 ; MatureChain: 27-59
AD**APAPSPT**SDATTLFV**PT**AFASLIALAFGFLF
>Medtr3g104110.1 _ transmembrane protein, putative _ LC _
chr3_47970572_47970739 _ 20130731 ; MatureChain: 18-55
TD**APAPSPASPA**TAI**SPS**FIVGCVTAATALVFHSSLRI
>Medtr1g028610.1 _ arabinogalactan peptide_like protein _ HC _
chr1_9704641_9703281 _ 20130731 ; MatureChain: 20-62
SMASSQYV**APAPAPAPT**SDGTSIDQAIAYVLMLVALLLTYIIH
>Medtr3g090830.1 _ arabinogalactan peptide_like protein _ HC _
chr3_41234931_41234172 _ 20130731 ; MatureChain: 28-63
QSI**APAPAPT**SDGTKIDQAIACVLMLVALVLTYIIH
>Medtr6g086350.1 _ arabinogalactan protein _ LC _
chr6_32440224_32440033 _ 20130731 ; MatureChain: 26-63
KIFDI**SPAPTPT**GGAQILDQVIACFLMLVALVITYMLH
>Medtr6g086365.1 _ arabinogalactan protein _ LC _
chr6_32461334_32459573 _ 20130731 ; MatureChain: 26-88
QNFDI**SPAPTPT**SDAQSLDQVIAYSLMVVALMGEYIMQIQLGTRVLHNLRKQNQWDSRKRNDI
>Medtr6g086390.1 _ arabinogalactan peptide_like protein _ HC _
chr6_32464229_32463703 _ 20130731 ; MatureChain: 27-64
QGFDM**SPAPSPT**SDAQILDQGIAYFLMLVALVITYMFH
>Medtr6g090535.1 _ arabinogalactan protein _ HC _
chr6_34411665_34410320 _ 20130731 ; MatureChain: 20-63
FVV**PS**INAQI**PAPAPAPT**SDGTTLDQGIACILMLLALVLTYIIH
>Medtr8g099245.1 _ arabinogalactan peptide_like protein _ HC _

chr8_41765566_41763694 _ 20130731 ; MatureChain: 32-75
QTL**APAPSPT**SDGSSVDQGIAYLLMLLALVLTYIIHSADISSTF
>Medtr8g083480.1 _ arabinogalactan protein _ HC _ chr8_43600392_43600914 _ 20130731 ; MatureChain: 26-69
HDF**SPSPTPAPAPAPAPS**SDGTAFDQGIAYFLMLVALLITYMFH
>Medtr3g464630.1 _ transmembrane protein, putative _ LC _ chr3_25992627_25993229 _ 20130731 ; MatureChain: 22-69
RDVPIVNSHMTLEERDYADPHPYPGFPPPKFLNEDPHTH**PSPTP**NLKV
>Medtr3g037730.1 _ Clavata3_ESR (CLE) gene family member MtCLE11 _ HC _ chr3_13874477_13873747 _ 20130731 ; MatureChain: 28-83
LRTKDLVQISHVLKGDAMEGLESKMDL**APSPS**MTFDANQSNKRTVKKGSNPIHNRS
>Medtr1g021638.1 _ salt stress response_antifungal domain protein _ LC _ chr1_6521275_6522643 _ 20130731 ; MatureChain: 22-313
TTDTANKFQYFCDQNNDGGNY**PT**NSTYHTNLNTLLSTLTSNKDINYGFYNSSYGNNTDKVNAIG
LCRGDVKLNDCQNCLKNSTVLLTQHCQNRKEAIGWYNDEECMLRYSNRSIFGLNEIG**PA**YFVWN
TNNATNEVEFNKVVNFFLDSLRNRAASGDSDLKYAVGSDEVG**PS**NNQTIYGLVQC**TP**DLSKTLC
DDCLVQSIKEISNCCNNRLGARIVR**PS**CNLRYETNSFFYQ**PTPS**D**SPSPSP**VPVPV**PS**FS**TP**PP
FAQNNTSSQDKGNTSRNVVPVILLMLILLCSLLSSF
>Medtr1g022160.5 _ glucan endo_1,3_beta_glucosidase_like protein _ HC _ chr1_6896988_6892247 _ 20130731 ; MatureChain: 21-417
DPFLGVNYGQVADNLP**PA**SATAKLLQTTAFEKVRLYGTD**PA**IIKSLANTGIGIVIGAANGDI**PS**
LASD**PS**FAKTWISTNVLPYY**PA**SNIILITVGNEVITSNDTNLINGML**PA**IQNIQKALDEASLGG
KIRVSTVHTMSVLKNSEP**PS**AGSFHDEYSTVLQGLLSFNKDTG**SP**FAINPYPYFAYKSDPGRAD
NLAFCLFQPNAGRVDANTKLNYMNMFDAQVDAVRSALDSMGFKDVEIVVAETGWPYKGDNDEAG
PSIENAKAYNGNLIKHLRSKVG**TP**LMPGKSVDTYIFALYDEDLKPGAGSEKAFGLYNTDQSMIY
DAGLSKQQST**TPTPT**S**SP**VV**SPTP**EVSK**TP**VI**PAPT**G**PS**S**TP**KGHNH**TP**NKAEFLNGHEASILQ
SLIMLTYLILMLF
>Medtr1g022160.1 _ glucan endo_1,3_beta_glucosidase_like protein _ HC _ chr1_6896957_6893217 _ 20130731 ; MatureChain: 21-481
DPFLGVNYGQVADNLP**PA**SATAKLLQTTAFEKVRLYGTD**PA**IIKSLANTGIGIVIGAANGDI**PS**
LASD**PS**FAKTWISTNVLPYY**PA**SNIILITVGNEVITSNDTNLINGML**PA**IQNIQKALDEASLGG
KIRVSTVHTMSVLKNSEP**PS**AGSFHDEYSTVLQGLLSFNKDTG**SP**FAINPYPYFAYKSDPGRAD
NLAFCLFQPNAGRVDANTKLNYMNMFDAQVDAVRSALDSMGFKDVEIVVAETGWPYKGDNDEAG
PSIENAKAYNGNLIKHLRSKVG**TP**LMPGKSVDTYIFALYDEDLKPGAGSEKAFGLYNTDQSMIY
DAGLSKQQST**TPTPT**S**SP**VV**SP**INQDVSK**SP**MNS**TP**KV**PSPTTP**YNSNIAWCVPKAGLTDVQLQ
ANLDYACGQGIDCSLIQPGGACFEPNTLANHAAYAMNLFYHTVGQNPLTCDFSQTATLTSNNPR
AEVR**TP**D**TP**LFHN
>Medtr1g022160.4 _ glucan endo_1,3_beta_glucosidase_like protein _ HC _ chr1_6896988_6892247 _ 20130731 ; MatureChain: 21-486
DPFLGVNYGQVADNLP**PA**SATAKLLQTTAFEKVRLYGTD**PA**IIKSLANTGIGIVIGAANGDI**PS**
LASD**PS**FAKTWISTNVLPYY**PA**SNIILITVGNEVITSNDTNLINGML**PA**IQNIQKALDEASLGG
KIRVSTVHTMSVLKNSEP**PS**AGSFHDEYSTVLQGLLSFNKDTG**SP**FAINPYPYFAYKSDPGRAD
NLAFCLFQPNAGRVDANTKLNYMNMFDAQVDAVRSALDSMGFKDVEIVVAETGWPYKGDNDEAG

PSIENAKAYNGNLIKHLRSKVG**TP**LMPGKSVDTYIFALYDEDLKPGAGSEKAFGLYNTDQSMIY
DAGLSKQQST**TPTPT**S**SP**VV**SP**INQDVSK**SP**MNS**TP**KV**PSPTTP**YNSNIAWCVPKAGLTDVQLQ
ANLDYACGQGIDCSLIQPGGACFEPNTLANHAAYAMNLFYHTVGQNPLTCDFSQTATLTSNN**PS**
KFSCGFCATFFYILSSDL
>Medtr1g022160.2 _ glucan endo_1,3_beta_glucosidase_like protein _ HC _ chr1_6896988_6892247 _ 20130731 ; MatureChain: 21-417
DPFLGVNYGQVADNLP**PA**SATAKLLQTTAFEKVRLYGTD**PA**IIKSLANTGIGIVIGAANGDI**PS**
LASD**PS**FAKTWISTNVLPYY**PA**SNIILITVGNEVITSNDTNLINGML**PA**IQNIQKALDEASLGG
KIRVSTVHTMSVLKNSEP**PS**AGSFHDEYSTVLQGLLSFNKDTG**SP**FAINPYPYFAYKSDPGRAD
NLAFCLFQPNAGRVDANTKLNYMNMFDAQVDAVRSALDSMGFKDVEIVVAETGWPYKGDNDEAG
PSIENAKAYNGNLIKHLRSKVG**TP**LMPGKSVDTYIFALYDEDLKPGAGSEKAFGLYNTDQSMIY
DAGLSKQQST**TPTPT**S**SP**VV**SPTP**EVSK**TP**VI**PAPT**G**PS**S**TP**KGHNH**TP**NKAEFLNGHEASILQ
SLIMLTYLILMLF
>Medtr1g022160.3 _ glucan endo_1,3_beta_glucosidase_like protein _ HC _ chr1_6896988_6892247 _ 20130731 ; MatureChain: 21-479
DPFLGVNYGQVADNLP**PA**SATAKLLQTTAFEKVRLYGTD**PA**IIKSLANTGIGIVIGAANGDI**PS**
LASD**PS**FAKTWISTNVLPYY**PA**SNIILITVGNEVITSNDTNLINGML**PA**IQNIQKALDEASLGG
KIRVSTVHTMSVLKNSEP**PS**AGSFHDEYSTVLQGLLSFNKDTG**SP**FAINPYPYFAYKSDPGRAD
NLAFCLFQPNAGRVDANTKLNYMNMFDAQVDAVRSALDSMGFKDVEIVVAETGWPYKGDNDEAG
PSIENAKAYNGNLIKHLRSKVG**TP**LMPGKSVDTYIFALYDEDLKPGAGSEKAFGLYNTDQSMIY
DAGLSKQQST**TPTPT**S**SP**VV**SP**INQDVSK**SP**MNS**TP**KV**PSPTTP**YNSNIAWCVPKAGLTDVQLQ
ANLDYACGQGIDCSLIQPGGACFEPNTLANHAAYAMNLFYHTVGQNPLTCDFSQTATLTSNN**PS**
YNSCIYTGGNA
>Medtr1g084820.2 _ carbohydrate_binding X8 domain protein _ HC _ chr1_37754447_37757434 _ 20130731 ; MatureChain: 22-571
SLVGFSYHERGDTLTSFLQHSKVSSSQIRAFVTDHWILSTLTNSKLLVDLYLNKSQVEKFITSK
PSAVSELKAQLVNFLPHLNIKSIIVSCGSECLLQNELPLIMHALKSIHSILSDLHISKEVKISV
AFPLQVLRKLNASQEHEIRRLLSFIKETKSFVMIEDNIDGELRMDDHFVQTIIKRANLAASVLP
CKDVPLVLTIKSSVI**PS**SIEVTQFSKRVSKYLEAKRIAALYVELHTTEDSSMKELKREEEGMFH
LSRREILSKFHRRKIIDNTN**SPT**NTVY**PT**N**PTP**VI**TPS**D**TPT**IIAV**PS**TNPVTI**SPT**N**PA**AMPV
TV**PS**T**TP**VVPL**APTTPT**I**TPAP**VFN**PA**T**TPT**TVPG**AP**PVTSYPPPVTSYPPPLGNVPVVNPQQP
P**PS**NTN**APS**IQGQSWCVAKTG**AP**QASLQSALDYACGNGADCSQIQQGASCY**SP**VTLQNHASFAF
NSYYQKK**PAPT**SCDFGGAAMLVSSN**PS**SGSCIY**PS**SSSSSSSTST**SP**MIS**SPAPPT**QSTSTSIP
P**PS**LT**TPSPS**I**PT**I**APPT**MST**APS**SI**PTAPPT**SSGTFG
>Medtr1g084820.1 _ carbohydrate_binding X8 domain protein _ HC _ chr1_37754391_37757434 _ 20130731 ; MatureChain: 22-633
SLVGFSYHERGDTLTSFLQHSKVSSSQIRAFVTDHWILSTLTNSKLLVDLYLNKSQVEKFITSK
PSAVSELKAQLVNFLPHLNIKSIIVSCGSECLLQNELPLIMHALKSIHSILSDLHISKEVKISV
AFPLQVLRKLNASQEHEIRRLLSFIKETKSFVMIEDNIDGELRMDDHFVQTIIKRANLAASVLP
CKDVPLVLTIKSSVI**PS**SIEVTQFSKRVSKYLEAKRIAALYVELHTTEDSSMKELKREEEGMFH
LSRREILSKFHRRKIIDNTN**SPT**NTVY**PT**N**PTP**VI**TPS**D**TPT**IIAV**PS**TNPVTI**SPT**N**PA**AMPV
TV**PS**T**TP**VVPL**APTTPT**I**TPAP**VFN**PA**T**TPT**TVPG**AP**PVTSYPPPVTSYPPPLGNVPVVNPQQP
P**PS**NTN**APS**IQGQSWCVAKTG**AP**QASLQSALDYACGNGADCSQIQQGASCY**SP**VTLQNHASFAF
NSYYQKK**PAPT**SCDFGGAAMLVSSN**PS**SGSCIY**PS**SSSSSSSTST**SP**MIS**SPAPPT**QSTSTSIP
P**PS**LT**TPSPS**I**PT**I**APPT**MST**APS**SI**PTAPPT**SSGTFGYG**TPPS**VLNSSN**PA**SGTMPDFGSD**SP**
PIVNTTSASHPRALKPFTGCIVLMIPFVTASLSMRL

>Medtr1g086390.1 _ polygalacturonase_glycoside hydrolase family protein _ HC _ chr1_38659765_38662894 _ 20130731 ; MatureChain: 25-477
RHHFHTKHKHYSHSHNSSSEISLP**PAP**L**PSPT**SS**AP**EE**APS**ASA**SPSPSP**EVASGLLDVRKFGA
IGDGITDDTESFKMAWDTACQSELDLNVIFVPPGFSFIVQSTIFTGPCKGGLVLKVDGTIM**TP**D
GPESWLKNNSRRQWLVFYRVNGMSLEGSGTIDGRGQKWWDLPCKPHKGPNGTTLPGPCD**SP**VAI
RFFMSSNLTVQGLRIKN**SP**QFHFRFDGCQSVHVESIFIT**APALSP**NTDGIHIENTNDVKIYNSV
VSNGDDCVSIGSGCYDVDIKNITCGPGHGISIGSLGNHNSRACVSNITVRDSVIRVSDNGVRIK
TWQGGSGSVSGVTFSNIHMDTVKNPIIIDQFYCLSKDCSNKTSAVFVSDIVYTSIKGTYDIRHP
PMHFACSDSIPCTNLTLSDIELL**PS**QGDMLNDPFCWNAYGNSETLTIPPVFCLLDGIPQSI**PA**N
DIDHC

>Medtr1g086880.1 _ plant invertase_pectin methylesterase inhibitor _ HC _ chr1_38897793_38898557 _ 20130731 ; MatureChain: 25-254
V**PS**THISTSV**PAPAPS**IHYVDYILPKRKPNAGIGLFENSVPLKKQNSASV**SP**DDSDFQFEDPNL
LGSIAADVAKNADPEIVKLCVNGEN**PA**LCAATISSLLKGPFDPLKALEIEVDFTLKQAKSVAAI
ITALLNDPNTDKKAMKALQICQIQYNSMLDAINETVELLGQHNVVDSFYKFSSVISYKTTCEDA
FVK**SP**GVEIPFSQDSRTLFDLGGNCLGIMNTLVNNHKF

>Medtr1g086890.1 _ plant invertase_pectin methylesterase inhibitor _ HC _ chr1_38899955_38900695 _ 20130731 ; MatureChain: 24-246
V**PS**TRIDDSISTTV**PTPTPS**SEVIDYIVPEQTDFGLNMPENSA**SP**DDFVFRLEDPNLLGTVSED
IAKNADPEIVKLCVDGE**SPS**LCAATISSLLKGPFDPLKALEIEVDLTLQQAKSVAAIITELLKD
PSTDKKAVKALEICQTQYKSMLDAINETVELLGQHNVVDSFYKFSSVISYKTTCEDAFVK**SP**GV
EMPFSRDSSTLFDLGGNCLGIMNTLVNNHKI

>Medtr1g116870.1 _ carbohydrate_binding X8 domain protein _ HC _ chr1_52786322_52784446 _ 20130731 ; MatureChain: 28-362
DANRSIRLLK**AP**KHRRIVINKNPKKLKIMKHFDFD**PS**IYSSNAQPYGIS**SP**LSLPPYESL**AP**GK
SPPYCVYPP**PS**SS**TPS**TTI**PTPT**SSQ**PT**Q**PSP**PYT**SP**DL**PS**Q**SP**PPG**PT**TV**TPSP**PENF**PTPTP**
EIV**PSP**PDNF**PTPTP**EIV**PSPPS**NIPG**SP**EPILNPPIIFPGPPG**PS**M**SP**PYFE**PAP**PYYE**PTP**P
FI**PSPT**GGSGSI**PSPPS**TFQ**SPS**GGTI**PSPT**VYQPPVVYPP**PS**VPPRSNT**AP**QASLWCVAKASV
PDPIIEEAMNYACWSGADCSSIQPNGPCFQPDSVFAHASYAFNSYWQRTKASGGTCEFGGTAVL
VSVD**PS**YDGCHFIYN

>Medtr2g032710.1 _ polygalacturonase plant_like protein _ HC _ chr2_12322300_12315408 _ 20130731 ; MatureChain: 24-463
RYHHKKPKKT**SPAPS**D**PSPPSPS**F**PS**DPYPYPNDPGE**SPS**NCVFDVRSFGAVGDGDADDTAAFR
AAWKAACAVDSGVLL**AP**ENYCFKITSTIFSGPCKPGLVFQIDGTLM**AP**DGPNCWPEADSKSQWL
VFYRLDQMTLNGTGIIEGNGEQWWDLPCKPHR**SP**DGKTVSGPCV**SPT**MIRFFMSSNLVLRGLKI
QN**SP**QFHVKFDGCQGVLIDELSIS**AP**KL**SP**NTDGIHLGNTRDVGIYNSLISNGDDCISIGPGCS
NVNVDGVTC**APT**HGISIGSLGVHNSHACVSNLTVRNSIIKESDNGLRIKTWQGGTGSVTGLTFD
NIQMENVRNCINIDQFYCLSKECMNQTSAVYVNNISYRKIKGTYDVR**TP**PIHFACSDTVACTNI
TLSEIELLPYEGELVDDPFCWNAYGRQETLTIPPLDCLREGQPETVVELSEYECNN

>Medtr2g032710.2 _ polygalacturonase plant_like protein _ HC _ chr2_12319301_12315408 _ 20130731 ; MatureChain: 24-463
RYHHKKPKKT**SPAPS**D**PSPPSPS**F**PS**DPYPYPNDPGE**SPS**NCVFDVRSFGAVGDGDADDTAAFR
AAWKAACAVDSGVLL**AP**ENYCFKITSTIFSGPCKPGLVFQIDGTLM**AP**DGPNCWPEADSKSQWL
VFYRLDQMTLNGTGIIEGNGEQWWDLPCKPHR**SP**DGKTVSGPCV**SPT**MIRFFMSSNLVLRGLKI
QN**SP**QFHVKFDGCQGVLIDELSIS**AP**KL**SP**NTDGIHLGNTRDVGIYNSLISNGDDCISIGPGCS

NVNVDGVTC**APT**HGISIGSLGVHNSHACVSNLTVRNSIIKESDNGLRIKTWQGGTGSVTGLTFD
NIQMENVRNCINIDQFYCLSKECMNQTSAVYVNNISYRKIKGTYDVR**TP**PIHFACSDTVACTNI
TLSEIELLPYEGELVDDPFCWNAYGRQETLTIPPLDCLREGQPETVVELSEYECNN
>Medtr2g035760.1 _ polygalacturonase QRT3_like protein _ HC _
chr2_15158916_15161418 _ 20130731 ; MatureChain: 25-497
DKNL**PA**RKISDGNYHVTMRKLQSFKTSLTRHDSFAS**TPS**SSF**APSPS**SQ**PA**EGMNNPRVYHVTS
YGAD**PT**GNSDSTEALLAAIADATNG**PS**EGYLMEGISNLGGAQINLEGGNYMIRRSLKLPVSGVG
NLMIHGGTIKASNDFPNDGYIIDLSTSSNENDGKN**SPS**SSYNFEYITLKDLLLDSNFRGGGISV
INSLRTNIDNCYITHFTTNGILVQSGHETYIRNSFLGQHITAGGDKNERNFSGTGINIQGNDNA
VTDVVIFSAAIGIMVTGQANTFSGVHCYNKATGFGGTGIYLKLPGLTQTRIVNSYMDYTSIVAE
DPVQLHISSSFFLGDANIVLKSMKGVLNGVTIVDNMFSGSNQGVEVIHLDKSNGPFHQIDQVTV
DRNVATGMNLKATVAKRSLQGNGTSWNVDFNNILLFPNLIKNVQYSLSSTGSSFPNHAIRNVSD
NRVVIETNEAVAANVFVAVDQSMSS
>Medtr2g035790.1 _ polygalacturonase QRT3_like protein _ HC _
chr2_15178610_15181077 _ 20130731 ; MatureChain: 24-488
FNEHVQLSHLMKKIAKHKVDIALSQP**PSPSPSPSPSPS**SHTKKSGRVVYPIEYGAD**PT**GVNESS
DAMMKAVEAAFDIDNLGLELLLGIRDLGGVIIDFQGGNYKISNPITF**PS**SSGNLVVKGGTLRAS
NTF**PT**DRYLVELCASSSKVLQNATTAYNNKLLQQTIGIYYEDITFRDILFDSSYRGGGIFIVDS
ARIRIDNCFFLHFNTEGIKVQSGHETFISSSFLGQHSTVGGDKGERQFSGTAIDLASNDNAITD
VAIFSAAIGIVVRGQANIITGVHCYNKATGFGGIGILLKLAGNSQTRIDNCYMDYNSIVMEDPV
QVHVTDGFFLGDANIVLKSIKGKVYGLNIVNNMFSGNPNNNVPIVKLDGGFSNIDQVVIDRNNV
IGMILRSTVGKLSVDGNGTKWVGDFSNVLVFPNRISHFQYSFHTLEGPKFVAHSVSNVSNNVVV
VESEKPVHGVVSYFVEQ
>Medtr2g042145.1 _ Thionin related (TAP1) _ HC _
chr2_18396381_18396835 _ 20130731 ; MatureChain: 22-113
DD**SP**L**PSPSP**HISISEIICFGKCGFQCNNLLGEIQSFVACFAGCGYTCLDVTSKSAYDCATSCA
ISKSINDNIDARGVNLIVNSCVKDCKNK
>Medtr2g042163.1 _ Thionin related (TAP1) _ HC _
chr2_18402300_18402754 _ 20130731 ; MatureChain: 22-113
DD**SP**L**PSPSP**QISISEIICFGKCGFQCSNLLGEIQSFVACFAGCGYTCLDVTSKSAYDCATSCA
ISNSINDNIDARGVNLIVNSCLEDCKNK
>Medtr2g095080.1 _ Ripening related protein family _ HC _
chr2_40617795_40616664 _ 20130731 ; MatureChain: 22-222
PFLTDAISSCNGPCRTLNDCAGQLICINGKCNDDPDVGTRICSQ**PSPSP**PP**PS**GGGGTCRTSGN
LQCDGQSYPQYRC**SP**PYVSSGTQALLTLNDFSAGGDGGGASQCDEIFHSNSERIVALSTGWYSG
GSRCGKFIRIRASNGRTVTAKVVDQCDSVNGCDDEHAGQPPCRNNIVDGSAAVWDELRLNKNDG
VVPVTWTMA
>Medtr2g095130.1 _ Ripening related protein family _ HC _
chr2_40628355_40627563 _ 20130731 ; MatureChain: 28-220
ISSCNGPCNTLNDCDGQLICINGKCNDDPDIGTHICTN**PSPSPPS**GGGGGTCQSSGTLQCKKKS
YPQYRC**SP**PVSSSTQASLTLNDFSEGGDGGG**PS**QCDEKYHDNSERVVALSTGWYNGGSRCGKMI
RITARNGRSVTAKVVDQCDSVNGCDKEHAGQPPCHNNIVDGSVSVWNALGLNTDDGVVPVTWSM
A
>Medtr3g435540.1 _ GDSL_like lipase_acylhydrolase _ HC _
chr3_11669121_11665026 _ 20130731 ; MatureChain: 21-364
KMVQFIFGDSLSDVGNNKYLSKSLAQASLPWYGIDIGNGLPNGRFSNGRTVADIIGDNMGLPRP
PAFLD**PS**LSEDVILENGVNYASGGGGILNETGSYFIQRFSLYKQIELFEGTQELIRSKIGKFEA
EKFFQEARYVVALGSNDFINNYLMPLYSDSWTYNDETFIDYLVGTLQEQLKVLHSLGARQLMVF

GLGPMGCIPLQRVLSTSGNCQEKTNKLALSFNKASSKVVNDLGKQLPNASYRFGDAYDVVNDVI
SN**PS**KYGFQNAD**SP**CCSFGRIR**PA**LTCI**PA**STLCKDRSKYVFWDEYH**PS**DKANELIATELIKKF
GFKRVDQAE**TPSPSP**EV**APSPS**DD
>Medtr3g040680.1 _ non_specific phospholipase C4 _ HC _
chr3_14343115_14337780 _ 20130731 ; MatureChain: 32-530
IEENPIKTIVILVMENRSFDHMLGWMKKLNPNINGVTGLESSQLSVSDPNSKRIFFKNEAQYVD
PDPGHSFQAIREQIFGSEDTSVD**PAP**MNGFAQQAYSMDNTTKMSQEVMNGFEPDHVAVYKTLVS
EFAVFDRWFASV**PA**STQPNRLYVHSATSGGATSNVASLLAKGYPQQTIFDSLHDNGFDFGIYYQ
NI**PA**TLFYRNLRKLKYIPKFHLYDLTFKHHAKEGKLPNYVVVEQRYTDTKLF**PA**NDDH**PS**HDVY
QGQMFVKEVYETLRA**SP**QWNETLFLITYDEHGGFFDHVA**TP**VRGV**PSP**DGIVGPEPFNFTFNRL
GVRV**PT**IAI**SP**WIEKGTVVHGPNG**SPTPT**SEYEHSSIAATVKKIFNLPKFLTKRDEWAGTFEGI
VQIRTEPRTDCPEQL**PTP**IKLRKGEANEDAKLSEFQQELIQLAAVLKGDNIFTSYPNTIGKDMS
VKQGKYYMDEAVKRFFEAGRYAKKMGVSDENIVKMK**PS**LTTRSSKSSNTKP
>Medtr3g065460.1 _ glucan endo_1,3_beta_glucosidase_like
protein _ HC _ chr3_29590167_29594105 _ 20130731 ; MatureChain:
19-491
DEEPFIGVNIGTDLSDMPH**PT**QVVALLKAQQIRHIRLYNADQAMLTALSKSGIQVVISVPNEEL
LAIGQSNSTASNWVSRNVLAYY**PA**TNITAICVGSEVLTTLPNVAKVLVNALNYIHSALVASNLD
RQIKVS**TP**L**PS**TMILDSFP**PS**QAFFNTSMNQVLKPMLDFLQSTQSYLMLNVYPYHDYMESNGVI
PLDYALFKPIPPNKESIDSNTLLHYSNVFDAVVDAAYFAMSYMNYTNIPVVVTETGW**PS**KGDSN
EPGATLANANTYNSNLIKHVLNKTG**TP**KLPGIGVSTYIYELYNEDKQPGALSEKNWGLFDSNGV
PVYVLQLTGSGAVLANDTKGETYCVAKDGADPKMLQAGIDWACGPGKVDC**SP**LLQGKPCYDPDN
VVAHANYAFDAYYHKMGKSTESCNFNDMATISTSD**PS**HGSCIFPGSLGFSN**APAPAPS**HFSGCT
LLRCELRIRSLLIVIGLLISEVVLL
>Medtr3g072410.1 _ LysM domain GPI_anchored protein _ HC _
chr3_32553629_32557165 _ 20130731 ; MatureChain: 22-412
KSTIEPCTTSDTCNSLLGYTLYTDLKVSELSSLFQIDPISLLTANSIDISYPDVEHHIL**PS**KLY
LKIPIQCSCIDGIRKSVSTNYKIR**PS**DTLSSIADSIYGGLVSSDQLREANSVTDPNVLDVGQNL
VVPLPCTCFNGSDNGL**PA**IYMSYVVQPLDSLNNIAARYFTTLTDLMNVNAMGTTGISAGDILAI
PI**PA**CASKFPKDSADFGLLVPNGSYAITAGHCVQCSCGPRNLNLYCM**PA**SLAVSCSSMQCKSSN
LMLGNVTVQQTSGGCNVTSCTYDGIVNGTIATTL**TPS**LQPRCPGSQEFPPLI**APPT**VVPRESLF
APAPSPSSLFDGTDGR**SP**KSSVV**PS**TGF**TPA**LG**PS**GISSGASAACSLVKPL**PT**LTLTLVLLFVK
LMIPVAL
>Medtr4g108690.1 _ GDSL_like lipase_acylhydrolase _ HC _
chr4_45012744_45015362 _ 20130731 ; MatureChain: 25-368
KVVQFIFGDSLSDVGNNMHLSRSLAQASLPWYGIDMGNGLPNGRFSNGRTVADIIGDSLGLPRP
PPVLDTSLTEKDILINGLNYASGGGGILNETGTYFIQKLSLDKQIELFQGTQKLIRSKIGKRAA
DKFFREAQYVVALGSNDFINNYLMPLYTDSWTYNDETFMDYLIGTLRRQLKLLHSLGARQLQLF
GLGPMGCIPLQRVLTTTGNCRESVNKLALSFNKASSELIDDLVKQLPNSNYRFGDAYDVVSDLI
SNPLKYGFQNSD**SP**CCSFGRIR**PA**LTCV**PA**STLCSDRSKYVFWDEYH**PS**DSANELIANELIKKF
GFLRDGQNA**APSPAPA**AI**APSP**ED
>Medtr4g116990.1 _ glycoside hydrolase family 18 protein _ HC _
chr4_48456229_48454178 _ 20130731 ; MatureChain: 22-385
AKSSKS**TPSPS**STTRVKGIYWIENPLFP**PS**SIDTSLFTHIFYAFV**SP**NKFTYKLEEEEEDSTTV
ATSLTTFTNTFKTK**TP**PI**PT**LLSIGGATSNSTLFAFIASD**PT**ARATFINSTIQVARTFGFDGID
FDWEF**PT**TTKEMNDLGELLFQWRRAISDEATSTSRPPLLLTAAVYFAVNFFLSGERRMYPVDSI
NKNLDWVNVMSYDLLGSGSNVTG**APS**GMFDSKSNVSVVSGLFSWIRGGV**AP**EKIVMGMPLYGKS
WKLQDPNVHGIG**AP**NVGPGPGVDGGMAYFQVVDFNKQMGAKVVYDKETGSVYSYSGSTWIGYDD

PFTVSVKVGFAQALKLGGYFFWAAGYDTSDWKVSTQASKAWRPE
>Medtr5g009100.1 _ GDSL_like lipase_acylhydrolase _ HC _
chr5_2093629_2091194 _ 20130731 ; MatureChain: 27-382
QPQPQ**PA**IP**PSPSP**PPL**APA**LFVIGDSSVDSGTNNFLATFARADRLPYGRDFDTHQ**PT**GRFSNG
RIPVDFLASRLGLPFV**PS**YLGQRGNVEDMIHGVNYASAGAGIIVSSGSELGQHISLTQQVQQFT
DTFQQLIISMGEDAAKTLISNSIVYISIGINDYIHYYLLNASNVDNLFLPWHFNRFLASSLMRE
IKNLYNLNVRKMVVMGL**AP**IGC**AP**RYMWEYGIQNGECVEPINDMAIEFNFLMRYIVEKLAEELP
DANIIFCDVYEGSMDILKNHDQYGFNVTSEACCGSGKYKGWLMCL**SP**EMACSNASNYIWWDQFH
PTDTVNGILAANIWNGEHAKMCYPMHLQDMVIQKAN
>Medtr5g078210.1 _ GDSL_like lipase_acylhydrolase _ HC _
chr5_33423323_33421258 _ 20130731 ; MatureChain: 22-353
AKLVTYIFGDSLTDVGNNNFLQYSLAKSNFPWYGIDYSGGQATGRFTNGRTIGDIISSKLGI**PS**
PPAYLSVPQNVDALLKGVNYASGGAGILNDTGLYFLQRLTFDDQIKSFKKTKVAITAKLGEDAA
NKHFNEATYFIGIGSNDYVNNFLQPFMADGQQYTHDEFIELLISTLDQQLKRLYQLGAQKMVFH
GLGPLGCI**PS**QRVKSKRGQCLKQVNEWIQQFNSKVQKLIIKLNRGLPNAKLVFADTYPLVLDLI
DN**PS**TYGFKVSNTSCCNVDTSIGGLCLPNSKLCKNRNEYVFWDAFH**PS**DAANAILAEKFFSSLF
SS**APS**A**APSPSP**
>Medtr5g081840.1 _ pectinesterase inhibitor domain protein _ LC
_ chr5_35079937_35079313 _ 20130731 ; MatureChain: 22-175
ISL**PTPSP**KLYKSVCKEPRQKEFEQRCLKLIEAYPKITLIKDYLTFCRSFLKTVAIKKAIKSQH
RVKEIVKKY**PS**SQPIKECGDDYNTVVSEVQGALREDPEMISLAVKYASDALDMCERSLANEKIV
NTSSIAALNHEMMLYTDIVVIAGGHL
>Medtr6g079630.1 _ chitinase _ HC _ chr6_29058520_29059610 _
20130731 ; MatureChain: 25-306
TRLSSSTNLIFREYIGAESNNIKFSDVPINPNVEFHFILSFGIDYDTSS**SPSPT**NGKFNIFWDT
KNLN**PS**QVSSIKSQNPNVKVALSLGGDSVEGGYAYFD**PS**SVESWLSNAVSSLTKIIKEYNLDGI
DIDYEHFKGNPNTFAECIGRLIKTLKANGVITFASI**AP**FDDDQVQSHYLALWKSYGHLIDYVNF
QFYAYDKGTSVSQFIDYFNKQSSNYNGGKVLVSFLSDGSGGL**SPS**DGFFKACQRLKSQQKLHGI
FVWSADDSMGNGFRFEKQSQALLAIH
>Medtr6g083270.1 _ Chitinase _ Hevein _ PR_4 _ Wheatwin2 _ HC _
chr6_31175338_31171217 _ 20130731 ; MatureChain: 23-457
QSANNVTATYQSFNLTQTKQGIGNKWNLNTSGVFCAAQDGNKPLSWRSKYGWAAFCGPVGPVGK
SACGKCLNVTNQDGGNSAVQQKSI**TPT**TQTVRIVDECSNGGLDLDMDVFDLLDTSGDGNAQGYL
LVDYEFVDCGDLLASTLPKNFSNSSTTV**SPTPAPTPSPT**FL**SP**NPLSNLTQEIHRATTSSHSNW
QMKRKVIIGTKKALIYEFMSNGSLDKFIYNKGPETIASLSWENLYQIAKGIARGLEYLHRGCTT
RILHFDIKPHNILLDENLCPKISDFGLAKLCPKQESIISMSDQRGTMGYV**AP**EVWNRHFGGVSH
KSDVYSYGMMLLEMVGGRKNIIADASHTSEIYFPHWVYNRLELGTNLRPDGVMDTEEDEIARRM
TIVGLWCIQTF**PS**DR**PT**MSKVIEMLEVTMNSLEMPPKPLHS**SPT**RSVSESS
>Medtr8g042910.1 _ pectinesterase_pectinesterase inhibitor _ HC
_ chr8_16610530_16607577 _ 20130731 ; MatureChain: 27-616
QTS**SPSPSPS**HSS**SPS**S**SPSPS**HSS**SPSPSPSPSPS**H**SP**PPP**PS**TSTSSSKACKSTLYPKLCRSI
LSSIRS**SPT**DPYHLGKFSIKQSLKQAKKLSKVFKDFLKKHHHSTSLNHAEIAALDDCSELNQLN
VDYLELVRDELKSADSSSFNNTDVLVEKVETYLSAVATNHYTCFDGLVVTKSNIANALAVPLSN
VTQLYSISLGLATQALSKNIKKRKTRKQGL**PT**KAYKVRQPLKKLIKLLHTKHSCSNNSSNCLRH
ERILKESENRGILLKEYVIVSLDGTENFTSIGEAIDA**AP**HNLKAEDGYFLIYVREGVYEEYVTI
TKHKKNILLIGDGINKTCITGNHSVIDGWTTFNSSTFAVSGERFIAVDITFRNTAGPQKHQAVA
VRNNADLSIFYRCSFEGYQDTLYVHSLRQFYRECDIYGTVDFIFGNAAAVLQSCNIFARKPMAN
QKNAITAQGRTDPNQNTGISIQNCRIEAADDLAVDLNSTKNYLGRPWKVYSRTVYLQSYIGDLI

