Feb 26, 2013 # Get list of genes from peptide file:

grep ">" YnKn\_dehydrins.faa | sed 's/|/ /g' | sed 's/>//' | awk '{print $1}'

# Download the promoters using query\_phytozome.py

# Add extra promoters from species not on Phytozome

# Replace the promoters for genes that are wrong

for line in `cat wrong\_promoters.txt`; do grep $line KS\_dehydrins\_promoters.fas; done

Cucsa.109360 YnSKn Eucgr.F01727 done Kn MDP0000629961 YnSKn Glyma17g24193 KS Glyma09g31740 YnKn Gorai.012G154800 YnSKn Potri.013G062400 done Kn ppa010975m.g YnSKn Thecc1EG025860 KS GRMZM2G169372 KS

# These genes are not SK2, but look very similar to SK2, placed in that category

Lus10003340.g Lus10022643.g

Feb 5, 2013 # Incorrect translation of a gene in maldo While trying to build a transfac matix of Y-segments, BioPython would spit out errors, i finally figured out that they were due to an X amino acid in the identified Y-segment of MDP0000265874. The problem was due to Phytozome missing a nucleotide in the following sequence: GACACAGATGAGTT, it should have been GACACAGATGAGTAT, which results in a proper translation of the nucleotide sequence, in its current state it translates with several stop codons. I will mannually replace the X wiht the correct Y.

For K-segment two genes contained and X: MDP0000196703 and MDP0000178973. The proper sequences were obtained by taking the nucleotide sequence from Phytozome, BLASTing on GenBank and locating the best Malus domestica match.

MDP0000196703, was modified >gi|382948205|gb|AFG33217.1| dehydrin 7 [Malus x domestica] MAGKSTMVPLACFIALAAQARALEVVGGEGLKQKIKEKLPGSTTTDTTYDTTYPGRHHQEKGMKDKNKDK LPEGHKDDPYYSTPHTTLTTTTYGVTTYMEEHH*EKKGIMDKINEKLPS*GHHWRL

MDP0000178973 No proper matches found, was removed from dehydrins.faa

Data downloaded from Phytozome on February 2, 2013

### Command used to count the number of unique dehydrins in Phytozome v9.0 release:

grep ">" scripts/dehydrins.fas|sed 's/|/ /g'|awk {'print $1'}| sort|uniq|wc -l

the total number of unique genes is 251 and the number of unique transcripts is 281

**Aquca\_038\_00119** has 4 different transcripts and has a dehydrin and *reticulon* domain, the dehydrin domain is 7-39.

The K-segments: >Aquca\_038\_00119|Aquca\_038\_00119.1 HSESLMEKIKEKIHG

Dehydrins were not detected in lower plants, except Physcomitrella

### Physcomitrella patens dehydrins

Selginella moellendorffii is closer to higher plants then Physcomitrella, however no dehydrins were found using either my search or the keyword ontology search on Phytozome.

Physcomitrella patens contains 5 potential dehydrins

Pp1s149\_49V6|Pp1s149\_49V6.1|Ppatens EKKGFVGKIKDMIHH not a dehydrin Pp1s442\_22V6|Pp1s442\_22V6.1|Ppatens PKKGLMTKIKEKLPG KKEGFMTKLKEKLPG annotated as dehydrin Pp1s421\_9V6|Pp1s421\_9V6.1|Ppatens PKKGMMEKIKEKLPG KKDGLMNKIKEKLPG dehydrin Pp1s442\_22V6|Pp1s442\_22V6.2|Ppatens second transcript of Pp1s442\_22V6 Pp1s211\_95V6|Pp1s211\_95V6.1|Ppatens HKKGIITKIKEKLHH not a dehydrin Pp1s201\_52V6|Pp1s201\_52V6.1|Ppatens KKHGLINKIKEKLPG dehydrin

Out of 5 putative dehydrins, 3 were probable dehydrins and 2 were not dehydrins the following sequences were added to the dehydrins.faa

