Introduction

The gene commonly known as LFY or LEAFY functions as transcription factor in the plant Arabidopsis thaliana; it regulates the floral development of the plant. The gene is located on chromosome 5. The accession number for the gene is NM_125579.1 and the protein accession number is NP_200993.1. The protein responds to environmental and internal signals and acts as a switch that transitions growth in the shoot apical meristem from vegetative to floral development. LFY is responsible for both negative and positive feedback loops for meristem genes, plant hormones, and further production of LFY(Li, W., Y. Zhou, X. Liu, P. Yu, J. D. Cohen, and E. M. Meyerowitz). When LFY is activated the expression of the downstream meristem genes AP1 and CAL are turned on to induce floral initiation. (Blazquez). LFY then turns on the expression of the homeotic ABC genes; these genes control floral genes. (Coen and Meyerowitz).

Although the protein LEAFY is not fully understood on a molecular level, the crystal structure of the protein as well as mutations within the gene can describe the function. The protein LEAFY and its orthologs contain a highly conserved N-terminus with an unknown function and the DNA binding C-terminus. The DNA binding domain of LEAFY is made up of a seven-helix fold and binds to DNA as a dimer; this gives evidence that the protein acts as a developmental switch due to its resemblance to other homeodomain transcription factors (Hamès). Furthermore, the lfy null mutant causes a significant delay in meristem transition as well as abnormal flowering and development of sepals and carpels; this provides evidence of the function of the protein (Huala). Collectively, the gene commonly known as LFY or LEAFY codes for a protein that is involved in the control of floral development; the protein functions by interacting with meristem homeotic genes.

Blast analysis

Using the BLAST tool on the NCBI website, a local pairwise alignment was utilized to determine possible paralogs and orthologs of the gene LFY. The program blastp with a non-redundant database was used with the algorithm parameter settings 1000 maximum target sequences, an expected threshold of 10 and a word size of 3. The BLOSUM80 matrix was used and models of XM/XP and uncultured/ environmental sample sequences were excluded. Table 1 displays the results of the possible homologs with the corresponding accession numbers, maximum score, e-value, and percent identity.

One of the results within the output has the possibility to be a paralog but it is more likely redundancy occurring for the gene LFY. The possible paralog is the protein listed leafy [Aabidopsis thailiana]and has a 99% identity to the gene of interest. The accession number of the protein is AAM27931.1; because the accession number is not Refseq the possibility of redundancy is high and this protein is unlikely a paralog. Furthermore, the likelihood of this protein being a pseudogene is low. Pseudogenes tend to have high identity with the gene of interest but also have detrimental sequence mutations that cause a loss of function. The protein does not maintain full function; therefore, it is ruled-out as a possible pseudogene.

Furthermore, the BLAST results do show indications of a vast amount of orthologs due to high percent identities and low e-value scores. Table 1 displays the possible orthologs with higher maximum scores listed first. All of the possible orthologs have an e-value of 0.0 and percent identities ranging from 99% to 64%. The LEAFY transcription factor of the species *Brassica juncea* (a species of mustard plant) appears to be the closest relative to the LFY protein of *Arabidopsis thaliana*. There is a 99% identity between the two alignments and therefore the sequence is highly conserved between the two species of plants. Similarly, the leafy protein of the species *Arabidopsis lyrata* has a high percent identity of 90% but a lower score than *Brassica juncea.* This protein likely has conserved domains but is a more distant relative of the *Arabidopsis thaliana* species.

Multiple sequence alignment

The possible homologs that were obtained and analyzed using the BLAST tool were inserted into two global multiple sequence alignment tools to observe global alignment and possible conservation within the residues. T-Coffee was chosen as the global multiple sequence alignment because it is highly accurate and able to incorporate heterogeneous types of information; it is also ideal for 2-100 sequences. Dialign was also chosen because it also is ideal for 2-100 sequences with conserved regions that are surrounded by regions that are not able to be aligned. The FASTA sequences obtained from the BLAST search were used as an input for global multiple sequence alignment. Figure 1 displays the sections of the T-COFFEE multiple global alignment output in which good alignment occurs. Figure 2 displays the output for the Dialign Pfam alignment. The conserved domains are not as easily identified as with the T-COFFEE alignment but the gaps and the indel regions of the alignment are evident.

The T-COFFEE program designates good alignment as highlighted pink areas; these sections are a possible indication of conserved domains between the homologs. The alignment indicates that there is good alignment at the beginning of the sequences and two more large sections throughout the alignment. Although the there are large sections of the amino acid sequence that have good alignment, there are only small sections where perfect alignment occurs. The largest fully conserved region of the alignment appears to be between residues 269 and 302. It is suspected that a conserved domain occurs between homologs at within this section of the sequence.

Indel regions, on the other hand, or regions of very low conservation appear multiple times throughout the sequence. The largest indel region appears to be in the middle of the sequences, roughly between amino acid residues 180 and 260. Within this region, a significant amount of insertions and deletions are observed; therefore, the alignment shows a vast amount of gaps and poor alignment.T-COFFEE designates poor alignment by highlighting the residues with blue and green. Within this region, a higher mutation rate is occurring and is likely driving the evolution of the protein throughout species. Insertions and deletions have the high likelihood of drastically effect the function of the gene due to frame-shift mutations. On the contrary, the

homologs of the LFY protein of Arabidopsis thaliana all maintain a common function in which is initiation of floral development. Another significant indel region occurs at the end of the sequences ranging from amino acid residue 425 to the end. This could be due to premature stop codons of the various plausible orthologs. A significant amount of gaps and misalignment is observed at this region. The LFY and LFY-like proteins of the species *Pistacia chinensis, Mangifera indica, Litchi chinensis, Dimocarus longan, , and Clausena lansium* appear to have the most gaps at the end of the sequence; it is possible that LFY and LFY-like proteins of these species either have deletions or premature stop codons.

Phylogenic Tree

The possible homologs that were determined from the BLAST analysis were inserted into the online phylogeny.fr program to create a phylogenic tree. To determine an output for the phylogeny tree the BLAST tool was used once again but this time the parameters were set to achieve more distant relative of the protein leafy of Arabidopsis thaliana. The program was set to blastp, the matrix BLOSUM6 was selected, the word size was decreased to 2, and the maximum hits were increased to 10,000. The distant tree of results link was selected to determine a possible ancestral relative of the previously selected hypothetical homologs. A distant relative and ancestor of the leafy protein was determined to be the marpoflo protein of *Marchantia polymorpha,* or commonly known as the liverwort plant.

The multiple global alignment program MUSCLE was chosen because it is compatible with FASTA, it is fast and accurate, is compatible with “A la Carte”, and compatible with all sizes of datasets. The program Gblocks was used for alignment curation because it is more thorough than just removing the positions with gaps; it is the more sophisticated cleaning programs out of the two. The program phyML was chosen for the construction of the phylogenetic tree because it can be customized and gives the maximum likihood. Lastly, the output group chosen was PHYLIP format because it gives the number of sequences and the length of residues. It also gives the sequence in blocks of 10 characters. The global multiple sequence alignment was curated to delete positions with gaps and any adjacent ambiguously aligned positions. This step increases the confidence that regions are correctly aligned; this is essential for phylogenic construction because a gap indicates a evolutionary event dispite the size of the gap. Within this step the amino acid residues decreased to 284 where 63% of the sequences were conserved.

The phylogenic tree with the LFY protein of *Arabidopisis Thaliana* and possible homologs is displayed in figure 3 and the evolutionary distance is displayed in figure 4. The confidence levels range from 0 to1.0 and the evolutionary distances range from 0 to 1.004. Generally, 70% or higher of confidence indicates a reliable phylogeny tree. Therefore, the majority of the groupings within the tree are likely to be accepted. It is determined with high confidence of that the LFY protein of Arabidopsis thaliana and the LEAFY transcription factor of Brassica juncea are closely related orthologs. The phylogeny tree can be split into two

different categories; close orthologs and distant orthlogs because the phylogenic tree has two distinct common ancestors. The Arabidopsis thaliana LFY protein is more closely evolutionarily related to leafy/leafy-like proteins of *Arabis alpina, Leavenworthia crassa, Brasssica rapa, Brassica napus, cochlearia acualis, Brassica juncea, and Arabidopsis lyrata.* On the contrary, the Arabidopsis thaliana LFY protein is more distantly evolutionarily related to leafy/leafy-like proteins of *Citrus sinensis, Cirtus reticulata, Clausena lansium, Litchi chinensis, Dimocarpus longan, Pistacia chinensis, and Mangifera indica*. The totality of the supposed orthologs is thought to diverge from the marpoflo protein of *Marchantia polymorpha* as indicated in the phylogeny tree.