Q**PS**GWLEWNGTVGLDTLFYGEFDNYGPGAITHNRVQWHGYNLM**TP**NQAWNFTVLNFTLGNTWLP
DTDVPYTEGLHLDD
>Medtr8g074335.1 _ Chitinase (Class Ib) _ Hevein _ HC _
chr8_31432870_31435038 _ 20130731 ; MatureChain: 23-309
EQCGSQANRAVCPNGLCCSKFGWCGTTDQYCGAGCQSQCRSSS**TPTPSTPTP**GTGGGDVGRLV
PSFLFDQMLKYRNDARCPGHGFYTYDGFIAATRSFNGFGTTGDDTTRKRELAAFLAQTSHETTG
GWSS**AP**DGPYAWGYCFVNERNAQEKRYYGRGPIQLTHDYNYGQAGKAINQDLINNPDLVSTN**PT**
VSFKTAIWFWM**TP**QGNK**PS**SHDVIIGRW**TPS**GADRSAGRVPGYGVITNIINGGLECGHGQDARV
NDRIGFYRRYCQILGV**SP**GDNLDCNNQRSFA
>Medtr8g074350.1 _ Chitinase (Class Ib) _ Hevein _ HC _
chr8_31442722_31445255 _ 20130731 ; MatureChain: 24-320
QQCGRQANGAVCANRLCCSQFGYCGNTADYCGAGCQSQCTSN**PTPTPTTPTPS**GGDVGSLISSS
MFDEMLKYRNDPRCAARGFYSYDSFITAARSFNGFGTTGDENTRKREVAAFLGQTSHETTGGW**P**
TAPDGPYAWGYCFVNERNP**PS**DYC**SP**GTWPC**AP**GKRYYGRGPIQLTHNYNYG**PA**GRAINQDLIN
NPDLVSSN**PS**VSFRTALWFWM**TP**QGNK**PS**SHDVITGRW**TPS**DADRSARRVPGYGVITNIINGGL
ECGRGQDPRVEDRIGFYKRYCQLLRTTTGDNLDCYNQRPFA
>Medtr8g087870.1 _ GDSL_like lipase_acylhydrolase _ HC _
chr8_36358124_36355685 _ 20130731 ; MatureChain: 24-351
KPQVPCLFIFGDSLSDSGNNNNLATDAKVNYRPYGIDF**PA**G**PT**GRFTNGRTSIDIITELLGFDH
FIPPYANTHGADIVQGVNYASGAAGIRNETGTQLGPNISMGLQLQHHTDIVSQIAKKLGYDKVQ
QHLNKCLYYVNIGSNDFLNNYFLPQHY**PT**KGKYTTDQYAAALVQELSTYLKAIHGLGARKFSLV
GLSLLGCVPHEISTHGKNDSRCIQEENNAALLFNDKLKPLVDHLNKELTDSKFIFINSAVIRLS
QLKLQDLVKCCKVGSNGQCIPNTKPCKARNLHPFFDAFH**PT**EVVNKLSANLAYN**APAPS**FAYPM
DIGRLVKL
>Medtr8g093460.1 _ pectinesterase_pectinesterase inhibitor _ HC
_ chr8_39045135_39042219 _ 20130731 ; MatureChain: 19-561
THHHNHNHHQ**PTPSPS**V**SPT**AGSEIQQACKATRFPQQCESSLANLPPN**PT**ALQLIQSAINLSST
NLVTAQSMVKAILDSSSSSRNRTVAATTCIEILTNSQNRISLSNDALTHGKIKDARAWLTAALV
YQYDCWNSLKYANDTHAVGEAMSFIDSLETLTSNALAMAFSYDVYGKDTSFWKP**PT**TERDGLWQ
ATGSGGGSVSSVGI**PA**KL**TP**DVTVCKGGEKGCYKTVQEAVNA**AP**DNGVDRKRFVIYIKEGVYEE
TVRVPLEKRNVVFLGDGIGKTVITGSANVGQPGMTTYNSATVAVLGDGFMAKDLTIENTAGPDA
HQAVAFRLDSDLSVIENCEFLGNQDTLYAHSLRQFYKSCRIVGNVDFIFGNSAAIFQDCQILVR
PRQLKPEKGENNAITAHGRTD**PA**QSTGFVFQNCLINGTEDYMALYHSNPKVHKNYLGRPWKEYS
RTVFIHSILEVLV**TP**QGWMPWSGDFALKTLYYGEFENSGAGSDLSQRVSWSSKI**PA**EHVSSYSA
ENFIQGGEWMQSSHLSAHHGRSHHSTHRRHE
>Medtr0341s0010.1 _ salt stress response_antifungal domain
protein _ LC _ scaffold0341_1358_435 _ 20130731 ; MatureChain:
21-307
ATTDATGKFRYICDQGNDDGNYTTNSTYHTNLKTLLSTLTSNKDINYGFYNSSYGVNTDKVNAI
GLCRGDVKPDDCQNCLQNSTALLTQLCQNRKEAIGWYDDEKCMLRSSHRLIFGINEMG**PA**MIAW
NLQNATKEVEDEFDKGLKNLLYNLKNLAASGESDLKYAVGSDVIGPNNNQTIFALVQC**TP**DLSE
TLCDECLDKSIKEIPNCCDNRIGIGGRFLR**PS**CYLRYETNSLFYQQTQD**SPS**SSSS**SPSPTPSP**
ASV**PS**FS**AP**PPFAENTTS**SP**GIYHIKLMKRS
>Medtr4g036590.1 _ transmembrane protein, putative _ LC _
chr4_13242800_13243567 _ 20130731 ; MatureChain: 22-255
IEVGTHHHHHPYHP**PTPTPAP**VQP**PS**SYGIHSLRLDHHKKHPPHH**PSTP**RK**AP**VQP**PSPSP**I**PS**
NNLDNPQIERHKPHPHHP**PTPTPAP**FQSKSSSDIHSLCVDRRKKHRPHH**PSSP**RK**TP**VHP**PSPS**
PI**PS**NNLDNPQIERHKPHPHHP**PTPTPA**SVQPKSSSDIHSLRVDRHKKHRSHH**PSTP**RK**AP**VHP

PSPSPISSNSLDNPQIERRRRLPHHP**PTPTPAP**IQ**TPS**SSFL
>Medtr3g012570.1 _ hypothetical protein _ LC _
chr3_3549498_3548605 _ 20130731 ; MatureChain: 22-297
EDTFNM**SP**EEFISRHK**PS**SAH**SP**HHRP**PA**HSSHHHHH**SPS**HG**PS**HHHH**SPSPAPS**RHQKSHAR**S**
PHHHNHSLSHNLIRHRHSHAH**SP**HHQSH**SPSP**HHKSH**APSP**HHHDH**SPSPSPSP**HHHHHSLSLT
HFSHHK**SPAPSP**HHKSH**APSP**HHYSSTH**SP**HNQSH**APSP**HHQSH**APSP**RHHQH**SPSP**I**PSP**IHH
Q**SPA**H**SP**HHHSHAH**SP**HHKSH**SPSP**HHHH**SPSPTPS**RHHKSHAH**SP**HHHHH**SPSPTPSP**HHQ**SP**
AS**SP**RHQYYAMSS**SP**PLTVV
>Medtr1g055415.1 _ pollen Ole e I family allergens _ HC _
chr1_24541448_24539849 _ 20130731 ; MatureChain: 22-343
EKLETLHHK**PATP**HH**SPT**K**SP**VHKPLA**SP**PHHNH**SPS**H**APS**HVH**TP**LHPPH**PA**KP**PT**HHHHQHQ
HH**SPSPTPS**HVHPPLHPRH**PA**KP**PT**HHHQQH**SPA**H**SP**IK**SP**VH**TP**LHPPH**PA**KP**PT**HHRHQHH**S**
PSPAPSHVH**TP**LHPRH**PA**KP**PT**HYHHH**SPA**H**AP**IKPPVHKPLLPPHSAK**SPT**HHHP**PA**H**APT**HT
HVSRNLIAVEGVVYVKSCNHTGVDTLKGA**TP**LPGAIVKLQCNNTKYKLVLKAKTNKKGYFYIGG
PKNITGYSTRHCNVVLDS**AP**KVLK**PS**NLHGGVIGALLKLVKRSMSKGVYVKLFSVGPFAFEPKC
HH
>Medtr4g017240.1 _ pollen Ole e I family allergens _ HC _
chr4_5341759_5343041 _ 20130731 ; MatureChain: 22-347
EDIFDALGDISDFINPFNAFSKESETLHK**PT**HAHSLHHQ**SPTPSP**NHHLH**SPSPSPSP**YHQLHA
S**SP**HHHYH**SPSPAPS**HHHQY**PTPSP**HHQPH**APSP**HHHSL**SPAPS**LPHQ**SPAPSP**HHQ**SPA**R**SP**H
HHHH**SP**I**PSPSP**HHHNH**SPSPSPSP**HHQ**SPTPSP**RNQSL**APS**SHHQSTIV**PAKSP**NYHHP**PA**H**S**
PSPLRS**TP**IRGSVMVVQGVVYVKSCKSSGLDTLKGA**TP**LLGAVVKLQCNSTKYKLDETDKDGHF
SLVGPKIITLFTAKQCNVVLVS**AP**HGPK**PS**NLHHGITGAILRPKRRFVFKGVPFILFATQPLAF
EPICPC
>Medtr3g012020.1 _ pollen Ole e I family allergens _ HC _
chr3_3176250_3174857 _ 20130731 ; MatureChain: 22-393
EELETFHHK**PSSP**LHP**PT**K**SP**IHKPLA**SP**PHHNH**SPS**H**APS**HVH**TP**LHPPH**PA**KP**PT**HHRHQHH
SPSL**APS**HVH**TP**LHPRH**PA**KP**PT**HHHHHHHH**SPA**H**AP**IK**TP**VH**TP**LHPPH**PA**KP**PT**HHHHQHH**SP**
SPAPSHVH**TP**LHHRH**PA**KP**PT**HHHHHH**SPA**H**AP**IK**PS**FH**TP**LHPPH**PA**KP**PT**HHHHQHH**SPSPA**
PSHVH**TP**LHPRHPKKP**PT**HYHHH**SPA**H**AP**IKP**PT**HHHHLP**PA**H**APT**H**TP**LVSRSLIAVEGVVYV
KSCNHTGVDTLKGA**TP**LFGAVVKLQCDNAKHKLVLKAKTDKKGYFYIGGPKNIAGYSTRHCNVV
LDT**AP**KVLK**PS**NLHGGLSG**AP**LLKPVKRPVSKGVSLKLYTVGPFAFEPKCHH
>Medtr4g066680.1 _ pollen Ole e I family allergens _ HC _
chr4_25222320_25223534 _ 20130731 ; MatureChain: 22-338
EDPVDALNDIFGLLDPFNAFSKESETLHK**PS**HAHSLHHESL**APSP**HHHLH**SPS**HSHIPHHQ**SPA**
PSRHHHHH**SPSPAPS**RHHQ**SP**P**PS**SHHKPH**APSP**HQH**SPSPAPSP**HHQSL**APSP**HHHHH**SPSPS**
PSPHHHNR**SPSPS**S**SP**RHQ**SPAPSP**HHQ**SPAPSP**RHQSTIV**PAKSP**NHHNP**PT**H**SPSP**L**SPTP**V
PRSVMVIRGVVYVKSCKYSGLDTLKEAKPLLGAVVKLQCNNTKYKLDETDKDGHFSLVGPKIIT
IYTAKQCNVVLVS**AP**HGLK**PS**KLHDGISGAILRPKRRFVSKGVPFILFATQPLAFEPNCPR
>Medtr4g088735.1 _ pollen Ole e I family allergens _ HC _
chr4_35327598_35325086 _ 20130731 ; MatureChain: 22-276
EELETL**TPT**YPPHT**TPAP**LHP**PA**N**AP**HHHHHHHIH**SPTPAPTPTPSPSP**IH**TP**LHPPYHS**AP**V**P**
AKP**PT**HGHHHHHPHP**PAPTP**VH**TP**V**APA**HPPLHPPVH**TP**VV**PT**HPPLHP**PAPA**HPPLH**PTP**LPR
SFIAVQGVVYVKSCKYAGVDTLLGATQILGAVVKLQCNNTKYKLVQKVQTTDKNGYFFIEGPKN
ITSYAAHKCNIVLIS**AP**NGLK**PS**NLHGGLTGAGLRPEKPFVAKGLPFILYTVGPLAFEPKCPR
>Medtr4g088740.1 _ pollen Ole e I family allergens _ HC _
chr4_35335943_35334462 _ 20130731 ; MatureChain: 22-402
EELESLH**TPT**QLHP**PAKSP**HHYHHN**PSPTPS**RSHVH**AP**LHPPHHTKP**PT**HRHHQHP**PA**H**AP**VQP

PTHHHH**SPA**H**AP**VHKHVPVHPPHHAKP**PT**HHRHQH**TPA**H**SP**IQP**PT**HRHH**SPA**H**AP**IHKHVPVH
PPHHAKP**PT**HHRHQHP**PT**H**SP**VQP**PT**HHHQP**PA**H**AP**VHKHVPVHPPRHAKP**PT**HHNHQRPLTH**A**
PIQP**PT**HHHP**PA**H**APA**HKHIPIHPPFHAHV**PT**K**SPTP**NPHPRHHH**TPA**L**APA**H**TP**LRSIFHPRS
LIAVQGVVYVKSCKHAGGDTLLGATSIIGAVVKLRCNNTKQYVVLKGKTDKNGYFYIKGSNNIS
SYAARKCNVVLVS**AP**NGLK**PS**NLHGGITGAVLKAKKSFVSEGHPFILYNV**AP**LAFEPKCSH
>Medtr4g124400.1 _ pollen Ole e I family allergens _ HC _
chr4_51457430_51458050 _ 20130731 ; MatureChain: 18-206
FIAFAQELYSGH**PA**YPPH**APAP**LHP**PA**N**AP**HHHHHHHN**PSPTP**PPFPHTSLHP**PT**KSSVHPRAN
TPHHYHSSSS**APT**NVH**TP**VHPH**PT**YKP**PA**HHHHHHHP**PA**H**APA**VNTLVV**PT**NPPLHTIV**PS**KP**P**
TPNHHHP**PAPA**RVR**TP**VV**PT**HPQLH**PTP**CPRSFNVVQGIVHVKSCEYDGLNSLLGATKLLG
>Medtr5g047890.1 _ pollen Ole e I family allergens _ HC _
chr5_20963820_20965177 _ 20130731 ; MatureChain: 20-342
FTEELETFHNK**PATP**LHP**PT**K**SP**VHKPHHNH**SPS**H**APS**HVH**TP**LHPPH**PA**KP**PT**HHHHQHQHH**S**
PSPTPSHVH**TP**IHPRH**PA**KP**PT**HHHHHHH**SPA**H**AP**IKPPVH**TP**LHHPH**PA**KP**PT**HHHHQHH**SPS**
PAPSHFH**TP**LHPRH**PA**KP**PT**HYHHH**SPA**H**AP**IKPPVH**TP**LLPPQSAKP**PT**HHHHHP**PA**H**APT**H**T**
PLLSRSLIVVEGVVYVKSCNHNGVDTLKGA**TP**LLGAIVKLQCNNAKHKLVLKAKTDKKGYFYIG
GPKNIVGYSTRHCNVVLDN**SP**KALK**PS**NLHGGLTGALLKPVKRSVSKGISLKLFTVGPFAFEPK
CHH
>Medtr6g042490.1 _ pollen Ole e I family allergens _ HC _
chr6_14675282_14674127 _ 20130731 ; MatureChain: 23-255
EELETLHHK**PTTP**LHP**PT**K**SP**VHKPLAKP**PT**H**AP**HHHHHH**SPS**H**AP**LPPPH**PA**KPLTHHHHQHQ
HH**SPAPSP**YHV**PTP**LQR**PA**KP**PT**LHHHQHP**PA**H**APT**HMPRVSRSSIAVEGVVYVKSCHHAGFDT
LKGATLLFGAVVKFQCHNAKYKFVLKAKTNKEGYIYIGSSKNISSYASGHCNVVLES**AP**NGLK**P**
SNLHGGLTGAHPKSVKRIVSKGVSLIRYTVDPLAFEPKCNH
>Medtr7g016950.1 _ pollen protein Ole E I_like protein _ HC _
chr7_5358846_5361469 _ 20130731 ; MatureChain: 21-465
SHHHHHHKKPLSAVVVGTVYCDTCFQQDFSMGNHFISGASVEVECKDGNEISKPRFKKQVKTN
EHGEFKIQLPFSVSKHVKRIKGCVVKLVSSNEPFCSIASAASSSSLHLKSRKQGLHIFSAGFFS
FKPLKQPNLCNQK**PS**VVQNTKKLLDSLKKTSFPPKID**PS**FPPPLQDPNP**PS**GVLPFLPPVPLVP
EIL**TP**ILPPEL**SP**LI**PS**GMTSEESKYKSTKTSKNLDEKNVINLDTFNLPPNPFL**SP**PLVPNNPL
QP**PT**S**TP**LVPNPLQPPPLVPNNPLQP**PSTP**HFPDPFH**PT**IPLLHQILPPIL**SP**LVPNGMTSDAS
KSKSTKTTKNVQSLDEKKATNLDSFNLPPNPFFPPPLLPNNPLQP**PSTPT**IPNPFQP**PTPTP**LV
PNNPFLPP**PS**GS**SP**LFPF**PS**VPGL**SPSP**P**PSSP**PGLAFPFPPLFPPPGSG**TPPA**STKNV**SP**
>Medtr8g092230.1 _ pollen protein Ole E I_like protein _ LC _
chr8_38542715_38541582 _ 20130731 ; MatureChain: 22-296
EDTSDTFNDIFDFDADDDALNAFSKEFETLHK**PS**HAHSLNHQ**SPA**S**SP**HHHFH**SPSPAPS**RHHQ
SLP**PS**LHQQPH**APSP**HHH**SPSPAPS**LPHQ**SPA**A**SP**HHQF**PAPSP**HHHLLPLLATNLPLRPLTTN
LPLPPLTTNLPLYLQNPQPITTL**PS**VMVVRGVVYVKSCNYSGFDTLKKA**TP**LLGPVVKLQCNNT
KYKLDETDKDGHFSLVGPKIITIYTAKQCDVVLVS**AP**HGLK**PS**NLHNGITGVILRPKRRFVSKG
VPFILFGTQPLAFE**PS**CPH
>Medtr4g019600.1 _ transmembrane protein, putative _ LC _
chr4_6171269_6171930 _ 20130731 ; MatureChain: 22-152
TTTNKPQISVQSIS**SPS**SN**APS**Q**SP**ISSQRTNEKESSIGRKFGKHQHDHIIT**APSPSP**LEGSIL
SHQKTSILDSQGHIHLVNHHHHHPFDKSMAGGGVILGGLATTFLVAVYCYIKATSKHNKLDNTT
TST
>Medtr1g039270.1 _ aspartic protease in GUARD CELL_like protein
_ HC _ chr1_14516688_14514095 _ 20130731 ; MatureChain: 20-469
ISFS**TP**LSYFQHLNVENAISETKLKPLKQQNHNTQQPQWKTKLFHRDNINLKKTTHKTRFISRI

NRDIKRVTFLLNRLNKNTQEQQTTTATEASFGSDVVSGTEEGSGEYFVRIGIG**SPA**IYQYMVID
SGSDIVWIQCEPCDQCYNQTDPIFN**PA**TSASFIGVACSSNVCNQLDDDVACRKGRCGYQVAYGD
GSYTKGTLALETITIGRTVIQDTAIGCGHWNEGMFVGAAGLLGLGGGPMSFVGQLGAQTGGAFG
YCLVSRGTGSSGSLEFGRQAMPVGAMWVPLIHNPFY**PS**FYYVSLSGLAVGGIRVPISEQIFQLT
DIGTGGVVMDTGTAITRL**PT**VAYNAFRDAFIAQTTNLPR**AP**GVSIFDTCYDLNGFVTVRV**PT**VS
FYFSGGQILTF**PA**RNFLI**PA**DDVGTFCFAF**APSPS**GLSIIGNIQQEGIQVSIDGTNGFVGFGPN
VC
>Medtr2g016180.1 _ papain family cysteine protease _ HC _
chr2_4899750_4901997 _ 20130731 ; MatureChain: 27-490
I**PS**EYSILAFDLNKF**PS**EEQVVELFQQWKKEHQKFYIHPEEAALRLENFKRNLKYIVERNAMRN
SPVGHHLGLNRFADMSNEEFKNKFISKVKKPISKRASNLHVKVESCDD**AP**YSLDWRKKGVVTGV
KDQGNCGSCWSFSSTGAIEGVNAIVTGDLISLSEQELVDCDTTNDGCEGGYMDYAFEWVINNGG
IDTEADYPYIGVGGTCNVTKEETKVVTIDGYTDVTQSDSALFCATVKQPISVGIDGSTLDFQLY
TGGIYDGDCSSNPDDIDHAVLIVGYGSDGNQDYWIVKNSWGTSWGIEGFIYIRRNTNLKYGVCA
INYMASF**PT**KESTSI**SPTSPPSPPSP**PP**PTPPSPTPS**KCGDFSYCTTEETCCCLYELFDFCLAY
GCCEYENAVCCTGTKYCC**PS**DYPICDTEDGLCLQNYGDLMGVAAKKKKMGKHKFPWTKYEQTKK
THYPLQLRRGAFATVR
>Medtr2g016270.1 _ papain family cysteine protease _ HC _
chr2_4947144_4949388 _ 20130731 ; MatureChain: 27-475
I**PS**EYSILAFDLNKF**PS**EEQVVELFQQWKKEHQKFYIHPEEAALRLENFKRNLKYIVERNAMRN
SPVGHHLGLNRFADMSNEEFKNKFISKVESCDD**AP**YSLDWRKKGVVTGVKDQGNCGSCWSFSST
GAIEGVNAIVTGDLISLSEQELVDCDTTNDGCEGGYMDYAFEWVINNGGIDTEADYPYIGVGGT
CNVTKEETKVVTIDGYTDVTQSDSALFCATVKQPISVGIDGSTLDFQLYTGGIYDGDCSSNPDD
IDHAVLIVGYGSDGNQDYWIVKNSWGTSWGIEGFIYIRRNTNLKYGVCAINYMASF**PT**KESTSI
SPTSPPSPPSPPP**PTPPSPTPS**KCGDFSYCTTEETCCCLYELFDFCLAYGCCEYENAVCCTGTK
YCC**PS**DYPICDTEDGLCLQNYGDLMGVAAKKKKMGKHKFPWTKYEQTKKTHYPLQLRRGAFATV
R
>Medtr4g084020.2 _ trypsin family protein _ HC _
chr4_32707882_32704601 _ 20130731 ; MatureChain: 29-456
DVDVVEFSYYG**APAPTP**KEQLYTELADGLRGSDSCVGSGSQVASQETYGTLGAIVRSRTGNREV
GFLTNRHVAVDLDYPNQKMFHPLP**PS**LGPGVYLGAVERATSFITDDLWYGIFAGTNPETFVRAD
GAFIPFAEDFNMNNVITSIRGVGDIGEVHRIDLQ**SP**INSLIGRQVIKVGRSSGLTTGTIMAYAL
EYNDEKGICFLTDFLVVGENQQTFDLEGDSGSLILLTGQNREKPRPVGIIWGGTANRGRLKLRV
GQPPENWTSGVDLGRLLDLLELDLVTTNETLQDSGQEQMNGSTAGIGSTVGES**SPT**VPIKEKLE
ESFEPFCLNMEHVPVEE**PS**TIVK**PS**LRPCEFHIRNEIETVPNVEHQFIRTSFAGK**SP**VHQSFLK
EDMQFKSLSELRNEPDEDNFVSLHLGEPEAKRRKHSNSSLSLKN
>Medtr8g075010.1 _ eukaryotic aspartyl protease family protein
_ HC _ chr8_31704142_31707373 _ 20130731 ; MatureChain: 24-527
HIFTFTMHHRYSEPVKKWSHS**APSPS**HRWPEKGSVEYYAELADRDRFLRGRRLSQFDAGLAFSD
GNSTFRISSLGFLHYTTIELG**TP**GVKFMVALDTGSDLFWVPCDCTRCSATRSSAFASALASDFD
LSVYNPNGSSTSKKVTCNNSLCTHRNQCLGTFSNCPYMVSYVSAETSTSGILVEDVLHLTQPDD
NHDLVEANVIFGCGQVQSGSFLDVA**AP**NGLFGLGMEKISV**PS**MLSREGFTADSFSMCFGRDGIG
RISFGDKGSLDQDE**TP**FNVN**PS**H**PT**YNITINQVRVGTTLIDVEFTALFDSGTSFTYLVD**PT**YSR
LSESFHSQVEDRRRPPDSRIPFDYCYDM**SP**DSNTSLI**PS**MSLTMGGGSRFVVYDPIIIISTQSE
LVYCLAVVKSAELNIIGQNFMTGYRVVFDREKLILGWKKSDCYDIEDHNNAIPIGQHSDKVP**PA**
VAAGLGDY**PT**TDSSRKSKYNSQHSSASQSTSLYSRTSLLTCFGFLISYILFCLHVL

>Medtr0721s0020.1 _ transmembrane protein, putative _ LC _

scaffold0721_6330_5643 _ 20130731 ; MatureChain: 28-156
DDGTFMSKLAKSL**SPTPS**GWSSNTSFCLWTGVTTQFIKFEPV**PT**YLELSFILFCIISSIAILKS
KKISKITKKLFFYIFLKSVNNKHILFELVSRWSIHRH**PS**RERLYSERETWLSIVERTSVLSKMQ
S
>Medtr5g024983.1 _ transmembrane protein, putative _ HC _
chr5_10105076_10108942 _ 20130731 ; MatureChain: 24-122
HLNIANCCFVEHVTELCTGVRSAYLD**PS**FVI**PS**TQSLSEGVTRSFVVINIVLRMLVIYLLFLNP
FFTIP**PSPTPS**L**SP**KDISSERLQMLSQRDIKRNGG
>Medtr3g028560.1 _ transmembrane protein, putative _ LC _
chr3_9109195_9112249 _ 20130731 ; MatureChain: 21-185
AAALVWHNHIPLNLVHRGILSNTVVGCLLGYCGAMETSQHLFISCDFYGSLWSQLLDSVKWHSL
WWLKASHVVFVFGSDLWWLRLLDFSLSLFDNFAMASSATTMKILSFLFVVVLAVNVVSAQDL**SP**
SL**APAPAP**DTGAVGSLTNPVAMMGALIVLSMLAIFKH
>Medtr3g064040.1 _ transmembrane protein, putative _ HC _
chr3_28794619_28795716 _ 20130731 ; MatureChain: 28-191
RPCRTFIISSYSIRN**PS**TNAFATITEIRSI**SP**LFINDNTKPFEILLDRPVQHQTHSQSASHPRG
PLGLGFSTDAYDFSSLRDRTKDILSVALALLFGVGCGALTAATMYLVWSVFTARHELRAAAYGD
FSDDEIE**SP**KKMGGYVKI**PA**TETA**APAPAAP**VKDSV
>Medtr2g090660.1 _ hypothetical protein _ HC _
chr2_38818689_38817390 _ 20130731 ; MatureChain: 25-243
RQNKFFSFFTHFKTTHNVEDPQLPQ**SPAPAP**EPGA**AP**EIESTNI**PS**G**PAP**EPQFLVETGNGYGL
YGIDSSQY**SPT**KE**TP**KTTLTDFEDELLNEDFNDNKSYKKGYPQTNFHNNNEVYTKNYNNEENKN
SYNSYNGKEFYNNNYERKGEGMSDTRFMENGKYYYNVNSENEKYNNVNGYESGRGSTENEGNYE
KNQYQNEFETMEEYEKQQEAQGYTY**TP**
>Medtr1g041410.1 _ neurogenic locus notch_like protein _ HC _
chr1_15490155_15486147 _ 20130731 ; MatureChain: 23-214
DFL**SP**LL**SP**IFEDVCKEVECGKGTCK**PS**KNSTFPFECECDHGWKKALDSIDDEGLKFLPCIIPN
CTLDYSCSK**APAPAP**EKAKKSNESIFDACHWVDCGGGSCKKTSTFNYSCECDTSYYNLLNITTF
PCFRECSIGMGCSDLGISMSNSSSSS**AP**PPVLNENSKNEAISILQGRSLWLVLVIMVMAKIQLQ
>Medtr2g005930.1 _ DUF1682 family protein _ HC _
chr2_384861_381814 _ 20130731 ; MatureChain: 26-468
DSHFEGFEAEDDDSEFEE**PS**ID**PA**SLR**SPPS**QFLSTDPNPNPIN**PTPSPSPPS**DLPKS**TPPS**TT
TFDFWDDDEFEGL**PT**QPHPDFQV**PT**TDPQSTDNTNTTASDNQNVKPQPRSFTVEIVCGSFLIMF
ALNYFTGKKENENIALSWASHFAAKDSIFEKNFSLLGIGDGGDD**TP**LLLKEGQTTFKFYASGRR
YCQGLLATMELKSRHDLIARIYNMVV**PS**KDEITFEVYMNDDAMDHVVFAMARKKAAKAMHKDYR
DLQRFATIM**TPPT**SRKWVSDDLAVISESREVASDLITDALIDQVFGDKSFEKFGKGLISVHFSD
NHPGIHKKVLLFRFVL**PA**AKNMADMTRLVALVPYYIDLIGRYKLSSQARSKTEAARQKVAQEVQ
KELRNIQQEAMQRRKAERKKMMEEAEAKLGAEAIRKKEAKDRARQMKKAMPRMKMSRGA
>Medtr3g095330.1 _ ECA1 gametogenesis related family _ HC _
chr3_43558929_43559330 _ 20130731 ; MatureChain: 20-133
TSLSSTKTLASRLELFDGSGPNNKCWETMLELQHCTGDIVTFFLNGQTHLGSGCCNALLTIAQE
CWGNLLTSLGLTVEEAEILRGFCARVASVNNSLL**PS**ITVD**APSPAP**INNY
>Medtr3g108760.1 _ surface protein, putative _ HC _
chr3_50234565_50235467 _ 20130731 ; MatureChain: 21-260
INCLDITKTLGQYPELSTFSKYLTETKLAEQINSGKAVTILALDNKAIASLSGKPIDAIKAVIG
THVIPEFYDEKKLFDIIGSHAQL**PT**LST**AP**GLAAKIYVSLINEGEMAFSSAVEGSTFDATLVQS
TEAEPGVVSILQVSQPIVKVGAS**APATPA**TLSK**PATPA**AVSTSSAGEVA**TPA**AV**PA**SGAGQVT**T**
PAA**SPS**VVIAE**SP**ESFGD**APAPAPS**ASSRATFRFIGAVIAFASIFVSL

>Medtr4g134220.1 _ nicotiana tabacum ORF protein _ HC _
chr4_56191208_56189400 _ 20130731 ; MatureChain: 20-333
SETNPPCPKAFSDLKESILLEDSNRWDYVDLPIFWVEDQNGFEGNGLPVL**AP**FQLAGPMEKLFW
LMVQWLPLPLNRTQNGFAAGLLTLAEHLRCASRGQS**AP**LLSLRIVG**PT**SLE**APS**SAVSTSSDNR
LKLKRL**APA**LVELSSQSKPKAKSSETSLSTVDLQEE**APT**LL**TPT**QFTALWPLASLNGSNTNLLG
FEKLLSSVLGPKAKEKGSFRLLKADVSAQTFVKIGFQAEKKLKEGDGVSFEGFPKWRTKPETVR
LQFEVLAKVDGDKVIPERVVQVNPVFVEDTVALSVLTNNGTMSKCLLFSLR**PSPSP**CN
>Medtr6g016930.1 _ leguminosin group486 secreted peptide _ HC _
chr6_6613033_6613533 _ 20130731 ; MatureChain: 29-166
YSFAYNVTIFIYNKVL**APTPA**DITVHCKSKDDDLGDHTLMPGESYVFSFK**PT**HLPFKNTLFFCS
FTWPGNPHRHYLDIYDEAHDECKHCSWDINVNGGCLNDGKCVPWKSIAFMESYNTSKWPGEKGL
HELAHGHPLT
>Medtr6g016970.1 _ leguminosin group486 secreted peptide _ HC _
chr6_6631713_6632147 _ 20130731 ; MatureChain: 29-144
ISFAEVKVTIVNKVL**APTPT**NITFHCKSRDDDLGFHTLVSEGSYAFTF**SP**NF**TP**WFSKTLFFCS
FTWPGNPQLHYLDIYDQVRDNCYRCRWTINKDGGCLNTHKCYKWNSVKLMDA
>Medtr6g089320.1 _ leguminosin group486 secreted peptide _ HC _
chr6_22823555_22823055 _ 20130731 ; MatureChain: 29-166
YSFGYRVTVFIYNKVL**TPTPT**NITVRCKSKDDNLGDHTLMPGETYEFSFK**PS**HLPFKNTLFFCG
FTWPGNPHRHYLDIYDQAHDHCKQCSWDIILEGGCFYDGKCVPWKSIEFMEAYNTSKWPGKKGS
HELAYGHPLT
>Medtr7g113740.1 _ signal peptidase complex subunit 3B_like
protein _ HC _ chr7_46878551_46875326 _ 20130731 ; MatureChain:
32-167
FN**SPSPS**AQVQVLNINWFQKQPNGNDEVSMTLNISGDLQSLFTWNTKQVFVFLAAEYETRKKPL
NQISLWDGII**PS**KEHAKFLIHTSNKYRFIDQGTNLRGREFNLTLHWHVMPKTGKMLADKIVMPG
YRLPKEYR
>Medtr0002s0150.1 _ NAD(P)H_quinone oxidoreductase subunit H _
HC _ scaffold0002_165003_167020 _ 20130731 ; MatureChain: 23-596
R**SPPSSP**QS**SP**F**PSPA**I**SPS**AD**SP**VAS**SP**VPLNS**APSPSPS**AFSG**SPA**V**APS**WIS**TPPS**QG**PSP**
TI**SPPA**A**APA**V**TP**WWDYLATMFTEAITVN**AP**EFLENIQIPQRASYIRVIMLELSRIASHLLWLG
PFMADLGAQ**TP**FFYIFRERELIYDLFEAATGMRMMHNYFRIGGVAADLPYGWIDKCLDFCDYFL
RGVVEYQQLITQNPIFLERVEGVGFISGEEAVNWGLSGPMLRASGIQWDLRKVDPYESYNQFDW
KVQWQKEGDSLARYLVRVGEMSESIKIIQQAIEKIPGGPYENLEVRRFKKEKNSEWNDFEYKFL
GKK**PSP**NFELSRQELYVRVE**AP**KGELGIYLVGDDSLFPWRWKIRPPGFINLQILPQLVKRMKLA
DIMTILDRVEVETINSFSKSELLKEVYGLISILPILTLLLGITIEVLVIVWLEREISASIQQRI
GPEYAGPLGLLQAIADGTKLLFKEDIL**PS**RGDISLFSIG**PS**IAVISVLLSFLVIPLGYHFVLAD
LSIGVFLWIAISSI**AP**IGLLMAGYSSNNKYSFSGGLRAAAQSISYEIPLTLCVLSISLRVIR
>Medtr3g028560.1 _ transmembrane protein, putative _ LC _
chr3_9109195_9112249 _ 20130731 ; MatureChain: 21-185
AAALVWHNHIPLNLVHRGILSNTVVGCLLGYCGAMETSQHLFISCDFYGSLWSQLLDSVKWHSL
WWLKASHVVFVFGSDLWWLRLLDFSLSLFDNFAMASSATTMKILSFLFVVVLAVNVVSAQDL**SP**
SL**APAPAP**DTGAVGSLTNPVAMMGALIVLSMLAIFKH
>Medtr3g064040.1 _ transmembrane protein, putative _ HC _
chr3_28794619_28795716 _ 20130731 ; MatureChain: 28-191
RPCRTFIISSYSIRN**PS**TNAFATITEIRSI**SP**LFINDNTKPFEILLDRPVQHQTHSQSASHPRG
PLGLGFSTDAYDFSSLRDRTKDILSVALALLFGVGCGALTAATMYLVWSVFTARHELRAAAYGD
FSDDEIE**SP**KKMGGYVKI**PA**TETA**APAPAAP**VKDSV