Pp1s442\_22V6|Pp1s442\_22V6.1|Ppatens MAAQYTQDQSTEFRPELDEPRRTTTTTTSTTGSGLENENFGGYGGVSENEPPRHKIHSED EPLPKPMSGTYLDEEGAKLDQDSDRSGLGASDLGRDEHIMPKPTSEGYPAGTPQSTEKYQ EHRDLEPTRLDESPKTEEFGAANASTGDFDRTSSDGLRTDKTPASEPRGFRSEDTAPTSG GYADPTSAFPGAPIDRRVEEPGYGYDQQTSEPTSLESGTQHSPKKEGFMTKLKEKLPGHH KTPESGVEHQGAGVEHDTTDAPPKKGLMTKIKEKLPGHHSTAPASTTTDV\* Pp1s442\_22V6|Pp1s442\_22V6.2|Ppatens MAAQYTQDQSTEFRPELDEPRRTTTTTTSTTGSGLENENFGGYGGVSENEPPRHKIHSED EPLPKPMSGTYLDEEGAKLDQDSDRSGLGASDLGRDEHIMPKPTSEGYPAGTPQSTEKYQ EHRDLEPTRLDESPKTEEFGAANASTGDFDRTSSDGLRTDKTPASEPRGFRSEDTAPTSG GYADPTSAFPGAPIDRRVEEPGYGYDQQTSEPTSLESGTQHSPKKEGFMTKLKEKLPGHH KTPESGVEHQGAGVEHDTTDAPPKKGLMTKIKEKLPGHHSTAPASTTTDV\*

Pp1s421\_9V6|Pp1s421\_9V6.1|Ppatens MAGYMGEERKPVSTYNAHQDEFGFKEGNDYPASGGGYGDHRHHRPDVPSSTPGEGYGRQG VEHGYGDQSEETFQDAPERLTGYGDTDNGVGGPKSGYGDSREGTAYDRATDEQTQFGVGG KENSYELEGSDPQLGGTDTSSYAAVDPQRLDSDRSPIRATETVPGGDGWGPEDTSRIHGA KKDGLMNKIKEKLPGHHNTAQGEVSDPNALPKKGMMEKIKEKLPGHDSGSADV\* Pp1s201\_52V6|Pp1s201\_52V6.1|Ppatens MADYNWNMEDSAGGFMGLMSKPDMSSNDGRQDESEERIDNQPGGSSPSPLELTKGHGDSD TGVLGPNSGRGASSEVTGADHRPSSFSESQFGLDAKTVSYGHAAAGSRNVNSHADVDGFC VSRPLCSTERNAGGDGCNLDNTTSTHGVKKHGLINKIKEKLPGHHHTAGRVIENDPPKKA RWIR\*

A gene annotated as dehydrin in Phytozome, contains a possible K-segment and an S-segment, similar to KS dehydrins, was not identified by my search: >Pp1s52\_261V6.1 MDKIKDMLHGHKKDDEHAHATPATATTTNVGTAEAGYSDPA*TTHHEEGKEKKGFLG*LGGHKKEGEEGKKHGFMGMGGGSSSSSSSDEEREGAAERNRLRR EKRAQRTAGKTAAGTVPVEGTGEKKHGFFG\*i

### Selaginella dehydrins

Close paper from 1997 {Close:1997kt} mentions that Selaginella dehydrins contain a different K-segment: EKKEGVMOKIKKK, searching for KIKKK in S. moellendorffii results in the following, the genes are not annotated, there were a few others, without the S-segment:

Looks like a KS dehydrin, with a typical SSSS..DSD at the end >437484|437484|Smoellendorffii MSMMEKISETLHIGGHKKEEEHKKEDKAHAAGEHKGHDHPAGVPCSGSGAVPPPAAAGYK KDEEHKAGEEKKHEGFMEKIKKKKKDRKERKEGEGKGSSSSSSDSD\* >446994|446994|Smoellendorffii MASMMEKISDKLHMGHKKDQVAHNTAPSHATPVQSSTPPSHNTAPGSYGTPQPGMATTTT STTTTQPQKEGLMDKIKKKKNQHKEKKKAGGGSSSSSSDSD\* >448256|448256|Smoellendorffii MAFRSVLVASKKNALGHLLVKKEFLATSPLLLNKEAGTAESMAHKVGDKVEKAGENLKET AKDSKEPAESGMREQVRSAAEKIGDATASAGDKIEHSMESEGGGEEKSKLNLEDKQIFGF VLFSRSQVLFAVLLILAALMASMMEKIGEKFGKKDDEHNKLGGQTHETHGLGQQTHGQGA GYGQEAGYGQHGQGVGHGQGVGHNPGSTPYGATTGTTGTGHNPGSTPYGATTGTNPTSGH TATGQPQKKEGLMDKIKKKKNQHKERKEGDKSSSSSESD\*

Somewhat similar protein from resurrection plant Selaginella lepidophylla was found on GenBank: >gi|2104947|gb|AAB57842.1| dehydrin-like protein [Selaginella lepidophylla] MASMMEKIGDKVHGNKDEQQQQHQYSAAGPQGHGTGGLGSDQHGYKGLGTGKDQHGYQGTGTGTGTDQHG YNAGAVGGGHDSQGAGFGSHTGATTGAVGTGEKKEGVMDKIKKKTHRNKGERKAGEGSSSSDSD