Determining Conserved Domains and Predicted Structure of the Arabidopsis thaliana Protein LFY

The domain FLO_LFY with the accession number pfam01698 was the only conserved domain that was revealed on the NCBI website. FLO_LFY is a family of proteins that are homologous of floricaula and LEAFY proteins. When mutations occur in these proteins the floral and leaf development of the plant is affected. This domain is found in 277 known species of metazoan, fungi, plants, and other eukaryotes. Unfortunately, the FASTA sequence of this domain could not be acquired. Therefore, the entire amino acid sequence of the protein LEAFY of Arabidopsis thaliana was inserted into the PDB (Protein Data Base) search. The protein data base is an online archive that withholds a vast amount of 3D structures of proteins as well as nucleotides and complex assemblies. Two complex structures of the Arabidopsis thaliana LFY protein were obtained from the PDB search. LFY of Arabidopsis thaliana in complex with DNA from AP1 promoter and LFY in complex with DNA from AG-I promoter were observed in the results. X-ray diffraction was used as the experimental process for determining both of the complexes with corresponding resolutions of 2.10Å and 2.30 Å . This is a powerful and rapid technique that is used to indentify crystalline material to determine unit cell dimensions. The modeling of the Arabidopsis thaliana protein can be accepted because x-ray diffraction provides definitive mineral structure and interpretation is clear-cut. ("X-ray Powder Diffraction (XRD)." Furthermore, both e-value of the two results are very close to 0 and have a high score of 901. Figure 5 displays the 3D image of Arabidopsis thaliana in complex with DNA at promoter AP1 and figure 6 is the same complex rotated 180^0.

Furthermore, three likely homologs were also observed in the results of the PDB. Table 2 summarized the 3 most likely homologs to the protein leafy protein. Crystal structure of moss leafy bound to DNA, crystal structure of plant glutamate cysteine ligase in complex with a transition state analogue, and crystal structure of plant glutamate cysteine ligase in comlex with Mg2+ and L-Glutamate are the top most possible homologs in PBD with the corresponding e-values 1.13946E-66 , 3.24128, and 3.24128. X-ray diffraction was the experimental processes for determining the three proteins with the corresponding resolutions of 2.32 Å, 2.18 Å, and

2.09Å . The possible homology indicated as the moss Leafy bond to DNA had the lowest e-value and the highest score compared to the other two possible protein homologues. With an e-value of $1.13x10^{-66}$ and a score of 641 it is likely that this protein complex is indeed a homologue of the Arabidopsis thaliana protein LFY. The following possible homologues obtains from the search have high e-values that are greater than 3 and low scores of 68. Therefore, the likelihood of these two complexes being homologues to the protein LFY of Arabidopsis thaliana is low.

The Phrye-2 server was used to create a homology model of the protein LFY of Arabidopsis thaliana. Phrye is an online server that functions as a protein homology recognition engine. This program can be used to predict which residues are involved in the function of the protein; this is achieved by obtaining the amino acid conservation scores. The top three homology models given for Arabidopsis thaliana LFY protein were the templates designated as C2vy2A, C2lpeA, and C3dezA. The corresponding confidence for the three templates was determined to be 100%, 56.2%, and 55.6% respectively and the corresponding percent identities were determined to be 100%, 17%, and 16%. The structure of the first template is predicted structure of the LFY protein of Arabidopsis thaliana in complex with DNA from AG-I promoter. The percent coverage of sequence is indicated to be 39% or 163 residues. The structure of the model template C2vy2A is displayed figure 5 and figure 6 displays the 3D structure rotated 180^0. The following two homology models in which are predicated homologous proteins are correspondingly kinase suppressor of ras 1, and orotate phospribosyltransferase.

The template of the second homology model indicated as kinase suppressor of ras 1 was selected and was then compared to the amino acid sequence of Arabidopsis thaliana LFY protein. Figure 9 displays the Phyre results of the second template kinase suppressor of ras1 alignment with Arabidopsis thaliana in complex with DNA from AG-I promoter. A total of 41 residues were aligned and only 17% identity occurs between sequences. Therefore, the query sequence and the template sequence have few similarities and many differences. The highlighted gray area indicates alignment in which only occurs within 6 of the amino acids located at residues 59, 66, 72, 76, 79, 86, and 87 of the query sequence. Conservation in structure can be observed towards the end of the alignment where alpha helices are displayed for both the query sequence secondary structure and the template secondary structure. These are indicated as green spirals and begin at residue 80 for the query sequence and at residue 138 for the template sequence. It can be inferred that this portion of both secondary structure function similarly; perhaps both of these structure are involved in binding mechanisms. Furthermore both sequences have small/polar, hydrophobic, charged, and aromatic amino acids in which supports the hypothesis that the sections of the proteins which are aligned are responsible for binding. On the contrary, the two sequences differ due to differences in residues. Also, a deletion occurs in the temple at residue 127 within the template sequence. The secondary structure of Arabidopsis thaliana LFY protein has beta strands whereas the template secondary structure of the kinase suppressor has only alpha helices. The beta strands are indicated as blue arrows. This is to be expected to the differences in functions of the two proteins.

Microarray analysis

The effect of the hormone salicylic acid (SA) on the expression of the protein LFY is displayed in figure 10. The expression of LFY is depressed when the leaves of the plant Arabidopsis thaliana are exposed to salicylic acid. The results were determined by using long-oligonucleotide microarrays to analyze the features of the Arabidopsis genome. The leaves were introduced to salicylic acid causing the plant undergo systemic acquired resistance (SAR). Six samples are shown within the expression chart; three of the samples are the leaf material control samples and the other three samples are the leaf materials that were exposed to SA. Two of the samples display a significant decrease in expression after being introduced to salicylic acid. The blue squares in the graph indicate the ranking of expression of the genes; all of the six samples are similarly ranked. This indicates that SA down regulates the protein LFY. The protein LFY in which initiates floral development is expected to be down regulated when introduced to the hormone salicyclic acid because this hormone is released when the plant is threatened by pathogens. It would not be beneficial to the plant to flower when the plant is being attacked. Therefore, systemic acquired resistance occurs, the protein LFY is repressed, and floral development is delayed.

The over expression of the protein LFY after seedlings of Arabidopsis thaliana were treated with ethylene is displayed in figure 11. Using the platform Affymetrix Arabidopsis ATH1 Genome Array it was determined that the expression of LFY is induced when 3 day old seedlings of Arabidopsis were introduced to the plant hormone ethylene. Eight samples were analyzed; four of the samples indicated on the left side of the graph were treated with ethylene and the four samples indicated on the right side of the graph were not. A significant increase in the expression of LFY can be observed for the four samples treated with ethylene on the left-hand side of the expression chart; samples 3 and 4 displays the increase in expression most significantly. Although the plant hormone ethylene is essentially responsible for the fruit ripening of plants, the results indicate that ethylene may also have an effect on floral development due to its ability to express the protein LF Y.

Lastly, figure 12 displays the significant decrease in expression of LFY when the seedlings of Arabidopsis thaliana are introduced to the plant hormone auxin. The platform Affymetrix Arabidopsis ATH1 Genome Array was used to determine the effects on the LFY gene when auxin was treated on 5 day old light-grown seedlings. The last 2 samples in figure 3 shows a significant reduction in the expression of LFY indicating that auxin may regulate floral development in plants. It is known that auxin and brassinosteroids integrate into one mechanism to promote the cell expansion in seedlings. Within the study “Interdependency of brassinosteroid and auxin signaling in Arabidopsis”, it was determined that an increase in auxin levels led to the brassinosteroid-stimulated growth response which reduces brassinosteroid effects on gene expression. This may indicate why there is a reduction in the expression of LFY. If this is true, then auxin and brassinosteroids may be involved in the regulation of floral development of flowering plants.

Conclusion

Collectively, using bioinformatics tools, biological knowledge of the Arabidopsis thaliana LFY protein was gained. It is known that LFY functions as a transcription factor to regulate and imitate floral development; when the protein is activated the downstream meristem genes AP1 and CAL are turned on to induce floral initiation. The protein LFY has a highly conserved C-terminus that is made up of a seven fold-helix that binds to DNA as a dimer. This is a strong indication that the protein functions as a developmental switch. When mutation occurs within the LFY transcription factor a drastic delay in meristem transitions occurs and abnormal flowering is observed. Using the local pairwise sequence alignment tool BLAST, homologs of the Arabidopsis thaliana LFY protein were determined. The protein was determined to have a vast amount of orthologs but no likely paralogs. The protein was determined to show the most identity with the leafy-like protein *Brassica juncea* the sequences are highly conserved.

With the global multiple sequence alignments used, conservation was observed throughout the species. Good alignment appeared throughout the species but only sort sections of the sequences were perfectly aligned. Poor alignment appeared mostly in the middle and the end of the sequence. It is likely that the function of the proteins does not depend on these portions of these sequences. The phylogeny tree supported the results of both the local pairwise alignment and the multiple sequence alignment. Both closely related orthologs and more distantly related orthologs were observed in the phylogenic tree. The tree indicates that the protein LFY of Arabidopsis thaliana is more closely evolutionarily related to leafy/leafy-like proteins of *Arabis alpina, Leavenworthia crassa, Brasssica rapa, Brassica napus, cochlearia acualis, Brassica juncea, and Arabidopsis lyrata.* Furthermore, it was determined that all the orthologs stemmed from the common ancestor indicated as the marpoflo protein of *Marchantia polymorpha.* The domain FLO_LFY was determined to be responsible for the conservation that is withheld through the species; when mutations occur in these proteins the floral and leaf development of the plant is affected.