>Medtr5g081820.1 _ transmembrane protein, putative _ LC _
chr5_35074890_35073090 _ 20130731 ; MatureChain: 22-224
L**SPTPSP**KLYERICKESGQKDFEQRCLKLIEAYPQITLIEDYLTFCRSFLKIVAIEKAVKSQQQ
VKEIMNKY**PS**SQPIKECGDDYNTVVIEVQGALREDPEMISMAVNYASDALDMCERSLADEKIVN
TSSIVALNHDMKLYTNIVVIAG**AP**VAQWIERPFPKRKVVGS**TPT**WRVSFFAQNALLFFFLKEKC
ILNDQLINSSF
>Medtr3g106510.1 _ transmembrane protein, putative _ HC _
chr3_49224097_49228278 _ 20130731 ; MatureChain: 23-575
HSSTSSSLSQLTSVKENEGNGVVFSRFDVFAA**AP**YEN**SP**LPLAAERTRRKDPLDGFNKYTSGWN
ISDHHYWASAAYTAVPVFSIAAVWFLGFGFCLLLLIVCYFCRKTESYGYSSTYYALSLILLILF
TFITLIGCAVLYIGQGSFHRSTTTTLQYVVVYQADSAVDKLRNVSDYLAQAKLVGIDRVFL**PA**NV
QTDIDAAETDINASAGTISDKTKENSDNIQDLLDSVRLALIIIAAVMLVLTFLGFLFSIFGMQV
LVYILVIAGWFLVTGTLILCGLFLILHNVTADTCVAMNEWIQY**PT**ANTALDDILPCVDKATAQE
TLLRSKEVTSELVNLVNQVITNVSNINF**AP**NF**TP**LYYNQSGPLMPLLCDPFRPDMTDRQCDSGE
VNISNATQVYGNFVCQV**SPS**EICMTQGRL**TPT**FYNQISAGINVGNALYNY**APS**LVELQDCTFVR
ETFTDIYNEHCPGLRHYSKLIYVGLIMVSFAVMFSLIFWGVYGRERRHRLYTQESKDSTLV**TPT**
RAH**APT**RR**APAPT**RR**APA**L**APAPT**RLAL**TPS**SHALEL**SP**YP
>Medtr3g030640.1 _ ribonuclease T2 family protein _ HC _
chr3_9713262_9712341 _ 20130731 ; MatureChain: 24-268
YS**TP**NNTSYYK**TPTPTSPS**LSHD**SPSP**YHFPLLKKGKKKPPPPPQPIV**TP**FEYLKIVQTW**PT**TF
CKVNKCINPP**PT**TWFTIHGLW**PS**NYSDPQPRLCTKEKIDWSTFSSLVSMTDLRKYWPRLDTAVR
NDDLFFWSEQWDNHGTCSSMHPPDFFNLAFKIYHKKELKTILQNEGIIPGGIKPETSQKIFDTI
ETGIGGFKPQIECLRVQNKDYLYQIKLCLDKTGDKYKDCPGPLIKCPMDVYFP
>Medtr4g107400.1 _ hipl2 protein precursor _ HC _
chr4_44415144_44409998 _ 20130731 ; MatureChain: 26-697
LPLCVDSK**AP**FTLNTTLKFCPYNNGSTCCNSIQDGQIQKQFQQMNVSDTACASLLKSILCARCD
PFSGQLFTVQS**TP**RSVPVLCNSAI**PA**NSSQSKALVHDFCSEVWDTCQTASIVN**SP**F**SPS**LQGQG
GGL**PA**NTNATKLNELWQSKNDFCKAFGGASNNQSVCFEGEPVALHKNETIIPPHGLCLEKIGNG
SYLNMVAHPDGSNRAFFSSQMGKVWLATIPEEGSGGQLELDES**SP**FVDLTDQVYFDTQFGMMGM
SFHPNFANNGRFFASFNCNKDKWSGCNGICSCNSNVNCD**PS**KIGTSNGAQPCQYQTVIAEYTAN
GTGSQ**PS**SAKSAK**PT**EVRRIFTMGLPFTSQHGGQILFGPNDGYLYFMMGDGGGTGDPYNFSQNK
KSLLGKVMRLDVDNI**PS**ASEVSKLGLWGSYSIPKDNPFSEDKDLEPEIWALGLRNPWRCSFDSE
K**PS**YFFCGDVGQDLYEEVDLITKGGNYGWRVYEGLYPFTTTE**SP**GGNTSLKSINPIPPIAGYNH
SQINKNEGSASITGGYVYRSMTDPCMFGRYLYADLYAGALWAATED**PS**NSGNFTTSKIPFSCAH
D**SP**IPCDS**APS**SSL**PA**LGYIFSFGQDNNKDVYILASSGVYRVVR**PS**RCSYTCSQEKASTTT**TPP**
SPSPSHASHLSSFYGYMFLQLSSLLLVLMCFM
>Medtr5g022810.1 _ ribonuclease T2 family protein _ LC _
chr5_9037425_9038278 _ 20130731 ; MatureChain: 22-229
DL**TPTPSP**NLRWPKTFCQTVKCSKDIQPL**PT**EFVMHGLW**PA**NRVISDPRGCLDKTNQKTIDIGN
FPLDLKEELDKVWPDLLVYEKSRLINIAFWDEQWKAHGSCSNMDIIDFFKLSLSIYKKIGSLKE
VLGKEGY**SP**GPQSHVEKQKIVDIIKKHTDGKA**SP**RIKCEKHDGKTYLHEIQVCVDKTENHNYTN
CR**TP**DLIDCEKDVYFP
>Medtr2g030825.1 _ late embryogenesis abundant (LEA)_like
protein _ LC _ chr2_11653982_11656243 _ 20130731 ; MatureChain:
22-595
A**TP**PGIAKDHEYGHASCKIKKYKHCYNLVHVCPKFCPNDCTVECASCKPICVGGTNN**PS**HDY**PT**
PSTPTPSHDY**PTPSTP**HNP**PS**SSSSSSSSSSSSSSS**SP**NSS**SP**PHNQHNSSS**TPS**SSSSSSSSS
SSSSSSSSSSS**SP**K**SPSTP**KNPP**PS**SS**TPS**SSSNSSSSSSSSSSSSSSSSSSSSSSSSSSS**SP**Q**S**

PSTPHS**SPPS**SS**TPS**SSSSNSSSSSSSSASSSSSSSATSSSSSSTN**SPS**TSHNPP**PS**SSS**TPS**S
SSSSSSSSSSSSSSSSSSSSHKTARCGNKNYPKCYNMEHACPNACRGGCEVDCNTCKPICKCDK
PGAVCQDPRFIGGDGITFYFHGKKDNNFCLVSDKNLHINAHFIGRRNENMKRDFTWVQSIVILF
DNHQLFLGAQKTATWDDSVDRLAISFDGEPITLHESEGAKWESSGVSFVRETSTNNIIVEVEGN
FRITAKVVPITEEDSRVHNYGITKDDCFAHLDLGFKFLSLSNEVSGVLGQTYKPNYVSRVNIGA
KMPIMGGGKEYETTSLF**SP**DCSVARFVGNDGFNNDIAMVENLSL**PS**MSCTSGIDGQGVVCRR
>Medtr2g030845.1 _ late embryogenesis abundant (LEA)_like
protein _ LC _ chr2_11667751_11669978 _ 20130731 ; MatureChain:
22-596
A**TP**PGIAKDHEYGHASCKIKKYKHCYNLVHVCPKFCPNDCTVECASCKPICVGGTNN**PS**HDY**PT**
PSTPTPSHDY**PTPSTP**HNP**PS**SSSSSSSSSSSSSSS**SP**NSS**SP**PHNQHNSSS**TPS**SSSSSSSSS
SSSSSSSSSSS**SP**K**SPSTP**KNPP**PS**SS**TPS**SSSNSSSSSSSSSSSSSSSSSSSSSSSSSSS**SP**Q
SPSTPHS**SPPS**SS**TPS**SSSSNSSSSSSSSASSSSSSSATSSSSSSTN**SPS**TSHNPP**PS**SSS**TPS**
SSSSSSSSSSSSSSSSSSSSSHKTARCGNKNYPKCYNMEHACPNACRGGCEVDCNTCKPICKCD
KPGAVCQDPRFIGGDGITFYFHGKKDNNFCLVSDKNLHINAHFIGRRNENMKRDFTWVQSIVIL
FDNHQLFLGAQKTATWDDSVDRLAISFDGEPITLHESEGAKWESSGVSFVRETSTNNIIVEVEG
NFRITAKVVPITEEDSRVHNYGITKDDCFAHLDLGFKFLSLSNEVSGVLGQTYKSNYVSRVNIG
AKMPIMGGGKEYETTSLF**SP**DCSVARFVGNDGFNNDIAMVEDLML**PS**MSCTSGVDGEGVVCRR
>Medtr2g079030.1 _ MATE subfamily protein _ LC _
chr2_33094342_33097785 _ 20130731 ; MatureChain: 23-544
YTFYVGGKDGWVVN**PS**ESYSQWTGRNRFQVNDILVFKYKKDSNSVLVVNKEDYDKCNKKYPIKK
FDDGDSEFQLDRSGPFYFISGKDDNCEKGQKLIVVVLAVRKI**SP**PYVPIIPPK**TPSP**IH**SP**PNV
PY**TP**PKV**PSP**GNQPPYVPI**TP**PK**TPSP**IINPPNVPY**TP**PK**TPSP**ENQPPYVPIIPPK**TPSP**IY**S**
PPN**TP**Y**TP**PKS**TPPSP**NQPPYV**PTP**PK**TPSP**IYNPPNVPY**TP**PK**TPSP**RNQPPYVPI**TP**PK**TPS**
PIH**TP**PNVPY**TPSPT**NQPPHVPI**TPS**TLYETYYIVLYHASLEHWAFEVLVFLAGLMPDSQITTS
LIAICANTEFIAYMITYGLSAAASTRISNELGAGHPERAKHAMRVTLKLSVVLGFCFALMIVFG
HDIWIRMFSS**SPT**IKHKFASI**SP**FLATSILLDSVQGVLSGVVRGCGWQHLVVYVNLATFYLIGL
PISCLLGFKTNLEYKGLWIGLICGLACQTGTLLLLTRYAKWTKLNLSGDKGKDQPVVVLTTECM
PIET**AP**LDAV
>Medtr3g062840.1 _ plant_F17O14_7 protein _ HC _
chr3_28400981_28399252 _ 20130731 ; MatureChain: 26-311
QIPFDNGISSKALDSVLQNYAFKALVKPKTGVPYDAKLPNSIIGVKVSTLRLRSGSLRTRGVQY
KEFQIPPGVVEQPYVERLVLVYHNLGNLSEHFYPLPGYTYL**AP**VLGLLSYSGVDLFANELPELD
VRASDKPILVKFLKVKS**AP**YGSVPKCVYFDLHGSVQFDILLPDNVCSTMEQGHFSIVVESN**APT**
PAPAAVAAAVGKGGRRKSNVWIIVGSVVGGCILLIMLSLLVVRLVRIKKGMKIQELEWTADSHE
TLQMKSIGGTK**AP**LAIGTRTR**PT**IENDYIP
>Medtr2g094170.1 _ transmembrane protein, putative _ HC _
chr2_40165150_40164257 _ 20130731 ; MatureChain: 25-149
D**SPPSPSPAPS**L**SPSPT**D**TPSP**YYP**PA**S**SP**PVS**SPPAPSP**LN**PSP**I**PAP**V**PSP**EDSTSLNHIDV
DEKTEDSSTEGGMSGSKKAGIAIGIIVAASVLMLAGMVYKKRQQNLRRNQYNFGVRRDIIL
>Medtr4g459420.1 _ transmembrane protein, putative _ HC _
chr4_21862479_21863379 _ 20130731 ; MatureChain: 23-166
AD**SPAPTPA**TNSSLN**SPSPTP**I**PTPSPA**N**SPPAPTPTPTPSP**HSD**SPPAPSP**DN**SPS**S**SPSPSP**
SS**SPAPSP**DEAADNNAISHTGIGEDGKSSGGGMSSGKKAGIAVGVIAAVGVVALGAMVVKKRRQ
NIQRSEYGYTARRELL
>Medtr4g059520.1 _ transmembrane protein, putative _ HC _
chr4_21918530_21919328 _ 20130731 ; MatureChain: 23-166
AD**SPAPTPA**TNSSLN**SPSPTP**I**PTPSPA**N**SPPAPTPTPTPSP**HSD**SPPAPSP**DN**SPS**S**SPSPSP**

SS**SPAPSP**DEAADNNAISHTGIGEDGKSSGGGMSSGKKAGIAVGVIAAVGVVALGAMVVKKRRQ
NIQRSEYGYTARRELL
>Medtr4g093780.1 _ transmembrane protein, putative _ LC _
chr4_37242467_37241798 _ 20130731 ; MatureChain: 22-202
RSLSS**SP**VA**TPSPA**G**SP**LAI**SPA**VS**SP**VPLMN**APSPSP**FAVNY**PTSP**P**PA**SFG**SPASP**VA**APA**V
TPLSIS**TPPS**Q**APSPA**I**APS**TSAN**SP**VAS**SP**VPVTS**SPSPSPS**AIN**SPPSP**PPHA**SPA**A**APA**I**T**
PSSIS**TPPT**K**APS**SIS**TPPT**K**APSTPT**NGATMNGFNVAGYAAAVVFIAALILM
>Medtr4g127180.1 _ transmembrane protein, putative _ LC _
chr4_52761334_52760825 _ 20130731 ; MatureChain: 24-169
V**AP**VN**SPA**V**SPT**SN**SPSPS**VVH**SPT**M**SPA**VN**APAPS**ATNTH**AP**V**PSPT**STN**APAPS**TASNN**APA**
PSATNTHG**PAPS**DTTSNH**APAPS**ATSTH**APAPS**ETSTAHAS**APS**VTKTH**APAPSP**HHSGSARLS
GYVGINVVVALVLGSFVF
>Medtr8g069283.1 _ transmembrane protein, putative _ LC _
chr8_28996145_28996955 _ 20130731 ; MatureChain: 22-121
Q**APAPAPT**QSL**SP**NS**AP**V**PS**SSTR**SPTPSPA**V**SPS**TSS**SP**G**PSP**NTS**SPPAP**GSN**SPA**SGPGGP
GLEQPGAE**SPS**AAFSISNTFVAVTAFAGIFLSMVLA
>Medtr8g069315.1 _ transmembrane protein, putative _ LC _
chr8_29008027_29008943 _ 20130731 ; MatureChain: 22-122
Q**APAPAPT**QL**SP**NS**APAPSPS**TK**SPTPSPA**V**SPS**TSS**SP**GS**SPSP**NTF**SPPAP**GSNV**PA**SGPGG
A**TP**GQPGAD**APS**AAFSITNTFVAVTAFAGIFVSMVLA
>Medtr7g032620.1 _ transmembrane protein, putative _ LC _
chr7_11453256_11452810 _ 20130731 ; MatureChain: 22-148
AE**APA**S**SP**KS**SP**KSSAKE**SP**KSSAKD**SPAPS**SEKK**SPAPS**SSSVKK**SP**KASGPVTGLDV**PT**ISD
APPLFVD**SPPA**Q**SP**VAEEPIVAVPPVSG**PA**AG**PTSPA**ADEASDASSLKVSAVVAVAAVGFFAF
>Medtr3g049390.1 _ transmembrane protein, putative _ LC _
chr3_18389858_18389424 _ 20130731 ; MatureChain: 29-144
QPV**PTPAPSP**N**SPSPAP**NLRP**PSPTPAP**VPPPN**SP**P**PAP**V**PSP**IS**SP**PPNPN**AP**EPE**PT**AAGQG
GSSGLSGGQKAGIVIGTLLGAAILGFIGMVCWKRRVNIRRNRYSDAARNIEL
>Medtr3g099520.1 _ hypothetical protein _ HC _
chr3_45630661_45631516 _ 20130731 ; MatureChain: 21-148
E**APS**T**SPT**A**AP**KASHA**APAP**KATA**TPPS**STT**TP**PKSSAT**SPT**S**SPAP**KVS**SPPSPTPT**SAE**AP**V
E**SPT**E**SPPAP**V**SPT**V**SPA**T**SP**VASG**PA**VSD**APA**E**APA**GSSAAASFRVSFVGGSVAAFVAAALLM
>Medtr4g032750.1 _ hypothetical protein _ LC _
chr4_11312661_11313422 _ 20130731 ; MatureChain: 24-253
FETTLQELDAFK**PSVPV**K**PVPV**K**PAPV**KPLKPPI**PVTPPTP**IKP**PAP**R**TPPTP**IKPR**VPVTP**PI
PVTPPTPIKP**PAP**R**TPPTP**IKSR**VPVTPPTP**VK**PPVPV**KP**PT**HV**TPAP**VKP**PAP**V**TPAP**VKPPV
SIK**PAP**VK**PTP**VKPGPF**VP**NPISVHI**PV**MPIPGIT**VP**I**TPA**FV**TPPA**QVDPHALV**TPAPA**EP**PS**
HVTL**SP**VEPPII**PT**NPLALAY**SP**LPLRLNHISRKLMVV
>Medtr4g093810.1 _ hypothetical protein _ HC _
chr4_37233312_37232519 _ 20130731 ; MatureChain: 22-119
Q**SPS**S**SPS**K**SPA**I**SPS**AH**SPA**A**SPPAP**VKN**SPSPSPS**AIN**SPPSP**P**PA**SSG**SPA**A**APA**V**TPS**SI
S**TPPA**E**APS**NGAALNRFTVAGSAAVVIFAAALMM
>Medtr4g093790.1 _ hypothetical protein _ HC _
chr4_37237974_37237609 _ 20130731 ; MatureChain: 22-121
Q**SPS**S**SPT**I**SP**VA**TPA**I**SPS**AD**SP**VA**SP**PVPVKT**SPSP**LASAVN**SPPA**SSNSL**APA**V**TPS**SIS**T**
PPSE**APSPS**DNSIAALNRFTFAVSAAAVVFAAALMM
>Medtr4g093825.1 _ hypothetical protein _ HC _
chr4_37253406_37252734 _ 20130731 ; MatureChain: 22-124

Q**SPS**S**SPT**I**SP**VA**TPA**I**SPS**AD**SPA**AS**AP**IPVKN**APSPSPS**AIN**SPPSP**P**PA**SSD**SPA**V**SPALT**
PSSIS**TPPS**EG**PS**ENGAALNRFTVAGSAAVVVFAAALIL
>Medtr4g093835.1 _ hypothetical protein _ HC _
chr4_37257618_37256885 _ 20130731 ; MatureChain: 22-122
Q**SPS**S**SPT**I**SP**VA**TP**PKSSQ**APSPS**AV**SP**VA**SP**PVPVKN**APSPSPPA**SSD**SPA**V**SPA**V**TPS**SIS
T**TPS**E**APSPS**DNSAAAFNRFTVAGSAAVMVFAAALMM
>Medtr4g095280.1 _ hypothetical protein _ HC _
chr4_39695831_39693836 _ 20130731 ; MatureChain: 22-207
Q**SPS**A**APT**T**SPSTPA**AT**TP**VS**SP**VA**APPTTPTTPAP**VA**SP**KS**SPPA**T**SP**KAA**APT**A**TPPA**AS**SP**
PAV**TP**VS**TP**P**PAP**VPVK**SPPTPAP**VS**SPPA**V**TP**VA**APT**T**TPA**V**PAPAPS**KGKKNKKKHG**APAPS**
PALLGP**PAPPA**G**AP**G**PS**EDAS**SP**G**PA**TTANDESGAETIRSLKVLGGLAMSWMAVVLFF
>Medtr5g013285.1 _ hypothetical protein _ HC _
chr5_4189954_4189079 _ 20130731 ; MatureChain: 23-122
Q**APSP**GPIMLH**TPSPA**S**SP**ESLPPE**SPS**Q**SPS**M**SPS**MNM**SPS**M**SP**PF**PT**DA**SPSPA**S**SPSPS**TG
ESMSDNPVAS**SPS**NAVVRRSSFFMLPFFAGAALLLA
>Medtr5g016250.1 _ hypothetical protein _ HC _
chr5_5787919_5789615 _ 20130731 ; MatureChain: 22-224
Q**SPS**S**APT**T**SPPT**VT**TPS**A**AP**VA**APT**KPK**SPAP**VA**SP**KS**SPPA**S**SPT**AATV**TPA**V**SPAAP**VPVA
K**SPA**AS**SP**VV**AP**VS**TP**PK**PAP**VS**SPPAP**VPVS**SPPTP**VPVS**SPPT**AS**TPA**V**TPS**AEV**PAAAPS**K
SKKKTKKGKKHS**APAPSPA**LEGP**PAP**PVG**AP**G**PS**LDAS**SP**G**PA**SAADESGAETIRCLQKVIGSL
ALGLASLVFMF
>Medtr6g021940.1 _ hypothetical protein _ HC _
chr6_7589553_7590203 _ 20130731 ; MatureChain: 22-145
KEVSSSA**SP**KS**SPSPS**S**AP**EASEKESE**APA**S**SP**KATTS**APTPAPS**TD**AP**VS**SP**EA**SPA**S**SP**EAD
EEIS**SPPSPSPA**DDLF**AP**GS**APA**SDDV**APAAAPT**ADEAAASSLRFSAAAATVVVAGFFAF
>Medtr6g039360.1 _ hypothetical protein _ LC _
chr6_14249346_14248151 _ 20130731 ; MatureChain: 23-190
R**SPPSSP**QS**SP**F**PSPA**I**SPS**AD**SP**VAS**SP**VPLNS**APSPSPS**AFSG**SPA**V**APS**WIS**TPPS**QG**PSP**
TI**SPPA**A**APA**V**TP**WSIS**TPPS**QG**PSPA**I**SP**IPVKS**SPSPSPS**AIN**SPSSP**PQAF**PAATPA**I**TPS**
AIS**TPPS**K**APS**TSTNAAAALNRFTVAGSTAVVVFVAALMM
>Medtr6g039690.1 _ hypothetical protein _ HC _
chr6_14317270_14316689 _ 20130731 ; MatureChain: 23-193
R**SPPSSP**KS**SP**F**PSP**VA**SP**PVPLNS**APSPSPS**ASSG**SPAAAP**VV**APS**WIS**TPPS**QG**PSPT**I**SPS**
AA**AP**DV**TPS**SIS**TPPS**Q**APSPA**I**SPS**ANSTAS**AP**VPVKS**SPS**I**SPS**AIN**SPSSP**PQA**SPAATPA**
I**TPS**SIS**TPPS**K**APS**TSTNDAALNRFTLAGSAAVVVFAAALMM
>Medtr8g069925.1 _ hypothetical protein _ HC _
chr8_29605296_29607773 _ 20130731 ; MatureChain: 23-621
QSHAA**APS**NLPITT**SP**FVAKQQPVVAAI**SPT**SNK**PT**ATVTAA**APS**KL**PT**TN**PS**E**PS**TLPLYKQQ
PIVAA**TPA**SNTNKQQPIVAA**TPA**SNTNKQ**PA**AT**SPT**SNKPITTVTAA**APS**KL**PT**TN**PSAPS**TLP
LDKQKPIVAA**TPA**SSTKQP**PA**AT**SP**ISNK**PS**TTISAA**APS**KL**PT**II**PAAPS**KL**PT**AA**SPT**SS**TP**
SAAKKPPIVVAT**SP**LSSKQ**PT**TTQP**PT**TAKS**SP**VK**SP**VPKVT**SPA**S**SP**VK**SP**VPKVA**TPT**S**AP**V
K**SP**VPKAT**TPT**S**AP**VKPPVPKVA**TPT**V**APA**KPLVPKVT**TPT**A**AP**VKPIVPKTT**TPT**S**AP**LKPPV
PK**APAP**K**PSP**VKPPVPKSKP**PT**T**AP**VK**SP**VPDPPKL**AP**VKLPVPKV**TPA**L**SP**K**TPSP**KIQPPPH
PPKK**AP**VSLPPLSIPPVSLPPLPLP**PA**SKPPKV**SPAPA**KVPK**APAPA**KE**APAPAPT**HKKK**APAP**
APDKE**TPAPAPT**HKKKK**AP**KS**SP**V**PSPA**ILP**PSPAPTPA**ID**TPS**S**APAPSP**EDD**AP**EPPPPHKH
KRRKHKHSKHKKHHALAL**AP**E**PT**SSSSTIIRR**SPPAP**LADDNTTMSSDEG**PSPAPSPS**ANGAQS
YQGQWRKMLATGGIAIAILLCVT
>Medtr0002s0900.1 _ hypothetical protein _ HC _

scaffold0002_343403_341926 _ 20130731 ; MatureChain: 19-250
FNAQ**APA**T**APT**KLP**PTTPT**AI**TP**VTTQP**PT**VVA**SP**PITSQPPVTV**AP**KS**AP**VT**SPAP**KI**APA**S**S**
PKVPPPQPPKS**SP**VS**TPT**LPPPLPPPPKI**SPTP**VQ**TPPAPAP**VKA**TP**V**PAPAPA**KQ**APTPAPA**T
SPPI**PAPTPA**IE**AP**V**PAP**ESSKHKRRRHRHKHRRHQ**APAPAPT**VIHK**SPPAPPT**DTTADSDT**AP**
APAPSLNLNG**APS**NHLQGGNIWTTIGFAITVFLAVTGYSF
>Medtr0902s0010.1 _ hypothetical protein _ HC _
scaffold0902_3830_3207 _ 20130731 ; MatureChain: 23-207
RS**SP**KS**SP**F**PSPA**I**SPS**AD**SP**VA**SP**PVPLNS**APS**S**SPT**ASSG**SPA**A**APA**V**TPS**SISIP**PT**Q**APS**
PTIF**TPS**A**AP**LAA**TPA**V**TPS**SIS**TPPS**Q**APTPA**I**SPS**ANSMASIS**AP**VPVKS**SPSPSPS**AIN**SP**
PSPPHA**SPA**A**APA**I**TPS**AIS**TPPS**K**APS**TFTNAGTALNRFTVAGSAAVVVFTAALMM
>Medtr3g007900.1 _ GASA_GAST_Snakin _ HC _ chr3_1236356_1234999
_ 20130731 ; MatureChain: 19-154
ATIE**APTPAPAPAPAPA**SQI**PS**DS**TP**E**TP**V**TPTTP**E**TPATPTTPPAP**V**TPTPPATPPAP**V**TP**D**P**
TDGTTEGSLKPEEC**AP**RCEDRCSKTHHRKPCLFYCKYCCAKCLCVPPGTYGNKEVCPCYNDWKT
KEGGPKCP
>Medtr5g077360.1 _ leguminosin group485 secreted peptide _ HC _
chr5_33028100_33027104 _ 20130731 ; MatureChain: 26-147
RNLVEVAMKQNDVDNVKYYAKKSLTNKGDSSSDCFIPGFPGIPGIPEIPGIPGIPGFPGIPGWP
G**TP**N**TP**N**TPASPSPS**T**TP**E**TP**D**SPTPSP**DD**SPAP**N**TPASP**N**PS**T**TP**E**TP**D**SPAPS**LDD
>Medtr5g077715.1 _ leguminosin group485 secreted peptide _ HC _
chr5_33167682_33166106 _ 20130731 ; MatureChain: 26-200
RNLVEVATKQNNVDGIKYCTTKNVKDNDDASSDCFIPGFPGIPFPKIPGFPDIPEIPFPKIPGI
PDIPGIPIPYIPGIPFPKIPGIPDIPGIPIPDIPGIPFPKIPGIPFPFPFPFPFPGWPDIPGW**P**
TPSTPK**APSPTP**E**SPSPS**E**SPSPS**ESSS**SPSPS**E**SPSSPS**L**SPT**YVM
>Medtr5g095990.1 _ GASA_GAST_Snakin _ HC _
chr5_41969296_41967991 _ 20130731 ; MatureChain: 19-149
ATIE**APTPAPAPSPSPSPAPT**STTASDS**TP**E**TPATPTTPPAP**V**TPTPPA**T**SPAP**V**TP**D**PT**DGTT
EGSLKPEEC**AP**RCEDRCSKTHHKKPCLFYCKYCCAKCLCVPPGTYGNKEVCPCYNDWKTKEGGP
KCP
>Medtr8g022490.1 _ leguminosin group485 secreted peptide _ HC _
chr8_7967035_7967849 _ 20130731 ; MatureChain: 26-177
RNLVEVATKQNNVDGIKYCTKTNGKENVDASSDCFIPGFPGIPGIPFPNIPGIPDIPGIPFPKI
PGIPDFPFPFPGIPGWPYIPGWPDIPGWPN**PS**IPK**APPSPTP**Q**SPSPSPS**E**SPSSPSPS**E**SPSP**
SE**SPSSPSPS**E**SPS**SS**SPSPT**YVM
>Medtr8g022550.1 _ leguminosin group485 secreted peptide _ HC _
chr8_7985783_7986582 _ 20130731 ; MatureChain: 26-174
RNLVEVATKQNNVDGIKYCTKTNGKYNIDASSDCFIPGFPGIPGFPFPKIPGIPDIPGIPFPKI
PGIPDFPFPF**PS**IPGWPDIPGW**PTPSTP**K**APSPTP**Q**SPSPS**E**SPSPS**E**SPSSPSPS**KF**PSPS**V**S**
PSSPSPSE**SP**N**SPSP**N**PT**YVM
>Medtr2g009150.1 _ transmembrane protein, putative _ LC _
chr2_1807719_1807363 _ 20130731 ; MatureChain: 25-118
Q**SPPS**LPRKTLSVVYRSTAD**SPS**TGS**SP**LE**PS**AA**SPSPS**DHPQQ**PA**DDQQSVLPFIHGIQIFLA
IYLSILCLLIIVVMVFLLFLFVKYIRSRCL
>Medtr2g009275.1 _ transmembrane protein, putative _ HC _
chr2_1854017_1854661 _ 20130731 ; MatureChain: 25-98
Q**SPA**S**SPT**KSQPPRKAI**SPSPA**AS**SP**PE**PS**AT**SPA**V**SPS**SISGP**PS**E**AP**G**PA**SSAVLNRVSVAA
ILIFVAAFIM
>Medtr2g009290.1 _ transmembrane protein, putative _ HC _

chr2_1860925_1861200 _ 20130731 ; MatureChain: 21-91
E**SPT**T**SP**KVSSLT**TPA**A**APSPS**SV**SP**L**SPS**SVS**SPPA**H**APAPAP**RKSGAVSHGFSFVGIFVVAL
GATALIL
>Medtr7g083830.1 _ transmembrane protein, putative _ HC _
chr7_32276392_32276057 _ 20130731 ; MatureChain: 29-111
VSSTTTE**PT**ISA**SP**GVLPYVT**SP**DISSFF**PTP**MSSSE**AP**YEAEAS**APAPAPS**SGEKSSSSTRLD
CVAAIVGIMLFSVFLSFIA
>Medtr2g450250.1 _ transmembrane protein, putative _ HC _
chr2_22169134_22168833 _ 20130731 ; MatureChain: 25-81
EV**AP**PK**PT**RKNNNNKDVGASTSAIGSTWTR**VPVPV**ASTVTTTQR**VP**ISDIS**TPTPT**L
>Medtr8g089615.1 _ transmembrane protein, putative _ LC _
chr8_37276848_37277165 _ 20130731 ; MatureChain: 24-105
AD**APAPA**ESDFLEDDDEYLVGTRKGDP**PS**ANGINIVAGPIGGPVHAGTFDNI**APAPS**AASTIYI
SSIAGTVVTASIAGFFYF
>Medtr6g045633.1 _ leguminosin proline_rich group669 secreted
peptide _ HC _ chr6_16530751_16529919 _ 20130731 ; MatureChain:
21-95
RNPKHQIHAQPNSGLVDYQP**PA**ADQAASFI**SPSPSPSP**PHGSGYNGHGGGYHR**PS**IGYSRFPPH
INPRRPVRPPP
>Medtr3g111910.1 _ transmembrane protein, putative _ LC _
chr3_52377332_52376402 _ 20130731 ; MatureChain: 24-224
QDGLKDLLGGGDLASTIDEAKKALGAASGEAGGQAALDDYAGGEEGQTAN**PA**DASLAGWIETDN
ATSNGTGS**SP**PKSAD**SP**IGAAEDL**PA**D**TP**EGG**APTPT**ADGSSNK**SP**QQ**APT**N**APS**NE**PS**K**TP**ES
APTN**APT**AASTN**APA**E**AP**KGNGANVAGSSK**AP**ES**AP**IV**SPTGAP**QGDLTDDDDDDQMV**APSGA**
PKGTSTDDD
>Medtr2g090575.1 _ early nodulin_like protein _ LC _
chr2_38712243_38710650 _ 20130731 ; MatureChain: 23-360
YNFNVGGNDGWGVK**PS**RHYNSWSERMRFQVNDTLYFKYHKGSDSVLEVNKQDYDSCNTKNPIHE
RDSGHSTFLLDKSGPFYFISGKVSNCLKGEKLHLVVL**SP**HHGKGHQG**PS**S**SP**FV**AP**VH**SPATSP**
AWI**APTPS**ATRWN**APSPS**AIGWN**APT**H**SPS**Q**SPA**WK**APSPS**ATSKT**APT**Q**SP**VWN**APSPS**ATGW
T**APA**H**SPT**Q**SPT**WSAS**SPS**ATGWT**TPT**H**SPT**Q**SPA**WH**APSPS**ATGWT**APA**H**SPT**KSSGWN**APTP**
SATGWT**APT**H**SPA**QSSAWN**APTPSPA**NIH**APTPSPT**DED**AP**K**PS**DIDND**SPAPSP**GHSGSRRLS
GFVGVSVVVALVLGSFAF
>Medtr2g090580.1 _ early nodulin_like protein _ HC _
chr2_38720886_38719533 _ 20130731 ; MatureChain: 25-200
YKFHVGGKHGWAVK**PS**AGYSHWAEKNRFQVNDTLYFKYNKGSDSVLVVNKQDFDSCNTKNPILK
LDDGDSTFKFDKSGPFFFISGIVENCQKGEKLIVVVL**SP**NHHY**TPPSPT**TV**APAPS**HSAEN**APS**
PSATGD**TPPT**T**SP**IDEN**SPAPSPA**HSGSDRFRGSVGVGVALVLASFVF
>Medtr3g092170.1 _ early nodulin_like protein _ LC _
chr3_42095471_42094282 _ 20130731 ; MatureChain: 21-277
SGYTSRVDGKEGWPVK**PS**SGYNVLTSGIKLLIHDNIYFKYNKEIDSVLVVNKQDHDSCNTKNPI
YKMEGGDSAFQLDKSGPFYFISGNVENCQKGRKLNVVAWFPHRRLMSLAAD**APSPS**MVQV**PA**M**S**
PTVN**APTP**NVIGWN**APAPSPA**DIH**APSPSPT**TNH**AP**V**PSPT**DNHAS**TP**N**PS**GNH**APAPS**ATNIQ
V**SPTPS**ATHKKCHRRRHWGLCFGSKCHRDSCSDI**APSP**GHSGSTRLSGSVGVNVVVALVLGSLA
F
>Medtr3g105930.1 _ early nodulin_like protein _ HC _
chr3_48854561_48853915 _ 20130731 ; MatureChain: 25-180
KELLVGGKIDAWKV**PS**SETDSLNQWAEKSRFKVDDHLVWKYDGGKDSVLQVNKEDYANCNSSNP

IEQYNDGNTKVKPDRPGPFYFISGAKGHCEQGQKLIVVVM**SP**KKRSIGV**SPAPSPA**ELEEG**PA**V
APTSS**AP**VLRTGLVTVLGLLAIYVGFLI
>Medtr4g130780.1 _ ENOD16_ HC _ chr4_54496082_54497288 _
20130731 ; MatureChain: 25-182
ESTDYLIGDSHNSWKVPL**PS**RRAFARWASAHEFTVGDTILFEYDNETESVHEVNEHDYIMCHTN
GEHVEHHDGNTKVVLDKIGVYHFISGTKRHCKMGLKLAVVVQNKHDLVLPPLITMPMP**PSPSPS**
PNSSGNKGGAAGLGFIMWLGVSLVMMMFLI
>Medtr7g090170.1 _ early nodulin_like protein _ HC _
chr7_35395785_35395287 _ 20130731 ; MatureChain: 25-148
KELLVGGKIDAWKV**PS**SEADSLNQWAEKSRFKVSDHLVWKYDGGKDLVLQYNDGNTKVKLDRPG
PFYFISGAKGHCEQGQKLIVVVM**SP**KKRSIGV**SPAPSPA**ELEEG**PA**V**APT**NS**AP**VLRTGL
>Medtr1g009910.1 _ fasciclin_like arabinogalactan protein _ LC
_ chr1_1613159_1612264 _ 20130731 ; MatureChain: 24-287
FDIQKLLDR**TP**EFSTFNKYLNETKLVGQINRRNTITVFALDNGAMSSVSDKMPEAIRAIMATHV
LLDYYD**PT**KLIGAMHKREPLTTLYQSSGIAVDQQGYLKVNRTSDGDLAIGSAVSG**AP**IDVKLVK
VVFAQPYNITIIQVAKPILFPGLETQTLG**APS**NAS**APA**AETNVDVSSVFK**APPT**KAKNAN**APS**A
AEEPITEVSD**SPSPS**DE**PS**E**SP**VE**APA**K**SPS**L**AP**GPGGDEAAADA**APT**SSSSRIVIGFVGAVMC
LASLLVVM
>Medtr1g009920.1 _ fasciclin domain protein _ HC _
chr1_1622595_1620447 _ 20130731 ; MatureChain: 24-227
FDIQKLLDR**TP**EFSTFNKYLNETKLVGQINRRNTLTVFTLDNGAMSSVSDKMPEAIRAIMATHV
LLDYYD**PT**KLIGAMRKRELLITLYQSSGIAVDQQGYLKVNRTSDGDLAIGSAVS**APPA**KAKNAS
APSAAEEPITEVSD**SPSPS**DE**PS**E**SP**VE**APA**K**SPS**L**AP**GPGGDEAAADA**APT**SSSSRIVVGFVG
AVMCFSSLLVVM
>Medtr1g009940.1 _ fasciclin domain protein _ LC _
chr1_1629663_1628800 _ 20130731 ; MatureChain: 24-287
FDIQKLLDR**TP**EFSTFNKYLNETKLVGQINRRNTITVFALDNGAMSSVSDKMPEAIRAIMATHV
LLDYYD**PT**KLIGAMRKRELLITLYQSSGIAVDQQGYLKVNRTSDGDLAIGSAVSG**AP**IDVKLVK
VVFAQPYNITIIQVAKPILYPGLETQTLG**APS**NAS**APA**AETNVDVSSVFK**APPA**KAKNAS**APS**A
AEEPITEVSD**SPSPS**DE**PS**E**SP**VE**APA**K**SPS**L**AP**GPGGDEVAADA**APT**SSSSRIAVGFVGAVMC
FASLLVVM
>Medtr1g009950.1 _ fasciclin domain protein _ HC _
chr1_1637789_1636871 _ 20130731 ; MatureChain: 24-268
LDITKLLGQNPDFAAFNKQLTETKLVDQINSRNTITVLAVSDGAMSAISGK**SP**QAIKAIMSTHV
VLDYFDEKKLSEAVGSGILLTTLFQASGQAKNQQGFLKVKLIGEGEMDFGSAVSG**AP**IDVALVK
TVVKQPYNISILQVAKPIIFPGVDSVSTAS**APT**AAKNAS**SPS**AAKAD**APT**AE**TPS**ESAT**APSPS**
KE**PA**TN**APA**E**APT**AEAAGPGGAAAD**AP**P**PS**SSSRTVVGLVGAMMCFASLLVVM
>Medtr2g007870.1 _ fasciclin domain protein _ HC _
chr2_1180291_1179140 _ 20130731 ; MatureChain: 24-354
IPNREFDSMLNTLRSRGYHLFCNAILTSDLRIDLLDPNSNATNSFTFF**APT**DSSLFALDMTQTA
SSYTDTLRYHIIPRRLTLSELRLLPNGYTL**PT**MLSTRRISFTRRSGSSSVTTVGGVEVAFPGLF
YGRHVTVHGLAGILNVRSVDFT**SPAPAP**VNPIH**SP**DHRHF**TP**RRIPH**SP**EKQNQTVLDPVPRSV
SFNVTGRQGGGSSHAVE**AP**VKP**PAP**E**PA**Q**SP**EIGRIHVHSSVNFGT**APS**A**AP**VL**SP**QHSDSGIS
FPPEGYSEAE**APAPAP**VGLEAVVQKKKKNRRVSLMEMSEKSEALDGVRKCESVAVGLKEHISDI
DGVGHMQCYAA
>Medtr2g017970.1 _ fasciclin domain protein _ HC _
chr2_5265732_5267208 _ 20130731 ; MatureChain: 24-280
Q**SPA**A**AP**KAAAK**PAPATPAPA**T**APA**KPLV**PS**LPQP**PS**SDSSSGQDIIKILRKAKSFNTLIRLLK

