

**Supplementary fig. S1. Gene trees and sequence comparisons. Data are shown first for KNOX, then for ARP, then for YABBY.**

## **KNOX**

### **(A, B) Gene trees of KNOX proteins.**

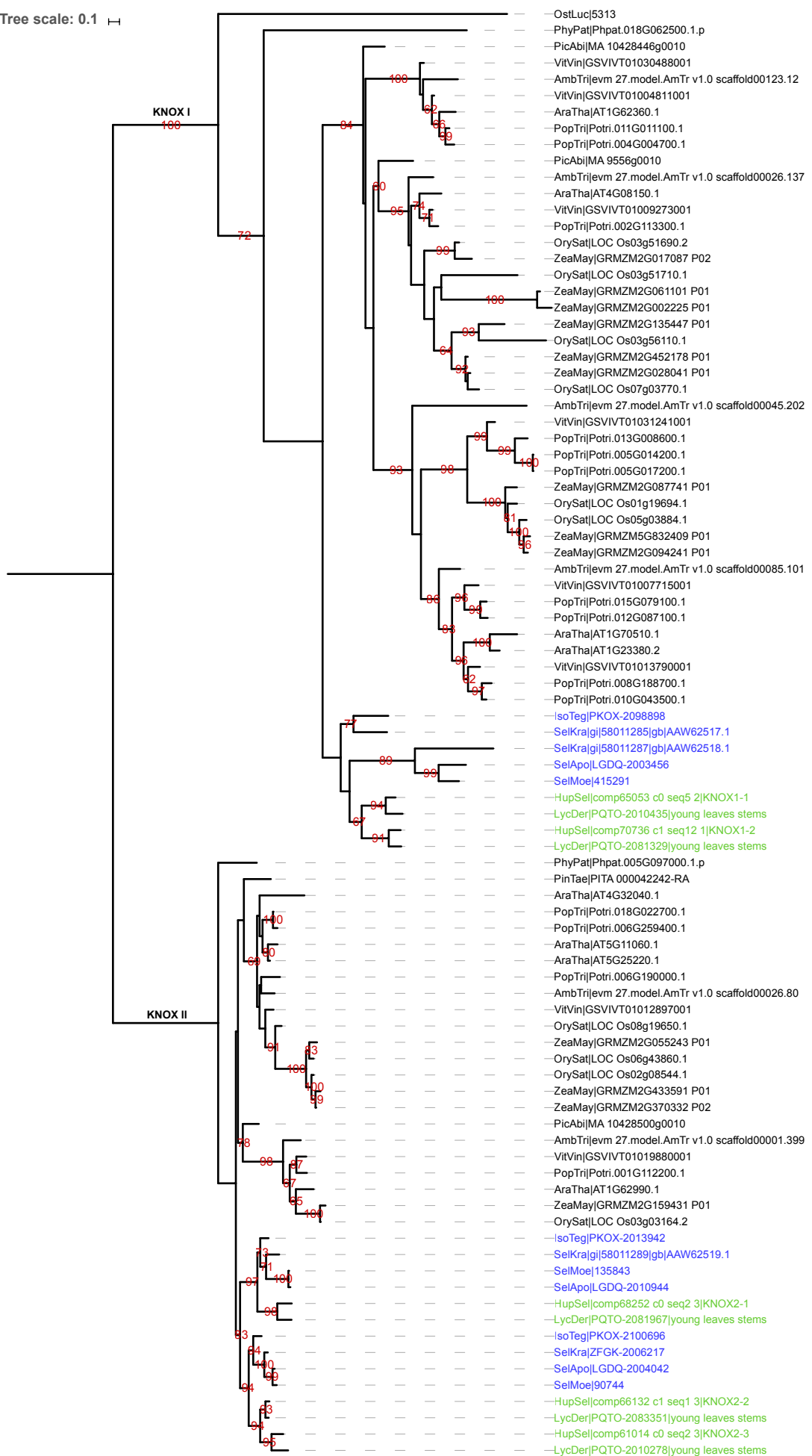
**(A)** Gene tree with sequences that contain all four characteristic domains of KNOX proteins (KNOX1, KNOX2, ELK and homeodomain).