The online PDB program provided structures of the protein LFY of Arabidopsis thaliana in complexes with DNA and also possible homologous proteins. Therefore, the prediction of the structure of LFY was not necessary. Instead prediction of which residues that are involved in the function of the protein could be made using the Phrye-2 server. It was determined that the protein LFY does not share a high amount of identity with other proteins. When the amino acid sequence of LFY was compared to the most likely protein homolog kinase suppressor of ras 1, it is was determined to have more differences than similarities. On the contrary, conservation within the structures was observed towards the end of the alignment where alpha helices are displayed for both the query sequence secondary structure and the template secondary structure. It is likely that the alpha helices play a role in binding.

Lastly, it was determined that the expression of LFY is affected by the three hormones, salicylic acid, ethylene, and auxin. When the leaves of the plant Arabidopsis thaliana are exposed

to salicylic acid the expression of LFY is depressed. Therefore, when the plant is threatened by pathogens, floral initiation may be delayed due to the release of salicylic acid. On the contrary, the expression of the LFY protein can be significantly reduced when seedlings are introduced to ethylene. The hormone ethylene is primarily responsible for the fruit ripening of plants; this indicates that ripening fruits down regulate floral initiation. Finally, the hormone auxin also reduces the expression of LFY indicating that auxin also indirectly influences a delay of floral development on the protein LFY.

References

- Alonso JM, Stepanova AN, Leisse TJ, Kim CJ et al. Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science* 2003 Aug 1;301(5633):653-7. PMID: 12893945

- Blazquez MA, Ferrandiz C, Madueno F, Parcy F (2006) How floral meristems are built. Plant Mol Biol 60: 855–870 [PubMed]

- Coen ES, Meyerowitz EM (1991) The war of the whorls: genetic interactions controlling flower development. Nature 353: 31–37

- C Hamès, D Ptchelkine, C Grimm, E Thevenon, E Moyroud, F Gérard, J.LMartiel, RBenlloch, F Parcy, and C W Mülle. " Structural Basis for LEAFY Floral Switch Function and Similarity with Helix-turn-helix Proteins." *National Center for Biotechnology Information*. U.S. National Library of Medicine, 11 Sept. 2008. Web. 26 Oct. 2014.

- Huala E., Sussex I. M. (1992). *LEAFY* interacts with floral homeotic genes to regulate Arabidopsis floral development. *Plant Cell* 4, 901-913

- Thibaud-Nissen F, Wu H, Richmond T, Redman JC et al. Development of Arabidopsis whole-genome microarrays and their application to the discovery of binding sites for the TGA2 transcription factor in salicylic acid-treated plants. *Plant J* 2006 Jul;47(1):152-62. PMID: 16824183

- Nemhauser JL, Mockler TC, Chory J. Interdependency of brassinosteroid and auxin signaling in Arabidopsis. *PLoS Biol* 2004 Sep;2(9):E258. PMID: 15328536

- W., Y. Zhou, X. Liu, P. Yu, J. D. Cohen, and E. M. Meyerowitz. "LEAFY Controls Auxin Response Pathways in Floral Primordium Formation."*Science Signaling* 6.270 (2013): Ra23. Web.

- "X-ray Powder Diffraction (XRD)." *X-ray Powder Diffraction (XRD)*. N.p., n.d. Web. 19 Dec. 2014.
- http://tcoffee.vital-it.ch/apps/tcoffee/do:regular
- http://blast.ncbi.nlm.nih.gov/
- http://dialign.gobics.de/chaos-dialign-submission
- http://www.phylogeny.fr/
- http://www.rcsb.org/pdb/home/home.do
- http://www.sbg.bio.ic.ac.uk/phyre2

Figures and tables:

Description	Accession	Max score	E-value	Percent Identity
LEAFY transcription factor [Brassica juncea]	ABF06559.1	794	0.0	99%
leafy [Aabidopsis thailiana	AAM27931.1	793	0.0	99%
LFY protein [Litchi chinensis]	AGR45584.1	691	0.0	65%
LEAFY-like protein [Dimocarpus longan]	ABP02007.1	690	0.0	65%
LEAFY-like protein[Cochlearia acaulis]	AAO73066.1	715	0.0	89%
LFY [Arabis	AEH43351.1	690	0.0	81%

alpine]				
LEAFY-like protein [Pistacia chinensis]	AGF33326.1	686	0.0	66%
LEAFY-like protein [Leavenworthia crassa]	AAO73067.1	687	0.0	83%
LEAFY [Clausena lansium]	ABF61861.2	681	0.0	64%
LEAFY [Citrus sinensis]	AAR01229.1	680	0.0	65%
LEAFY-like protein [Pistacia chinensis]	AGF33327.1	678	0.0	67%
LEAFY-like protein [Mangifera indica]	ADX97319.1	678	0.0	65%
Leafy [Arabidopsis lyrata]	AAM27942.1	678	0.0	90%
Protein LEAFY [Brassica rapa]	NP_001288996.1	671	0.0	81%
BnaCnng24550D	CDY53155.1	671	0.0	87%