TTQIINQINSQLVTTKNGGLTIL**AP**DDGAFSQLKAGYFNSLGEHQQKELIQFHVLPVYVSSSNF
DSLSNPVLTLASD**SPS**GYHMNVTAYGNNVNISTGPVNATLTGIVYSDKTLAIYHVDKVLIPLDF
SKPKAL**APAPS**TAK**AP**KADKDSSSEDGDQGETTKATSGANNLISHQGTMFVSLFVGLVAAITIS
I
>Medtr2g017960.1 _ fasciclin domain protein _ HC _
chr2_5268762_5270188 _ 20130731 ; MatureChain: 24-277
Q**SPA**A**AP**K**AP**EK**PAPATPAPA**T**APA**KPLV**PS**LPQ**SPS**SDSSSGQDIIKILRKAKSFNTLIRLLK
TTQIINQINSQLVTTKNGGLTIL**AP**DDGAFLQLKAGYFNSLGERQQKELIQFHVLPVYVSSSNF
DSLSNPVLTLASD**SPS**GYHMNVTAYGNNVNISTGPVNATLTGIVYSDKTLAIYHVDKVLIPLDF
SKPKAL**APAPS**TAK**AP**KADKDSSSEDGDQGESTKATSGANNLISLQGTMFVSLLVAAVTMFG
>Medtr2g017950.1 _ fasciclin domain protein _ HC _
chr2_5273119_5274452 _ 20130731 ; MatureChain: 29-295
QI**SP**VQ**SP**QAS**TPAP**GFNTVPLVPV**TPS**G**APTP**NILPK**SPS**IDIIQILKKAKRFSVLIRLLKTT
QLINQLNSQLVTSSSSDSSSGGLTLF**AP**EDNAFSKLKPGFLNSLSDRHKVELLQFHTLSSFISI
SNFDTLTNPVQTQAGDDAKRLQLNVTTSGGSHVSMTTGTVNATVTGTVYTDNKLAIYQVDKVLV
PLDLVL**PA**K**APAPAP**VSKGDSAKPDDRSKSSSAKDDDDENNDL**PA**EASGAGYSNVDGMWLALVF
GMVLLAGEAML
>Medtr2g087810.1 _ fasciclin_like arabinogalactan protein 16
precursor _ HC _ chr2_36956273_36953488 _ 20130731 ;
MatureChain: 24-446
SPTSQINSNSILIALLDSHYTELTYLIEKANLLQTLENTVSKHNITIF**AP**NNVALERNLDPDFK
HFLLQPGNILSLQTLLLSHVI**PT**RIQSGSTRFKSISDYHLHLETNTTTGEFMVNQENVTHPNDI
TRPDGVIHGIQKILIPRAVVDDFNNRRSLSSITAVKPEG**AP**EADTRKLKKS**SP**EKPG**SP**PEIPI
YEAL**AP**G**PS**L**APAPAP**GPGGPHHHFNGDLQVKDFIKTLLHYGGYKEMADILVNLTSLATEMSRL
VSEGYVLTVL**AP**NDEAMAKLTTEELSEPG**SP**EEIMYYHIIPEYQTEESMYNAVRRFGKVRYETL
RLKYKVVAQEADGSVKFGDGDGSGYLFDPDIYTDGRISVQGIDGVLFPMKEEKEVVPVEQVKPI
SMMGQPRKGVVEHRRGKLLETACWMLGTFRNHSRFTSCQ
>Medtr2g090765.1 _ fasciclin_like arabinogalactan protein _ HC
_ chr2_38919962_38918307 _ 20130731 ; MatureChain: 27-249
Q**APAPAPS**GPINITTIFEKAGQYNFLIRLLNETQQLTQIQTQLNSTSEGFTIF**APT**DNAFQNL**P**
SGAINDLSDQQKVQLILYHV**TP**KYYSLSDFLTVSNPVRTQASGKEGNWGLHFTSQGNQVNVSTG
VVTVPINNALRQQFPLAVYQLDKVLL**PS**ELFGAK**SPS**S**SPAP**KSSK**TPS**K**TPS**SSDVEGD**APSP**
ASSKKDDSAAGRNVGFGFVAGLGLICMGVLL
>Medtr2g093970.1 _ fasciclin domain protein _ HC _
chr2_40045832_40044669 _ 20130731 ; MatureChain: 26-293
AL**SPA**T**SP**IQ**SPPT**QPIS**SPT**KPLV**PT**FPPNDNT**TP**E**TPTPT**TDIVQILKQANSFNIFLRLMKT
TQLINQLNSQLLTIKSGGLTIL**AP**DDSAFSELKPGFLNSLSNEKKLELLQFHVISDFVSSSNFD
TLTNPVRTLAGNKPGKVELNVISYGGNVNISTGSVNTTINGIIYTDKHLAIYKVGKVLI**PS**EF**S**
PAKKIVAEV**PA**L**APAPA**IADAK**APTP**EKEK**PS**SSEDSTTSQVV**PA**VTSSAMRIDMCGSWVALVV
GIVFIGGFYNKN
>Medtr2g093990.1 _ fasciclin_like arabinogalactan protein _ HC
_ chr2_40051115_40049995 _ 20130731 ; MatureChain: 27-250
ASE**SPAPSPS**S**APT**DIIRILKKAGGFTTLIRLLQTTQVATQINAQLLNSNNGLTLF**AP**NDNSFS
TLKPGFLNSLNDQQKNELIQFHEL**PS**FVALSNFDTLSNPVRTQAGDDPERLALNITSSGTQVNL
TTGVVNATVGGSVFSDHQLAIYQVDKVLLPKDFFIPK**SP**P**PAPAP**EKSKDSKKKSAEG**PA**SLDD
NDKSSAMSLKDKNGIMLIFAVYMVIAATLFSL
>Medtr3g460780.1 _ fasciclin_like arabinogalactan protein _ HC
_ chr3_23879139_23880737 _ 20130731 ; MatureChain: 30-421

HNITRILAKHPELSTFNHYLSLTHLASQIDSRTTITVCAVDNAAMEDLLSKHLSINTIKNILSL
HVLLDYFGAKKLHQITNGTALAATMYQATGT**AP**GSSGFVNITDLRGGKVGFGAENNDGALTASY
VKSVEEIPYNISVIQISSVL**PS**AAAE**APTPAPS**QQNLTVIMSKHGCKIFAETLSTFHEAFNTFT
DAIDGGLSVFC**PA**DDAFKAFLPKFKNLTMSGKIDLLEFHGVPIYNSIPMLKSNNGLMHTLATEG
ASNYDFTVQNDGEEITLKTKRVTAKIIDTIIDEQPLAIYTISKVLLPQELFKGE**APSPSPSPAP**
E**PA**AAD**AP**EPPKKGKKKKKKVAAD**APA**DDSD**APA**D**SP**DDDAADDNADDSDGAVRFNGIVGMVLA
LVFGFLFL
>Medtr4g050170.1 _ fasciclin_like arabinogalactan protein 16
precursor _ HC _ chr4_17747502_17752074 _ 20130731 ;
MatureChain: 31-466
FSSCSSSSTVQINSNSILVALLDSHYTELAELVEKAMLLQTLEDTVGKNNITIF**AP**KNQALERD
LDPNFKTFLLEPRNLKSLQTLLMSHII**PT**RINGSVSSKTGSTRHRTLSLEHKIVIQSNETSQQW
LVNGAKIVHLNDVTRPDGVIHGIERLLIPRSVQDDFNRRRSLVSIAAIKPEG**AP**EVDPRTHRLK
KPPPPQNPG**SPPA**LPIFDAL**AP**G**PS**L**APAPAP**GPGGPHHHFNGEAQVKDFIQTLIHYGGYNEMA
DILVNLTSLATEMSRLVSEGYVLTVL**AP**NDEAMAKLTTDQLSEPG**SP**EQIMYYHLIPEYQTEES
MYNAVRRFGKVRYDTLRLPHKVDAEEADGSVKFGHGDGSAYLFDPDIYTDGRISVQGIDGVLFP
HEEKEEVVDKAVTQERKGK**PA**KVVAKQRRGKLLETACWMLGTFGQHSRFNCQ
>Medtr4g059720.1 _ fasciclin_like arabinogalactan protein _ HC
_ chr4_22034570_22035620 _ 20130731 ; MatureChain: 27-250
QK**SPAPAPS**TDT**APT**DIIRILKKAGGFTTLIRLLQTTQVSTQINSQLLNSNGGLTLF**AP**NDNSF
SSLKPGFLNSLSDEQKNKLIQFHLL**PT**FVSMSNFDTLSNPVRTQAGDDPYRLGLNVTSSGNQVN
MTTGIVNVTVGGTVYTDHQLAVYQVDKVLLPRDFFVAKP**PAPAPAP**EKTKSSKKKSSDSDGGVA
GDDDSGAVNVKQRRLMLVPVAVATIIVAIYSW
>Medtr4g099150.1 _ fasciclin domain protein _ LC _
chr4_41062867_41062127 _ 20130731 ; MatureChain: 20-246
IEDHARDLLAATDEMQRANYFTFVMLIKM**SP**PDTRLEGNVTFLMPNDRMLANVTLQEESVSKFL
LRHSI**PSP**LLFDTLKQF**PS**GTTV**PS**LLPNCIMRISNNGRKNFVVNNVKII**SP**NICVSGSSIRCH
GIDGVLSQTCTLENNHNHGVPIPPPQYNNTNTSCEA**SP**PI**PSPS**F**PSP**PYTGDNINPPIWI**APS**
PTISNGERHEYSG**SP**RWFFYDAYLIFIVCLMFSFL
>Medtr5g098060.1 _ fasciclin_like arabinogalactan protein _ HC
_ chr5_42933281_42935396 _ 20130731 ; MatureChain: 22-262
KTV**SPPSPPA**E**SPTPAPAPAPTP**DFVNLTELLTVAGPFHTFLQYLQSTKVLDTFQNQANNTEEG
ITIFVPKDSSFASLKK**PS**LSKLKDDEIKQVILFHALPHFYSLADFKNLSQTAS**TPT**FAGGDYTL
NFTDNSGTVKINSGWSITKVTSAVHATDPVAIYQVDKVLLPEAIFGTDIPPVL**APAPTP**EI**APA**
AD**SPT**EQSADSKSS**SPS**S**SP**DRSSSYKIVSYGIWGNLVLATFGLVVVIL
>Medtr7g075453.1 _ fasciclin_like arabinogalactan protein _ HC
_ chr7_28251817_28250054 _ 20130731 ; MatureChain: 22-418
HNITEILSSNPDYSQYNNFLSQTKLADEINSRQTITVLVLNNAAFSSITSSHPLSVVKKILSLL
ILLDYFDNTKLHQLTNGTTLSTTLFQTTGNAVNNIGSVNITDLKGGKVGFGSA**AP**GSKLDSSYS
KSVKQIPYNISVLEIT**AP**II**AP**GILT**AP**P**PS**SSVNLTALIEKAGCKTFASLISSNGLIKTFQST
ADKGLTIF**AP**NDEAFKAKGVPDLTKLSNAELVSLLQYHAAAKYLPIGSLKTTKDPISSLATNGA
GKFDYTVTTAGDSVTLHTGVDSSRIADSILDS**TP**LAIYSVDSVLL**PS**ELFAT**SPSPAP**E**PA**G**AP**
SPTPAS**APSPA**D**APTP**L**PASPPAPA**GE**SP**DG**APA**D**APS**TAAEKSTGKSAGVNVKATGVFTVAVT
ALSAFVVSFVFMS
>Medtr7g101080.1 _ fasciclin_like arabinogalactan protein _ HC
_ chr7_40783637_40782599 _ 20130731 ; MatureChain: 28-256
Q**PA**I**SPAPS**GPLNITKVLEKAGQFTTFIKLLKATQVSDRINSQLNNSNQGLTIF**APT**DNAFSSL
KSGTLNSISTQNQLQLLQFHIL**PT**LYTISQFQTASNPLHTQAGNSDDGEYPLNVTTSGNQVNVT

TGVIDTTVSNTIYSDNQLAVYQVDQVLLPMALFGQG**PT**A**APA**E**APAPT**KPEKSVRASD**AP**KGSS
D**SPA**DDSSAVGLNGYIVNGATLFVAVFANVVVSCLWM
>Medtr7g106760.3 _ fasciclin_like arabinogalactan protein 16
precursor _ HC _ chr7_43481765_43477708 _ 20130731 ;
MatureChain: 24-448
LP**PT**TTVQINSNSILVALLDSHYTELAELVEKALLLQKLEEAVGNHNITIF**AP**RNQALERDLDP
EFKRFLLEPRNLRSLQTLLMSHIL**PS**RIASHHWP**PT**HLHHHTLSNHRLHLTTN**PT**GAKTVDSAE
ILKPDDVIRPDGVIHGIQRLIIPRSVQEDFNRRRNLRDISAILPEG**AP**EVDPRTNRLKK**PAP**V**P**
AG**AP**PVLPIYDAL**AP**G**PS**L**APAPAP**GPGGPRHHFNGERQVKDFIQTLLHYGGYNEMADILVNLT
SLATEMGRLVSEGYVLTVL**AP**NDEAMAKLTTDQLSEPG**AP**EQIIYYHIIPEYQTEESMYNAVRR
FGKISYDTLRLPHKVVAQESDGSVKFGHGDSSAYLFDPDIYTDGRISVQGIDGVLLPPEEENED
ESS**TP**IRRVK**TP**LVKVASKPRRGKLLQLACSMVGSFGRVCS
>Medtr7g106760.1 _ fasciclin_like arabinogalactan protein 16
precursor _ HC _ chr7_43481515_43477779 _ 20130731 ;
MatureChain: 24-448
LP**PT**TTVQINSNSILVALLDSHYTELAELVEKALLLQKLEEAVGNHNITIF**AP**RNQALERDLDP
EFKRFLLEPRNLRSLQTLLMSHIL**PS**RIASHHWP**PT**HLHHHTLSNHRLHLTTN**PT**GAKTVDSAE
ILKPDDVIRPDGVIHGIQRLIIPRSVQEDFNRRRNLRDISAILPEG**AP**EVDPRTNRLKK**PAP**V**P**
AG**AP**PVLPIYDAL**AP**G**PS**L**APAPAP**GPGGPRHHFNGERQVKDFIQTLLHYGGYNEMADILVNLT
SLATEMGRLVSEGYVLTVL**AP**NDEAMAKLTTDQLSEPG**AP**EQIIYYHIIPEYQTEESMYNAVRR
FGKISYDTLRLPHKVVAQESDGSVKFGHGDSSAYLFDPDIYTDGRISVQGIDGVLLPPEEENED
ESS**TP**IRRVK**TP**LVKVASKPRRGKLLQLACSMVGSFGRVCS
>Medtr7g106760.2 _ fasciclin_like arabinogalactan protein 16
precursor _ HC _ chr7_43481765_43478342 _ 20130731 ;
MatureChain: 24-448
LP**PT**TTVQINSNSILVALLDSHYTELAELVEKALLLQKLEEAVGNHNITIF**AP**RNQALERDLDP
EFKRFLLEPRNLRSLQTLLMSHIL**PS**RIASHHWP**PT**HLHHHTLSNHRLHLTTN**PT**GAKTVDSAE
ILKPDDVIRPDGVIHGIQRLIIPRSVQEDFNRRRNLRDISAILPEG**AP**EVDPRTNRLKK**PAP**V**P**
AG**AP**PVLPIYDAL**AP**G**PS**L**APAPAP**GPGGPRHHFNGERQVKDFIQTLLHYGGYNEMADILVNLT
SLATEMGRLVSEGYVLTVL**AP**NDEAMAKLTTDQLSEPG**AP**EQIIYYHIIPEYQTEESMYNAVRR
FGKISYDTLRLPHKVVAQESDGSVKFGHGDSSAYLFDPDIYTDGRISVQGIDGVLLPPEEENED
ESS**TP**IRRVK**TP**LVKVASKPRRGKLLQLACSMVGSFGRVCS
>Medtr8g087450.1 _ fasciclin domain protein _ HC _
chr8_36147816_36149210 _ 20130731 ; MatureChain: 23-292
Q**SPA**A**AP**K**AP**EK**PAPT**K**PAPA**K**PTPSTPAPAPA**KPLV**PA**LPQ**SPT**TNPDASGNQDIIKILRKAK
SFNTLIRLLKTTQIINQVNAQLVTTKNGGLTIL**AP**DDGAFSELKAGYFNSLGERQQKELIQYHV
LPEYVSSSNFDSLSNPVLTLASD**SP**QGFQINVTAYGNSVNISTGVVDTTITGIVYADKTLAIYH
VNKVLIPLDFIKPKAK**APAPA**IAK**AP**KAEKENSSADDDDQTQAHKDSSDAISLNNMHGISVLVS
IGVSLLAAGVTTML
>Medtr8g087460.1 _ fasciclin domain protein _ HC _
chr8_36150681_36151957 _ 20130731 ; MatureChain: 29-316
QL**SP**IQ**SPT**TS**SPSP**PL**PS**TTAS**TP**L**PA**TTAT**APSP**GLSTVPLV**PTTPA**G**APSPT**ITVPKG**PT**I
DIINILQKAKRFSVLIRLLKTTQLINQLNSQLVS**SPS**GSGGLTIF**AP**EDSAFSKLKAGFLNSLT
DRQKVELLQFHSLASFVSISNFDTLTNPVQTQAGDDARLQLNVTTYGGSQVSMATGAVNATVTG
TVYTDSKLAIYQVDKVLMPLDLVL**PA**K**APA**L**APAPA**KGLLPKAGKTNSSVADDGSGAGSDDGDG
KDL**PA**DISAAGSVMWVNAVVVFGMGLVGGVVL
>Medtr8g087470.1 _ fasciclin domain protein _ HC _
chr8_36154070_36155160 _ 20130731 ; MatureChain: 29-316

QL**SP**IQ**TPS**TS**SPSP**PL**PS**TTA**SP**PL**PA**TTAT**APSP**GLSTVPLV**PTTPT**G**APSPT**ITVPKG**PT**I
DIINILQKAKRFSVLIRLLKTTQLINQLNSQLVS**SPS**GSGGLTIF**AP**EDSAFSKLKAGFLNSLT
DRQKVELLQFHSLASFVSISNFDTLTNPVQTQAGDDARLQLNVTTYGGSQVSMATGAVNATVTG
TVYTDSKLAIYQVDKVLMPLDLVL**PA**K**APA**L**APAPA**KGLLPKAGKTNSSVADDGSGAGSDDGDG
KDL**PA**DVSAAGSVMWVNAVVVFGMGLVGGAVL
>Medtr1g083260.1 _ formin_like 2 domain protein _ HC _
chr1_37055541_37051517 _ 20130731 ; MatureChain: 26-909
QRTHILSVGALINAAGSFHVHTLQDNYYMKEQENEKKQVQKISGLDENEEKQGFIVEKFRSLLG
LKSFHKRV**PS**KSNGDSDSDSDQFL**TPSPSPS**QNIEAEVE**APAPAPTPS**QVMHFHPHSYHQKHHF
HWNQPPKKLHHDDRGRTKRILVAVFVSVGVAAFVISLGLILFCRKKFTNHKKKKPKRTMPLCSS
NTKGKTKGKVSLNPGLDLFYLDALGEDVEQHACTLTKTSDNNVSSSFTKEIVSVHEEELVIKNE
HECVDKIVHEDCDSSEDESFHSFVDSQSNTRLSNASAGSLSDTQSLLL**SP**QNSFSLLPNQL**PSS**
PQNTNDSHQPPY**SP**KQKDQDIENETFVQCPQTSNS**SP**PPPPPPPP**PTP**PLKMPLFTLHSLTTSSR
VSSH**SP**LSLTSHTLS**SP**VNSETSSRSNL**SP**EKDSF**SPS**SSN**PT**K**SP**PPPPCPPPFPRGNSNKNA
K**TP**PPPPYQFPQ**SP**LGKDG**TP**LAKLKPLHWDKVRA**AP**NRTMVWDKLRSSSFELDEEMIESLFGY
NLQSSINNDESKSK**TPSPS**KHVLDPKRLQNITILSKALNVTAEQVCDALMQGKGLSLQQLEALV
KMV**PT**KEEEGKLFNYKGNINELGSAEKFVRAVLCVPFAFQRVETMLYKETFDDEVVHLRNSFSM
LEEACKELRSSRLFLKLLEAVLKTGNRMNVGTIRGGARAFKLDALLKLADVKGTDGKTTLLHFV
VQEIVRSEGIRVSDSIMGKISQKSNKNRTEEEKEEDYRKMGLELVSGLSTELYNVKKTATIDLD
VLASSVSNLYDGMVRLKQLVENELHEDEMCHNFVMSMKLFLQYVDGNLKELRGDEDRVIARVKE
ITEYFHGDVSKEDNPLRIFVIVRDFMGMLDNVCKELRRSK**TP**R**TP**NPL**AP**FR
>Medtr4g081410.2 _ formin_like 2 domain protein _ HC _
chr4_31560742_31567076 _ 20130731 ; MatureChain: 26-990
VSSIEKKETQDGLVRELFDSASGLLDEHTAKVFRTTCCEDFIHLKKEVEYHDLCLPLELFASTN
KVSSTVRPFAQTDIQKLLNACHPQIKEIFLHYLRKNNLLLHVLGEEDDSKIWHVTNTGYLFSTS
SIPRRNPGRVLLQHISEP**PS**LG**PT**VG**SPTPS**L**TPSP**E**PS**LP**PSP**E**PS**L**SPSPAPAP**LPPKPL**SP**
PL**SPA**SFFPKL**TPPA**AADIS**APPS**SDTSGKEDNHSNKTTVVLSVVITISVIFIAAALFFLCFRK
AGRRRQNDERPLLSLSMNDYSFG**PS**NHAFGNTTKGEKLGFQSSSNNLGDNKKTSLQGNQSMGAF
AVVG**SP**FELNPPPGRVGTIHSGMPPLRPPPGRMNPLPHEP**PS**F**TP**FGNTAVSAA**AP**PPPRQSGV
ASARP**PATPSP**PQSLLAGAKPVPPRPPLH**PA**LPGAK**PSP**PPP**PAP**LGAKPGPPPPPP**PAPPS**AK
PGPPPPPP**PAPS**GAKPGPRPPPPPPPKSGV**AP**PRPPIGPKAGGPKATENAEAGAEGGADTSKAKL
KPFFWDKV**PA**NSDQSMVWNQIKSGSFQFNEEMIETLFGYNAVNKNNGQRQKESSSSQD**PSP**QYI
QIVDKKKAQNLLILLRALNVTMEEVCDALYEGNEL**PS**EFLQTLLKM**APT**SDEELKLRLFNGDLS
QLG**PA**DRFLKAMVDI**PS**AFKRMEVLLFMCTFKEELTTTMESFAVLEVACKELRNSRLFHKLLEA
VLKTGNRMNDGTYRGGAQAFKLDTLLKLSDVKGTDGKTTLLHFVVQEIIRSEGIKAARAAKDSQ
SLSNIKTDELHETEDHYRELGLEMVSHLSTELENVKRGSVLDADSLTATTIKLGHGLVKAKDIL
NKNLKNVEDDRGFRETVESFVKNAEADVKKLLEDEKKIMALVKSTGDYFHGNATKDDGLRLFVV
VRDFLIMLDKVCKEVRDAQKK**PA**KPIKQETSRGLSSSDTR**PSPS**DFRQRLF**PA**IAERRIDDDSS
DEE**SP**
>Medtr4g081410.1 _ formin_like 2 domain protein _ HC _
chr4_31561450_31566634 _ 20130731 ; MatureChain: 20-984
VSSIEKKETQDGLVRELFDSASGLLDEHTAKVFRTTCCEDFIHLKKEVEYHDLCLPLELFASTN
KVSSTVRPFAQTDIQKLLNACHPQIKEIFLHYLRKNNLLLHVLGEEDDSKIWHVTNTGYLFSTS
SIPRRNPGRVLLQHISEP**PS**LG**PT**VG**SPTPS**L**TPSP**E**PS**LP**PSP**E**PS**L**SPSPAPAP**LPPKPL**SP**
PL**SPA**SFFPKL**TPPA**AADIS**APPS**SDTSGKEDNHSNKTTVVLSVVITISVIFIAAALFFLCFRK
AGRRRQNDERPLLSLSMNDYSFG**PS**NHAFGNTTKGEKLGFQSSSNNLGDNKKTSLQGNQSMGAF
AVVG**SP**FELNPPPGRVGTIHSGMPPLRPPPGRMNPLPHEP**PS**F**TP**FGNTAVSAA**AP**PPPRQSGV
ASARP**PATPSP**PQSLLAGAKPVPPRPPLH**PA**LPGAK**PSP**PPP**PAP**LGAKPGPPPPPP**PAPPS**AK

PGPPPPPP**PAPS**GAKPGPRPPPPPPKSGV**AP**PRPPIGPKAGGPKATENAEAGAEGGADTSKAKL
KPFFWDKV**PA**NSDQSMVWNQIKSGSFQFNEEMIETLFGYNAVNKNNGQRQKESSSSQD**PSP**QYI
QIVDKKKAQNLLILLRALNVTMEEVCDALYEGNEL**PS**EFLQTLLKM**APT**SDEELKLRLFNGDLS
QLG**PA**DRFLKAMVDI**PS**AFKRMEVLLFMCTFKEELTTTMESFAVLEVACKELRNSRLFHKLLEA
VLKTGNRMNDGTYRGGAQAFKLDTLLKLSDVKGTDGKTTLLHFVVQEIIRSEGIKAARAAKDSQ
SLSNIKTDELHETEDHYRELGLEMVSHLSTELENVKRGSVLDADSLTATTIKLGHGLVKAKDIL
NKNLKNVEDDRGFRETVESFVKNAEADVKKLLEDEKKIMALVKSTGDYFHGNATKDDGLRLFVV
VRDFLIMLDKVCKEVRDAQKK**PA**KPIKQETSRGLSSSDTR**PSPS**DFRQRLF**PA**IAERRIDDDSS
DEE**SP**
>Medtr8g027995.1 _ formin_like 2 domain protein _ HC _
chr8_10328257_10323250 _ 20130731 ; MatureChain: 23-1071
NEPRRFLHQPFPQNNNL**SPSSPPSSP**P**PSP**NPKYPFST**TP**PNTSSSSS**TP**PFF**PT**Y**PSTP**PP**PS**
PSSFASF**PA**NISSLTIPQTQKPKSSSSKLLAVAITAVIAAVAVVAISAFIYCRRSRNKRFLADD
KTLRSDSSIRLFPRDGGVATIAKSRNVSSTSSEFLYLGTIANSRADELPDPRGAGGGGRNPRKM
D**SP**ELQPLPPLMRQGSMFDEGNGGATVTVGEDDEEEFY**SP**RGSLNGNGSGSRRVFAGISAENLV
GRSSSESTSSSSSYSSSSA**SP**DRSHSISL**SP**PVSL**SP**RRSQPK**SP**ENVV**TPAPT**QPLLVSDVGR
SSLSSSRASSNRHVQSCSSMSS**SP**EKIFAGECK**SPS**L**SP**LNL**SPT**KNLDGSFVKVEKTQSCNEE
GSS**SP**RLSNASSGKSSSSSSSAFTL**PSP**EKMMTMMNLHSNHGLDQ**SPT**ISDVSDRFRH**SP**LSSL
PL**SPT**LLS**SP**ERDIMSTQPPPPPPQ**PA**SRKHWEIPDLL**TP**IAE**SPA**ILNQNGVSQRKHWEIPVL
S**TP**I**TPS**NRVS**AP**PPPPPPPPPPPPLTMPMKQRKQWEV**PSPTTP**VGQQVVCRPPELKP**PS**RPFV
LQ**TPS**NTLV**SP**VELP**PS**FEENEEVSKPKLKPLHWDKVRASSDREMVWDHLRSSSFKLNEEMIET
LFVVN**TP**NPKPKDA**TP**RSVL**TPPS**HEDRVLDPKKSQNIAILLRAVNVTVEEVCEALLEGVTDTL
GTELLESLLKM**APS**KEEERKLKEHKDD**SP**NKLGSAEKFLKAVLDVPFAFKRVEAMLYIANFESE
VEYIRKSFQTLEVACEELRNSRMFLKLLEAVLKTGNRMNVGTNRGDAHAFKLDTLLKLVDVKGA
DGKTTLLHFVVQEIIRTEGARHSDTSTNQ**TPS**ATLIDDAKCRRLGLQVVSSLSSDLANVKKAAT
MDSEVLTSEVSKLSKGITHIAEIVKLNQTVGSDETVRKFAESMNKFMRMAEEEILRIQAQESVA
LSLVKEITEYFHGNLSKEEAHPFRIFMVVRDFLTVLDRVCKEVVNINERTMISSAHRFPVPVNP
MLPQPLPGLHGKRHYSSSDDDSS**SP**
>Medtr1g007880.2 _ ABIL1_like protein _ HC _ chr1_609028_606935
_ 20130731 ; MatureChain: 21-342
GKIATQPLPRMASNYDEVFMHQTLLFDDSLTDLKNLRTQLYSAAEYFELSYTNDDQKQILVETL
KDYAVKALINTVDHLGSVAYKVSDLLDEKVTEVFGEDLRLSCIEQRIKTCQGFMDHEGHTQQSL
VIS**TP**KHHKRYILPVGETLHGTNSTKSKYIGCHLDDEDDWHHFRNAVRATIRE**TPT**STSSKGN**S**
PSPSLQPQRVGAFSFT**SP**NMAKKDLEKRTV**SP**HRFPLSRTGSMSSRST**TP**KTGRST**TP**NSSNRA
TT**SPS**NARVRY**PS**EPRKSASMRLSSDVNNIRDIDQH**PS**KSKRLLKSLLSRRKSKKDDTLYTYLD
EY
>Medtr1g021630.1 _ cysteine_rich receptor_kinase_like protein _
LC _ chr1_6507333_6502386 _ 20130731 ; MatureChain: 21-676
ATTDATGKFRYICDQGNDDGNYTTNSTYHTNLKTLLSTLTSNKDINYGFYNSSYGVNTDKVNAI
GLCRGDVKPDDCQNCLQNSTALLTQLCQNRKEAIGWYDDEKCMLRSSHRLIFGINEMG**PA**MIAW
NLQNATKEVEDEFDKGLKNLLYNLKNLAASGESDLKYAVGSDVIGPNNNQTIFALVQC**TP**DLSE
TLCDECLDKSIKEIPNCCDNRIGIGGRFLR**PS**CYLRYETNSLFYQQTQD**SPS**SSSS**SPSPTPSP**
ASV**PS**FS**AP**PPFAENTTS**SP**GKSNNSTTIGIALGVPIAIVVMVFIFICIYLRLRKPKKTCEEVQ
EEEEDDIIDITTEQLQLDFNTIRIATNDFSDSDKLGKGGFGAVYKGRLFNGQEVAVKRLSMNSG
QGDSEFKNEVFLVAKLQHRNLVRLLGFCLEGRERLLVYEFVCNKSLDYFIFDQAKRAQLNWGKR
YLIIILGIARGILYLHQDSRFRVIHRDLKASNILLDEHMNPKIADFGMARLFGVDQTQENTNRIV
GTYGYM**AP**EYVMHGQFSVKSDVFSFGILVLEIVSGAKNSSIRDGENTEYLSSFAWRNWREGTAT
NIIDSTLNNDSRNEILRCIHIGLLCVQENIVNR**PT**MASIVVMLNSDSVTLPMPFEDLKSLSDSS

AKESVNGASNTEPFPR
>Medtr1g021635.1 _ cysteine_rich receptor_kinase_like protein _ LC _ chr1_6517643_6514070 _ 20130731 ; MatureChain: 22-674
DVDQGKNFQYFCDQNNDGGIYTTNSTYHTNLNTLLSILTSNKEINYGFYNSSYGINSDKVNAIG
LCRGDLKPNDCRNCLQNSTVFLTQRCQNRKQAIGWYDDDRCMLRYSSRSIFGLYDTRPYYEAWS
LKTAINEDEFDKVRKNLLDNLRNRAASGDSDLKYAVGSDEVGPNNNQTIYGLAQC**TP**DLFKTFC
DDCLVQSINEIAICCNNRMSARVVR**PS**CYLRYETDSLFYQ**PT**QD**SPS**SL**SPSPT**SV**PS**LA**AP**PP
FANNTTSYPGKSNNYGSTIGIAIGVPIALVAMVFIFICIYLKVRKPKKRFEEVQEEEDDDKIEI
TEGLQFHFNTIRIATNDFSDSDKLGQGGFGVVYKGRLSNGLEIAIKRLSMNSGQGDLEFKNEVF
FLAKLQHRNLVRLLGFCLEGSERLLIYEFVHNKSLDYFIFDQAKKAQLNWERRYTIILGIARGI
LYLHEDSRVRVIHRDLKASNILLDKRMNPKIADFGMARLFGVDQTQENTNRIVGTYGYM**AP**EYA
RHGQFSTKSDVFSFGILVLEIVSGTKNSYIRDGENTEYLSSFAWRNWKEGTAANIID**PT**LNNDS
LNEIMRCIHIGLLCVQENVASR**PT**MASVVVTLN**SPS**VTLPIPLQ**PA**FHIGPQDMKSSGHSSAQE
SVNGASNTQLFPR
>Medtr1g069340.1 _ receptor_like Serine_Threonine_kinase ALE2 _ HC _ chr1_29953415_29947213 _ 20130731 ; MatureChain: 25-756
KLVSMSVSFASSE**PS**KVWLVKSSSGTSSAHM**PS**QPFQ**APS**K**TP**GPKHPHHPRQYHRVKPYAV**SP**
SPSEGQECGLSCRDPLTT**TP**FG**SP**CGCVFPLKVGLLLDV**AP**LVVFPVLRELEIELAFGTYLKQS
QVRIMGVSADIQNQERTIVDIYLVPLGEKFSNTTVVLISRRFWHKKVPLNRSLFGDYTVLYTNY
PGM**PSSP**PHGTITGSGPL**PS**GSAAGILPFTANFINKNEEMTLRTIIIISASSIILFLGLVGAFF
IIFKLRKLRR**PS**GAVHPPFTSSLNKSSAMESMLSSRITSSSSMSHTSNLATSALSVKTFSLSEL
EKATNKFSSQRLLGEGGFGRVYHGRLDDGTDVAVKQLRRDIHQSGDREFIVEVEMLCRFHHRNL
VKLIGICTEGHKRCMVYELIRNGSVESHLHGVDRINHPLDWEARKKIALGAARGLAYLHEDSNP
RVIHRDFKASNVLLEDDF**TP**KVSDFGLAREATEGSHSI**PT**RVVGTFGYV**AP**EYAMTGHLLVKSD
VYSYGVVLLELLTGRKPVDMSQPLGEENLVVWARPLLKSREGLEQLVD**PT**LAGTYDFDEMTKVA
AVASMCVHLEVTKRPFMGEVVQALKLIFNDNDGMDRYSRKESSDQESDFRGDLSDSSWRCDEPE
DITCRLGFRQPLASSFITMDYSSGPLEELENRHFVASGFVADDMSLASRLGNRSGPLRTSRRSN
LSFYRFNGSQSD**PA**VL**PS**KRVRNEGYWP
>Medtr1g105615.1 _ cysteine_rich receptor_kinase_like protein _ HC _ chr1_47604247_47607685 _ 20130731 ; MatureChain: 28-506
QDPNFLYQVCSKNNFTSNSTYQTNLKTLFSSLSSKAKTDQNIQFFNNTVAGKNHSETVYGLYMC
RVDI**PS**NLCTSCVGNATQRLSTDKECTVSVAAVMWYDECMVRYSNKSFFSTVATGPGYVL**PSPT**
NMTNQGSFNRLLYDTLNKTSDEASSSSKNFATREAKINIFQNLYCLAQCTQDIEERECRSCLYG
LINSDLPRCCAGTQGGRVLYPNCVVRFEIYPFYRSLTLP**PTPAPA**GLV**PSP**NSAGKSRIAIIVA
IPIVVASTIVFAVCFYYRRRKARTSRSNILLQENFGRESTTLEGLQFDLATIATATNNFSHENK
IGKGGFGEVYKGTLRDGRDIAVKRLSTSSTQGSIEFKNEILLIAKLQHRNLVALIGFCLEEQEK
ILIYEYVPNGSLDHFLFGAQQQKLSWSERYKIIGGTALGVLYLHDYSRLKVIHRDLK**PS**NILLD
EHMNPKISDFGMARIVDIGQDRAKTTRIVGT
>Medtr1g105615.2 _ cysteine_rich receptor_kinase_like protein _ HC _ chr1_47604242_47607699 _ 20130731 ; MatureChain: 28-506
QDPNFLYQVCSKNNFTSNSTYQTNLKTLFSSLSSKAKTDQNIQFFNNTVAGKNHSETVYGLYMC
RVDI**PS**NLCTSCVGNATQRLSTDKECTVSVAAVMWYDECMVRYSNKSFFSTVATGPGYVL**PSPT**
NMTNQGSFNRLLYDTLNKTSDEASSSSKNFATREAKINIFQNLYCLAQCTQDIEERECRSCLYG
LINSDLPRCCAGTQGGRVLYPNCVVRFEIYPFYRSLTLP**PTPAPA**GLV**PSP**NSAGKSRIAIIVA
IPIVVASTIVFAVCFYYRRRKARTSRSNILLQENFGRESTTLEGLQFDLATIATATNNFSHENK
IGKGGFGEVYKGTLRDGRDIAVKRLSTSSTQGSIEFKNEILLIAKLQHRNLVALIGFCLEEQEK
ILIYEYVPNGSLDHFLFGAQQQKLSWSERYKIIGGTALGVLYLHDYSRLKVIHRDLK**PS**NILLD
EHMNPKISDFGMARIVDIGQDRAKTTRIVGT