**(B)** Gene tree with sequences that contain at least one of KNOX1, KNOX2 or Homeobox domain. Bootstrap support is only shown for values >60. Species from class Isoetopsida are shown in blue (*Selaginella* sp. and *Isoetes tegetiformans*) whereas sequences from class Lycopodiopsida (*Huperzia selago* and *Lycopodium deuterodensum*) are shown in green.

### **(C) KNOX amino acids important for trafficking via PD**

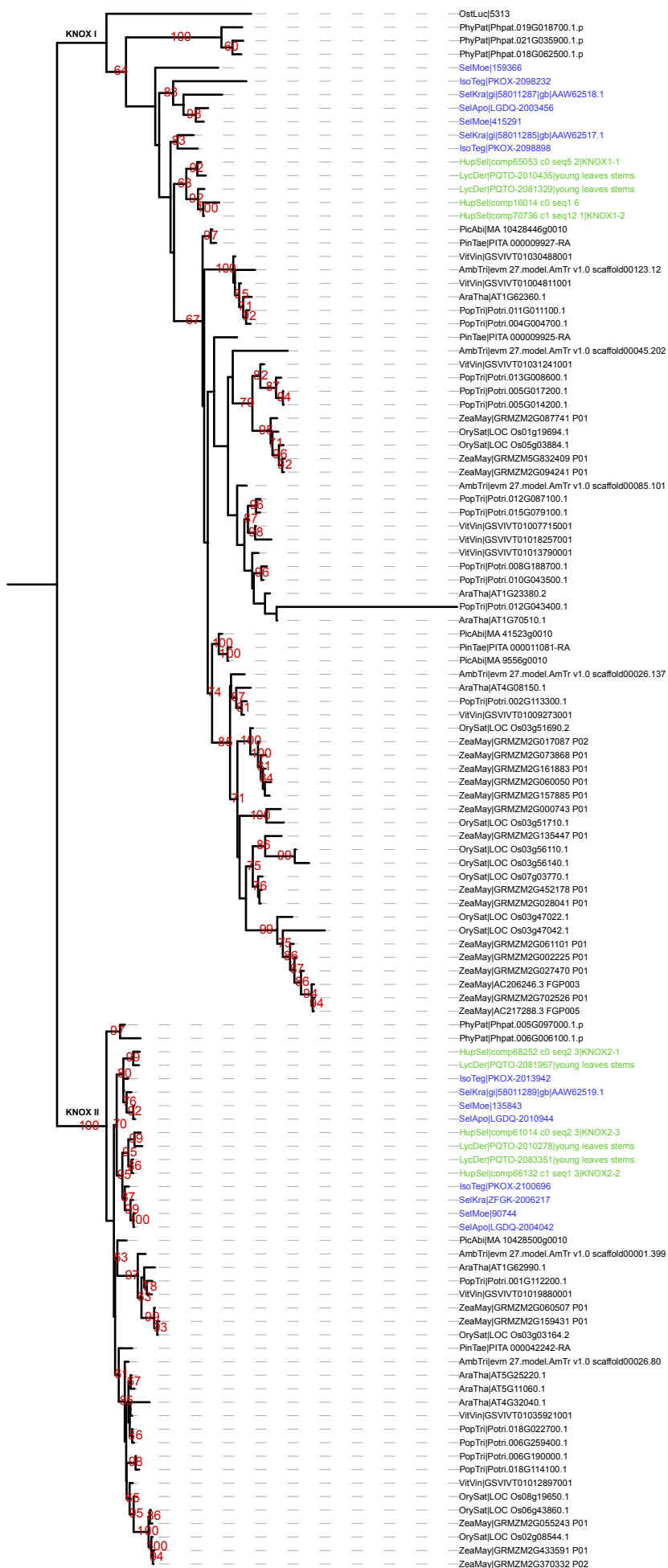
**A**

Tree scale: 0.1



B

Tree scale: 1



# C

**KNOX amino acids important for trafficking via PD.** The homeobox of KNOX proteins is necessary and sufficient for trafficking through PD. Chen et al. (2014) determined that the R30 residue of helix alpha-1 (strong positive charge) and the L53 residue of helix alpha-3 are important for selective intercellular trafficking of homeodomain proteins in plants. The homeodomain parts of the lycophyte KNOXes were added to the Chen et al. (2014) figure (everything above the line is from Chen et al. (2014); sequences from lycophyte KNOXes are below the line). Class II lycophyte KNOXes are highlighted in yellow.