Table 1: The table above displays the possible homologs selected from the BLAST results. The corresponding accession numbers, maxiumum scores, e-values, and percent identities are indicated in the table.

```
T-COFFEE, Version_11.00.8cbe486 (2014-08-12 22:05:29 - Revision 8cbe486 -
 Build 477)
Cedric Notredame
SCORE=909
*
 BAD AVG GOOD
*
[Arabidopsis      :  88
[Brassica         :  88
[Cochlearia       :  87
[Leavenworthia    :  89
[Brassica_1       :  88
[Arabis           :  87
[Arabidopsis_1    :  87
[Streptanthus     :  87
[Pistacia         :  89
[Pistacia_1       :  89
```

```
[Mangifera          :  89
[Litchi             :  87
[Citrus             :  87
[Dimocarpus         :  87
[Clausena           :  88
cons                :  90

[Arabidopsis          1 MDPEGFTSGLFRWNPTRALVQAP-P--PVPP-PLQQ   32
[Brassica             1 MDPEGFTSGLFRWNPTRALVQAP-P--PVPP-PLQQ   32
[Cochlearia           1 MDPEGFTSGLFRWNTTRAMVQHQ-PPPQVPPPPSQQ   35
[Leavenworthia        1 MDPEGFTSGLFRWNPTRATVQAL-P--PVPP-PLQQ   32
[Brassica_1           1 MDPEGFTSGLFRWNPTRAMVQQPPP--PVPPPPQQQ   34
[Arabis               1 MDPEGFPSGLFRWNPTRPLVQAP-PQPQVPP-PPQQ   34
[Arabidopsis_1        1 MDPEGFTSGLFRWNPTRAMVAAP-P--PVPP-QPQQ   32
[Streptanthus         1 -------SGLFRWNSTRALVQQP-P--PVPP-PQQQ   25
[Pistacia             1 MDPEAFTASLFKWDPR-GLV--P-PQTRV-LEPVVP   31
[Pistacia_1           1 MDPEAFTASLFKWDPR-GLV--P-PQTRV-LEPVVP   31
[Mangifera            1 MDPEAFTASLFKWDPR-GVV--P-PQTRM-LEPVAP   31
[Litchi               1 MDPEAFTASLFKWDPR-TVV--P-PPARL-LEGVTT   31
[Citrus               1 MDPEAFTASLFKWDPR-VVV-AP-PPARVQLEQVSQ   33
[Dimocarpus           1 MDPEAFTASLFKWDPR-TVV--P-PPARL-LEGMAT   31
[Clausena             1 MDPEAFTASLFKWDPR-VVV-AP-PPARLQLEQVSQ   33

cons                  1         :.**:*:.   *   *   :        36

[Arabidopsis         33 QPVT--------PQT-A---AFGMRLGGLEGLFGPY   56
[Brassica            33 QPVT--------PQT-A---AFGMRLGGLEGLFGPY   56
[Cochlearia          36 SPVT--------PQT-A---AFGMRLGGLEGLFGPY   59
[Leavenworthia       33 QPAT--------VQS-A---AFGTRLGGLEGLFGVY   56
[Brassica_1          35 PPAT--------PQT-A---AFGMRLGGLEGLFGPY   58
[Arabis              35 SPAT--------PHT-AAAGAFGMRLGGLEGLFGAY   61
[Arabidopsis_1       33 QPAT--------PQT-R---AFGMRLGGLEGLFGAY   56
[Streptanthus        26 PPET--------PQT-V---AFGMRLGGLEGLFGAY   49
[Pistacia            32 -LPAAISA--ATAFS-V---VRPRELGGLEELFQAY   60
[Pistacia_1          32 -LPAAISA--ATAFS-V---VRPRELGGLEELFQAY   60
[Mangifera           32 -PPVPLSA--ATAFS-V---VRPRELGGLEELFQAY   60
[Litchi              32 -PSAPLVG--SAAYP-M---VRPRDLGGLEELFQAY   60
[Citrus              34 PPAVPLGAAAAAAYSAL---VRPRELGGLEELFQAY   66
[Dimocarpus          32 -PSAPLVG--SAAYS-M---VRPRDLGGLEELFQAY   60
[Clausena            34 PPAVPLGAAAAAAYSAL---VRPRELGGLEELFQAY   66

cons                 37    .    .  .   .    *****  **   *    72

[Arabidopsis         57 GIRFYTAAKIAELGFTASTLVGMKDEELEEMMNSLS   92
[Brassica            57 GIRFYTAAKIAELGFTASTLVGMKDEELEEMMNSLS   92
[Cochlearia          60 GIRFYTAAKIAELGFTASTLVGMKDEELEDMMNSLS   95
[Leavenworthia       57 GIRFYTAAKIAELGFTASTLVGMRDEELEEMMNSLS   92
```

```
[Brassica_1          59 GVRFYTAAKIAELGFTASTLVGMKDEELEDMMNSLS   94
[Arabis              62 GIRFYTAAKIAELGFTASTLVDMRDEELEEMMNSLS   97
[Arabidopsis_1       57 GIRFYTAAKIAELGFTASTLVGMKDEELEEMMNSLS   92
[Streptanthus        50 GIRFYTAAKIAELGFTASTLVGMKDEELEDMMNSLS   85
[Pistacia            61 GIRYYTAAKIAELGFTVNTLVNMKDEELDEMMNSLS   96
[Pistacia_1          61 GIRYYTAAKIAELGFTVNTLVNMKDEELDEMMNSLS   96
[Mangifera           61 GIRYYTAAKIAELGFTVNTLVDMKDEELDEMMNSLS   96
[Litchi              61 GIRYYTAAKIAELGFTVNTLLDMKDEELDEMMNSLS   96
[Citrus              67 GIRYHTAAKMAELGFTVNTLLDMKDEELDEMMNSLG  102
[Dimocarpus          61 GIRYYTAAKIAELGFTVNTLLDMKDEELDEMMNSLS   96
[Clausena            67 GIRYHTAVKIAELGFTVNTLLDMKDEELDEMMNSLG  102

cons                 73 *:*::**.*:******..**:.*:****::*****.  108


[Arabidopsis         93 HIFRWELLVGERYGIKAAVRAERRRLQEEEEEESSR  128
[Brassica            93 HIFRWELLVGERYGIKAAVRAERRRLQEEEEEESSR  128
[Cochlearia          96 HIFRWELLVGERYGIKAAVRAERRRLQEEEEDDSSR  131
[Leavenworthia       93 HIFRWELLVGERYGIKAAVRAERRRLQEEEEEESSR  128
[Brassica_1          95 HIFRWELLVGERYGVKAAVRAERRRLLEEEEQESSR  130
[Arabis              98 HIFRWELLVGERYGIKAAVRAERRRVQEEEEEEASR  133
[Arabidopsis_1       93 HIFRWELLVGERYGIKAAVTAERRRLQEEEEEESSR  128
[Streptanthus        86 HIFRWELLVGERYGIKAAVRAERRRLQEVEEEESSR  121
[Pistacia            97 QIFRWELLVGERYGIKAAVRAERRRLDE----EDSR  128
[Pistacia_1          97 QIFRWELLVGERYGIKAAVRAERRRLDE----EDSR  128
[Mangifera           97 QIFRWELLVGERYGIKAAVRAERRRLDE----EDSR  128
[Litchi              97 QIFRWELLVGERYGIKAAVRAERRRLED----EDSR  128
[Citrus             103 HLFRWELLVGERYGIKAAVRAERRRLDE----DDLR  134
[Dimocarpus          97 QIFRWELLVGERYGIKAAVRAERRRLED----EDSR  128
[Clausena           103 HLFRWELLVGERYGIKAAVRAERRRLED----EDLR  134

cons                109 ::************:**** *****: :    : *  144


[Arabidopsis        217 GMDNGNG--GSGLGTERQREHPFIVTEPGEVARGKK  250
[Brassica           217 GMDNGNG--GSGLGTERQREHPFIVTEPGEVARGKK  250
[Cochlearia         222 GMDNGNG--GIGLGTERQREHPFIVTEPGEVARGKK  255
[Leavenworthia      217 GMDNSNG----GLGTERQREHPFIVTEPGEVARGKK  248
[Brassica_1         217 DGNGG-G-GGGVLGIERQREHPFIVTEPGEVARGKK  250
[Arabis             217 E---GNG--VGVLGMERQREHPFIVTEPGEVARGKK  247
[Arabidopsis_1      225 GMDNGNGGGGGGLGTERQREHPFIVTEPGEVARGKK  260
[Streptanthus       217 GMDNCNG-GGGGLGIERQREHPFIVTEPGEVARGKK  251
[Pistacia           204 -----N-----AEGSERQREHPFIVTEPGEVARGKK  229
[Pistacia_1         204 -----N-----AEGSERQREHPFIVTEPGEVARGKK  229