>Medtr1g105650.1 _ cysteine_rich receptor_kinase_like protein _
HC _ chr1_47634426_47637645 _ 20130731 ; MatureChain: 30-661
QN**PT**HLYQICSMNKTTSNSTYKSNIITLFSYLSSNATTNNEFYNTV**PS**RNISNSVYGLFMCRGD
VS**SP**LCIQCVANATQRLSSDADCSLSKQAVIWYDECMVRYSNTSFFSTMATRPGVFMMNSLNIT
NQETFMPLLFDTMNKTADNAANSSVGAKKYATKEASISGFQTLYCMAQCTEDLSQQDCRTCLSD
AIGALPQCCDGKQGGRVLF**PS**CNVRYELYPFYRNL**APSPSPAPS**A**APA**LVP**PS**TS**TPT**LGGSSG
ISSGTIVAIVVPISVATLLLIVGVCFLSKRAWKKKHDSAAQDPKTETDISTVESLRFDLSTLEE
ATNKFSEANKLGEGGFGEVYKGSL**PS**GQEIAVKRLSKHSGQGGEQFKNEVELVAQLQHRNLARL
LGFCLEREEKILVYEFVANKSLDYILFDPEKQRLLDWTRRYKIIGGIARGIQYLHEDSRLKIIH
RDLKASNILLDGDMNPKISDFGMAKLFGVDQTQGNTSRIVGTYGYM**SP**EYAMHGEFSIKSDVYS
FGVLVMEIISGKKSNSFYETGVADDLVTYAWKLWKNG**TP**LELVDHTVRESY**TP**NEAIRCIHIGL
LCVQEDPEDR**PT**MATVVLMLDSFTVTLPVPKQ**PA**FFLHSGTDSNM**PT**IQISQSTTN
>Medtr2g016620.1 _ LRR receptor_like kinase _ HC _
chr2_5124647_5122806 _ 20130731 ; MatureChain: 25-437 -- SKETCHY
TSSS**SPS**LPLP**PTPSPTSPPSSPT**TSKTTLDPKQITALQSLNI**PT**SKDPCTQ**PS**YHNATLCDSS
KPFRHLISLTLSNCTSHLSLSFTALKSLSTLTSLNLINCPI**SP**FHFPQELITSLKTFTFINSFN
KISGVWLSQLHNLTDLTISNVQVKASGPYVILGHMNYLNSLTISNANLTGFLPGHIHSNLTHVD
FSGNQLKGNIPFSITMLESLESLNLSSNKLVGEI**PS**SIGDLISLKNLSLASNSFSGSIPDSFSS
VPDLVHLDLSSNQLNGTIPKFISQLKNIKYLNLANNNLHGAVPFNQSFIKGLEVFKVGGNSNLC
YNHSILSSKLKLGISHCDKYGKPV**SPPA**AKDSSEDDSSDDDYDDSDDDDGSSKHKKDQHHGPNK
FVLGVAIALSSIVFLIVFLIVCSKCCRGR
>Medtr2g039290.3 _ receptor_like Serine_Threonine_kinase ALE2 _
HC _ chr2_17223213_17211976 _ 20130731 ; MatureChain: 20-1225
QGSEGSII**SPSPA**FLPGVHPEGE**AP**GPIHNGQSWRSTAS**SPS**DPDGSVI**SPSPA**ILPMDPFSSE
APSLLHPNGTILQPPVALP**PA**TS**AP**PPQKIKGIESTV**SPSPSPS**TKSI**SP**PYKS**APAPS**TAERN
LP**PS**IQPIPPQMK**TPT**V**SP**PIS**TPTAP**DPVA**TP**PGNLPKT**SP**ISQPIDHGSLPPNVDNRNESKS
HNPEPV**SPAP**VA**TPS**TNEPKV**SP**ISHSTNNGSFPPPQPM**SPAP**VFNIPKHLPVNQSTEPRSLP**P**
TVHRRNSSISHTLEPVSQ**AP**VAEP**PA**NFPKNSSVSQ**PS**QHGSVPPNVHNTTIHKGHIH**TP**EPVM
PPPISTFPVDPPLIHPVI**PAAP**PNEV**PAP**VI**SPTLTPS**RSFNGKNGGEPVS**AP**LNKKPK**SPPA**I
VN**SPA**Q**APS**VNKARPFHH**AP**EPLT**SP**PK**SP**FNKEDH**SPA**S**SPS**TTFHKHQHTRNTIT**SPAPA**SS
YFV**SPPT**SKHQDKPIP**PS**FL**PT**NRRRHN**APAP**MNRGSAD**SP**FTFPIQ**SP**VSQV**SPAPSPS**FKTF
PHSTKIPFHPPKI**SP**QRSFSK**SP**KKPILPRVQALPPPPPNEDCLSFVCSEPYTN**SP**PGVPCMCI
WPMRVGLRLNVPLYTFFPLVSELASEISSGVFMKQSQVRIMGANAATDQPDKTDALIDLVPLGE
QFDNTTAFLTSDRFWHKKVVIKSSYFGDYEVLYISYPGLP**PSP**PLP**PS**SVNMIDGGPYSNNGNN
GRTIKPLGVDIQKRQHRSGLSKGIIAIIALSSFLAIVLCSAAVFALIKFRDHVSESQ**PT**S**TP**RV
FP**PS**LTK**TP**GTAG**PS**NAGASASTSFRSSIAAYAGSAKTFSMNEIEKATDNFH**PS**RILGEGGFGL
VYSGNLEDGSKVAFKVLKREDHHGDREFLSEVEMLSRLHHRNLVKLIGICTELSFRCLVYELIP
NGSVESHLHGVDREK**SP**LDWSARIKIALGAARGLAYLHEDS**SP**HVIHRDFKSSNILLENDF**TP**K
VSDFGLARTAADEDNRHISTRVMGTFGYV**AP**EYAMTGHLLVKSDVYSYGVVLLELLTGRKPVDF
SQPPGQENLVAWARPLLTSREGLEVIID**PS**LGSNVPFDSVAKVAAIASMCVQPEVSDRPFMGEV
VQALKLVCNECDEAKEAGSTSSNKDGSSSDFYTVTEQLPDNFQSHSAAANYDFGVDIENGLSTS
EIFSSSARFERQVSGSFRRHSYSGPLRTGRSKRLWHIIRKLSGGSVSEHGDSLR
>Medtr2g039290.1 _ receptor_like Serine_Threonine_kinase ALE2 _
HC _ chr2_17223213_17211980 _ 20130731 ; MatureChain: 20-1447
QGSEGSII**SPSPA**FLPGVHPEGE**AP**GPIHNGQSWRSTAS**SPS**DPDGSVI**SPSPA**ILPMDPFSSE
APSLLHPNGTILQPPVALP**PA**TS**AP**PPQKIKGIESTV**SPSPSPS**TKSI**SP**PYKS**APAPS**TAERN
LP**PS**IQPIPPQMK**TPT**V**SP**PIS**TPTAP**DPVA**TP**PGNLPKT**SP**ISQPIDHGSLPPNVDNRNESKS
HNPEPV**SP**GSFLH**PS**LA**SP**P**PSPS**AKSL**SP**PYKAV**PAPS**TAERNFPPPMQPIPPQSK**AP**IVRPP

IS**TP**IAQ**AP**VAIP**PA**TKT**SPT**SQPIDHGSLP**PS**VDKRNESKSHNLEPV**SP**GSFLQPPVALQP**PT**
S**AP**PPQKTKGSESSI**SPSPS**QSTKPL**TP**PDKAV**PAPS**TAERSFPPPMRPIPPQRK**APAP**VA**TP**P
GNLPKT**SP**ISQPIEHGSLPPKVDKRNESRNHNLEPV**SPAP**VA**TPS**TNEPKV**SP**ISHSTNNGSFP
PPQPM**SPAP**VFNIPKHLPVNQSTEPRSLP**PT**VHRRNSSISHTLEPVSQ**AP**VAEP**PA**NFPKNSSV
SQ**PS**QHGSVPPNVHNTTIHKGHIH**TP**EPVMPPPISTFPVDPPLIHPVI**PAAP**PNEV**PAP**VI**SPT**
L**TPS**RSFNGKNGGEPVS**AP**LNKKPK**SPPA**IVN**SPA**Q**APS**VNKARPFHH**AP**EPLT**SP**PK**SP**FNKE
DH**SPA**S**SPS**TTFHKHQHTRNTIT**SPAPA**SSYFV**SPPT**SKHQDKPIP**PS**FL**PT**NRRRHN**APAP**MN
RGSAD**SP**FTFPIQ**SP**VSQV**SPAPSPS**FKTFPHSTKIPFHPPKI**SP**QRSFSK**SP**KKPILPRVQAL
PPPPPNEDCLSFVCSEPYTN**SP**PGVPCMCIWPMRVGLRLNVPLYTFFPLVSELASEISSGVFMK
QSQVRIMGANAATDQPDKTDALIDLVPLGEQFDNTTAFLTSDRFWHKKVVIKSSYFGDYEVLYI
SYPGLP**PSP**PLP**PS**SVNMIDGGPYSNNGNNGRTIKPLGVDIQKRQHRSGLSKGIIAIIALSSFL
AIVLCSAAVFALIKFRDHVSESQ**PT**S**TP**RVFP**PS**LTK**TP**GTAG**PS**NAGASASTSFRSSIAAYAG
SAKTFSMNEIEKATDNFH**PS**RILGEGGFGLVYSGNLEDGSKVAFKVLKREDHHGDREFLSEVEM
LSRLHHRNLVKLIGICTELSFRCLVYELIPNGSVESHLHGVDREK**SP**LDWSARIKIALGAARGL
AYLHEDS**SP**HVIHRDFKSSNILLENDF**TP**KVSDFGLARTAADEDNRHISTRVMGTFGYV**AP**EYA
MTGHLLVKSDVYSYGVVLLELLTGRKPVDFSQPPGQENLVAWARPLLTSREGLEVIID**PS**LGSN
VPFDSVAKVAAIASMCVQPEVSDRPFMGEVVQALKLVCNECDEAKEAGSTSSNKDGSSSDFYTV
TEQLPDNFQSHSAAANYDFGVDIENGLSTSEIFSSSARFERQVSGSFRRHSYSGPLRTGRSKRL
WHIIRKLSGGSVSEHGDSLR
>Medtr2g039290.2 _ receptor_like Serine_Threonine_kinase ALE2 _ HC _ chr2_17223213_17211976 _ 20130731 ; MatureChain: 20-1459
QGSEGSII**SPSPA**FLPGVHPEGE**AP**GPIHNGQSWRSTAS**SPS**DPDGSVI**SPSPA**ILPMDPFSSE
APSLLHPNGTILQPPVALP**PA**TS**AP**PPQKIKGIESTV**SPSPSPS**TKSI**SP**PYKS**APAPS**TAERN
LP**PS**IQPIPPQMK**TPT**V**SP**PIS**TPTAP**DPVA**TP**PGNLPKT**SP**ISQPIDHGSLPPNVDNRNESKS
HNPEPV**SP**GSSDGNFGFHHPGSFLH**PS**LA**SP**P**PSPS**AKSL**SP**PYKAV**PAPS**TAERNFPPPMQPI
PPQSK**AP**IVRPPIS**TP**IAQ**AP**VAIP**PA**TKT**SPT**SQPIDHGSLP**PS**VDKRNESKSHNLEPV**SP**GS
FLQPPVALQP**PT**S**AP**PPQKTKGSESSI**SPSPS**QSTKPL**TP**PDKAV**PAPS**TAERSFPPPMRPIPP
QRK**APAP**VA**TP**PGNLPKT**SP**ISQPIEHGSLPPKVDKRNESRNHNLEPV**SPAP**VA**TPS**TNEPKV**S**
PISHSTNNGSFPPPQPM**SPAP**VFNIPKHLPVNQSTEPRSLP**PT**VHRRNSSISHTLEPVSQ**AP**VA
EP**PA**NFPKNSSVSQ**PS**QHGSVPPNVHNTTIHKGHIH**TP**EPVMPPPISTFPVDPPLIHPVI**PAAP**
PNEV**PAP**VI**SPT**L**TPS**RSFNGKNGGEPVS**AP**LNKKPK**SPPA**IVN**SPA**Q**APS**VNKARPFHH**AP**EP
LT**SP**PK**SP**FNKEDH**SPA**S**SPS**TTFHKHQHTRNTIT**SPAPA**SSYFV**SPPT**SKHQDKPIP**PS**FL**PT**
NRRRHN**APAP**MNRGSAD**SP**FTFPIQ**SP**VSQV**SPAPSPS**FKTFPHSTKIPFHPPKI**SP**QRSFSK**S**
PKKPILPRVQALPPPPPNEDCLSFVCSEPYTN**SP**PGVPCMCIWPMRVGLRLNVPLYTFFPLVSE
LASEISSGVFMKQSQVRIMGANAATDQPDKTDALIDLVPLGEQFDNTTAFLTSDRFWHKKVVIK
SSYFGDYEVLYISYPGLP**PSP**PLP**PS**SVNMIDGGPYSNNGNNGRTIKPLGVDIQKRQHRSGLSK
GIIAIIALSSFLAIVLCSAAVFALIKFRDHVSESQ**PT**S**TP**RVFP**PS**LTK**TP**GTAG**PS**NAGASAS
TSFRSSIAAYAGSAKTFSMNEIEKATDNFH**PS**RILGEGGFGLVYSGNLEDGSKVAFKVLKREDH
HGDREFLSEVEMLSRLHHRNLVKLIGICTELSFRCLVYELIPNGSVESHLHGVDREK**SP**LDWSA
RIKIALGAARGLAYLHEDS**SP**HVIHRDFKSSNILLENDF**TP**KVSDFGLARTAADEDNRHISTRV
MGTFGYV**AP**EYAMTGHLLVKSDVYSYGVVLLELLTGRKPVDFSQPPGQENLVAWARPLLTSREG
LEVIID**PS**LGSNVPFDSVAKVAAIASMCVQPEVSDRPFMGEVVQALKLVCNECDEAKEAGSTSS
NKDGSSSDFYTVTEQLPDNFQSHSAAANYDFGVDIENGLSTSEIFSSSARFERQVSGSFRRHSY
SGPLRTGRSKRLWHIIRKLSGGSVSEHGDSLR
>Medtr3g075440.1 _ LRR receptor_like kinase family protein _ HC _ chr3_34342818_34340379 _ 20130731 ; MatureChain: 26-628
KLTLH**PS**DTKALSTLQNNLGLNLDTTTNNLCNKEGVFCERRLTNNESYALRVTKLVFKSRKLSG
IL**SPT**IGKLTELKEISLSDNKLVDQI**PT**SIVDCRKLEFLNLANNLFSGEV**PS**EFSSLIRLRFLD

ISGNKLSGNLNFLRYFPNLETLSVADNHFTGRVPVSVRSFRNLRHFNFSGNRFLEGVPLNQKLL
GYEDTDNT**AP**KRYILAETNNSSQTRPHRSH**SP**GA**APAPAPAAP**LHKHKKSRRKLAGWILGFVAG
AFAGILSGFVFSLLFKLALILIKGKGKGSG**PA**IYSSLIKKAEDLAFLEKEDGLASLEKIGQGGC
GEVYKAELPGSNGKMIAIKKIIQPPKDAAELAEEDSKLLHKKMRQIKSEIDTVGQIRHRNLLPL
LAHISRPDCHYLVYEFMKNGSLQDMLHKVERGEAELDWLARHKIALGIAAGLEYLHTSH**SP**RII
HRDLK**PA**NVLLDDEMEARIADFGLAKAMPDAQTHITTSNVAGTVGYI**AP**EYHQILKFNDKCDIY
SFGVMLGVLVIGKL**PS**DDFFTNTDEMSLVKWMRNVMTSENPKEAIDARLLGNGFEEQMLLVLKI
ASFCTMDNPKERPDAKNVRIMLYQIKH
>Medtr3g093710.1 _ receptor_like kinase _ HC _
chr3_42815002_42818320 _ 20130731 ; MatureChain: 27-635
ADLNSDRQALLEFASAVPH**AP**RLNWNESSSICTSWVGVTCNSNHTRVVGIHLPGIGLTGSIPEN
TIGKLDALRVLSLHSNGLGGNL**PS**NILSI**PS**LQFAHLQKNNFSGLI**PS**SV**SP**KLVALDISFNSF
SGSI**PS**AFQNLRRLTWFYLQNNSISGPIPDFNL**PS**LKYLNLSNNKLNGSIPNSIKTF**PS**SAFVG
NSLLCGPPLLNYCSSI**SPSPSPSPA**STQIQKATVAHKKSFGVAAILALVIGGIAFLSLLALVFF
LCFLKKKNNKRSGILKGKSSSCAGKAEVSKSFGSGVQAAEKNKLFFFEGSSYTFDLEDLLKASA
EVLGKGSYGTAYKAVLEEGVTMVVKRLKEVMVGKKEFEQQLDIVGRIGRHPNVMPLRAYYYSKD
EKLLVYSYMPEGSLFFLLHGNKGAGR**TP**FDWNSRVKVALGAAKGIAFIHTEGGQKFTHGNIKST
NVLITEEFDSCISDVGLPPLMN**APA**TMSRTNGYR**AP**EVTDSKKITQKSDVYSFGVLLLELLTGK
VPMRYPGYEDVVDLPRWVRSVVREEWTAEVFDEELLRGQYVEEEMVQMLQIALACVAK**TP**DMRP
RMDEAVRMIEEIKNPEFKNRTSSESEYSNVQ**TP**
>Medtr3g093930.1 _ leucine_rich receptor_like kinase family
protein _ HC _ chr3_42916167_42919917 _ 20130731 ; MatureChain:
35-989
VN**PS**LNDDVLGLIVFKADIKDPKGKLTSWNEDDESACGGSWVGVKCNPRSNRVVEVNLNGFSLS
GRIGRGLQRLQFLRRLYLGNNNLTGSINANIATIDNLRVLDLSNNNLSGVVPDDFFRQCGSMRV
VSLARNRFSGNV**PS**SLGSCAAIATIDLSFNQFSGNVPKGIWSLSGLRSLDMSDNLLEGEVPEGV
EAMKNLRSISLARNSFSGKIPDGFGSCLLLRSIDFGDNSFSGSV**PS**DLKELVLCGYFSLHGNAF
SGDVPDWIGEMKGLQTLDLSQNRFSGLVPNSLGNIWSLKTLNLSGNGFTGNLPESMVNCTNLLA
LDVSQNSLSGDL**PS**WIFRWDLEKVMVVKNRISGRAK**TP**LYSLTEASVQSLQVLDLSHNAFSGEI
TSAVSGLSSLQVLNLSYNSLGGHI**PA**AIGDLKTCSSLDLSYNKLNGSI**PS**EVGGAVSLKELSLE
NNFLIGKIPISIENCSSLKTLILSKNRLSGSI**PS**AVASLTNLKTVDLSFNNLTGNLPKQLSNLP
NLITFNLSHNNLKGEL**PA**GGFFNTI**SPS**SVSGNPFICGSVVNKKCPVKLPKPIVLN**PT**NF**SP**DS
GPG**SPTPT**LAHKRNILSISALIAIGAAAFIVIGVIGITVLNLRVRSTTSR**SPA**ALAFSAGDEYS
R**SPT**TDANSGKLVMFSGEPDFSSGAHALLNKDCELGRGGFGAVYQTVLGDGRSVAIKKLTVSSL
VKSQEDFEREVKKLGKVRHQNLVELEGYYWTSSLQLLIYEFVSRGSLYKHLHEGSGESFLSWNE
RFNVILGTAKALSHLHHSNIIHYNIKSTNILIDSYGEPKVGDYGLARLLPMLDRYVLSSKIQSA
LGYM**AP**EFACKTVKITEKCDVYGFGVLVLETVTGKRPVEYMEDDVVVLCDMVRGALDEGRVEEC
IDERLQGKFPVEEVIPVIKLGLVCTSQV**PS**NRPEMGEVVTILELIRC**PS**GSEGQEELSG
>Medtr4g074080.1 _ receptor_like kinase _ HC _
chr4_28154907_28162503 _ 20130731 ; MatureChain: 28-674
QKDDSQPLLALKSSVDIHNKLPWPEKKNDDVCTWVGVKDCYKGKVRKLVLEYFNLTGKLDSNIL
NRLDQLRVLSFKGNSLSGQIPNLSNLVNLKSLYLNDNDFSGQFPVSVSVLHRVKVIVLSGNRIS
GEI**PA**SLVKVPRLYVLYLQDNLFTGSVPRFNQTGLKYLNVSNNKLSGEIPVTAALNRFNASSFS
GNLELCGEQIHRKCKSSTVLPPL**APSPS**V**SP**IGGNGKTTSSKSNRTKLIKIIGGSVGGLVLLIC
LIIIMWLICKNRRKRVGSSAARRGKSSVDVAEGENVVGGEGEGRGSNYEAKQGGFAWESEGIGK
LVFCGAGDREMGYSLEDLLKASAETLGRGIMGSTYKAVMESGFIVTVKRLKDARYPGLEEFRAH
IDLLGKLRHPNLVPLRAYFQAKEERLLVYDYFPNGSLFSLVHGTKTSSGGKPLHWTSCLKIAED
LATGLLYIHQNPGMAHGNLKSSNVLLGADFESCLTDYGLTVFLNPDTMEE**PS**ATSFFYR**AP**ECR

SFQRPQTQ**PA**DVYSFGVLLLELLTGK**TP**YQDLVQAHGSDIPRWVRSVREEETESGDD**PA**SSGNE
ASEEKLQALLNIAMACVSVVPENR**PT**MREVLKMIRDARGEAHVSSNNSSDH**SP**GRWSDTVQSLP
RDEHLSI
>Medtr4g085480.1 _ Serine_Threonine kinase, plant_type protein,
putative _ HC _ chr4_33344910_33346941 _ 20130731 ; MatureChain:
29-460
LEHIATSEYEALGLATEPKRETLEIIIGGGGGGS**APAPSP**ESSCPPPPPPPCP**PT**LSRLDRARR
VLLKFKSNIYD**PT**GYTSNWNENTDTCNFHGILCGIFPNTNDRAVAGLDFNQKKFQGKNCDTIPL
KGILDQIEELTFFHVNSNSFYGSIPKEITNYKYFYELDLSNNKLVGEFPKEVLDSKQLVFLDLR
FNQLTGSV**PS**QLFKKDLDVIFINNNKFSQCLPENFGS**TPA**RYLTFAHNNFVGEIPRSIGNANKT
LTEVLFLGNKFEGCLPYEIGYLKKATVFDVSQNSLTGPI**PA**SFGCLEKIQFLNLAHNKFYGTVP
ESVCVLSGIKNNGNLSLAGNYFTKLE**PA**CWSLLKSKILDVSNNCIPGLPNQRSKQECYEFQCKI
KPCSNPQSLSYVPCKPHWGNKQNN**APA**SQEMATEPVTYKSLNPHRLRP
>Medtr4g113100.1 _ LRR receptor_like kinase _ HC _
chr4_46475886_46479891 _ 20130731 ; MatureChain: 26-655
EPVQDKQALLAFISQ**TP**HSNRVQWNASDSVCNWVGVQCDATNSSVYSLRL**PA**VDLVGPLPPNTI
GRLTNLRVLSLRSNGLTGEI**PT**DFSNLTFLRSIYLQKNKFSGEF**PA**SLTRLTRLTRLDLSSNNF
TGSIPFSINNLTHLSGLFLENNTFSGSL**PS**ITANLNGFDVSNNNLNGSIPKTLSKFPEASFAGN
LDLCGPPLKTSC**SP**FF**PAPAPSP**DNIP**PA**DKPKKKSKKLSTGAIVAIVVGSILFLAILLLLLLL
CLRKRRRR**TPA**KPPKPVVAARS**APA**EAGTSSSKDDITGGSAEAERERNKLVFFDGGIYSFDLED
LLRASAEVLGKGSVGTSYKAVLEEGTTVVVKRLKDVVVTKKEFEMQMEILGKIKHDNVVPLRAF
YYSKDEKLLVYDYMAAGSLSALLHGSRGSGR**TP**LDWDNRMRIALGASRGVACLHASGKVVHGNI
KSSNILLKGPDNDASVSDFGLNPLFGNG**SPS**NRVAGYR**AP**EVLETRKVTFKSDVYSFGVLLLEL
LTGK**AP**NQASLGEEGIDLPRWVQSVVREEWTAEVFDAELMRFHNIEEEMVQLLQIAMACVSIVP
DQR**PS**MQDVVRMIEDMNRGETDEGLRQSSDD**PS**KGSEGH**TP**PPESR**TP**PRSR**TP**
>Medtr4g126270.1 _ receptor_like Serine_Threonine_kinase
ALE2_like protein, putative _ HC _ chr4_52511256_52521195 _
20130731 ; MatureChain: 24-980
TQG**SP**GLIIF**PSPA**NLHGYPFPIE**TP**GSSHQRDSHRTI**APSSPS**ISNGPFSHPPIGS**SP**PYEVV
PSPSTVQGNV**TP**PKK**APA**VRPPVSTVT**AP**GSV**PSP**LS**TPS**RIIHPVKGG**AP**VA**AP**LYK**TP**KPL**P**
AKIH**SP**VL**TPS**VSNYKHHHKRNIITV**PAPAPS**YIV**SP**PISKPRDRAIP**PS**L**SP**KTGGQRHHVPL
PWKSV**SP**LS**SP**IHSNKSQF**APAPSPS**IKFGSH**PT**KI**SPS**GSSSK**SP**KTLPPPPPNEDCLSTACT
EPYVT**SP**LG**AP**CRCVWPLRVGLRLSISLYTFFPLVSELASELASGIFMKQSQVRIMGANAANQQ
PDKTVVIIDLVPLGEKFDNTTAFFTSDMFWHKKVVIKASYFGYYDVLYVNYPGLP**PSP**PLP**PS**S
ITVIDGGPYTTYGKNGSTIKPFGVDIQKKQNRGDLGKGFIVIIAFSVFVAVVLCTAAVWFMFKF
RVHVSQRASIPR**PSPPS**LTKASVTATGSLIGDGRVGSVSSSFEHGI**PA**YKGSARTFSMDDIEKA
TDNFHASRILGEGGFGLVYSGVLGDGTKVAVKVLKSKDHQGDREFLAEVEMLSRLHHRNLIKLI
GICAEEDSFRCLVYELIPNGSLESHLHGVEWEKRALDWGARMKIALGAARGLSYLHEDS**SP**CVI
HRDFKSSNILLEDDF**TP**KISDFGLAQTATDEESRHTSMRVVGTFGYV**AP**EYAMTGHLLVKSDVY
SYGVVLLELLTGRKPIDMSQ**AP**CQENLVAWACPFLTSREGLEAIID**PS**IGLDVPFDSVAKVAAI
ASMCVQQEVSNRPFMSEVVQALKLVCNETEEAKKDGGSRSFSQEHEDLSVVDIERGLSASELFG
SSSRIERVEHEKSRRKSYSGPMGNGRSKQLWQIMRKLSGAEGARSKHFIM**AP**KAIA**SP**IPDSLY
PTLSVFTLAIGLILTASFFIYEATSSRKTRNLAQELTTGAIASVFLGFGSLFLLLACGVYV
>Medtr5g042440.1 _ LysM_domain receptor_like kinase _ HC _
chr5_18654547_18659757 _ 20130731 ; MatureChain: 31-632
QNNTGYTCPNNNNNNNNTYPCQTYVYYKA**TP**PNYLDLATISDLFQLSRLMISK**PS**NIS**SPSSP**L
LPNQPLLIPLTCSCNFINTTFGSISYSNITYTIKPNDTFFLVSTINFQNLTTY**PS**VQVVNPNLV
ATNLSIGDNAVFPIFCKCPDKTKTNSSFMISYVVQPHDNVSSIASMFGTSEKSIVDVNGERLYD

YDTIFVPVTELPVLKQ**PS**TIV**PSPAP**RGNSDDGDDDDDKSGIVKGLAIGLGILGFLLILVIVFW
FYREVLFKKEKKGKGLYFGDKGYKGNDEKKKKMDVNFMANVSDCLDKYRVFGFDELVEATDGFD
ERFLIQGSVYKGEIDGQVYAIKKMKWNAYEELKILQKVNHGNLVKLEGFCIEPEESNCYLVYEY
VENGSLYSWLHEDKNEKLNWVTRLRIAVDIANGLLYIHEHTRPKVVHKDIKSSNILLDSNMRAK
IANFGLAKSGINAITMHIVGTQGYI**SP**EYLADGIVSTKMDVFSFGIVLLELISGKEVIDEEGNV
LWASAIKTFEVKNEQEKARRLKEWLDRTMLKETCSMESLMGVLHVAIACLNRD**PS**KR**PS**IIDIV
YSLSKCEEAGFELSDDGFGSERLVAR
>Medtr5g068210.1 _ LRR receptor_like kinase _ HC _
chr5_28839823_28842750 _ 20130731 ; MatureChain: 26-604
DLISDKYSLLEFSSTLPHALRLNWNNS**TP**ICTSWIGITCNQNETNVISIHLPGIGLKGAIPNNS
SLGKLDSLRILSLHSNELSGNL**PS**NILSI**PS**LQYVNLQHNNFTGLI**PS**SISSKLIALDLSFNSF
FGAIPVFNLTRLKYLNLSFNNLNGSIPFSINHFPLNSFVGNSLLCG**SP**LKNCSTI**SPSPSPSPS**
TTRNQKSTTSKKFFGVASILALSIGGIAFLSLIVLVIFVCFLKRKSNSSEDIPIGKTKNEDSIS
KSFESEVLEGERNKLLFFEGCSYSFDLEDLLKASAEVLGKGSYGTTYKAKLEEGMTVVVKRLRE
VLVGKKEFEQQMEVVGRIGRHPNVLPLRAYYYSKDEKLLVCDYMLGGSLFSLLHGNRGEGR**TP**L
NWNSRMKIALGAAKGIASIHKEGGPKFIHGNVKSTNVLVTQELDGCIADVGL**TP**LMNTLSTMSR
SNGYR**AP**EVIESRKIATQKSDVYSFGVILLEMLTGKIPLGYSGYEHDMVDLPRWVRSVVHEEWT
AEVFDEEMIRGGEYVEEEMVQMLQIALACVAKVVDNR**PT**MDEVVRNMAEIRHPELKKSTSSESE
SNV
>Medtr7g082310.1 _ LRR receptor_like kinase _ LC _
chr7_31550481_31552209 _ 20130731 ; MatureChain: 24-439
QDEANYMSQLLKAL**TPTPS**VKLW**PS**SSLSGTLPNLTYIDLHNNSLTGSL**PS**MFALFSLETIYLG
HNNFTSIPGHCFQLLLGMQTLNLSNNLNLKPWLFPEAEDLGYSELMHTLDLEATNILGPL**PS**DV
FDWFPRLHTVSLSHNNIRGTLPLSLGKSVVRFLSQAWLNNNAFTGTIPNMSNSTHLFDLQLHSN
GLIGLV**PS**SLFSL**PS**LTNISLDNNNLEGPIPMFHKRVKATWESNNFCRSNVGPCDPQVMVMLEI
FAALGHPEFSRIKGNDVCTDGVFLRCRRGKIVSVDFRGQYLNGAI**SPA**FSNLTSLVNLTLTNNN
FTGSIPKSLTTLPQLQLLDVSRNNLSGQIPKFSSKVKLITRGNAFLGLNVSRQGEGEKAAASRN
GG**PS**KTKVLIPLWIVEITNTDHQAILNFDFLL
Medtr7g104360.1 _ purple acid phosphatase superfamily protein _
HC _ chr7_42290323_42293494 _ 20130731 ; MatureChain: 27-427
YNRPPPRKTIFVPHDHDDS**SP**QQVHISQVGQDKMRISWITE**SPTPA**TVHYG**PSPS**ANALSATGI
TTSYHYALYESGEIHNVVIGPLRPNTVYYYRLGDSEKTYNFKT**APA**HFPIMFGVVGDLGQTEWT
VSTLKHLGDSNYDMLLLPGDLSYADFLQNLWDSFGRLVEPLASQRPWMVTTGNHDVEKIPVVHE
EPFTAYNARWQMPFEESGSDSNLYYSFDVSGVHVIMLGSYTDF**AP**DSSQYKWLQGDLQKINRGK
TPWVVVLVH**AP**WYNSNQAHQGEAESVDMKTAMEGLLYNALVDVVFTGHVHAYERFTRVYKDKGD
NCGPVHITIGDGGNREGLATRYQDPKPEISIFREASFGHGVLEVVNASHALWSWHKNDNEEPVV
SDSVWLTSLSSN**PA**CKA
>Medtr7g115740.2 _ lectin receptor kinase _ HC _
chr7_47870184_47872190 _ 20130731 ; MatureChain: 26-592
ISFNYEGFKYDDVKLEGDASLLDSYIQLTSTSRYQSNAFSVGRVTFFEPLQLWDKTSRKITDFT
TKFSFVIYSNETGFGDGLAFFIADPELPFVGYTKEGGGLGLVDKYQILNSTEHSFLAVEFDTHQ
NGWDPYGIHVAATGLMFEMNTIKSWSFNSSLEIQDDKLS**SP**IST**TPSPSP**IPNSLKISSNKRTV
WAGLGVGVGIASIFLILGCFCFLMWKRAKEKKEDSNFDMKMDDEFQKGTGPKKFCYNKLVSATN
NFEETQKIGQGGFGGVYKGYLKDIDSNVAIKRISRESKQGIKEYATEVKIISQLRHRNLVQLIG
WCHMKKDFLLIYEFMQNGSLDSHLYRGKSILTWQVRYNIAMDLASALLYLHEEWEQCVIHRDIK
SSNIMLDDSFNAKLGDFGLARLVDHEKVSQSTTIIAGTMGYI**AP**EYFTTGKATKESDIYSFGIV
SLELASGRKPIDRKAKEGQVAIFDRVWDLYRLGRLLEVVDTKLGGAFDEEQMERLVVIGLWCAN
PNYSCR**PS**VRQVIQVLKFEASLPILSEEMFESTY**PT**TTMSTIFDPVSF**PS**EAYNS