[Mangifera          204 -----N-----GEGSERQREHPFIVTEPGEVARGKK  229
[Litchi             207 -----NG---GSGGTERQREHPFIVTEPGEVARGKK  234
```

```
[Citrus              210 -----YG--GGCGGMERQREHPFIVTEPGEVARGKK  238
[Dimocarpus          207 -----NG---GSGGTERQREHPFIVTEPGEVARGKK  234
[Clausena            209 -----YG--GGCGGMERQREHPFIVTEPGEVARGKK  237

cons                 253          * ********************  288


[Arabidopsis         251 NGLDYLFHLYEQCREFLLQVQTIAKDRGEKCPTKVT  286
[Brassica            251 NGLDYLFHLYEQCREFLLQVQTIAKDRGEKCPTKVT  286
[Cochlearia          256 NGLDYLFHLYEQCREFLLQVQTIAKDRGEKCPTKVT  291
[Leavenworthia       249 NGLDYLFHLYEQCREFLLQVQTIAKDRGEKCPTKVT  284
[Brassica_1          251 NGLDYLFHLYEQCREFLIQVQTIAKDRGEKCPTKVT  286
[Arabis              248 NGLDYLFHLYEQCREFLLQVQTIAKDRGEKCPTKVT  283
[Arabidopsis_1       261 NGLDYLFHLYEQCREFLLQVQTIAKDRGEKCPTKVT  296
[Streptanthus        252 NGLDYLFHLYEQCREFLLQVQTIAKDRGEKCPTKGT  287
[Pistacia            230 NGLDYLFHLYEQCRDFLIQVQNIAKERGEKCPTKVT  265
[Pistacia_1          230 NGLDYLFHLYEQCRDFLIQVQNIAKERGEKCPTKVT  265
[Mangifera           230 NGLDYLFHLYEQCRDFLIQVQNIAKERGEKCPTKVT  265
[Litchi              235 NGLDYLFHLYEQCRDFLIQVQNIAKERGEKCPTKVT  270
[Citrus              239 NGLDYLFHLYEQCRDFLIQVQNIAKERGEKCPTKVT  274
[Dimocarpus          235 NGLDYLFHLYEQCRDFLIQVQNIAKERGEKCPTKVT  270
[Clausena            238 NGLDYLFHLYEQCRDFLIQVQNIAKERGEKCPTKVT  273

cons                 289 **************:**:***.***:******** *  324


[Arabidopsis         287 NQVFRYAKKSGASYINKPKMRHYVHCYALHCLDEEA  322
[Brassica            287 NQVFRYAKKSGASYINKPKMRHYVHCYALHCLDEEA  322
[Cochlearia          292 NQVFRYAKKSGASYINKPKMRHYVHCYALHCLDEEA  327
[Leavenworthia       285 NQVFRYAKKSGASYINKPKMRHYVHCYALHCLDEEA  320
[Brassica_1          287 NQVFRYAKKSGASYINKPKMRHYVHCYALHCLDEEA  322
[Arabis              284 NQVFRYAKKSGASYINKPKMRHYVHCYALHCLDEEA  319
[Arabidopsis_1       297 NQVFRYAKKSGASYINKPKMRHYVHCYALHCLDEDA  332
[Streptanthus        288 NQVFRYAKNSGASYINKPKMRHYVHCYALHCLDEEA  323
[Pistacia            266 NQVFRYAKKAGASYINKPKMRHYVHCYALHCLDEEA  301
[Pistacia_1          266 NQVFRYAKKAGASYINKPKMRHYVHCYALHCLDEEA  301
[Mangifera           266 NQVFRFAKKAGASYINKPKMRHYVHCYALHCLDEEA  301
[Litchi              271 NQVFRYAKKAGASYINKPKMRHYVHCYALHCLDEEA  306
[Citrus              275 NQVFRYAKKAGASYINKPKMRHYVHCYALHCLDEEA  310
[Dimocarpus          271 NQVFRYAKKAGASYINKPKMRHYVHCYALHCLDEEA  306
[Clausena            274 NQVFRYAKKAGASYINKPKMRHYVHCYALHCLDEEA  309

cons                 325 *****:**::************************:*  360


[Arabidopsis         323 SNALRRAFKERGENVGSWRQACYKPLVNIACRHGWD  358
[Brassica            323 SNALRRAFKERGENVGSWRQACYKPLVNIACRHGWD  358
[Cochlearia          328 SNALRRAFKERGENVGSWRQACYKPLVNIACRHGWD  363
[Leavenworthia       321 SNALRRAFKERGENVGSWRQACYKPLVNIACRHGWD  356
```

```
[Brassica_1          323 SNALRRAFKERGENVGSWRQACYKPLVNIACRHGWD  358
[Arabis              320 SNALRRAFKERGENVGSWRQACYKPLVDIACRHGWD  355
[Arabidopsis_1       333 SNALRRAFKERGENVGSWRQACYKPLVNIACRHGWD  368
[Streptanthus        324 SNALRRAFKERGENVGSWRQACYKPLVNIACRHGWD  359
[Pistacia            302 SDALRRVFKERGENVGAWRQACYKPLVGIAARQGWD  337
[Pistacia_1          302 SDALRRVFKERGENVGAWRQACYKPLVGIAARQGWD  337
[Mangifera           302 SDALRRVFKERGENVGAWRQACYKPLVGIAARQGWD  337
[Litchi              307 SNALRKAFKDRGENVGAWRQACYKPLVAIAARQGWD  342
[Citrus              311 SNALRRAFKERGENVGAWRQACYKPLVAIAARQGWD  346
[Dimocarpus          307 SNALRKAFKERGENVGAWRQACYKPLVAIAARQGWD  342
[Clausena            310 SNALRRAFKERGENVGAWRQACYKPLVAIAASQGWD  345

cons                 361 *:***:.**:******:********** **. :***  396


[Arabidopsis         359 IDAVFNAHPRLSIWYVPTKLRQLCHLERNNAVAAAA  394
[Brassica            359 IDAVFNAHPRLSIWYVPTKLRQLCHLERNNAVAAAA  394
[Cochlearia          364 IDAVFNAHPRLSIWYVPTKLRQLCHLERNNAVAAAA  399
[Leavenworthia       357 IDAVFNSHPRLSIWYVPTKLRQLCHMERNNEVAAAT  392
[Brassica_1          359 IDAVFNAHPRLSIWYVPTKLRQLCHLERSNAVAAAS  394
[Arabis              356 IDAVFNAHPRLSIWYVPTKLRQLCHLERNNAVATAA  391
[Arabidopsis_1       369 IDAVFNAHPRLSIWYVPTKLRQLCHLERNNAVAAAA  404
[Streptanthus        360 IDAVFNAHPHLSIWYVPTKLRQLCHLERNNAVAAAA  395
[Pistacia            338 IDAIFNAHPRLAIWYVPTKLRQLCHAERNSV-TASS  372
[Pistacia_1          338 IDAIFNAHPRLAIWYVPTKLRQLCHAERN-------  366
[Mangifera           338 IDAIFNAHPRLAIWYVPTRLRQLCHAERN-------  366
[Litchi              343 IDAIFNAHPRLAIWYVPTKLRQLCHAERN-------  371
[Citrus              347 IDAIFNAHPRLGIWYVPTRLRQLCHAERN-------  375
[Dimocarpus          343 IDTIFNAHPRLAIWYVPTKLRQLCHAERNT------  372
[Clausena            346 IDSIFNAHPRLAIWYVPTRLRQLCHAERN-------  374

cons                 397 **::**:**:*.******:****** **.         432


[Arabidopsis         395 ALVG-GISCTGSSTSGRGGCGGD-DLRF  420
[Brassica            395 ALVG-GISCTGSSTSGRGGCGGD-DLRF  420
[Cochlearia          400 ALVG-GISCTGSSASGRGGCGGDEELRY  426
[Leavenworthia       393 VLVG-GISCTGTSASGHGECGGE-----  414
[Brassica_1          395 ALVGNGISCTGSSASG------------  410
[Arabis              392 ALVG-GISCTGSSASGRGGCGGD-ELRF  417
[Arabidopsis_1       405 ALVG-GISCTGSSTSGRGGCGGD-DLRF  430
[Streptanthus        396 ALVG-GISC------------------  403
[Pistacia            373 SVSGG----------------------  377
[Pistacia_1          367 --------------------------  366
[Mangifera           367 --------------------------  366
[Litchi              372 --------------------------  371
[Citrus              376 --------------------------  375
[Dimocarpus          373 --------------------------  372
[Clausena            375 --------------------------  374
```

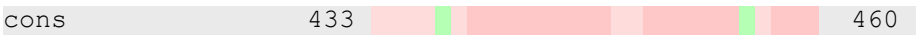

FIGURE 1 (ABOVE): T-coffee output of conserved regions of the multiple sequence alignment of possible homologs. Sections of poor alignment and indel regions are deleted from the figure.

```
Dialign-Pfam


 Name:  protein          Len:  421
 Name:  LEAFY            Len:  421
 Name:  LFY              Len:  389
 Name:  LEAFY-like       Len:  389
 Name:  LEAFY-like       Len:  427
 Name:  LFY              Len:  418
 Name:  leafy            Len:  418
 Name:  LEAFY-like       Len:  384
 Name:  LEAFY-like       Len:  418
 Name:  LEAFY            Len:  398
 Name:  LEAFY            Len:  399
 Name:  LEAFY-like       Len:  384
 Name:  LEAFY            Len:  399
 Name:  leafy            Len:  431
 Name:  protein          Len:  416
 Name:  BnaCnng24550D    Len:  412

//

protein          MDPEGFTSGL FRWNP-TRAL VQ-APPPVPP PLQ--QQPVT PQTAAFGMR-
LEAFY            MDPEGFTSGL FRWNP-TRAL VQ-APPPVPP PLQ--QQPVT PQTAAFGMR-
LFY              MDPEAFTASL FKWDP-RTVV -P-PPARLLE GVTTPSAPLV GSAAYPMVRP
LEAFY-like       MDPEAFTASL FKWDP-RTVV -P-PPARLLE GMATPSAPLV GSAAYSMVRP
LEAFY-like       MDPEGFTSGL FRWNTTRAMV QhqPPPQV-P PPPSQQSPVT PQTAAFGMR-
LFY              MDPEGFPSGL FRWNPTRPLV QA-PPQPQVP PPPQQSPATP HTAAAGAFGM
leafy            MDPEGFTSGL FRWNPTRALV QA-PPQPQVP PPPQQSPATP HTAAAGAFGM
LEAFY-like       MDPEAFTASL FKWDP-RGVV -P-PQTRVLE PVVPLPAAIS AATAFSVVRP
LEAFY-like       MDPEGFTSGL FRWNP-TRAT VQ-ALPPVPP PLQ--QQPAT VQSAAFGTR-
LEAFY            MDPEAFTASL FKWDP-RVVV AP-PPARLQL EQVSQPPAVP LGAAAAAAYS
LEAFY            MDPEAFTASL FKWDP-RVVV AP-PPARVQL EQVSQPPAVP LGAAAAAAYS
LEAFY-like       MDPEAFTASL FKWDP-RGVV -P-PQTRMLE PVAPPPVPLS AATAFSVVRP
LEAFY            MDPEAFTASL FKWDP-RVVV AP-PPARVQL EQVSQPPAVP LGAAAAAAYS
leafy            MDPEGFTSGL FRWNP-TRAM VA-APPPVPP QPQQ--QPAT PQTRAFGMR-
protein          MDPEGFTSGL FRWNPTRAMV QQ-PPPPVPP P-PQQQPPAT PQTAAFGMR-
BnaCnng24550D    MDPEGFTSGL FRWNP-TRVM VQ-APTPIPP PQQQSPa--T PQTAAFGMR-


protein          --LG-----G LEGLFGPYGI RFYTAAKIAE LGFTASTLVG MKDEELEEMM
LEAFY            --LG-----G LEGLFGPYGI RFYTAAKIAE LGFTASTLVG MKDEELEEMM
LFY              RDLG-----G LEELFQAYGI RYYTAAKIAE LGFTVNTLLD MKDEELDEMM
LEAFY-like       RDLG-----G LEELFQAYGI RYYTAAKIAE LGFTVNTLLD MKDEELDEMM
LEAFY-like       --LG-----G LEGLFGPYGI RFYTAAKIAE LGFTASTLVG MKDEELEDMM
LFY              R------LGG LEGLFGAYGI RFYTAAKIAE LGFTASTLVD MRDEELEEMM
leafy            R------LGG LEGLFGAYGI RFYTAAKIAE LGFTASTLVD MRDEELEEMM
LEAFY-like       R-----ELGG LEELFQAYGI RYYTAAKIAE LGFTVNTLVN MKDEELDEMM
LEAFY-like       --LG-----G LEGLFGVYGI RFYTAAKIAE LGFTASTLVG MRDEELEEMM
LEAFY            ALVRPRELGG LEELFQAYGI RYHTAVKIAE LGFTVNTLLD MKDEELDEMM
```

```
LEAFY            ALVRPRELGG LEELFQAYGI RYHTAAKMAE LGFTVNTLLD MKDEELDEMM
LEAFY-like       R-----ELGG LEELFQAYGI RYYTAAKIAE LGFTVNTLVD MKDEELDEMM
LEAFY            ALVRPRELGG LEELFQAYGI RYHTAAKIAE LGFTVNTLLD MKDEELDEMM
leafy            --LG-----G LEGLFGAYGI RFYTAAKIAE LGFTASTLVG MKDEELEEMM
protein          --LG-----G LEGLFGPYGV RFYTAAKIAE LGFTASTLVG MKDEELEDMM
BnaCnng24550D    --LG-----G LEGLFGPYGV RFYTAAKIAE LGFTASTLVG MKDEELEDMM


protein          NSLSHIFRWE LLVGERYGIK AAVRAERRRL QEEEEEESSR RRHLLLSAAG
LEAFY            NSLSHIFRWE LLVGERYGIK AAVRAERRRL QEEEEEESSR RRHLLLSAAG
LFY              NSLSQIFRWE LLVGERYGIK AAVRAERRRL EDED----SR RRNLLSGD--
LEAFY-like       NSLSQIFRWE LLVGERYGIK AAVRAERRRL EDED----SR RRNLLSGD--
LEAFY-like       NSLSHIFRWE LLVGERYGIK AAVRAERRRL QEEEEDDSSR RRHLLLSAAG
LFY              NSLSHIFRWE LLVGERYGIK AAVRAERRRV QEEEEEEASR RRHLLLSAGG
leafy            NSLSHIFRWE LLVGERYGIK AAVRAERRRV QEEEEEEASR RRHLLLSAGG
LEAFY-like       NSLSQIFRWE LLVGERYGIK AAVRAERRRL DEED----SR RRHILSGD--
LEAFY-like       NSLSHIFRWE LLVGERYGIK AAVRAERRRL QEEEEEESSR RRHLLLSAAG
LEAFY            NSLGHLFRWE LLVGERYGIK AAVRAERRRL EDED----LR RRHLLSSD--
LEAFY            NSLGHLFRWE LLVGERYGIK AAVRAERRRL DEDD----LR RRHFLSSD--
LEAFY-like       NSLSQIFRWE LLVGERYGIK AAVRAERRRL DEED----SR RRHILSGD--
LEAFY            NSLGHLFRWE LLVGERYGIK AAVRAERRRL DEDD----LR RRHFLSSD--
leafy            NSLSHIFRWE LLVGERYGIK AAVTAERRRL QEEEEEESSR RRHLLLSAAG
protein          NSLSHIFRWE LLVGERYGVK AAVRAERRRL LEEEEQESSR RRHLILSAAG
BnaCnng24550D    NSLSHIFRWE LLVGERYGIK AAVRAERRRL QEEEEEESSR RRHLLLSAag


protein          NGGSGLGTER QREHPFIVTE PGEVARGKKN GLDYLFHLYE QCREFLLQVQ
LEAFY            NGGSGLGTER QREHPFIVTE PGEVARGKKN GLDYLFHLYE QCREFLLQVQ
LFY              SGGT----ER QREHPFIVTE PGEVARGKKN GLDYLFHLYE QCRDFLIQVQ
LEAFY-like       SGGT----ER QREHPFIVTE PGEVARGKKN GLDYLFHLYE QCRDFLIQVQ
LEAFY-like       NGGIGLGTER QREHPFIVTE PGEVARGKKN GLDYLFHLYE QCREFLLQVQ
LFY              VLGM----ER QREHPFIVTE PGEVARGKKN GLDYLFHLYE QCREFLLQVQ
leafy            VLGM----ER QREHPFIVTE PGEVARGKKN GLDYLFHLYE QCREFLLQVQ
LEAFY-like       --GS----ER QREHPFIVTE PGEVARGKKN GLDYLFHLYE QCRDFLIQVQ
LEAFY-like       GLGT----ER QREHPFIVTE PGEVARGKKN GLDYLFHLYE QCREFLLQVQ
LEAFY            CGGM----ER QREHPFIVTE PGEVARGKKN GLDYLFHLYE QCRDFLIQVQ
LEAFY            CGGM----ER QREHPFIVTE PGEVARGKKN GLDYLFHLYE QCRDFLIQVQ
LEAFY-like       --GS----ER QREHPFIVTE PGEVARGKKN GLDYLFHLYE QCRDFLIQVQ
LEAFY            CGGM----QR QREHPFIVTE PGEVARGKKN GLDYLFHLYE QCRDFLIQVQ
leafy            GLGT----ER QREHPFIVTE PGEVARGKKN GLDYLFHLYE QCREFLLQVQ
protein          GGGGVLGIER QREHPFIVTE PGEVARGKKN GLDYLFHLYE QCREFLIQVQ
BnaCnng24550D    GSGM----ER QREHPFIVTE PGEVARGKKN GLDYLFHLYE QCREFLLQVQ


protein          TIAKDRGEKC PTKVTNQVFR YAKKSGASYI NKPKMRHYVH CYALHCLDEE
LEAFY            TIAKDRGEKC PTKVTNQVFR YAKKSGASYI NKPKMRHYVH CYALHCLDEE
LFY              NIAKERGEKC PTKVTNQVFR YAKKAGASYI NKPKMRHYVH CYALHCLDEE
LEAFY-like       NIAKERGEKC PTKVTNQVFR YAKKAGASYI NKPKMRHYVH CYALHCLDEE
LEAFY-like       TIAKDRGEKC PTKVTNQVFR YAKKSGASYI NKPKMRHYVH CYALHCLDEE
LFY              TIAKDRGEKC PTKVTNQVFR YAKKSGASYI NKPKMRHYVH CYALHCLDEE
leafy            TIAKDRGEKC PTKVTNQVFR YAKKSGASYI NKPKMRHYVH CYALHCLDEE
LEAFY-like       NIAKERGEKC PTKVTNQVFR YAKKAGASYI NKPKMRHYVH CYALHCLDEE
LEAFY-like       TIAKDRGEKC PTKVTNQVFR YAKKSGASYI NKPKMRHYVH CYALHCLDEE
LEAFY            NIAKERGEKC PTKVTNQVFR YAKKAGASYI NKPKMRHYVH CYALHCLDEE
LEAFY            NIAKERGEKC PTKVTNQVFR YAKKAGASYI NKPKMRHYVH CYALHCLDEE
LEAFY-like       NIAKERGEKC PTKVTNQVFR FAKKAGASYI NKPKMRHYVH CYALHCLDEE
LEAFY            NIAKERGEKC PTKVTNQVFR YAKKAGASYI NKPKMRHYVH CYALHCLDEE
leafy            TIAKDRGEKC PTKVTNQVFR YAKKSGASYI NKPKMRHYVH CYALHCLDED
protein          TIAKDRGEKC PTKVTNQVFR YAKKSGASYI NKPKMRHYVH CYALHCLDEE
BnaCnng24550D    TIAKDRGEKC PTKVTNQVFR YAKKSGANYI NKPKMRHYVH CYALHCLDEE


protein          ASNALRRAFK ERGENVGSWR QACYKPLVNI ACRHGWDIDA VFNAHPRLSI
```

```
LEAFY            ASNALRRAFK ERGENVGSWR QACYKPLVNI ACRHGWDIDA VFNAHPRLSI
LFY              ASNALRKAFK DRGENVGAWR QACYKPLVAI AARQGWDIDA IFNAHPRLAI
LEAFY-like       ASNALRKAFK ERGENVGAWR QACYKPLVAI AARQGWDIDT IFNAHPRLAI
LEAFY-like       ASNALRRAFK ERGENVGSWR QACYKPLVNI ACRHGWDIDA VFNAHPRLSI
LFY              ASNALRRAFK ERGENVGSWR QACYKPLVDI ACRHGWDIDA VFNAHPRLSI
leafy            ASNALRRAFK ERGENVGSWR QACYKPLVDI ACRHGWDIDA VFNAHPRLSI
LEAFY-like       ASDALRRVFK ERGENVGAWR QACYKPLVGI AARQGWDIDA IFNAHPRLAI
LEAFY-like       ASNALRRAFK ERGENVGSWR QACYKPLVNI ACRHGWDIDA VFNSHPRLSI
LEAFY            ASNALRRAFK ERGENVGAWR QACYKPLVAI AASQGWDIDS IFNAHPRLAI
LEAFY            ASNALRRAFK ERGENVGAWR QACYKPLVAI AARQGWDIDA IFNAHPRLGI
LEAFY-like       ASDALRRVFK ERGENVGAWR QACYKPLVGI AARQGWDIDA IFNAHPRLAI
LEAFY            ASNALRRAFK ERGENVGAWR QACYKPLVAI AARQGWDIDA IFNAHPRLGI
leafy            ASNALRRAFK ERGENVGSWR QACYKPLVNI ACRHGWDIDA VFNAHPRLSI
protein          ASNALRRAFK ERGENVGSWR QACYKPLVNI ACRHGWDIDA VFNAHPRLSI
BnaCnng24550D    ASNALRRAFK ERGENVGSWR QACYKPLVDI ACRHGWDIDA VFNAHPRLSI
```

FIGURE 2 (ABOVE): The figure above displays the output of the multiple sequence alignment with the program Dialign Pfam.