>Medtr7g115740.1 _ lectin receptor kinase _ HC _
chr7_47870184_47872190 _ 20130731 ; MatureChain: 26-668
ISFNYEGFKYDDVKLEGDASLLDSYIQLTSTSRYQSNAFSVGRVTFFEPLQLWDKTSRKITDFT
TKFSFVIYSNETGFGDGLAFFIADPELPFVGYTKEGGGLGLVDKYQILNSTEHSFLAVEFDTHQ
NGWDPYGIHVGINFNSMVSQKTKPWLIDIRNKKTYYYCKIQYNSSAYTLKVSFTGNIVNDKPVK
TYLSYNVDLRDYLPERVIFGFSAATGLMFEMNTIKSWSFNSSLEIQDDKLS**SP**IST**TPSPSP**IP
NSLKISSNKRTVWAGLGVGVGIASIFLILGCFCFLMWKRAKEKKEDSNFDMKMDDEFQKGTGPK
KFCYNKLVSATNNFEETQKIGQGGFGGVYKGYLKDIDSNVAIKRISRESKQGIKEYATEVKIIS
QLRHRNLVQLIGWCHMKKDFLLIYEFMQNGSLDSHLYRGKSILTWQVRYNIAMDLASALLYLHE
EWEQCVIHRDIKSSNIMLDDSFNAKLGDFGLARLVDHEKVSQSTTIIAGTMGYI**AP**EYFTTGKA
TKESDIYSFGIVSLELASGRKPIDRKAKEGQVAIFDRVWDLYRLGRLLEVVDTKLGGAFDEEQM
ERLVVIGLWCANPNYSCR**PS**VRQVIQVLKFEASLPILSEEMFESTY**PT**TTMSTIFDPVSF**PS**EA
YNS
>Medtr8g016330.1 _ receptor_like Serine_Threonine_kinase ALE2 _
HC _ chr8_5511011_5505857 _ 20130731 ; MatureChain: 18-856
NQQLHEFSDTNLPHFRTRSHLSIAVP**PT**VSSSHVRSSIAQPPDKSSSGVPRSIAFPHSK**SP**DK**A**
PRKVWKHGSLG**SP**ISHHKHHHHSRIKNENR**TP**GTN**SP**IQ**APS**YSSQG**PS**VFKAQPPFS**SP**KSKF
IH**APAPSPA**FWSGHLDVPY**PSP**RI**SP**LG**SP**LNKIK**TP**P**PA**YTLVLPPPPPNKDCLSVTCSEPLT
Y**TP**PG**SP**CGCVWPLQVKLRIGISIYKCFPLVSKLADEIAASAMLNHTQVRIMGADAANQQLDKT
NIIINLVPKGVKFDNTAAFLIYKKFWHREILIDDSLFGAYEVLYVHYPGLP**PSPPS**I**PS**GVSSI
DDGPNPGRDNNGMMMKPLGVALPNKEKEGSNGRMIFIIVLSSITAFVLFLGLAWLCLLKYSCCT
HQHEHVSDSLMSTSSKQLRVSGSLNHGIMSGSGSRSFNSGMIAYTGSAKNFTLNDLEKATNNFD
TSRILGEGGFGLVYKGVLNDGRDVAVKILKREDRRGGREFLAEVEMLSRLHHRNLVKLIGICIE
KQTRCLVYELVPNGSVESHLHGADKESDPLDWNARMKIALGAARGLAYLHEDSNPYVIHRDFKS
SNILLGHDFTAKVSDFGLARTALEDGNKHISTHVMGTFGYL**AP**EYAMTGHLLAKSDVYSYGVVL
LELLTGRKPVDLSQ**PA**GQENLVTWVRPLLTSDEGLQTIIDPFVKPNISIDTVVKVAAIASMCVQ
PEVSQRPFMGEVVQALQLVCSEFEETNYVRSGSFQKESLVTNVEGKFFEVSDERVEFSEYQKTL
HAYQSGDEKLRLSDSELLSTSGQEFESFRRQSTSGPITTEKKRHFWKNLRGLSKGSTSEHGFST
KLWPGSH
>Medtr8g072590.1 _ Serine_Threonine kinase domain protein _ LC
_ chr8_30683239_30686420 _ 20130731 ; MatureChain: 22-659
VH**AP**PPNKATKLI**PT**KLFA**TP**DGTIYFVANYENGRTRILWSFSTG**SPT**YSSYQ**AP**GATDFLECE
DDWSLYMQDEYYGKLIILQSIGEVVDL**AP**MISYKGEATIGSKKITSFQVDAKTGSVSTNSKNFA
GLRNLNASKPLLINIYRKDLFLKYDG**PT**SGSGFWNLTVAEFDAVLLCQHLTTFHIEDLNFKMPY
PCKKKQKVFKLNKNFLLESLISESSHGAYHGKDTLSM**PA**SDRMIQLQPNYDRFFNNHDGNMAMP
PTPFFPQQNDYKRKDKLRQPLTEISDLPGHAYLNKKSGW**PTPSPT**MFVILLVVVSHYCYLVVKGI
KYKYIPKDTNREVSMNFNEGVDGEIIGELFVSKKEIGRGRRRTNATAVLHDGQSVAVKRLLKSR
HSVALNEIKKLVSDHHQNIVRLYGVEYDEDFIYLALERCTCNLNDLVQVESGKDTTEYLWKKND
H**PSP**LLLKLMRGIVAGVVHLHKLGIIHGNLKPQNVLIIKDRSLSVKLSDMAITRHVPGKSVFAK
SYCTGWH**AP**EQQQGTETRAVDIFSLGCILFFCLTKGSHPFGDDHLWRESNILNDRKDLSLVEFI
PEAEDLISCLLNPDQNLRPNAAEVLQHPFFRNSQKRVSFLLDTGDKTELIDDPIVVKEHQIF
>Medtr0341s0020.1 _ cysteine_rich receptor_kinase_like protein
_ LC _ scaffold0341_9012_4034 _ 20130731 ; MatureChain: 21-663
ATTDATGKFRYICDQGNDDGNYTTNSTYHTNLKTLLSTLTSNKDINYGFYNSSYGVNTDKVNAI
GLCRGDVKPDDCQNCLQNSTALLTQLCQNTKRSNWMNATKEVEDEFDKGLKNLLYNLKNLAASG
ESDLKYAVGSDVIGPNNNQTIFALVQC**TP**DLSETLCDECLDKSIKEIPNCCDNRIGIGGRFLR**P**
SCYLRYETNSLFYQQTQD**SPS**SSSS**SPSPTPSPA**SV**PS**FS**AP**PPFAENTTS**SP**GKSNNSTTIGI
ALGVPIAIVVMVFIFICIYLRLRKPKKTCEEVQEEEEDDIIDITTEQLQLDFNTIRIATNDFSD

SDKLGKGGFGAVYKGRLFNGQEVAVKRLSMNSGQGDSEFKNEVFLVAKLQHRNLVRLLGFCLEG
RERLLVYEFVCNKSLDYFYFLLGMSNADQAKRAQLNWGKRYLIILGIARGILYLHQDSRFRVIH
RDLKASNILLDEHMNPKIADFGMARLFGVDQTQETQIELWELSNGYM**AP**EYVMHGQFSVKSDVF
SFGILVLEIVSGAKNSSIRDGENTEYLSSFAWRNWREGTATNIIDSTLNNDSRNEILRCIHIGL
LCVQENVASR**PS**MASVVVMLNSDSVTLPMPLE**PA**FHMDWSDFQDTNP**PS**SAQELSVNGASNTEL
FPC
>Medtr1g071720.1 _ Lipid transfer protein _ HC _
chr1_31856691_31855319 _ 20130731 ; MatureChain: 27-190
QIS**TP**CTTSMISSF**TP**CANFITGSTNYNGLI**TPS**SSCCDSLQSMMSTSMDCACLLITANVPFQL
PPINRVLSFFLPQSCNLNGLHAQCKASG**SP**L**PAP**G**PA**IFGSNV**PS**LPPINA**SP**LSSQVDETIEV
AD**SP**KYVIQHSAI**APAPT**KL**AP**MKHSRPRKLQEFNK
>Medtr1g071720.2 _ Lipid transfer protein _ HC _
chr1_31856691_31855319 _ 20130731 ; MatureChain: 27-189
QIS**TP**CTTSMISSF**TP**CANFITGSTNYNGLI**TPS**SSCCDSLQSMMSTSMDCACLLITANVPFQL
PPINRVLSFFLPQSCNLNGLHAQCKASG**SP**L**PAP**G**PA**IFGSNV**PS**LPPINA**SP**LSSQDETIEVA
D**SP**KYVIQHSAI**APAPT**KL**AP**MKHSRPRKLQEFNK
>Medtr1g103420.1 _ Lipid transfer protein _ HC _
chr1_46817645_46816110 _ 20130731 ; MatureChain: 27-217
QIN**TP**CN**PS**TISTLF**TP**CMGFLTGSSANGT**SPT**TECCGALKSLTSSGMNCLCLLVTASVPFKIP
INRTLAISLPRACNMPGVPVQCKASA**SP**L**PAP**GPVALG**PS**H**SPA**S**SPS**GFI**PTPSP**QASSDL**PA**
SPTS**SP**L**AP**QQDTNVPLL**TPPSPS**VSTTGSGRSSL**TPS**SAITSYNV**SPT**VLLIALGFVALKYY
>Medtr1g103450.1 _ Lipid transfer protein _ HC _
chr1_46824422_46825180 _ 20130731 ; MatureChain: 25-154
QSNCTDALMSLMPCLGYVRGN**SPTPT**AGCCTQLANVVKSQPECLCLIIGSDLGSSLGINKTLAL
AL**PA**ACNVE**TP**PVSQCETVANP**PA**GSRTVP**PT**AGGSSGVNSDSKMGQPEPRRKIHKPFWLKDFH
LK
>Medtr2g008580.1 _ Lipid transfer protein _ HC _
chr2_1540285_1539002 _ 20130731 ; MatureChain: 25-207
ARD**APSPS**ADCSTIVVIMADCLSFVSNDSTITK**PS**GACCSGLKTVLKT**SPT**CLCDSLKNSANLG
VVLNVTKAATL**PA**ACGLS**AP**PLSNCGCEFFCVYRPCWCCYSWYVSLCHLQIIGGSI**SPPSPPA**A
H**AP**GGT**TPS**T**APA**AT**TPA**AT**TPA**E**APS**NVKSAASTLLPISAGSLIVCLLSLFLGL
>Medtr3g079210.1 _ Lipid transfer protein _ HC _
chr3_35630248_35627567 _ 20130731 ; MatureChain: 20-206
FHGDGVSAAALAQ**SPAP**ETAVL**APSPA**DDGCLMALTNMSDCLTFVEDGSKLTKPDKGCCPELAG
LIDGNPICLCKLLGSNTADSFGIKINVNKALKL**PT**ICGVT**TP**PVSACSAIGVPVSLP**PS**LSGDA
M**SP**NMAM**SP**KGLAL**SPS**TSSDSSGVL**SPS**GSKIGGASSIQASALTLIFALSTLSVTIFF
>Medtr3g085210.1 _ lipid transfer protein _ HC _
chr3_38516744_38512453 _ 20130731 ; MatureChain: 25-192
AEDLASKCGSVVQKVIPCLDFATGK**APTP**KKECCDAANSIKATDPECLCYIIQQTHKG**SP**ESKS
MGIQEDKLLQL**PT**VCHVNGANISDCPKLLGLSAN**SP**DAAIFKNASKAN**PTPS**SAAATTA**TPTTP**
TPASTGGSSNLRPVMINNVMTVILAIVLAAV**PA**GFISIYT
>Medtr3g106740.1 _ Lipid transfer protein _ HC _
chr3_49328816_49329721 _ 20130731 ; MatureChain: 25-130
NCP**PSP**K**PTPSPTP**VP**PSP**KGQCPKDTLKLGVCADLLGLVNVVIGNP**PS**GSKCCALIKGLADLE
AALCLCTALKANVLGINLNVPITLSLLLSACQKTVPPGFQCP
>Medtr4g101780.2 _ Lipid transfer protein _ HC _
chr4_42100716_42109572 _ 20130731 ; MatureChain: 24-148
QSTDI**PS**CATNLIPCADYLNSTKP**PS**SCCDPIKKTVETELTCLCNLFY**AP**GLLATFNINTTQAL

ALSRNCGVTTDLTTCKHNGS**APAPT**SGG**SPPA**TTGGNKAAASRVSFTGFSFILLLASMMFN
>Medtr4g101780.1 _ Lipid transfer protein _ HC _
chr4_42100716_42102691 _ 20130731 ; MatureChain: 24-141
QSTDI**PS**CATNLIPCADYLNSTKP**PS**SCCDPIKKTVETELTCLCNLFY**AP**GLLATFNINTTQAL
ALSRNCGVTTDLTTCKHNGS**APAPT**SGG**SP**PGKKFLHILLNLIFTSCIHTLHQS
>Medtr4g109250.1 _ Lipid transfer protein _ HC _
chr4_45346628_45345071 _ 20130731 ; MatureChain: 21-177
APTPTTSCRSIMNDMMVECLPYFVDHNNNSQPCCSAFESVAATASGCICDIHMDINNFPMNVTK
MMKL**PA**VCGLSLPCHFD**AP**ESYAFENLISSSSQ**SPS**QS**SP**KPQPNH**APSSPA**SVPVQVADNPGS
FYPFSFGFLAMIIFILFFKVDEDEIAHPN
>Medtr4g115360.1 _ Lipid transfer protein _ HC _
chr4_47647810_47644558 _ 20130731 ; MatureChain: 28-234
CGTCGKPPKKHKPGKKPIVKPPIVKPPVTIP**PT**LPIPPVLPHLPIP**PT**LPIPPVLPHLPIP**PT**L
PIP**PT**LPIP**PT**LPIP**PT**LPIPPVLPHLPVPPVLNP**PS**SGGS**TPSPTSP**CPPKNK**PA**KDTCPIDT
LKLGACVDLLGGLVHIGLGDPVVNKCCPVLQGLAEIEAAACLCTTLKLKLLNLNIYVPLALQLL
LTCGK**TP**PPGYTCSL
>Medtr4g115360.2 _ Lipid transfer protein _ HC _
chr4_47647727_47644963 _ 20130731 ; MatureChain: 28-234
CGTCGKPPKKHKPGKKPIVKPPIVKPPVTIP**PT**LPIPPVLPHLPIP**PT**LPIPPVLPHLPIP**PT**L
PIP**PT**LPIP**PT**LPIP**PT**LPIPPVLPHLPVPPVLNP**PS**SGGS**TPSPTSP**CPPKNK**PA**KDTCPIDT
LKLGACVDLLGGLVHIGLGDPVVNKCCPVLQGLAEIEAAACLCTTLKLKLLNLNIYVPLALQLL
LTCGK**TP**PPGYTCSL
>Medtr5g070360.1 _ Lipid transfer protein _ HC _
chr5_29804370_29802224 _ 20130731 ; MatureChain: 23-190
ALTHQ**APAPS**VDCTNLVLTMADCLSFVTNGSTTTKPEGTCCSGLKSVLKT**APS**CLCEAFKSSAQ
FGVVLNVTKATSL**PA**ACKVS**APS**ATKCGLSEVTE**APA**S**APA**GGL**SP**QSSTA**SPT**SSGAASGLNG
PVSEL**SP**V**PAPSP**GNTASGLFPISMGSLLVCLLVATMSLF
>Medtr7g073100.1 _ Lipid transfer protein _ HC _
chr7_27296169_27296615 _ 20130731 ; MatureChain: 27-148
VITCRDAAITLMGCLPYVAH**PTPSP**PQNCCAAVLDVTGQAITREDRQAVCSCLKGLMNGIPGLD
LTALASLPKVCGANIGYEI**SP**DMDCSKYVSHHQPYFSISYGVYICACFKNQKKQVIDL
>Medtr8g446830.1 _ Lipid transfer protein _ HC _
chr8_18323481_18325978 _ 20130731 ; MatureChain: 26-195
DSSKEKQECTAQLTGLASCLPYVEGEGK**TPAP**DCCDGLKTLLKTNKKCLCVIIKDRNDPDLGGI
VINVTLALNL**PT**VCN**APA**NISKCPELLHMDPNSAEAQVFYTLNNGSSNI**SPS**A**APSPS**DGASSQ
ARSTTQKNDAFRKEKRLIGLEFLAIGVLVWFLFGSAAGNFFI
>Medtr1g012710.1 _ protease inhibitor_seed storage_LTP family
protein _ HC _ chr1_2608833_2607695 _ 20130731 ; MatureChain:
25-151
CGSCK**PTPSP**KPKPKPK**PSP**PP**PS**S**TP**CP**PTPS**T**TPPTPS**TSQKC**PS**DTLKLGVCADVLGLVNV
IVG**SPA**SSKCCALLQGLVDLDAAICLCTAIKANVLGINLNVPITLSLLLSACEKSV**PS**GFQCS
>Medtr1g097720.1 | Lipid transfer protein | HC | chr1:44059772-
44059512 | 20130731
MAQVKIGTGLVLVILVMLCAG**AP**M**APS**RCTNVLVNL**SP**CLDYITGKSS**TPT**SGCCTQLANVVKS
QRLCQVLDGALNHCYVSCTKDK*
>Medtr1g101320.1 | Lipid transfer protein | HC | chr1:45765339-
45765713 | 20130731
MELSTIRSYNVLRIIGILILLSS**SP**FVFAYIECSTVSQLFSSCSVFINYG**TP**D**PSP**GSRCCDAM

SGLSIIANSGGGNKQSVCRCVMSLIQNYIQNATAIGTLPGLCGISLGFTIIPNSNCLVYD*
>Medtr1g101360.1 | Lipid transfer protein (LIM1/2/3/MEN-8) | HC
| chr1:45783929-45783546 | 20130731
MAGPVSMRCQVALVLVLVVALGTKMEMGEAQTTC**PT**QLSNLNVC**AP**FVVPG**SP**NTN**PSP**DCCTA
LQSTNPDCLCNTLRIASQLTSQCNL**PS**FGCVLN*
>Medtr1g103490.1 | Lipid transfer protein | HC | chr1:46837798-
46839766 | 20130731
MAQVKISKGLVLVILAMLCADAMAQSSCTNVLVNL**SP**CLDYITGKSS**TPT**SGCCTQLASVVKSQ
PQCLCQVLDGGGSSLGIKVNQTQALAL**PS**ACNVQ**TPPT**SQCKTAN**SPA**GARTV**PS**TDDGSSDGN
SIKLSIPRLFVVFAATYLCYNFQDILAY*
>Medtr2g008560.1 | Lipid transfer protein | HC | chr2:1536456-
1535394 | 20130731
MASKFSLILCIIFGICALDFTNGASS**SPS**GECSKLVLAMTDCFSFLTNGSTLTQPEGSCCNGLK
TIVNT**APS**CLCGAFKGNAHLGVVLNVSKALTLPFACKVS**APS**ISNCGLPNASAA**AP**GVSI**SP**W**P**
AS**SPT**TSAE**APA**A**APPS**GKSAASTLLPISVGSLLVCLLSLFSGL*
>Medtr3g046350.1 | Lipid transfer protein | HC | chr3:15314030-
15314383 | 20130731
MSGRNYVPLFVSVMVLCMVMTTLHASQIDDISCSEAISSLLPCLPFLEGSL**PATPS**SDCCTGAT
NLFNKANTI**PA**RKSVCQCLQNA**SP**KLWIHSERAKQLPKLCHINLFFPIDKCNS*
>Medtr3g046443.1 | Lipid transfer protein | HC | chr3:15334099-
15334482 | 20130731
MLGKMSQNIFVLVMVLCMLMATLHARKIDDDISCSAESI**PA**VFMALYQCLPFLEG**SPPATPS**SD
CCVGATNLFQKANTTYSRRYICQCIKNIVSAGVPLVSERAKQFPQLCHISEQVPIDPKIDCNS*
>Medtr3g046515.1 | Lipid transfer protein | HC | chr3:15378643-
15379005 | 20130731
MAGKKFISLSMLVMILVMLVTKFDARQIDDVSCTSALFSLLPCLPFLQGVG**PATPT**SYCCAGAN
DLNQKADSTQSRRDVCNCLK**PA**ASRFGVKSDRSTQLPKLCNITLNVPFD**PS**VDCNA*
>Medtr3g046530.1 | Lipid transfer protein | HC | chr3:15389565-
15390140 | 20130731
MAGKKFISLSMLVMILGMLVTKFDARQIDDVSCTSALFSLLPCLPFLQGVG**PATPT**SYCCAGAN
DLNQKADSTQSRRDVCNCLK**PA**ASRFGVKSDRSTQLPKLCNITLNVPFD**PS**VDCNTVQ*
>Medtr3g046540.1 | Lipid transfer protein | HC | chr3:15393902-
15394477 | 20130731
MAGKKFISLSMLVMILGMLVTKFDARQIDDVSCTSALFSLLPCLPFLQGVG**PATPT**SYCCAGAN
DLNLKADSTQSRRDVCNCLK**PA**ASRFGVKSDRSTQLPKLCNITLNVPFD**PS**VDCNAVQ*
>Medtr3g046580.1 | Lipid transfer protein | HC | chr3:15406938-
15410926 | 20130731
MAGKKFISLSMLVMILGMLVTKFDARQIDDVSCTSALFLLLPCLPFLQGVG**PATPT**SYCCAGAN
DLNQKADSTQSRRDVCNCLK**PA**ASRFGVKFDRSTQLPKLCNITLNVPFD**PS**VDCNTIVSQKPQP
PKWQPKDNLDGIGSSKAFE**AP**VGNNDVEDI**PTP**HEVATDLPVQTDITVSADAQHSDPQNVNVSA
AIQHSDDDVIQSAAAHHSDDV**SP**QNVTKGQHDFEFIPYLHDDIRKDTL**PS**TTIHVLEEISDTAE
AQEEDITAAAEPLFTDVMELEAQTQTHSTNVELFTEEANNV**PA**RV**AP**LQHHDETQTHVL**SP**VKQ
MTLDISSIV**AP**EGSKIDPVLQKDLDFMHTWLSKAAATEGVSEITLPVRCGSGTTQEKAIGKERL
EVDEGGELLPLHKTWK*
>Medtr3g046590.1 | Lipid transfer protein | HC | chr3:15414469-
15415045 | 20130731
MAGKKFISLSMLVMILGMLVTKFDARQIDDVSCTSALFLLLPCLPFLQGAG**PATPT**SYCCVGAN
DLNQKADSTQSRRDVCNCLK**PA**ASRFGVKSDRSTQLPKLCNITLNVPFD**PS**VDCNTVQ*

>Medtr3g056640.1 | Lipid transfer protein | HC | chr3:22662869-
22662286 | 20130731
MMVKKFISFSMVVMVLGMLVTTLDARQIDDVSCTSALFSLLPCMPFLQGVG**PATPT**SYCCAGAN
DLNQKANTTQIRRDVCNCLK**PA**ASRFGVNPDRSKQLPHLCNITLSVPFD**PS**VDCNTVQ*
>Medtr3g464110.1 | Lipid transfer protein | HC | chr3:25757731-
25758252 | 20130731
MLRKKIICLSMLVMVLGMLVATLDARQIDDVSCSSALFSLLPCLPFLQGVG**PATPT**SYCCAGAN
DLNQKANTTQIRRDVCNCLK**PS**ASRXFSLLPCLPFLQGVG**PATPT**SYCCAGITDLNQKANTTQI
RRDVCNCLK**PA**ASRFGVNPDRLKQL**PT**LCNITLNVPFD**PS**VDCNT*
>Medtr3g083760.1 | Lipid transfer protein | HC | chr3:37789330-
37787885 | 20130731
MMVSNNSVTALLFLLLAGFVSSDLTEDRKDCADKLVTLASCLPYVGGSAN**TPT**IDCCTNLKQVL
NNTKKCICILIKDSNDPKLGFPMNATLAVQLPNACHI**PS**NISECVGNGIDLEPRSVLLLKVSDL
LHL**SP**K**SP**EAKVFEGLGNSTKTNSS**TP**ISSGSAEKGSSSSSEEKSGGGLGRRWLVAEVVCAILP
FLFISHFFILT*
>Medtr3g106750.1 | Lipid transfer protein | HC | chr3:49331065-
49331877 | 20130731
MASNTKLFSATILVLSLLAYSTFTEAKGSC**SPSP**KPKPK**PSP**QGHCPKDTLKLGVCADVLGLVN
VVVGN**PA**SGSNCCAIIKGLADLDAALCLCTAIKANVLGINLNVPLTLTWILGACQKTIPPGFQC
A*
>Medtr4g027800.1 | Lipid transfer protein | HC | chr4:9812808-
9811604 | 20130731
MANMKIACVVLMMCMIV**AP**MADAAISCGTVTSALGPCIGYLKGGPG**PSPA**CCGGVKRLNGAAAT
TPDRQAACNCLKQAAGAISGLNTAAASALPGKCGVNIPYKISTSTNCATIRA*
>Medtr4g027840.1 | Lipid transfer protein | HC | chr4:9831935-
9832433 | 20130731
MASMKVACVVLMMCMIV**AP**MADAAISCGTVTSALGPCIGYLKGGPG**PSPA**CCGGVKRLNGAAAT
TPDRQAACNCLKQAAGAISGLNTAAASALPGKCGVNIPYKISTSTNCATVREGVPPFGVLYAAL
PCFTQEAVFRT*
>Medtr4g069210.1 | Lipid transfer protein | HC | chr4:26015723-
26017178 | 20130731
MEACNKVMIVGMLFAIANAMFANGELTVCNLTRKERMACESYIVHHD**PSPT**CCYAITKADFQCF
CEYKKNGWIVFYGILPRITLEIPVKYCDPLHLHVDLWAMPLDDFEHDVADCHTGDVVGEVPQSS
CFERGG**PS**VPVSIVKETREYVVGSGELGRVNGPKVPWLI**SPS**KVDVVPIMER**TPS**VVASVGQTL
MSLATSMWMLQCGSYLL*
>Medtr4g077180.1 | Lipid transfer protein | HC | chr4:29568981-
29569791 | 20130731
MATSNTCLTLTILVVFGTLLYNTNQVSGQCGGSL**PA**LISECSKFVQKSGPKI**APSP**GCCAAIRS
FDVPCACKLITKEAEKFVSVPKAISVARSCGVKL**PA**GMQCGSIRIPPKAI*
>Medtr4g101280.1 | Lipid transfer protein | HC | chr4:41833903-
41833046 | 20130731
MASKTCSSLAIFLTINLLFFSLVSACGSYSCN**PTP**N**PTP**KPKPRPNPNPNPN**PTPS**SGTCPRDA
LKLGVCANVLSGLLNLTLGKPPV**TP**CCSLLNGLVDLEAAACLCTALKANILGINLNLPISLSLL
LNVCSRKVPHDFQCA*
>Medtr4g101310.1 | protease inhibitor/seed storage/LTP family
protein | HC | chr4:41841657-41842329 | 20130731
MNSSKSSTLVLFSIINMLFFAMANGCFFCPKPNPNPNPFPYPN**PSPS**TKSCPRDALKLGVCANL
LNGPIGAVIG**SP**PEHPCCSILEGLVDLEVAVCLCTAIKANILGIDINIPISLSLILNACEK**TPP**

TDFQCS*
>Medtr4g101330.1 | Lipid transfer protein | HC | chr4:41847596-
41846675 | 20130731
MASKTCSSLAIFLTINILFFTLVSSCGTCGSGPNPNPKPKHK**PSP**NHSGGSSHSGGS**SP**YGGS**T**
PSGGS**SPS**GGS**SPS**GGSGTASCPRDALKLGVCANVLNGLLNVTLGQPPV**TP**CCTLLNGLVDLEA
AVCLCTALKANILGINLNLPISLSLLLNVCSKQ**AP**RDFQCY*
>Medtr4g113405.1 | Lipid transfer protein | HC | chr4:46617359-
46618354 | 20130731
MSNVFVISALLFALMLFKTSNAFPYTIS**TP**RTRYCRTIISYMMVECLPYFINDDNSQQPN**TP**CC
IAVQSIAANDTNNCFCDIIIDNDDD**SP**MDLTKATNL**PT**ICGV**SP**PCHAN**APSP**G**PS**Q**AP**EDVTI
LYWFLIGLAIYFVVALLLWFLIPFCL*
>Medtr5g006940.1 | Lipid transfer protein | HC | chr5:1143135-
1141543 | 20130731
MDHFASLYRLTVVLAVVTAMA**APA**YAQIT**TP**CNMSMISSTI**SP**CLSFLTNSSGNGT**SPT**ADCCN
AIKTLTSGSKDCMCLIATGNVPFALPINRTLAISLPRACNLPGVPLQCKTSG**SP**L**PAP**G**PA**SFG
PSL**SPA**S**TPS**L**SP**QASSIL**PSP**V**TPS**L**SP**QPETTNPL**SPS**ANPDI**PS**A**TP**GSGRSDL**TPS**SAGS
SSYSLLLSVLVMGFSILKHY*
>Medtr5g006950.1 | Lipid transfer protein | HC | chr5:1146399-
1147714 | 20130731
MAHSKMNMNLVLVVIAMMCAGATAQSSCTNVLVSL**SP**CLNYITGNSS**TPS**SGCCSNLASVVSSQ
PLCLCQVLGGGASSLGISINQTQALALPGACKVQ**TPPT**SQCKTTNAA**SPA**D**SPA**GTEAE**SP**NSV
PSGTGSKS**TPS**TGDGSSSGNSINLSIPLFLILAAAYASVF*
>Medtr5g011980.2 | Lipid transfer protein | HC | chr5:3537211-
3534586 | 20130731
MDNCRMYIMRMMVVLLIIILVININVISIVDAQMLPPCAGDKMLPCTDYLNSTHPPDICCNPIKE
IFEATHDYTCFCQIS**TP**GLLESFGVKLALAVKVVNSCGVKFD**PT**SCKAS**AP**GL**SPS**LMQP**PA**TR
GSDGGGAGMITFTGHYFILFIWACMLFH*
>Medtr7g072810.1 | Lipid transfer protein | HC | chr7:27158009-
27158386 | 20130731
MASSMLIKVTCLSVMCLVLAIPLANAGPYCRDVVETILPCIEYIT**TP**GASTL**PAP**CCNGMKSLN
GEPQYVCRCLKETFFVLPGLNLAALAALPKNCGVNLPYQI**TP**DMNCDKYISHYH**PS**FFNFI*
>Medtr7g072900.1 | Lipid transfer protein | HC | chr7:27220239-
27220592 | 20130731
MASSMLLKVTCLAMICLVLGIPLANA**APS**C**PA**VAQTL**TP**CLPHVSNPG**PSP**PQPCCNRVKTLNS
QAKTTQDRHHVCGCLKSLMAGIPGLNL**PA**FASVAKDCGVDIGYII**SP**NTDCSK*
>Medtr7g072930.1 | Lipid transfer protein | HC | chr7:27227863-
27228312 | 20130731
MATSMFVKVTCLTVICLVLGISMTNAALSCPQVQLTVVPCLGYLRNPG**PS**V**PAP**CCNGLRGLNN
QAKT**TP**ERQSVCRCLKTTAQSLSGLNV**PA**LATLPKKCGVNLPYKISTAIDCNTYVHLSLNQ**PS**F
LISYTLYILLRKFPHYFQNSL*
>Medtr7g072960.1 | Non-specific lipid-transfer protein, putative
| LC | chr7:27235957-27236400 | 20130731
MATSMFVKVTFLIVICLVLGISMTNAALLCPQVQLTVVPCLGYLRN**PSPS**V**PAP**CCNGIRALNN
QAKT**TP**DLSGLNFAAAAAVPRKCGVNLPYKI**SPA**IDCNTYVHLSQNQ**PS**FFISYIYIHFVKFCQ
IKKLA*
>Medtr7g072993.1 | Lipid transfer protein | HC | chr7:27249081-
27249494 | 20130731
MASSMFVKVTCLAMICLVLGIPLANA**APS**CPEVQQTL**AP**CVPYVTHPGPPI**SP**PPPCCNAVKTL

NGQSKTTQDRRDVCGCLKSMMGGIPGLNL**PA**IASLPKDCGVDIGYII**SP**NMDCNKYISHHQ**PS**F
SISYVHYYA*
>Medtr7g073130.1 | Lipid transfer protein | HC | chr7:27304401-
27304784 | 20130731
MASSMLVKVTCLAMICLVLGIPLASA**APS**C**PA**VQQTLTTCLLYATNPHGPPPEPCCNGIKTLHG
QSQ**TP**LDRRDVCGCLKSMMTNLKLNL**PA**VATLPKECGVDLGYVI**SP**DMDCSKYVHL**SPS**TLFF*
>Medtr7g073150.1 | Lipid transfer protein | HC | chr7:27312621-
27313022 | 20130731
MASSMFVKITFLAMICLVLG**TP**LANAALSCGQIQLTV**AP**CIGYLR**TP**G**PS**V**PAP**CCNGIRSVYY
QAKTTADRQGVCRCLKSTTLSLPGLNL**PA**LAG**APA**KCGVNVPYKV**APS**IDCNTYFSLTH**PS**FCH
LYFVV*
>Medtr7g073170.1 | Lipid transfer protein | HC | chr7:27322859-
27323614 | 20130731
MASSMLFKVTCLAVICLVFGIPLANADLSCGQVLSTLYPCLGYIRNPGASV**PAP**CCNGIRIVND
EAKNTSDRQSVCRCLKSTIVLPGINLDALANL**PT**NCGVNLPYKI**TP**DIDCNKIPY*
>Medtr7g082590.1 | Lipid transfer protein | HC | chr7:31659374-
31658860 | 20130731
MMNMKVVCALMLMVVLMVEVATIAEAQNCDPNELLPCAGAIIFNS**TPS**AECCSKLREQ**TP**CFCE
YIRDPDYSQYVN**SP**RAREVASACNVVIPNNC*
>Medtr7g094650.1 | Lipid transfer protein related | HC |
chr7:37733712-37734227 | 20130731
MKKIEIFKIMAMALTLLAIVPKIESEGR**PTTPT**NQRPLCASQFALVNYACGRLPF**TP**GVP**PAP**L
EPP**PAP**DDDGGDDDDGGDDDGGNDEGHKNHHRDHGHGHGHGHKHGHKHRRHQTAEQENCCRWAR
EVDNQCVCELLVRLPPFLVRPLHLYTLNIGEDCEITYSCGGPI*
>Medtr7g095230.1 | Lipid transfer protein | HC | chr7:38109826-
38108564 | 20130731
MAQRSIEMLLSMLFVMVALRGVTVAQTDSNCANVLISL**SP**CLDYITGQTS**TPS**SGCCSQLASVV
GSQPQCLCEVVDGGASSIAASLNINQTRALALPMACNIQ**TP**PINTCPGSTTSSSL**PAPA**GVSIS
NIPN**SPS**GYFTSTTGSSGSIRGSTSSYRTSSSVKLQCSLLVLVIIANLTFTFMTMT*
>Medtr0071s0070.1 | Lipid transfer protein | HC |
scaffold0071:47776-47200 | 20130731
MAGKKFISLSMLVMILGMLVTKFDARQIDDVSCTSALFLLLPCLPFLQGAG**PATPT**SYCCVGAN
DLNQKADSTQSRRDVCNCLK**PA**ASRFGVKFDRSTQLPKLCNITLNVPFD**PS**VDCNTVQ*
>Medtr1g014120.1 _ plastocyanin_like domain protein _ HC _
chr1_3115335_3114127 _ 20130731 ; MatureChain: 22-184
AKDILLGGKTDAWKV**PS**SESDSLNKWASSVRFQVGDHLILKYEAGKDSVLQVSKEDYDSCNISK
PIKHYNDGNTKVRFDHSGPYYYISGEKGHCEKGQKLTVVVMSLKGGSRPIVAF**SPSPSPA**EVEG
PAASAV**APAPT**SGAAVLQGGGVFVAVGVFVAMWLF
>Medtr1g105130.1 _ plastocyanin_like domain protein _ HC _
chr1_24984621_24985495 _ 20130731 ; MatureChain: 23-155
ADHVVGDEKGWTVDFNYTQWAQDKVFRVGDNLVFNYDNTKHNVFKVDGKLFQSCTF**PS**ENEALS
TGKDVIQLKTEGRKWYVCGKANHCAARQMKLVINVLEEG**APSPS**SSAHSIVSSIFGVIMVATIA
IATFF
>Medtr1g105120.1 _ plastocyanin_like domain protein _ HC _
chr1_24990496_24991335 _ 20130731 ; MatureChain: 23-155
ADHVVGDEKGWTVDFNYTQWAQDKVFRVGDNLVFNYDNTKHNVFKVDGKLFQSCTF**PS**ENEALS
TGKDVIQLKTEGRKWYVCGKANHCAARQMKLVINVLEEG**APSPS**SSAHSIVSSIFGVIMVATIA
IATFF

>Medtr1g090420.1 _ plastocyanin_like domain protein _ LC _
chr1_40507541_40508971 _ 20130731 ; MatureChain: 26-220
TQFIVGDSAGWVIPPF**PT**YYTNWTNSHFIREGDSLEFDFNARFYNLIQVSQSEYEHCTALEPLK
VFNS**SP**VNFPLKERGIYYFICSVSNYCTLGQKVIINVHQIPPQNP**PTPS**A**SP**PQHQVPKI**SP**QL
SPNGS**AP**Q**PS**GGTSNP**PAP**INV**PSPTP**VGGNVGCP**PTPS**SIHGVKSNIDVALLVCAMFGTFLGF
WMM
>Medtr1g104800.1 _ plastocyanin_like domain protein _ HC _
chr1_47218566_47220241 _ 20130731 ; MatureChain: 23-185
AVYKVGDSAGWTTLGNIDYKKWAATKNFQLGDTIIFEYSAKFHNVMRVTHAMYKSCNAS**SP**IAT
FTTGNDTIKITNHGHHFFFCGVPGHCQAGQKVDINVLKVSVAA**SPAPS**S**SPS**ALA**SPA**EATV**PA**
SNV**PAPSPS**NA**AP**QKFIALKMMLAFLAMQFLAVNF
>Medtr2g083250.1 _ plastocyanin_like domain protein _ HC _
chr2_34952102_34954240 _ 20130731 ; MatureChain: 24-261
YEFVVGGQKGWS**APS**DPNANPYNQWAEKSRFQVGDSLVFNYQSGQDSVIQVTSQQDYENCNTDA
SSEKSSDGHTVIKLIKSGPHYFISGNKNNCLQNEKLLVIVLADRTNKNSNQTT**SPPSPSPS**V**AP**
SPSPLSSHSSDAL**TP**IPP**PSP**LNGSS**TPPSP**VLDGS**SP**P**PSP**LDGSTL**TP**PPVQQVGS**SP**PPLG
TDVTNPI**TPT**Q**SP**VSEPPPPNAASSILVSFGCSVGALMVSLLVFSK
>Medtr3g099580.1 _ plastocyanin_like domain protein _ HC _
chr3_45650942_45649884 _ 20130731 ; MatureChain: 24-186
TVHTVGDKSGWAIGSDYNTWASDKTFAVGDSLVFNYGAGHTVDEVKESDYKSCTTGNSISTDSS
G**PT**TIPLKKAGTHYFICAVPGHCTGGMKLSVKVKASSSASS**APS**A**TPSPS**GKG**SPS**DG**TPA**ATT
TT**TPS**TQSASSSTSI**SP**IVALFFTVSWILISYILV
>Medtr3g099680.1 _ plastocyanin_like domain protein _ HC _
chr3_45702791_45703579 _ 20130731 ; MatureChain: 24-220
TNHIVGDGLGWTVDSDYTTWASDKTFVVGDSLVFNYEAGWHTVDEVRESDYQSCTTRNSISTDG
SGATTIPLKKAGTHYFICAVPVHCISGGMKLSVKVQDSSSSSSSSSSSSS**APS**A**APSPS**GKG**SP**
SSDD**TPA**ATTTTTTI**PT**IA**APSPS**GKGL**PS**DD**TPA**ATTTT**TPT**MQSASSATCI**SP**IVAFFFIVS
WILIN
>Medtr3g099980.1 _ plastocyanin_like domain protein _ HC _
chr3_45886634_45887385 _ 20130731 ; MatureChain: 24-218
TNHIVGDGLGWTVDSDYTTWASDKTFVVGDSLVFNYEAGWHTVDEVNESDYNSCTTRNSISTDG
SGATTIPLKKAGTHYFICAVPVHCISGGMKLSVKVQDSSSSSSSSSSA**APS**A**APSPS**GKG**SPS**S
DD**TPA**ATTTTTTI**PT**IA**APSPS**GKGL**PS**DD**TPA**ATTTT**TPT**MQSASSATCI**SP**IVAFFFIVSWI
LIN
>Medtr4g078410.1 _ plastocyanin_like domain protein _ HC _
chr4_30315365_30314687 _ 20130731 ; MatureChain: 23-187
KEFHVGGKDGWVVN**PS**EDYNQWARTHRFRVNDTLHFKYVKGNDSVLVVKKEDYDSCNTNNPKQK
LDNGNSKFKLSDSGFYYFISGNADNCKHDEKMIVQVMAVRPNV**TP**NVTAVP**PS**QP**PA**SA**SP**PKI
PLTYVD**SPAPSPS**KASSVGVVVWVVLMLFGGYVGFVY
>Medtr4g081100.1 _ plastocyanin_like domain protein _ HC _
chr4_31394809_31394041 _ 20130731 ; MatureChain: 30-222
YDFIVGGQKGWSV**PS**DSNNPFNQWAEKSRFQVGDSLVFNYQSGKDSVLYVKSEDYASCNTG**SP**I
TKFSDGHTVFKLNQSGPHFFISGNKDNCLKNEKVTVIVLSDRSNNNNSSNTNQTSNA**TPPSP**QS
S**SPPSPAPS**NQEGQ**SP**PPDTNQ**TP**P**PT**ATSDHDHPPRNGAASNFVSLAGSVGTFMASALILSKY
L
>Medtr4g114870.1 _ plastocyanin_like domain protein _ HC _
chr4_47332652_47334953 _ 20130731 ; MatureChain: 26-370
QTRHVVGDTTGWTI**PT**NGASFYTNWASNKTFTVGDTLVFNYASGQHDVAKVTKTAYDSCNGANT

LFTLTN**SPA**TVTLNETGQQNFICAVPGHCSAGQKLSINVVKASA**SP**VS**APTPS**A**SP**PKA**TPAPT**
PV**PA**K**SPAPT**KAA**TPAPAPS**TTA**SPTPAPAPA**TGRVTYTVGDTIGWII**PS**NGTAAYTTWASGKS
FKVGDILVFNFQLNAHNVEEVTKEKYDSCNST**SP**IATFSNPPVRVTLNKTGTHYYICGVPGHCS
AGQKLSINVGSGSS**SPA**T**SPSPS**AS**SPSPS**TGA**TPPS**ASG**SPSP**G**SP**V**TPS**SQ**SP**GGSV**SP**PPE
NSGAASLGVAGLFVTVLSIAATFFC
>Medtr4g124280.1 _ plastocyanin_like domain protein _ LC _
chr4_51319065_51320408 _ 20130731 ; MatureChain: 23-238
YTYNVGAKDGWTVK**PS**QDYNYNFWASNIRFQINDTLFFKYQKGSDSVLVVNKQDYDSCNINNPI
HNMDNGDSSFLLDKSDHYYFISGKDLNCVNGEKFNLVVL**SP**HHHHYHEHHG**PS**L**SPA**V**AP**VHP**P**
TSPSPWN**APTP**DAHGTAV**PTPS**ARDMTTLTSSGVHNGN**APA**I**APA**SNDHGHN**SPAPSPA**RSDST
RLTGSVGVIVMVVLVLGSFTFFHG
>Medtr4g130800.1 _ plastocyanin_like domain protein _ HC _
chr4_54504136_54505468 _ 20130731 ; MatureChain: 27-270
TDYLVGDSENSWKFPL**PT**RHALTRWASNYQFIVGDTITFQYNNKTESVHEVEEEDYDRCGIRGE
HVDHYDGNTMVVLKKTGIHHFISGKKRHCRLGLKLAVVVMV**AP**VLS**SP**PPP**PSPPTP**RSS**TP**IP
HPPRRSL**PSPPSPSPSPSPSPSPSPSP**RS**TP**IPHPRKR**SPASPSPSPS**LSK**SPSPS**E**SPS**L**APS**
PSDSVASL**APS**S**SPS**DE**SPSPAPSPS**SSGSKGGGAGHGFLEVSIAMMMFLIF
>Medtr5g088990.1 _ plastocyanin_like domain protein _ HC _
chr5_38643152_38645471 _ 20130731 ; MatureChain: 31-240
YKNYTVGDSLGWFDNLEK**PT**VNYQKWVAKKQFSLGDFLIFNTDNNHSLVQTYNFTTYKQCDYDD
AQDKDTIQWSSVD**PS**NTDIHPVTVAVPLLKEGATYFFSSDYDGEQCKNGQHFKINVTHGQGLPK
SLQK**PS**ED**SPSPASP**ISGDDES**AP**DTNV**PS**NFNNPKEDSDDEKTSDKDKDKEESSSLSMLKYVK
FHNKLYGCLVLLGTFFFF
>Medtr6g013170.1 _ plastocyanin_like domain protein _ HC _
chr6_4145489_4142683 _ 20130731 ; MatureChain: 27-175
GPKLHKVGGSKGWKENVNYTTWSSQEHVYVGDWLKFVFDKRYYNVLEVNKTGYDYCIDMTFIRN
LTRGGRDVVQLTEAKTYYFITGGGYCFHGMKVAVDVQEH**PTPAPSPS**LSDTAKSGGDSIL**PS**MY
TCFGIIVANVVYVSLVLVGIL
>Medtr6g083240.1 _ plastocyanin_like domain protein _ HC _
chr6_31158404_31157789 _ 20130731 ; MatureChain: 25-176
KTFTVGDGLGWTI**PS**VGSQFYVNWATNKTFTSRDTLVFNYQLNAHAVAWVTKSDFDSCNGTNAL
AVMTKP**PA**KVFFNLSTVGQQYFICTFPGHCSAGQKLSIFVDEVYVL**PT**S**SPTAP**V**PSP**VQE**PAP**
TATT**SPAPA**TGMYFTLLLILIFFF
>Medtr7g086090.1 _ plastocyanin_like domain protein _ LC _
chr7_33397154_33395401 _ 20130731 ; MatureChain: 24-243
KDFVVGDERGWKLGVDYQYWAANKVFRVGDTLTFNYVGGKDNVVRVNGSDFQSCSIPWR**AP**VLT
SGHDTILLTTYGRRWYISGAAHHCNLGQKLFINVQPPQFGW**SP**V**PSPS**A**SPSPSP**V**PTP**EA**APP**
SN**AP**WAASVQTSEITS**SP**V**PSPSPTPA**HEA**APS**SN**AP**WAAGAQTSEITW**SP**V**PSPTPSP**EV**APP**
SN**AP**WTAARRSLLPKKLFKMIHRNLIAV
>Medtr7g086100.1 _ plastocyanin_like domain protein _ LC _
chr7_33400207_33399140 _ 20130731 ; MatureChain: 24-231
KDFVVGDESGWTLGVDYQAWAANKVFRLGDTLTFKYVAWKDNVVRVNGSDFQSCSVPWA**AP**VLT
SGHDKIALTTYGRRWYISGVANHCENGQKLFINVLPKQDGWY**PAPSSPS**A**SPSPSP**V**PAP**EA**AP**
PSN**AP**WAASVQTSELTW**SP**V**PSPSPTPAP**EA**APPS**N**AP**WVASVQASEINW**SP**V**PSPTPAP**QA**AP**
PSN**AP**WT**AP**RRSLLQK
>Medtr7g086140.1 _ plastocyanin_like domain protein _ HC _
chr7_33411040_33410034 _ 20130731 ; MatureChain: 24-161
TDHIVGDDKGWTVDFDYTQWAQDKVFRVGDNLVFNYD**PA**RHNVFKVNGTLFQSCTFPPKNEALS

TGKDIIQLKTEGRKWYVCGVADHCSARQMKLVITVLAEG**APAPSP**P**PS**SDAHSVVSSLFGVVMA
IMVAIAVIFA
>Medtr7g086160.1 _ plastocyanin_like domain protein _ HC _
chr7_33416902_33415916 _ 20130731 ; MatureChain: 24-162
ATDYIVGDDKGWTVDFDYTQWAQDKVFRVGDNLVFNYD**PS**RHNVFKVNGTLFQSCTFPPKNEAL
STGKDIIQLKTEGRKWYVCGVADHCSARQMKLVITVLAEG**APAPSP**P**PS**SDAHSVVSSLFGVVM
AIMVAIAVIFA
>Medtr7g086190.1 _ plastocyanin_like domain protein _ HC _
chr7_33424961_33423956 _ 20130731 ; MatureChain: 24-161
TDHIVGDDKGWTVDFDYTQWAQDKVFRVGDNLVFNYD**PA**RHNVFKVNGTLFQSCTFPPKNEALS
TGKDIIQLKTEGRKWYVCGVADHCSARQMKLVITVLAEG**APAPSP**P**PS**SDAHSVVSSLFGVVMA
IMVAIAVIFA
>Medtr7g086220.1 _ plastocyanin_like domain protein _ HC _
chr7_33436997_33436074 _ 20130731 ; MatureChain: 23-160
ADHIVGDDKGWTVDFNYTQWTQDKVFRVGDNLVFNYDNTKHNIFKVNGTLFKDCTFPPKNEALS
TGKDIIQLKTEGRKWYVCGVADHCSAHQMKFVITVLAEG**APAPSP**P**PS**SNAHSIVSSMFGVVMV
AIVAMATIFA
>Medtr7g086230.1 _ plastocyanin_like domain protein _ LC _
chr7_33442422_33441805 _ 20130731 ; MatureChain: 24-176
KDFVVGDEKGWTTLFDYQTWTANKVFRLGDTLTFNYVGGKDNVVRVNGSDFKSCSVPLT**AP**VLT
SGQDKIIITTYGRRWYISSVTDHCENGQKLFITVQPKQDGW**SP**V**PSPSPSPS**LDLV**TP**E**APPS**N
APW**PA**SSVPRRSLLPKKLFQIFNRD
>Medtr7g086280.1 _ plastocyanin_like domain protein _ LC _
chr7_33452672_33452025 _ 20130731 ; MatureChain: 24-176
KDFVVGDEKGWTTLFDYQTWTANKVFRLGDTLTFNYVGGKDNVVRVNGSDFKSCSVPLT**AP**VLT
SGQDKIIITTYGRRWYISSVTDHCENGQKLFITVQPKQDGW**SP**V**PSPSPSPS**LDLV**TP**E**APPS**N
APW**PA**SSVPRRSLLPKKLFQIFNRD
>Medtr8g007020.1 _ plastocyanin_like domain protein _ HC _
chr8_1180101_1181060 _ 20130731 ; MatureChain: 22-187
TDFTVGDANGWTQGVDYTKWASGKTFKVGDNLVFKYGSFHQVNEVDESGYKSCSTSNTIKSYDD
GDSKVPLTKAGKIYFIC**PTP**GHCTSTGGMKLEVNVVAAST**TPTPS**G**TP**P**PT**K**SPS**T**TPSAPS**ET
NST**TPSP**PKDNGAFSVSNGVSLLMGSFFVSAMILGLMG
>Medtr8g007035.1 _ plastocyanin_like domain protein _ HC _
chr8_1186900_1188335 _ 20130731 ; MatureChain: 22-199
TDFTVGDANGWNLGVDYTKWASGKTFKVGDNLVFKYGSSHQVDEVDESDYKSCTSSNAIKNYAG
GNSKVPLTKAGKIYFIC**PT**LGHCTSTGGMKLEVNVVAAST**TPTPS**G**TP**P**PT**K**SPS**T**TPS**T**TPS**T
TPST**TPSAPS**ETNST**TPSP**PKDNGAVGVSNGVSLLIGSFFVSAMILGLMG
>Medtr8g463180.1 _ plastocyanin_like domain protein _ HC _
chr8_22214467_22212599 _ 20130731 ; MatureChain: 22-303
FKFFVGGKDGWTLN**PS**ENYNQWAGRNRFQISDTIVFKYKKGSDSVLEVKKEDYEKCNKTNPIKK
FEDGETEFTLDRAGPFYFISGKDQNCENGQKLTLVVI**SP**R**TP**KS**SPSPS**AGGL**SPPSPSPT**TT**T**
PSPSG**SPPSP**VAIP**PA**S**SP**V**PT**SG**PT**AS**SPSP**VVS**TPPA**GGPMAS**SPSP**VVS**TPPA**GGPMAS**SP**
SPAGGP**PA**L**SPA**GG**PS**TAAGGP**PAP**GPGGAAT**SP**GPGGAGASGPGGAAA**AP**G**SP**GSNST**APS**GN
SGSFV**APS**NFLVYSVTLAVGALFLSY
>Medtr8g094990.1 _ plastocyanin_like domain protein _ LC _
chr8_39691782_39690137 _ 20130731 ; MatureChain: 23-277
YKFNVGGNHGWAVKSSRHYYNNWATRTRFRINDILFFKYNNGFDSVLVVNKHDYDSCNIKNPIH
KMSDGDSTYKFDKVSLFYFISGNLVNCQNGQKLKVVVY**SP**RHHHG**PS**L**SPA**V**AP**VH**SPS**S**SPS**W

N**SPA**QP**PA**RN**APSP**NV**APT**HSTTQPPVWN**APSPS**A**APA**R**SPT**QP**PT**WNAS**SPS**A**TP**PR**SPT**QP**P**
TWN**AP**P**PS**DIIWT**APA**R**SP**VQP**PA**WN**APT**QLPRN**SPSP**DNESSSNEDDDDDTFNILSICNNIE
>Medtr8g095013.1 _ plastocyanin_like domain protein _ LC _
chr8_39703739_39705150 _ 20130731 ; MatureChain: 23-299
YKFNVGGNHGWAVKSSRHYYNNWATRTRFRINDILFFKYNKGSDSVLVVNKHDYDSCNIKNPIH
KMHDGDSIYKFDKVGLFHFISGNLVNCQNGQKLKVAVY**SP**RHHHHH**SPS**L**SPT**V**AP**VH**SPS**L**SP**
SWN**SPA**R**SPT**Q**PS**ARN**APSPS**A**APT**R**SPT**Q**SPA**WN**SPSPS**A**APA**R**SPT**QP**PA**WN**APS**SSV**APS**R
SPTQP**PA**WN**APS**SSV**APT**R**SPT**QP**PA**WN**APSPS**AIVWT**APA**H**SP**VQ**SPA**WN**APS**SSA**APT**Q**SP**R
N**APSP**NNESISNEDDDDDDSF
>Medtr8g095020.1 _ plastocyanin_like domain protein _ LC _
chr8_39710343_39711774 _ 20130731 ; MatureChain: 21-315
SSYMFNVGGRNGWGVRR**SP**EHYNAWSSRTRFQINDTLRFKYNKGSDSVLVVNNQNYDSCDTKNL
IYKMDDGESTFSLNKTGPFYFISGVNCQNGEKFKVVVI**SP**HHNHEHQG**PS**S**SP**MV**AP**VY**SPAPS**
PSWN**SPT**Y**SPA**QP**PA**WN**APSPS**FAGWT**APA**Q**SPS**WN**APSPS**ET**AP**VR**SPS**Q**SPT**WN**APSPS**EA**A**
PVH**SPT**N**SP**PVN**APSPS**EV**AP**VQFAKN**SP**VVN**AP**E**PS**ATKEN**APSPS**ATEKTD**SPPS**AQSE**TP**P
PSTNND**SPAP**PPNQSDSTRLSGYVNGGLVVALVLGSFTF
>Medtr0334s0010.1 _ plastocyanin_like domain protein _ HC _
scaffold0334_3468_6498 _ 20130731 ; MatureChain: 26-200
ADHTVGGTTGWSV**PS**GASFYSDWAASNTFKQNDVLVFNFAGGHTVAEVSKADFDNCNINQNGLV
ITTG**PA**RVTLNRTGDFYFICTIQGHCSSGQKLSVKVSAS**TPSPPSPTPPT**S**TPPT**S**TPPT**SGT**T**
PTPPTNGG**TPSPSSPT**GPDA**TPPSP**GSATTLVATFPVLIAVIMNLLV
>Medtr1g098580.1 _ cupredoxin superfamily protein, putative _
HC _ chr1_44394069_44391772 _ 20130731 ; MatureChain: 22-237
TTILVDGSSEWKN**PT**VSIGDSITFKHKQNYNLYIFKNQKAFNLCNFTQANLLTD**PS**TTSCSYTW
H**PS**RVGFFYFTFSNDSLKACQDSQKLAIKV**TPT**KAS**AP**EAS**SP**M**PTTP**G**PS**SGGDIQS**SPS**FPW
PFHPHQGS**SP**G**PAPTP**EAS**SP**ITVPLVPYKGSGDGMPFINSN**PA**VPL**PT**GEVDSATIHPLATSG
HQGQVMIGLVGFHAAVHIMALLLL
>Medtr1g098580.2 _ cupredoxin superfamily protein, putative _
HC _ chr1_44393901_44392417 _ 20130731 ; MatureChain: 22-211
TTILVDGSSEWKN**PT**VSIGDSITFKHKQNYNLYIFKNQKAFNLCNFTQANLLTD**PS**TTSCSYTW
H**PS**RVGFFYFTFSNDSLKACQDSQKLAIKV**TPT**KAS**AP**EAS**SP**M**PTTP**G**PAPTP**EAS**SP**ITVPL
VPYKGSGDGMPFINSN**PA**VPL**PT**GEVDSATIHPLATSGHQGQVMIGLVGFHAAVHIMALLLL
>Medtr1g098580.3 _ cupredoxin superfamily protein, putative _
HC _ chr1_44394023_44391916 _ 20130731 ; MatureChain: 22-233
TTILVDGSSEWKN**PT**VSIGDSITFKHKQNYNLYIFKNQKAFNLCNFTQANLLTD**PS**TTSCSYTW
H**PS**RVGFFYFTFSNDSLKACQDSQKLAIKV**TPT**KAS**AP**EAS**SP**M**PTTP**G**PS**SGGDIQS**SPS**FPW
PFHPHQGS**SP**G**PAPTP**EAS**SP**ITVPLVPYKGSGDGMPFINSN**PA**VPL**PT**GEVDSATIHPLATSG
HQGQVPILFSHSCFTKVCTH
>Medtr3g027772.1 | transmembrane protein, putative | HC |
chr3:8755526-8755723 | 20130731
QDL**SPS**L**APAP**DPDAGAAGSVTNSVAMIGASIILSMLAIFKN*
>Medtr3g028260.1 | transmembrane protein, putative | HC |
chr3:8962413-8963001 | 20130731
QDL**SPS**L**APAP**GPDAGVAGSVAGSVAMIGASIVLSMLAIFKN*
>Medtr3g028340.1 | transmembrane protein, putative | HC |
chr3:8988222-8988674 | 20130731
QDL**SPS**L**SPAP**GPDVGAASSATSYVAMIGASIVLSMLAIFKN*
>Medtr3g028450.1 | transmembrane protein, putative | LC |

chr3:9015269-9015961 | 20130731
QDL**SPS**L**SPAP**GPDAGAAGYVTNSVTMVGVSIVLSMLAIFKH*
>Medtr3g028460.1 | transmembrane protein, putative | HC |
chr3:9019158-9019743 | 20130731
QDL**SPS**S**APAP**GPDVGACGSVTSSVAMIGASIVLSMLAILKN*
>Medtr3g028465.1 | transmembrane protein, putative | HC |
chr3:9023312-9023905 | 20130731
QDL**SPS**L**APAP**GPDAGAAGSVTSSVAMVGISVVLSMLAIFKH*
>Medtr3g028480.1 | transmembrane protein, putative | HC |
chr3:9026022-9026583 | 20130731
QDL**SPS**L**APAP**GSDAGVAGSVTSSVAMIGVSVVLSMIAIFKH*
>Medtr3g028495.1 | transmembrane protein, putative | HC |
chr3:9034496-9034963 | 20130731
QDL**SPS**L**APAP**GPDAGAAGSVASSVAMIGVSIVLSMLAIFKH*
>Medtr3g028540.1 | transmembrane protein, putative | HC |
chr3:9101047-9101784 | 20130731
QNL**SPS**L**APAP**GPDAGAAGSVTSSVAMIGVSIVLSMLAIFKH*
>Medtr3g028550.1 | transmembrane protein, putative | HC |
chr3:9108686-9108956 | 20130731
QDL**SPS**L**APA**SGPDVGAAGSVTSSVAMIGASIVLSMLALLKN*
>Medtr5g008650.1 | transmembrane protein, putative | HC |
chr5:1861708-1861505 | 20130731
RDIPLDHQYHQNHVE**SPAP**GPNSSDAVSFGSILGATLFSFIAYYLHIHA*
>Medtr7g092350.1 | transmembrane protein, putative | HC |
chr7:36592637-36592855 | 20130731
ATEV**PA**A**APA**GAGAGGAAA**AP**G**PS**SGASADIPIVASLIGASVVSFFALFH*
>Medtr8g036500.1 | transmembrane protein, putative | LC |
chr8:13503881-13503663 | 20130731
HDHHHDHGT**APAP**G**PS**KSH**AP**K**PS**KSDAASLGSIFGASIFSFVAYYLHFHV*
>Medtr4g087890.1 | formin-like 2 domain protein | HC |
chr4:34465056-34468637 | 20130731
HSSNHHNRRILHQPLSF**PS**ISLP**PT**QP**PS**S**SP**QTQPKPQQTQPKLPFSSISSS**SP**PQ**TP**FF**PS**Y
Y**SP**PLP**PSP**PFLATF**PA**NISSLLLPQHHTRTHNHRHAAAIVISLSLLSLVILSISTVFAFHRHR
HSHKTSSTTVNDDNASRSDSLRLFPPNTATSDSVDQTTNDKSSSMSELFNLGTITTLDDTKATA
ESSCNGNSNDGFPPPYRYVTD**SP**ELHPLPPLPRHNVRTWKNEPKKKEEEEEEEEEEEEEKKEKER
VHEEFY**SP**KG**SPS**GNKQQH**SPS**L**SPS**S**SP**VVTVAVAATSSRSFNVFHYDKFGSKSFTSRTASYP
LSYSLSR**SPS**LNL**SP**IESVQSFPPINPV**SPS**FSSESC**SP**MPMEDFGLKWDGNDTQVSKM**AP**PVP
PPLPPRLWE**TP**VVVSQDGNGDVSVENEENLKPKLKALHWDKVKASSDRAMVWDQLR**PS**SFQLNE
DMIESLFMANNSNSSGNSALASNPKDNARHQIIHA**SP**MPPENRVLDPKKSQNIAILLRALNVTI
DEVCEALREGNCDTLGTELLESLLKM**APT**EEEKSKLKEFKDE**SP**FKLG**PA**EKFLKVMLDIPFAF
KRMDAMLYIANFDSELEYLKKSFDTLKVACEELKNSRMFMKILEAVLRTGNRMNVGTDRGDAQA
FKLDTLLKLVDIKGTDGKTTLLHFVVQEIVRTECSHVSRASNHSVNNPEYTLQDEVDSMKLGLQ
VVSGLSGELANVKKAAVMDSDALSSDVSKLAKGIKKVAEVVKLNEE**SP**LKETNQKFSEAMKGFL
ERGEEEISRIMAQEKNALSSVKDITEYFHGNSAKEEAHRFRIFMVVRDFLSILDGVCKQVGKAN
ERTLVGSRQYVM**PA**VTTL**TP**IFPEFNGKQTSDSSESD*
>Medtr4g095780.1 | formin-like 2 domain protein | HC |
chr4:39940525-39944800 | 20130731
TTLADRRILHQPLLI**PA**TS**AP**PPEF**SP**PPPDN**TPPSP**DIPFFNEL**PT**GPPPPPVANNQNL**PSPS**

GSNARIAN**PT**ATK**PT**K**PA**KKVAIAVSVAFSLFVIFSVVAFFLYKHRIKH**SP**ETETRKLGGENSN
RILEASTLVPP**PS**TSSFLYIGTVVEPNG**TPA**NEPNRFNKLNPIGNHDR**PSP**ELHPLPPLTKSLV
VDSH**SPPA**VSSSSSSSEEDNRETRETAFH**SP**RESSSLNLSHEESYY**TP**VSRRSHG**SPTAP**VTA**T**
PVVPYSKRT**SP**KSRLSGS**SP**EMKRVMI**PS**IKH**APASPS**TLFQHG**SP**RRPKFS**SP**P**PAP**NLTHLH
SNDSRDSLTLPPPRPPPPPPPPPRVVSVSTSSVPRNYQWTRSMKHGEDSASGNSVNVKKALKEDD
EIEMDEAKPKLKALHWDKVRATSDRATVWDQLKSSSFQLNEDMMETLFGCNSLNS**AP**KPKEMSV
TRKPVF**PT**VELETRVLDPKKSQNIAILLRALNVTRDEVSEALLDGNPEGLGAELLETLVKM**APT**
KEEEIKLKNYDGDLSKLG**PA**ERFLKQVLDIPLAFKRVEAMLYRANFETEVIYLKKSFQTLEAAS
EELKNSQLFLKLLEAVLRTGNRMNVGTNRGDAKSFKLDTLLKLADIKGTDGKTTLLHFVVQEII
RSEGTTGESASENVQNQTNSQFNEDEFKKKGLQVVAGLSRDLGNVKKAAGMDSDVLSSYVIKLE
MGLDKVRSVLQYVKPDMQGNFFNSTEIFMKDAEEKILKIKADEIRALFLVKEVTEYFHGDTTKE
EAHPFRIFMIVRDFLNILDQVCKEVGRMQDRTVTGSSRSFRIAASASLPVLSRYRARQDTSSDD
ESL**SP***

>Medtr2g018340.1 | fasciclin domain protein | HC | chr2:5778670-5779378 | 20130731

SLNLS**SP**ITSSSS**SPS**SSFLNLTKILYSSHTFFKAASEFHSLGIDSEIDTRYSTTVFVPDDKAF
ANATVSKRYESLSDDNKYFVLKCHMLKEYLP**PA**VLRKIANDMHLQDTVATEIMGQATYRINITV
MVNGSVAVSNNIVRALVTRTLYDR**SP**IAVYAVSKVLMPKEL**PA**LITSDVT**AP**NVYCFKFSVVLI
LLVLWI*

>Medtr2g437700.1 | fasciclin-like arabinogalactan protein | HC | chr2:14853943-14855880 | 20130731

FNLTSLLSTFPELSTFTSLLSS**TP**PLTADLSHRTSLSLLAVPNTYLTTDPHLHNL**SPS**ALADIL
RYHILLQFLSWSDLQHLP**PS**GKLVTTLFQTTGRATNNFGSVNITH**SP**NTNTVTIH**SPAP**YSSSN
ATVLSQLKMLPYNLTIFTVDSLLIPYGFDLMASETR**PS**ILLNITKTLIDAHNFNVAASMLSASG
VVNEFEAGEGGSGITLFIPVDDAFADLP**PS**VSLQSL**PA**DQKAVVLKAHVLRAYYPLGSLQSTAN
PLQ**PT**LATEAMGAGSFTLNISTFNGSVAINTGIVQAIITQTVFDQNPIAIFGVSKVLLPREIFG
KNPIVSAK**SP**PESS**AP**PPYEDAS**SPT**GFDVQ**PS**HLS**SP**PGFGEDVSSDVASVYGSECCTSDFML
TGCGFKSSRLHLAEEFARLARFA*

>Medtr4g053380.1 | fasciclin-like arabinogalactan protein | HC | chr4:19402736-19404065 | 20130731

TTITLSQL**SPSPA**A**AP**IH**PT**L**PAPTTPA**A**AP**KPLV**PS**LPE**APS**DS**TP**DSSGAVDIVGILRKAKS
FNVLIRLMKTTQLINQLNSQLLATKSGGLTIL**AP**DDNAFSQLKAGFLNSLSDGQKLELLQFHVI
SDYVSSSNFDTLTNPVRTLAGAKPGKVELNVISYGGSVNISTGEVNTTIDGIIYTDKRLAIYKV
GKVLLPMDFFSVAK**APT**KGLSL**AP**E**PS**TM**TP**KADKEKPL**SP**DSSDSSVK**PT**NDNSGAVKVNVYG
RWASLVFGAVLVTMMPQ*

>Medtr4g059790.1 | fasciclin-like arabinogalactan protein | HC | chr4:22059038-22060252 | 20130731

TTTSLSQL**SPA**T**AP**IQ**PT**L**PAPTTPA**A**AP**KPLV**PS**LPE**SPS**DS**TP**DTAGTVDIVGILRKAKAYN
VFIRLMKTTQLINQLNSQLLATKTGGLTIL**AP**DDSAFSGLKAGFLNSLSDGQKLELLQFHVISD
YVSSSNFDTLTNPVRTLAGDKPGKVELNVVSYGGSVNISTGEVNTTINGIIYTDKRLAIYKVGK
VLLPMDFFSVAK**AP**GK**APS**L**AP**E**PS**TDTAK**AP**KPDKDTSSDSSQVIN**PT**VNKSGSIKIVYGKWM
SVGLVLPFAAMIQI*

>Medtr4g059840.1 | fasciclin-like arabinogalactan protein | HC | chr4:22076175-22074615 | 20130731

QL**SPA**T**AP**IQ**PT**L**PAPTTPA**A**AP**KPLV**PS**LPE**SPS**DS**TP**DTAGTVDIVGILRKAKAFNVFIRLM
KTTQLINQLNSQLLATKTGGLTIL**AP**DDSAFSGLKAGFLNSLSDGQKLELLQFHVISDYVSSSN
FDTLTNPVRTLAGDKPGKVELNVVSYGGSVNISTGEVNTTINGIVYTDKRLAIYKVGKVLLPMD
FFSVAK**AP**GK**APS**L**AP**E**PS**TDSAK**AP**KPDKDTSSDSSQVIN**PT**EDKSSSVKIVYGKFVSLGLVL

AFVMIMMQT*
>Medtr4g084160.1 | fasciclin domain protein | HC |
chr4:32804307-32803564 | 20130731
LPRQTIFEAADILYYSGFDSMSLTLELAESLLEH**SPS**ATIF**APS**DSAFKKSGQ**PS**LDLLLFHFV
ILPLPQQSLRRL**PA**GTKL**PT**MLTGQSLTVTTSSSDRVTSVNNIKIIG**SP**IYDNGVLFVYGIDRF
LD**PS**FQYTGPNKK**PS**SNSNSSFEQALETLKTTGYSEMGMNLIGDFEDHHWLFRRDVVPYKFIWN
DVVQNNVVQDNASVVVDVFSDEDQNLFDPGEL*
>Medtr5g098420.1 | fasciclin-like arabinogalactan protein | HC |
chr5:43093083-43091217 | 20130731
AAATSGNHNITRILAKHPGFSTFNHYLTITHLADEINRRQTITILAIDNAAMSSLLDKHLSITT
LKNVLSLHVLVDYFGAKKLHQITDGTTLVSSMFQATGSAAGTAGYVNITNLKGGKVGFGAEDND
GLHSYYVKSVEELPYVISVLQISNPLSSADAE**APT**A**APS**DIDLIGIMSKQGCQSFADLLRVSKA
L**PT**FKESVDGGLTVFC**PT**DTAVSGFSAKYKNLTDSQKVSLLLYHGVPVYQSLQMLKSNNGVMNT
LATEGHNKYDFTVQNDGEDVNLETKVNTANIVGTLIDQDPFVVYKISKVLMPRELFKGVKEEKD
L**APA**E**SP**KAAKSKAKKKKA**APA**ADEDAD**SPA**DGPDADADSDDQKAADNENGVNGLNQGLRFIMV
FFSLFIGALVL*
>Medtr6g007913.1 | GASA/GAST/Snakin | HC | chr6:2139621-2140861
| 20130731
FDEDFKKVNYVK**PSAPT**VVIPIITVP**PS**Q**PA**K**TPA**S**SPAP**P**PT**SKTK**PTP**IVKSWKDCIPLCIY
RCKLHLRKMVCITACMTCCDRCKCVPLDQTYGNRDKCGKCYTDMLTHHDKVKCP*
>Medtr1g088970.1 | transmembrane protein, putative | HC |
chr1:39956709-39955720 | 20130731
KTMHISTISA**APT**TLPE**AP**LF**SPA**M**SP**DMEPLF**PSP**GGVAY**SPS**DSSL**PT**I**PSSPSP**PNPDIST
HQGPVSAFP**PS**ESMP**PA**LG**PS**SQGSSLLMYSILHLAIMLVCIIQLHGKMSLIIQSLTVRKEAFN
DSACCALHIKI*
>Medtr7g073280.1 | transmembrane protein, putative | HC |
chr7:27366686-27373900 | 20130731
SFLKDLRKLIDSTKDSLIPEEKKLDPNKTAGEEKKNDVNIPPV**PTP**QPLPKVENNNGENQKETN
NKITN**TP**PPV**PA**T**AP**PPPVLVT**AP**PPPV**PA**T**AP**PPLPKKDEGKGQVEEKGKNEGIKLAHSTTND
TCEGLHTCRDDGDMVACISKMDSKNFVVLLQNRGGGTIKVKLRSDLESNLGDIVVDKNKTEKVT
IKQIKSESTELTLDAGKGDCVLHVTVV**TP**VPEASFFLRL**PS**FDKIL**TP**VNGAYFLIFTVIVFAV
TWACCCIFKKKPRDEIPYQELEMALPESASATVVESAEGWDQGWDDDWDDNVAVK**SP**VVRHAGS
ISANGLTSRSSNKDGWEDNWDD*
>Medtr1g077790.1 | plastocyanin-like domain protein | HC |
chr1:34753995-34753085 | 20130731
YQFKVGDLNAWGI**PT**SANPQVYAKWSKFHNFTLGDSLLFLYP**PS**QDSLIQVTQESYKSCNTKDP
ILYMNNGNSLFNITSHGDFYFTSGENGHCQKNQKIHISVGGTGNVDAEAN**SPS**SSL**PA**S**APS**SQ
TVFGSIPV**APS**SSN**SP**H**PT**STFHVFIIGSLYALFLALM*
>Medtr2g088990.1 | plastocyanin-like domain protein | HC |
chr2:37537716-37538597 | 20130731
TNHIVGGPIGGWDTNSNLQSWTSSQQFSVGDNLIFQYPPNHDVVEVTKADYDSCQQTNPIQSYN
DGATSIPLTSTGKRYFICGTIGHCSQGMKVEIDTLAAQV**SPASP**VAA**APS**IAD**SP**MISII**PS**A**A**
PAESTVSSAE**SP**EAS**SP**LFEAQVE**SPT**L**SP**MI**PS**TEFL**APSSP**IAQHSQDVSASSTEKGNLQAF
ISIVLSLVVVFMAF*
>Medtr2g093980.1 | fasciclin domain protein | HC |
chr2:40048746-40047457 | 20130731
ATATAT**APA**QAA**SP**KH**APAP**KAA**SPT**STKPLV**PT**LPDTSDS**TP**DDITKILKKAKTFTILTRLLR
TTQIVDNLNSQLISAKSGGLTIL**AP**DDSAFSHLKAGFFNSLNENKKIELLQFHILPQFVDSNNF