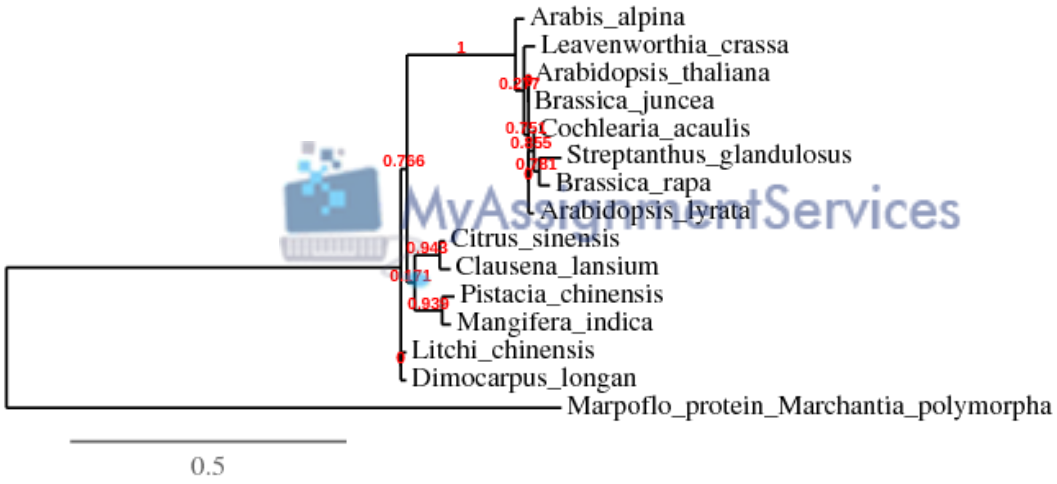

Figure 3: Above is the generated phylogenic tree of the possible homologs created using the phylogeny.fr browser. The corresponding confidence values are indicated in red.

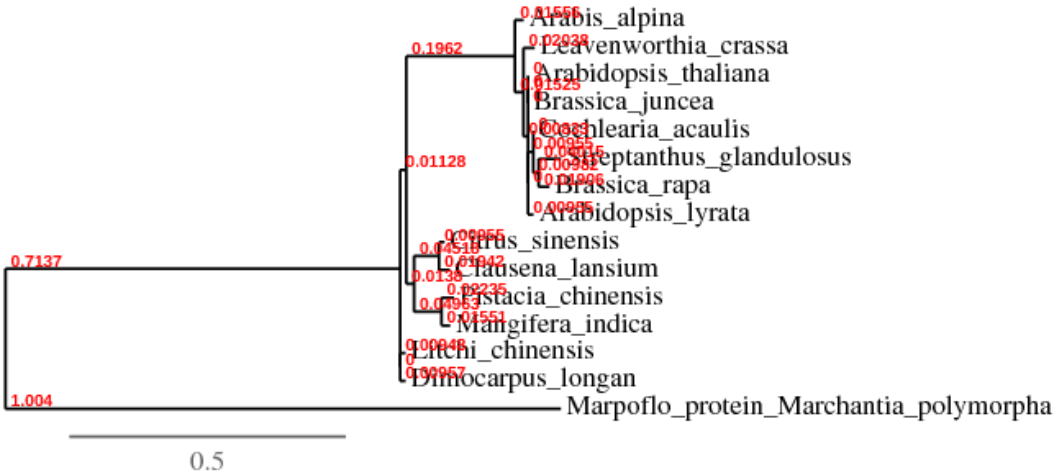

Figure 4: Above is the generated phylogenic tree of the possible homologs created using the phylogeny.fr browser. The corresponding evolutionary distances are indicated in red.

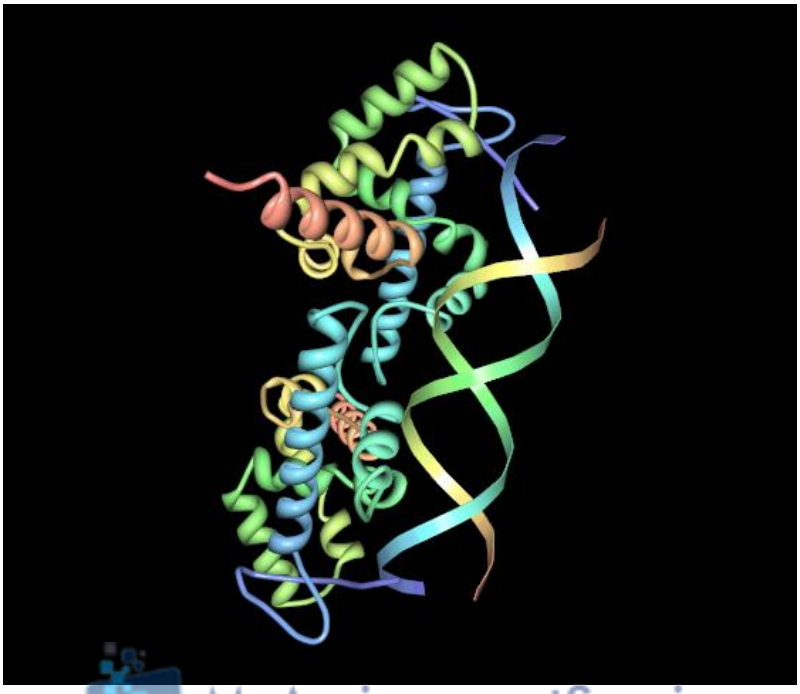

Figure 5(above): Protein structure of Arabidopsis LFY in complex with DNA from AP1 promoter determined with X-Ray diffraction. 3D structure was obtained from the protein database online archive.

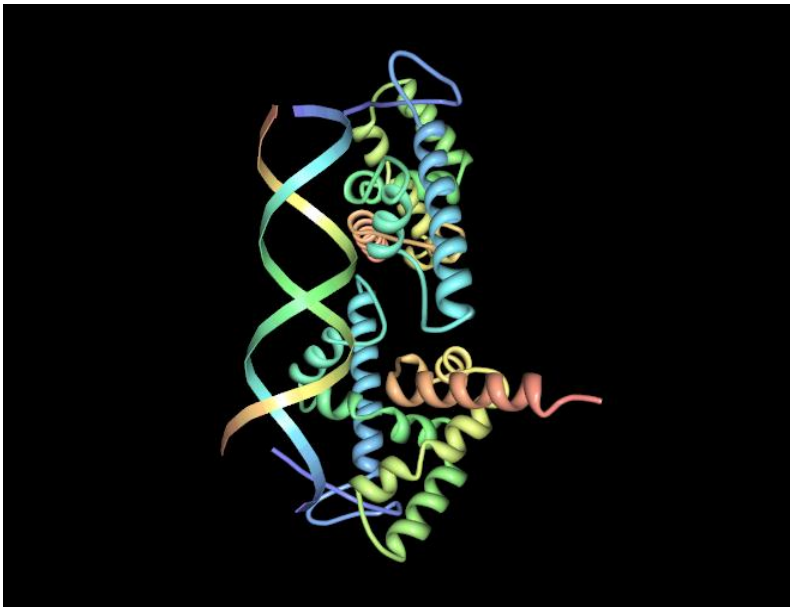

Figure 6(above): The 180 degree rotation of the Arabidopsis thaliana LFY protein in complex with the DNA AP1 promoter.

Name	E-value	PDB identifier	Experiment for protein structure	Resolution
Crystal structure of Moss Leafy bound to DNA	1.13946E-66	4BHK	x-ray diffraction	2.32 Å
Crystal structure of plant glutamate cysteine ligase in comlex with a transition state analogue	3.24128	2GWC	x-ray diffraction	2.18 Å
Crystal structure of plant glutamate cysteine ligase in comlex with Mg2+ and L-Glutamate	3.24128	2GWD	x-ray diffraction	2.09 Å

Table 2: Above is the top three homology models of the LFY protein determined using PDB

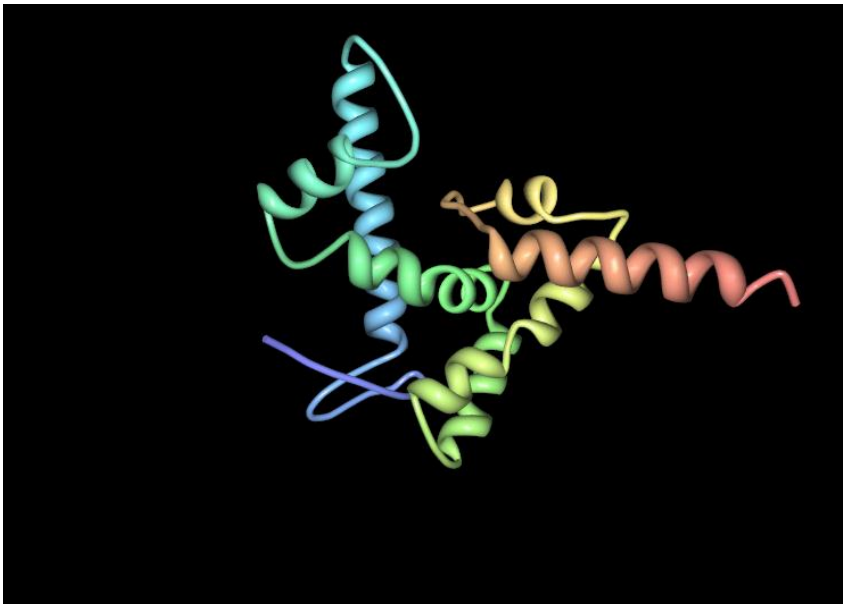

Figure 7: (Above) is the the protein structue of the templete C2vy2A obtained using the Phyre online program of predicted homology.