DSLSNPVETVAGKDPLKLPLNIESFGTSVSLSTGVVNASVTGVVYQDNKLAIYRLDKVLLPLDF
FGTK**APA**A**AP**VAEAI**AP**KADKTKSSSEEDEDDTTTQDKKSSGANLLGIQGTAYISIGVAFVAVA
MLWS*
>Medtr2g101300.1 | plastocyanin-like domain protein | HC |
chr2:43547176-43547831 | 20130731
ATATAT**APA**QAA**SP**KH**APAP**KAA**SPT**STKPLV**PT**LPDTSDS**TP**DDITKILKKAKTFTILTRLLR
TTQIVDNLNSQLISAKSGGLTIL**AP**DDSAFSHLKAGFFNSLNENKKIELLQFHILPQFVDSNNF
DSLSNPVETVAGKDPLKLPLNIESFGTSVSLSTGVVNASVTGVVYQDNKLAIYRLDKVLLPLDF
FGTK**APA**A**AP**VAEAI**AP**KADKTKSSSEEDEDDTTTQDKKSSGANLLGIQGTAYISIGVAFVAVA
MLWS*
>Medtr3g099570.1 | blue copper-like protein | HC |
chr3:45643508-45642850 | 20130731
TNHIVGDGLGWTVGPDYNTWTSDKTFAVGDSLVFNYVAGHTVDEVKESDYKSCTTGNSISTDSS
GATTIPLKEAGTHYFICAIPGHCTFGMKLFVKVK**PS**SA**APS**A**TP**L**PS**GKG**SPS**DR**TPA**ATTTTT
TPTIQSASSATTI**SPT**VAFFFTVSWILIHHILV*
>Medtr4g066110.1 | plastocyanin-like domain protein | LC |
chr4:24910220-24910812 | 20130731
IDHIVGSRNGWIIPVDGHSFYSDWASNITFKENDVLVFNFVTGRHTVVELNQTYFENCNVNQNI
QFLDT**SPSP**VRFTLNRTGVFYFTCSIPGHCASGQKLIVNVSAS**SPA**LSQG**PSSPT**SSVSSDIHI
DLVATFSILIAAVAVNFLF*
>Medtr4g067200.1 | plastocyanin-like domain protein | LC |
chr4:25360406-25359814 | 20130731
IDHIVGSRNGWIIPVDGHSFYSDWASNITFKENDVLVFNFVTGRHTVVELNQTYFENCNVNQNI
QFLDT**SPSP**VRFTLNRTGVFYFTCSIPGHCASGQKLIVNVSAS**SPA**LSQG**PSSPT**SSVSSDIHI
DLVATFSILIAAVAVNFLF*
>Medtr4g081100.1 | plastocyanin-like domain protein | HC |
chr4:31394809-31394041 | 20130731
YDFIVGGQKGWSV**PS**DSNNPFNQWAEKSRFQVGDSLVFNYQSGKDSVLYVKSEDYASCNTG**SP**I
TKFSDGHTVFKLNQSGPHFFISGNKDNCLKNEKVTVIVLSDRSNNNNSSNTNQTSNA**TPPSP**QS
S**SPPSPAPS**NQEGQ**SP**PPDTNQ**TP**P**PT**ATSDHDHPPRNGAASNFVSLAGSVGTFMASALILSKY
L*
>Medtr5g006040.1 | plastocyanin-like domain protein | HC |
chr5:739373-738212 | 20130731
AVHKVGDSSGWTIIGSIDYKKWAATKNFQIGDTIVFEYNSQFHNVMRVTHAMYKSCNGS**SP**LTT
FSTGKDSIKITNYGHHFFLCGIPGHCQAGQKVDINVLNVSASA**APT**K**SPS**ALA**SP**VPVASTQ**AP**
SPNNA**SP**LIVAKGAFGIIGLAMTVLVFSIST*
>Medtr8g086360.1 | plastocyanin-like domain protein | LC |
chr8:35820481-35819459 | 20130731
HKFKVGGKDGWTVKASGHYEVWASRIKFLVSDTLNFKYNKLVDSLLMVNKQAYDSCNVTNPIRK
MHGGDSTFLLDKPGHFYFISGNVKHCVKGEKLSLVVLSHQEHHG**PS**L**SP**V**PA**N**APT**SGVHDGIA
LVSSGHHMV**APAP**HHDHSGFTRLSGSFVVCVVLALILDSFVF*
>Medtr2g058670.1 | polygalacturonase 11c, putative | HC |
chr2:24244330-24242502 | 20130731
QSGVLDIAKFGGKPNSDIGMALTAAWKEACASTTAAKIVI**PA**GTYQLNGIELKGPCK**AP**IELQV
DGTIQ**APA**D**PS**VIKGTEQWFKFLYMDHLTLSGKGVFDGQGASVYKKAQ**PA**AAWSGKGGNSKNFM
NFGFNFVNNSLVHGVTSKDSKNFHVMVFGCNNITFDSFTIT**AP**GD**SP**NTDGIHMGKSTGVKILN
TNIGTGDDCVSIGDGSKQITVEGVKCGPGHGLSIGSLGKFTTEENVEGITIKNCTLTATDNGVR
IKTWPD**AP**GTITVSDIHFEDITMTNVKNPVIIDQEYCPWNACSKKN**PS**KIKLSKITFKNVKGTS

GTAEGVVLICSSAVPCDGVELNNVDLKFNG**APT**TAKCTNVKPIVTGT**AP**VCQ**AP**GA**APA**ASTTA
SPAAGK**APA**GK**SPA**K*
>Medtr2g058840.1 | polygalacturonase 11c, putative | HC |
chr2:24297949-24299598 | 20130731
QSGVLDIAKFGGKPNSDIGMALTAAWKEACASTTAAKIVI**PA**GTYQLNGIELKGPCK**AP**IELQV
DGTIQ**APA**D**PS**VIKGTEQWFKFLYMDHLTLSGKGVFDGQGASVYKKAQ**PA**AAWSVFLVQFQQNF
GFNFVNNSLVHGVTSKDSKNFHVMVFGCNNITFDSFTIT**AP**GD**SP**NTDGIHMGKSTGVKILNTN
IGTGDDCVSIGDGSKQITVEGVKCGPGHGLSIGSLGKFTTEENVEGITIKNCTLTATDNGVRIK
TWPD**AP**GTITVSDIHFEDITMTNVKNPVIIDQEYCPWNACSKKN**PS**KIKLSKITFKNVKGTSGT
AEGVVLICSSAVPCDGVELNNVDLKFNG**APT**TAKCTNVKPIVTGT**AP**VCQ**AP**GA**APA**ASTTA**SP**
AAGK**APA**GK**SPA**K*
>Medtr2g012670.1 | strubbelig receptor family 3 protein | HC |
chr2:3236861-3230480 | 20130731
DTDPVDVAAINSLYVAMN**SP**PLQGWKPVGGDPCLELWQGVDCVFTNITAIRLGGLNLGGELGSN
LDF**PS**IIDIDLSNNHIGGAISFTLP**PT**LRTLSLSGNKLNGSIPDALSLLTQLSNLDLANNNLTG
QL**PS**SMGSLSSLTTLLLQNNQLVGTLFVLQGLPLQDLNIENNLFSGPIPPNLLSIPNFSKNGNP
FNTTII**PSP**PVAA**APSP**VAIGR**SP**EE**SP**WHVAY**SPA**DFTASMPGKVKKSFLAEHVIWIAGAGLL
LFIALGICLLMVWCCKRKPKNKNPQKLDVEAFPKTLHK**PT**CDATVFETTNQDGKAEKTYRLNEV
PNRRTNSIPKVPDQKEVYVNKVSATSEYNNVSK**PS**LLQPPPHSLLSIPGEKVIVN**PA**ATTKATE
RQVMTSSVKIYTVASLQQYTNSFSQENRIGEGTLGSVYRAELPDGKMLAVKKLDATTFKDQNDE
PFLQLVSSISKIKHANIAKLVGYCAEYNQRLLIYEYCNNGTLHDALQGDDEHCIKFPWNARIKV
ALGAARALEYLHENFRPPIVHRNFRSANVLLNEKFEVRVSDCGLDHLLSSGTAGQLSGRLLTAY
GYS**AP**EFESGSYTQQSDVFSFGVVMLELLTGRKSYDRSRPRAEQFLVRWAIPQLHDIDALSKMV
DPRLNGSYSMKSLSRFADIVSSCIQREPEFR**PA**MSEIVQDLLLTM*
>Medtr3g071480.1 | LRR receptor-like kinase | HC |
chr3:32070026-32066014 | 20130731
LCCSLNDEGKALLKFKEGIFSDPFDALSNWVDDEVGVDPCNWFGVECLDGRVVVLNLKNLCLEG
NLAHELGSLVHIKSIVLRNNSFYGIIPEGIVRLKELEVLDLGYNNFSGPLPKDIGSNISLAILL
LDNNDLLCGFSHEINELVLISESQVDEKQLISARKLPGCTGRSTKWHNRRSKKGLRRLLQSG**AP**
REDPRNRAAIIPD**TPSPSPSPSP**F**PSPSPSPSPS**SSE**TP**QIVKK**PASP**DRNVSD**SPSP**L**PTP**GS
VPQLKSNSNNHHVAIVGGIVGGAAFILILSIVIYLFKTNKVATVKPWATGLSGQLQKAFVTGVP
KLKRSELEAACEDFSNVIGT**SP**IGNIYKGTLSSGVEIAVASVTVTSLKDWSKTSEVQFRKKIDT
LSKMNHKNFVNLLGFCEEDEPFTRMVVFEY**AP**NGTLFEHLHVKEAEHLDWATRLRVAIGTAYCL
QHMHQLDPPFAHSDLNTSSVQLTDDYAAKISDLSFLNEIASADIKAAAKKHTDATLASNIYSFG
IILLEIVTGRVPYSMGKDDSLEEWASRYLQGDQPLKEIVD**PT**LASFQEEQLVQIGALIKSCVNA
DQEQR**PT**MKQICERLREITKI**SP**EVAVPKL**SP**LWWAELEIASFDAS*
>Medtr4g046113.1 | strubbelig receptor family protein | HC |
chr4:16201157-16193578 | 20130731
LTD**PT**DVAALNSLHTSLG**SP**LLPGWVSSGGDPCGEGWQGIQCNGSFIQKIVLNGANLGGELGDN
LATFVSISVIDLSNNNIGGSI**PS**NL**PA**TMRNFFLSANLLTGSI**PT**SLSALTGLSDMSLNNNHLT
GEIPDAFQSLTQLINLDLSSNNLSGELP**PS**VENLSSLTTLRLQDNQLSGTLDVLQDLPLKDLNV
ENNQFAGPIPPKLLSIPNFRQAGNPFNDNSTATR**APPS**R**SP**VT**AP**PG**APPAAP**FFPV**PS**SGSGR
V**PT**KQADG**PT**VAIGSKSGKSNKHTKRVVLIVIGSVLAFIIFVLALVLFIPRCGRRERVDRRSRR
HQIGAYGGERQQ**TPS**SLGAIVLP**PS**QTEKV**PA**RDVSRPNDVRQEEPRKVWAVPNAQDKQEKDVQ
RMTAIPRDVLRPNDNRQEEPRRVWAVPIPNAHDKQEKDVQRMATIAKPVDHEIDIS**TP**EVYSVP
PPPPPPPPPPPP**PS**I**PT**KKGIVE**PT**TSHSL**PT**KRVIVE**PT**TSHRGTTVD**PS**LRSS**SPPT**FAKCF
TIASLQQYTNSFSQENLIGGGMLGTVYRAELPDGKLLAVKKLDKRASVHQKDDEFLELVNSLDR
IRHTNIVELIGYCSEHGQRLLIYEYCSNGSLYDALHSDDEFKASLSWNARIRMALGAARALEYL

HEQCQPPVVHRNLKSANVLLDDDLSVRVSDCGL**AP**LIASGSVTQLSGNLQSAYGYG**AP**EFESGI
YTYQSDVYSVGVVMLELLTGRQSHDRTRPRGEQFLARWAIPKLHDIDALSKMVD**PS**LNGVY**PA**K
SLSNFADIISRCLQTEPEFR**PA**MSEVVLYLLNMMKRESQKNDSNEK*
>Medtr1g069235.1 | photosystem I reaction center subunit IV A |
HC | chr1:29816886-29819066 | 20130731
NSTTSSASSRSTMLMFPLKSNNIGSSSSRLVVRATDEA**APAAPAAPAAPA**ADA**APTP**KPKPPPI
GPKRGSKVKILRKESYWYKGTGSVVAVDQDPKTRYPVVVRFQKVNYANVSTNNYALDEIEEVV*
>Medtr3g064510.1 | expansin A1 | HC | chr3:29068579-29070728 |
20130731
F**SPS**TSHYS**SPSPPATP**FNE**SPPTSPS**STYSEWLSAHATYYSVSDDTRDGVDGACGYGDTHRDG
YGITGAAALSETLFVRGQICGGCFELRCLEEDVPFDKRWCVSGSSVVVTATSFC**AP**NYGFDAES
DGGYCNPPKQHFVLPVEAFEKIAIWKGGNMPVHYRRIKCIREGGMRFTITGSGIFNSVLISNVA
GIGDIVGVKVKGSRTGWIPMGRNWGQIWHVNALLQNQPLSFEVTSSDGVTITSYNV**AP**KNWTFG
QTFEGKQFKS*
>Medtr4g046767.1 | pollen Ole e I family allergens | HC |
chr4:16591718-16592799 | 20130731
EELESLHE**PTP**LH**PS**AN**AP**LHHRHNHNL**SPT**TAH**TP**LH**PS**H**PA**KP**PT**RHHHQHP**PA**H**AP**LQ**PS**S
HQHP**PS**H**AP**IHHHQPRHPVKP**PT**HHHHQHP**PA**H**AP**VQP**PT**HRHHHHP**PA**HPPVHQQHP**PA**H**APA**
QIAVEGVVYVKSCKHAGVDTLSNATSLNGVEGAVVKLQCNNTKHNVVRKGKTDKNGYFYIKGPK
DISIFAVHKCNVVLVS**AP**NGLK**PS**NINGGITGARIKHKKSFVSKAHNLILYNVKPLAFEPKCTN
*
>Medtr2g049670.1 | carbohydrate-binding X8 domain protein | HC |
chr2:22307554-22304641 | 20130731
TAIRLREDSSYEFSTTQSDSIPIVN**PTTP**GTGGNPY**PT**IN**PTSP**QPPDTSTGGQNP**PSP**DTDTT
SPTNPYSNP**PT**ST**SP**YSNP**PA**STNPYSNP**PA**STNPYSN**PTSPT**TT**TPTSPT**V**TPTTP**GTGSSGG
GGGGGGQWCVASESAAETTLKVALDYACGYGADCSQLQQGGACYDPNTLKDHASYAFNDYYQKN
PAPTSCVFGGVASLTSKD**PS**HGNCHFSSSKTTSM**SPPT**YV**SPPT**TM**TPPTP**MTM**TPPS**SM**TPPS**
MTMPDPDGGSSSVYG**SP**PGG**SP**NMATSTSYSILLLLTTSLYATLHVQNYV*
>Medtr7g111600.1 | carbohydrate-binding X8 domain protein | HC |
chr7:45815016-45812250 | 20130731
FSTFSGARFLQKSKISKSQIQFLVTNHKIGRKLQDTN**PAPT**IITV**PS**TNPVTTV**SPT**NPGA**TP**V
TV**PS**T**TPPS**VPL**SPT**N**PA**N**SP**VPV**TP**ITVPGGT**TP**VNSYPP**PSP**LSGGTGTVPVTNPP**PS**ST**SP**
PSSSGGSWCVAKPG**TP**QSTLQTALDYACGTKGTDCSQINQGGICYNPNSLQNHASFAFNSYYVK
N**PA**ATSCDFGGVATITNTN**PS**SGTCIF**PS**SSGGGAGAGASGGSSTTS**SP**VGFGPQS**SP**LDSSHS
TGLRPLLSCMVVVTLLVGGRLGM**SP***
>Medtr8g085390.1 | carbohydrate-binding X8 domain protein | HC |
chr8:35381752-35384118 | 20130731
QGGSASWCVVRSDASFNALQTALDYACGAGADCLPLQPDGLCFLPNTIQAHASYAFNSYYQKRA
R**AP**GSCDFSGTSTIAQTD**PS**YGSCVY**PS**STSGAGGPN**TPT**TSVPMSNTNMS**SPA**TS**SP**IFGGL**S**
PGLS**SP**FNDNSR**APS**KELAKWFLLFFSSLLILSIIS*
>Medtr1g012630.1 | Lipid transfer protein | HC | chr1:2578840-
2578208 | 20130731
HDCASCK**PTP**I**SP**PP**PS**K**TP**KACPPP**PS**T**TP**KA**SP**PP**PS**TTA**SP**PPKAS**TP**P**PSTP**KA**SP**PP**PS**
M**PT**A**SP**PPKAS**TP**P**PSTPT**A**SP**PP**PS**T**TPT**A**SPPTPS**TAQKC**PS**DTLKLGVCADVLGLVNVIVG
SPASSNCCTLIQGLADLDAAVCLCTAIKANVLGINLNVPVTLSLLLSACQKSVPNGFQCS*
>Medtr1g012690.1 | Lipid transfer protein | HC | chr1:2592421-
2591957 | 20130731
CGSCK**PTPTSP**PP**PS**T**TP**KA**SP**PPKAS**TP**P**PS**T**TP**KS**SPPTPS**TSQKC**PS**DTLKLGVCADVLGL

VNVIVG**SPA**SSKCCTLIQGLADLDAAVCLCTAIKANVLGINLNVPVTLSLLLSACEKSVPNGFQ
CS*
>Medtr1g012700.1 | Lipid transfer protein | HC | chr1:2600229-
2599711 | 20130731
CGSCK**PTPPSP**PP**PS**K**TP**KACPPP**PS**T**TP**KA**SP**P**PT**AI**TPPS**T**TP**KS**SPPTPS**TAQKC**PS**DTLK
LGVCADVLGLVNVIVGI**PA**SSNCCTLIQGLADLDAAVCLCTAIKANVLGINLNVPVTLSLLLSA
CQKSVPNGFQCS*
>Medtr2g026775.1 | Lipid transfer protein | HC | chr2:9742223-
9744709 | 20130731
HSGRKPPPPRFR**TP**PP**PT**L**PS**TKPPPHSM**TP**PKQ**TP**MK**APPTPT**RLQPIL**SPS**TQPPH**PT**IPP
PKKP**PT**I**SPSTPTTP**PQTL**TPS**KQPIRPM**TP**PPKNP**PT**T**APPT**RY**TPPT**LP**PS**KQPPHSM**TPPT**
KQ**AP**MK**APPTPT**TLQPIL**SPT**QPPR**PT**T**SP**PKI**TP**I**TPPTPT**KSQPPKT**TPA**KP**PT**LT**TP**P**PT**L
TPSGQPPRPVTQQPRNP**PT**STL**PT**LR**PSTP**PPHSLT**SP**PKT**APT**K**APPTPS**TLPPRPM**TP**PPKK
P**PT**I**SPSTPTTP**P**PT**MFN**SP**PPRTI**APSP**KMP**PT**T**SPPTP**IT**SPS**KLP**PT**QPPR**PT**T**SP**PKI**TP**
I**TPPTPT**TSQPPKTTLAKP**PT**LT**TP**P**PS**L**TPS**GQPPRPVTQQPRNP**PT**S**TPPTPS**TL**PT**LR**PST**
PPPHSLT**SP**PKT**APT**K**APPTPS**TLPPRPM**TP**PPKKP**PT**I**SPSTPTTP**P**PT**MFN**SP**PPRTI**APSP**
KMP**PT**T**SPPTP**I**TPPS**KLPH**TP**PPRAM**TPSP**KI**TP**I**TPPTP**ITSQPPKT**TPA**K**SPT**L**TPS**GQ**PT**
RPVTQ**PS**R**PTTP**PPKL**TPTSPT**S**TPT**T**TPPT**FPP**PTPPS**RPF**TP**PPKL**TPTTP**PPPPM**SPPSTP**
PPNIG**SP**PIFT**TP**PPKL**PS**I**TP**PF**TPTPPSSP**PPPPPPPPMFP**PS**IPPPKIV**SP**PILM**TP**PPKL
PST**TP**PPP**PS**IPRNCPMGNLHVCANLLNIVIGRPQNQPCCSLINGLADFEASVCLCAAIKTNSI
PGVIRINHSIALNTLISRCGRKMPNGFACS*
>Medtr5g011960.1 | Lipid transfer protein | HC | chr5:3531965-
3531118 | 20130731
QVLPPCGKPLLPCIEYANSTSHSIQDIYPPDICCTAIKDVFDATQETCFCQLVY**TP**GLFEAFGV
KFTVGYRILRTCGVKFDTSFCNAS**SPT**LPLSSGKP**PAATP**IGDEGGAGRIALTGLCFIMFIWPF
FLFG*
>Medtr7g014880.1 | RALF | HC | chr7:4453529-4456208 | 20130731
ESAQEKQKCAEQLTDLTTCLPYLGGSAN**SPT**SDCCSGLIQSTKNNKKCICIIIKDRDDPDLGLK
INITLALGL**PS**LCN**TP**DNFSQCSSLLHLDPKSSEAQAFNQLRQNSNGANSI**SPA**T**SPS**AEGSSQ
HSRNQGTDETVTTKNG**AP**FKGKSLLYSLVAGLLVLYF*
>Medtr7g083580.1 | Lipid transfer protein | HC | chr7:32163030-
32163979 | 20130731
SDVDKKGCSFRTRAAILLECAEYIQILGPEIP**PS**YACCAVMKTADIPCLCKHIPRNIEVIISMK
KFVDAAHTCGSEIPPPGGMCGSYIIP**PSSP**EEPPVR**PSP**PPPESE**SPPSP**EE**PSP**PEEPPVQH**S**
PPPPESE**SPSSP**VE**PSP**PE**AP**VVQH**SP**PPPVHN**SPPSP**VRR**SPPTSP**VR**SP**ISR**SP**PK**SP**IRRP
PR**SP**RR**SP**HMP*
>Medtr4g084950.1 | transmembrane protein, putative | HC |
chr4:33169619-33168622 | 20130731
D**TP**CPYPCYPP**PT**GSGTV**TPT**N**PTPS**VST**APPAP**PQSGLPYPP**PS**GNNNYPYN**PTP**PYGNGGD
DGNNNNNNNGGVYG**AP**PPPDPILPYYPYYYRLPPNKPDDSSTSSITVEKKFIRMIATTIMLSL
FLVFGFV*
>Medtr1g022340.1 | transmembrane protein, putative | LC |
chr1:7067015-7065822 | 20130731
QDALQSLIDGAKQSGISQDTLDDAQSALGDASVQQAAEGALADESLADWVQQADELTSKGS**TPP**
AVE**APA**NL**PA**DEEEDIAESSKK**SPT**K**AP**KL**AP**NS**AP**DK**AP**KL**AP**NS**AP**DK**APPA**KPPELPRKYK
APSPAPQAH**SP**NAA**SP**QA*
>Medtr1g112220.1 | transmembrane protein, putative | HC |
chr1:50753076-50753390 | 20130731

QY**PT**A**SP**KASVVITIGTG**PS**IVNSL**SP**I**TPPS**V**SPS**LP**PS**SIS**SPPA**H**APAP**HKSGAASHGFSF
AIGTFVVALVAAALII*
>Medtr2g020370.1 | transmembrane protein, putative | HC |
chr2:6775699-6774908 | 20130731
SYSQSDPIPEL**PT**L**PT**TSSATKTD**PSPS**SI**SP**FQNL**SP**EI**AP**LL**PSP**GGAL**PTPT**GSDI**PT**I**PS**
N**PSP**PNPDDVI**AP**GPFYAF**AP**YGSIQATSNGHRSVAFDIATAAFAGLAALFSLQYMRV*
>Medtr2g026875.1 | transmembrane protein, putative | LC |
chr2:9781196-9781735 | 20130731
K**PAP**KRPPPPPLQVTC**PS**LPPPVH**APSTPPSTP**PIMV**SPS**TQPPVVDPF**TP**P**PSTPPSTPPT**MV
APATPPPMV**APS**T**TPT**S**TPPSTPPT**IV**APSTP**PPMVD**PS**TLPC**TPS**FMV**AP**L**AP**PPVAEPL**TP**P
PSTPATMV**APA**TQPPMDD**PA**TQLPLK*
>Medtr3g460760.1 | hypothetical protein | HC | chr3:23867366-
23868679 | 20130731
QN**APA**S**SP**KSSVTAKP**PS**SVSV**SPT**N**SPASPA**K**SPT**L**SPPS**Q**TPA**V**SPS**GSAS**TP**P**PA**T**SPPA**K
SPAVQP**PS**SV**SPA**I**SPS**NNVSS**TP**PVS**SPA**S**SPPT**AAV**SP**VS**SP**VE**APS**VS**SP**PEASSAGI**PS**S
SA**TPA**D**APA**ATL**PS**SK**SP**GT**SPA**SS**SP**ETSQG**PA**AADDSGSRSSFG**AP**VVLSGLALWISLSF*
>Medtr7g032400.1 | transmembrane protein, putative | HC |
chr7:11381022-11380648 | 20130731
AE**APA**S**SP**KAS**AP**VAEAE**SP**KTSE**APT**TFSD**SPPA**Q**SP**VVVDVVPPVSGPGAASGASSLKVSAV
VAVAAAGYFAF*
>Medtr7g032430.1 | transmembrane protein, putative | HC |
chr7:11396461-11396026 | 20130731
AE**APT**S**SP**KAS**AP**VAEAE**SP**KTSE**APT**TFSD**SPPA**Q**SP**VVVDVVPPVSGPGAASGASSLKVSAV
VAVAAAGFFAF*
>Medtr7g032470.1 | transmembrane protein, putative | HC |
chr7:11404486-11404117 | 20130731
AE**APT**S**SP**KAS**AP**VAEAE**SP**KTSE**APT**TFSD**SPPA**Q**SP**VVVDVVPPVSGPGAASGASSLKVSAI
VAVAAAGFFAF*
>Medtr5g043870.1 | hypothetical protein | LC | chr5:19262859-
19262173 | 20130731
EELETALYKPYGFK**PS**I**PT**K**PAP**IKPIVPVKPDPVK**SP**VIVKPPVPV**TPPTP**IK**PS**LPL**TPPAP**
I**SPPTP**V**TPPSP**V**TPPTP**V**TPTP**I**TPPSP**V**TPAP**V**TP**L**AP**V**TPTP**VK**PAP**VK**PSP**FTRNPVSFQ
IPVMPVPV**TP**IIV**TPPSP**V**TPS**TIEPPVPVTS**AP**VEPPFPV**TPSP**VE**SP**L**PA**TF**APAPS**YLPLS
LH**PT**QFSRRLMAFQS*
>Medtr5g035650.1 | transmembrane protein, putative | HC |
chr5:15513610-15514041 | 20130731
QYGSGGSDYDAN**APS**A**SP**RSLSY**PT**TIIVLLPFVLTFLAAKLNV*
>Medtr5g035610.1 | transmembrane protein, putative | HC |
chr5:15501113-15501772 | 20130731
HDGHVH**SPA**E**APS**SFASSLNCHSVIGGFVPILIAILFARNGL*
>Medtr5g032670.1 | hypothetical protein | HC | chr5:14074645-
14075917 | 20130731
Q**AP**GA**APT**QP**PS**A**TPTPPTPAP**VA**TPPT**A**TPPT**A**TPPT**A**TP**P**PA**A**APTPATPAPA**T**SPPAPTPT**
SD**APTP**DSTSS**SPPAP**GPGG**PAP**GPGSTD**APPSPS**AAFSINKPIMAATALSAAIFAIAF*
>Medtr5g004980.1 | transmembrane protein, putative | HC |
chr5:230877-229070 | 20130731
QQ**APS**T**SP**NS**SPAPPTPPA**N**TPPTTP**QA**SP**PPVQS**SP**PPVQS**SP**PPLQS**SP**P**PA**QS**TP**PPVQS**S**
PPPVQQ**SP**P**PTP**L**TP**PPVQS**TP**P**PASP**P**PASP**PPF**SP**P**PATP**P**PATP**P**PATP**P**PA**L**TPTP**LS**SP**

PAT**TPAPAPA**KLKSK**APA**L**AP**VL**SPS**D**APAP**GLSSL**SPS**I**SPS**GTDDSGAEKLWSHKMVGLVFG
CAFLSLLF*

>Medtr3g089570.1 | pollen Ole e I family allergens | HC
|chr3:16299572-16301945 | 20130731

TTVEDPHSVFSDSLCPHRST**SPS**HGSS**SPPS**HGSS**SPPS**HGGGYY**TPTPPS**TGCGY**SP**PHD**PS**T
S**TPS**HNQ**TPSTPS**NP**PS**SGGYYN**SPPSTPT**DPPVTL**TPPSPSTP**IDPG**TPT**V**PTP**PFL**PSPSP**F
PGTCNYWRTHPGIIWGILGWWGNMGNAFGVTNIPGF**SP**GLTLPQALSNTRTDGLGALYREGTAS
FLNSLVNNKFPYTTDQVKDRFASSLHSNKAAATQAHLFKMANEGKMKPRP*

>Medtr0522s0010.1 | hypothetical protein | LC |
scaffold0522:1211-549 | 20130731

EEFDTALYELDGFKFK**PAP**VKPIVPVK**PAP**VNPPVIVKPPVPV**TPPTP**VK**PS**VPL**TPPAP**I**SPP**
TPV**TPPAP**V**TPPTP**V**TPPS**LI**TPPT**TV**TPPT**AI**TPPSP**V**TPTLAP**VK**PSP**F**TP**NPISIDVPVVM
PVPVNHVPI**TPT**LV**TPPAP**VTLPPIEPIVPINV**AP**VE**SP**VV**PT**YPPLD**APSPS**Y**SP**LLARPLSR
RLMNFHY*

>Medtr3g082080.1 | carbohydrate-binding X8 domain protein | HC |
chr3:37027399-37028253 | 20130731

INHGRGSRSYSIKSWKKLKFSSFGYSDLDSSYG**TPS**SQPLPPFNSL**AP**Q**PT**SLPPY**SP**VPNPPH
DVFRPP**PSP**KTIVTM**SP**PPP**PSPS**KHV**PSP**PKSVIGPPQVYLPPIVYPPP**PS**THKKPPQYAIWC
VAK**PT**VPDPIIQVAMDYACGSGADCKSVQPNGICFQPNTVLAHASYAFNSYWQNTKIGGGTCDF
GGTAMLVTVD**PS**KFALRTIFDYLC*

>Medtr5g044530.1 | carbohydrate-binding X8 domain protein | HC |
chr5:19551318-19554463 | 20130731

ALYCVCKDGVGDQNLQKAIDYACGAGADCTQIQQNGPCFQPNTIKDHCNYAVNSYFQKKGQAQG
ACDFAGMA**TPS**Q**TPPT**SSTSSCAY**PS**SGNTGGGTTT**TPA**TN**TP**VGI**SPPS**TV**TPTTPT**GT**TPS**T
GTG**TPT**GTGTGTGTGTGTG**TPT**GTGTGTGMG**TPT**GTGTGTGIGTGTGTGTGTGTTTGGPNVFGI
SPTTSSSGNSGFSDPNGEGVQLKKCTYVLLTSLVFTIWLAAAIRD*

>Medtr3g048280.1 | carbohydrate-binding X8 domain protein | HC |
chr3:16119332-16121045 | 20130731

GYSETNKEAHGHVTQVNPQSHKEKLLEFK**AP**ISTTQRDIT**TP**ITTIPNLV**PT**IST**SP**ILNPDSN
PDTY**SPS**STVPIT**APS**TSN**SP**VSSGASWCIA**SPS**ASQRSLQVALDYACGYGGTDCSAIQPGGSC
YNPNSVHDHASFAFNKYYQKNPVPNSCNFGGNAVLTNTN**PS**KASTIYDHGFKLQSDTTM*

Appendix IV. *Medicago truncatula* proteins identified by Ma *et al.* 2017 not predicted to be AGP.

>Medtr7g072730.1 | Lipid transfer protein | HC | chr7:27134096-27135394 | 20130731
MASSMLVKITCMSMLCLVLVIPLSNAGMTCNEVTDTLYPCAGYAT**TP**GDDPPPAGCCGGLKDIK
DKATT**TP**ERQSVCECLKTNVLRIPGVNPDTVAALPEKCDVPLPYQIKADFDCSTYISLTLNPLF
SISYTNIHTHIIPFSAFSLRSNLGKLVQQLAGLVLNE*
>Medtr2g026930.1 | Lipid transfer protein | HC | chr2:9802063-9802665 | 20130731
MGSKGVGSIFIIFLNLLLVSTLVASH**TP**RRPPPPPCLSTPPPVD**TPS**TPPHVV**TPS**TPPPSTPP
PSTPPSTPP**SP**SPPII**TPS**APPPSTPPPTTPPSTPPSIPRTPPSTPPPI**PV**TPPQNCNLLNLNI
CAKVLNNVVVLNPRNNRCCTLISGLVDLDAAVCVCAALKANIIGISVNINADLKIILNSCGVN**T**
PAGFTCRR*
>Medtr2g026920.1 | Lipid transfer protein | HC | chr2:9795993-9796890 | 20130731
MGSKGVGSIFIIFLNLLLVSTLVASH**TP**RRPPPPPCLSTPPPVD**TPS**TPPPVV**TPS**TPPPSTPP
PSTPP**SP**SPPII**TPS**APPPSTPPPTTPPSTPPSIPRTPPSTPPPI**PV**TPPQNCNLLNLNICAKV
LNNVVVLNPRNNRCCTLISGLVDLDAAVCVCAALKANIIGISVNINADLKIILNSCGVNTPAGF
TCRR*
>Medtr4g094958.1 | LRR receptor-like kinase | HC | chr4:39379840-39375516 | 20130731
MKTTLSLLLTLLSLTLTLNQCFSLTNDGVLLLSFKYAVLNDPLLVLSNWNYSDQ**TP**CSWNGVSC
SII**TP**NTNND**TP**FRVTGLSLPNSQLVSSI**PS**DLGTIEHLQILDLSNNSINGSLSSNFFQPNSEL
CFLNFSNNLLTGEVPESLTELRNLQFLNFSDNAFTGKLPNNLSNMQNLTVASFKNNYFTGFLPK
DLRTLQILDLSSNLLNGSLTQDFGGDSLRYLNVSYNRFSGEIPREFAEKI**PS**NATVDLSFNNLT
GEIPE**SP**VLLNQETKVFSGNSDLCGEPMKNPCSI**PSSPS**SNPQGSSPPALAAMPKNFDND**SP**QS
QTTESSEKKQSGLRKGT*IIGIVIGDFVGIGILAMVFVYVY*KLKRKKDAENAIKNEVATARSENS
SSTLETKGFTRWSCLRKRTEDEESSETQSSSDSDVEISQKNVDAENQKQGENKAGTGSGTGTLV
IVDGERELEVETLLKASAYILGATGSSIMYKAVLEDGTSLAVRRIGENGVERFKDFENQVRVIA
KLVHPNLVRVRGFYWGHEEKLIIYDFVPNGCLANVRYRKVGS**SPS**HLPWEIRLKIAKGVARGLT
YLHEKKHVHGNLK**PT**NILLGNDMEPKIGDFGLERIVTGDTSYSKAGGSARIFGSKRSSASRDSF
QDLTCG**PSPSPSPS**SIGGV**SP**YH**AP**ESLRNLKPHPKWDVYSFGVMFLELLTGKIVVLDDMGQG**P**
AVLVEDKNRALRMVDVAIRADMEGKEDALLAYFKLGYSCVTNVPQKRPQMKEVLQVLEK**TPS**TI
SSSYYYSH*
>Medtr6g080660.1 | plastocyanin-like domain protein | LC | chr6:29008927-29007242 | 20130731
MRPVWAVKAMFVVVLASILFRCVCGGNHTVGGASAWDLESNMQDWASTESFNVGDDLVFTY**TP**L
YDVIEVNQQGYNTCTIANAISTHNTGETVIHLTESGTRYFVCGRMGHCQQGLKLEVKVQAQSNN
TGTSPPPSS**AP**RPHSPPSPPNGDSPPSPPHSDSPPSPPDAHAK**AP**CDCSCADEGRLGVPLITLV
IILAFAWSSFFVHLSAIN*
>Medtr6g023760.1 | plastocyanin-like domain protein | HC | chr6:8345768-8347436 | 20130731
MGCKKMILLALLLATLITKEVLATQHNVGGSQGWDPSSDFDSWSSGQTFKVGDQLVFKYTSMHS
VVELSDESAYKKCDISTPLNSLSTGKDVVKLDKPGTRYFTCGTLGHCDQGMKVKITVGNGNGSS
STASS**SPS**SSSS**SPS**SSSSSTSPPSTTSSASQYFASLMLIVTLSFATIFSLF*
>Medtr4g007250.1 | plastocyanin-like domain protein | HC | chr4:1058046-1054244 | 20130731

MAVSYSFLALSFFFFLFTLSPLPVSATDHIVGANRGWNPGINYTLWANNHTIYVGDYISFRYQK
NQYNVFLVNQTGYDNCTLDSAVGNWSSGKDFILFNKSMRYYFICGNGQCNNGMKVSVFVHPLPS
PPPSSSQHNHSSPNSAAPMVLEYLGHKFLMLSFVFVMFGYVLV*
>Medtr6g007897.1 | gibberellin-regulated family protein | HC |
chr6:2123510-2126214 | 20130731
MLQKMVSKSIFLLGIFLVLATKVYSYDEDLKIVVNYVN**PT**APP**SP**IVTPP**SP**VKAPPTPPLVKS
PPIVK**AP**SPPLVKTPPYQSPPVK**P**TPPIVKSPPSPPLVKSPPYQSPPIVK**AP**SPPLVK**P**TPPIV
KSPPSPPLVKTPPYQSPPIVKAPPTPPPIVKTPPYQSPPIVKPPV**AP**SPPP**TP**IVKSWKDCIPL
CGYRCQKHSRQNTCIRACMTCCDRCKCVPPGTYGNREKCGKCYTDMVTHGNRPKCP*