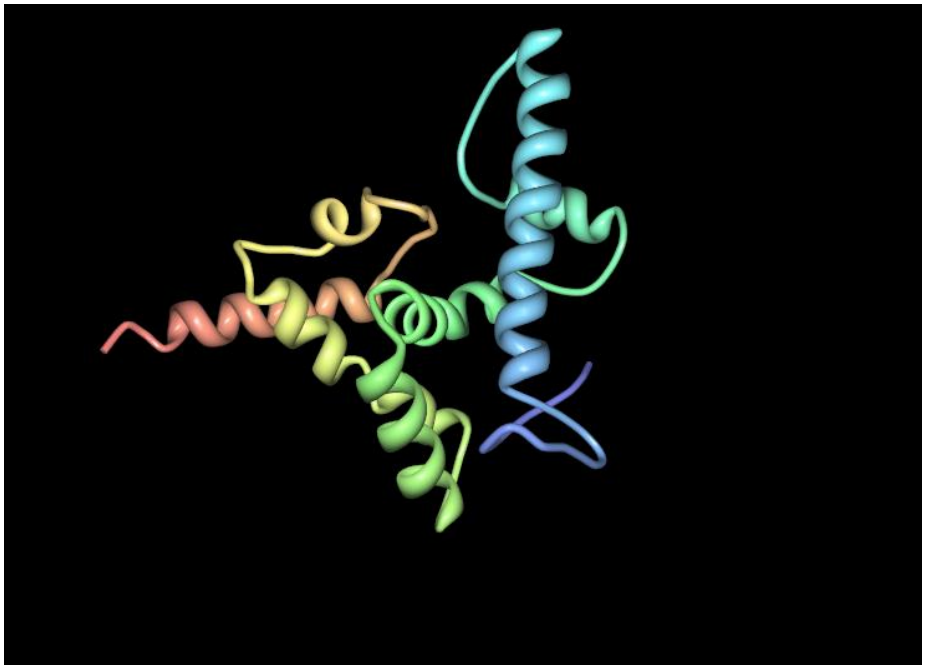

Figure 8 (above) is the protein structure of template C2vy2A rotated 180°.

Phyre2

Job Description	homology_modeling_for_LFY_protein_		
Confidence	56.17%	Date	Sun Nov 16 20:51:43 GMT 2014
Rank	2	Aligned Residues	41
% Identity	17%	Template	c2lpeA_
PDB info	**PDB header:**signaling protein	**Chain:** A: **PDB Molecule:**kinase suppressor of ras 1;	**PDBTitle:** solution nmr structure of the ksr1 ca1-ca1a domain
Resolution	UNK		
Model Dimensions (Å)	**X**:28.541 **Y**:26.142 **Z**:22.212		

Show / Hide SS confidence

Show / Hide Conservation and Alignment quality

Insertion relative to template
Deletion relative to template
Catalytic residue from the CSA

Figure 9 displays the Phyre results of the second template kinase suppressor of ras1 alignment with Arabidopsis thaliana in complex with DNA from AG-I promoter.

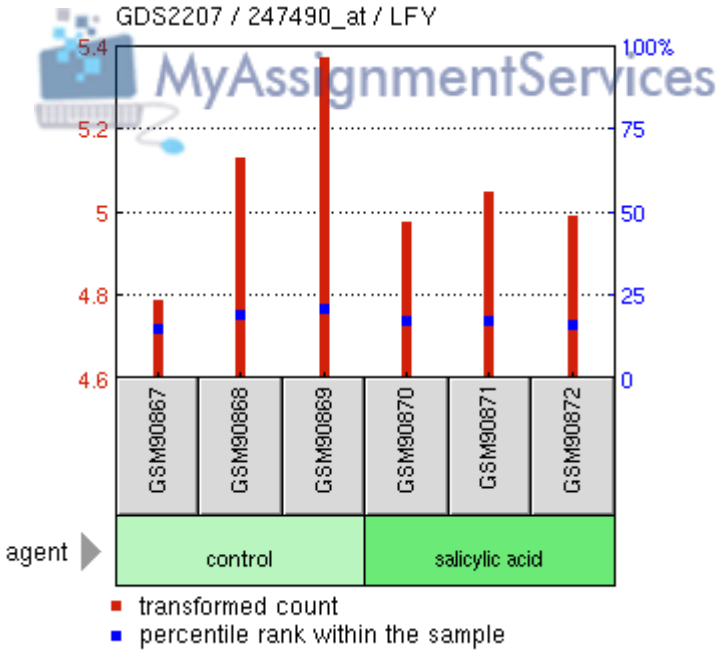

Figure 10: The figure above display the expression of LFY with and without the introduction of SA.

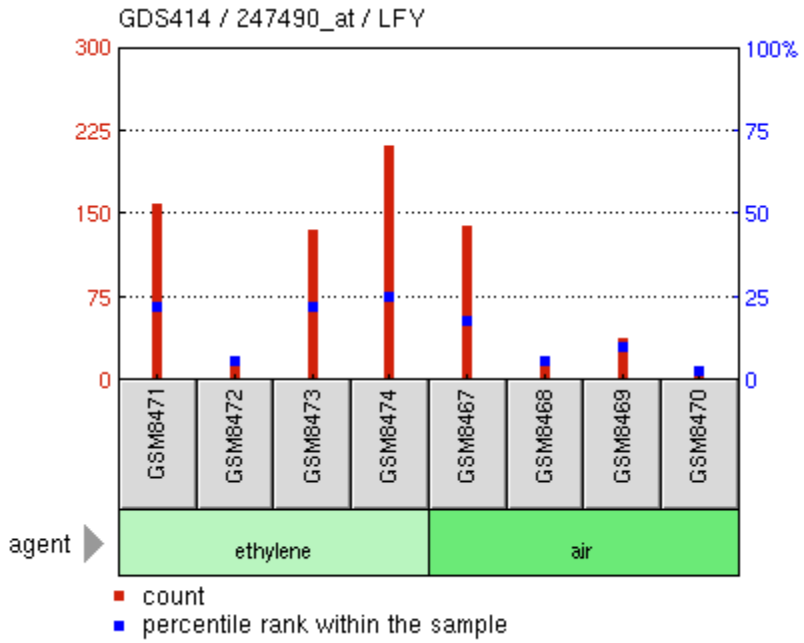

Figure 11: The figure above displays that expression of LFY when introduced to the protein auxin compared to the control.

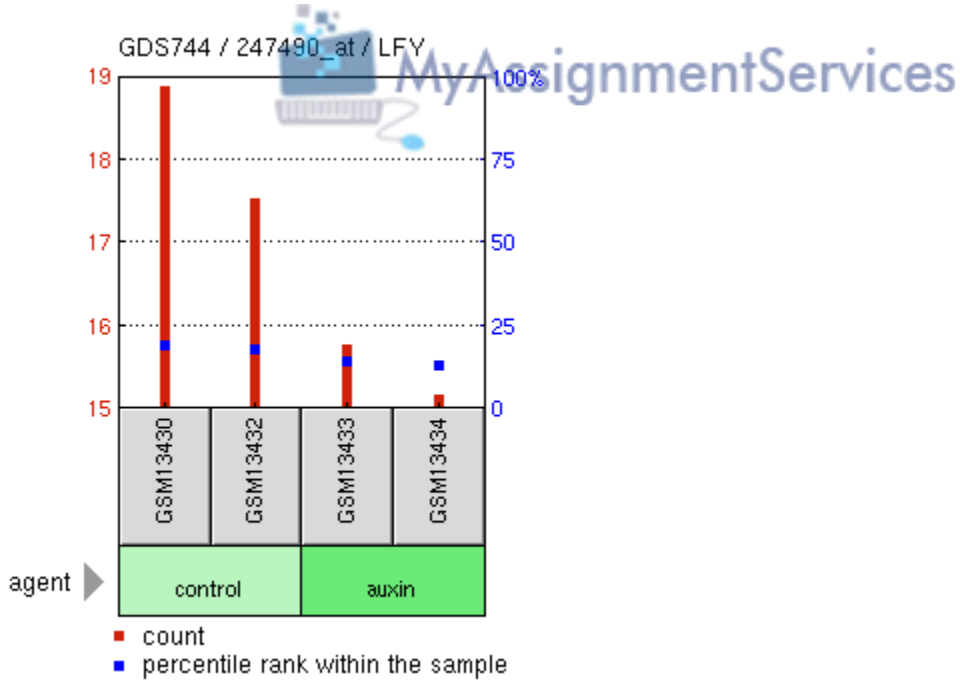

Figure 12 : The figure above displays that expression of LFY when introduced to the protein auxin compared to the control.