None of the lycophyte KNOXes contains the R30 (but neither does Arabidopsis KN1). However, all lycophyte class II KNOXes contain the L53 (with one exception that has M53 instead). Lycophyte class I KNOXes do not have the L53 (with one exception from Isoetes).

Altogether, this gives the impression that the Lycopodiales/Isoetales KNOXes are not good at traveling through PD, or that the features required for traveling through secondary PD are different in Lycopodiales/Isoetales. Note that the Physco KNOXes that can travel do have the L53 residue (Chen et al., 2014). Amino acid sequences can be found in Supplementary table S2 (Supplementary Material Online).

```

                                     *      *      R30      *      *      L53      *
                                     *      *      *      *      *      *      *
AtKN1      --SSLKQELSKKKKKKGKLPKEARQQLLSDWVQHYKWPYPSETOKVVAIAESTGLDLDKQINNWFINQRKRHWKPS
ATKNAT1    --SSLKQELSKKKKKKGKLPKEARQQLLTWVELHYKWPYPSESEKVALAESTGLDQKQINNWFINQRKRHWKPS
AtSTM      --GSLKQEFMKKKKKKGKLPKEARQQLLDWVSRHYKWPYPSEOKKLAAESTGLDQKQINNWFINQRKRHWKPS
LeT6       --GSLKQEFMKKKKKKGKLPKEARQQLVDWVLRHIKWPYPSESOKLAAESTGLDQKQINNWFINQRKRHWKPS
PhyscoKN2  --GELKAEFNVVRRKKGKLPESARTILKDWVNRHSHPYPSEMEKQYLRICGLNLKQINNWFINERKRHWKPS
PhyscoKN5  --GELKAEFNVVRRKKGKLPESARSILKDWVNRHSYWPYPSEMEKQYLQKLCGLNLKQINNWFINERKRHWKPS
PhyscoKN4  ADEELKKMLR--LKYGKLPENARQQLKDWVSRHSYWPYPSEMEKAYLRQRCGLNLKQINNWFINQRKRHWKPS
AtKNAT2    --SSLKLEFSKKKKKKGKLPKEARQQLLDWVNVHNKWPYPTEGDKIISLAETGLDQKQINNWFINQRKRHWKPS
AtKNAT3    --VDIREEILRKRRAKGLPGDITSVLKAWVQSHSKWPYPTEEDKARLVQETGLQLKQINNWFINQRKRHWKPS
ChlamGSM1  --VQVAASLAQRPKVCKLPPAATQQLKDWVDDNFVWPYPSEEDKKQLGEAALNNTQINNWFINQRKRHWKPS
ChlamHDG1  --GGGGSGPGSKKGRNLPKNAVTAALKQVVAHIVHPYPSEDEKEVLCAGHGLDLDLQINNWFINARVRIWKPPL
AtBLR      --GFPDHHAPVWRPHRGLEPRAVTVLRAWLFDHFLHPYPTDTDKLMIAKQGLSRNQVSNWFINARVRIWKPPL
HupSel|KNOX1-1  --GSLKQEFMRKKKKKGKLPRESRQQLLNWVSSHIKWPYPSEVEKASLAESTGLDQKQINNWFINQRKRHWKPS
HupSel|KNOX1-2  --GSLKQEFMKKKKKKGKLPESRQQLLDWVSNHIKWPYPSEAEKASLAESTGLGQKQINNWFINQRKRHWKPA
HupSel|KNOX2-3  --NDVREEILRKRRAKGLPGDITSVLKSWMHAHASKWPYPSEEEKVKLVQETGLELQKQINNWFINQRKRWNHNN
HupSel|KNOX2-2  --NDVREEILRKRRAKGLPGDITSVLKAWVHAHASKWPYPSEDEKARLVQETGLELQKQINNWFINQRKRWNHNS
HupSel|KNOX2-1  --ADVREEILRKRRAKGLPGDITSVLKAWVHAHASKWPYPTEDEKTRLVQETGLELQKQINNWFINQRKRWNHQH
SelApo|LGDQ-2003456  --KGLTQEYLKKKKKKGKLPKESRQQLLDWVQAQHQHWPYPNENOKASLAOCTGLEPKQINNWFINQRKRWNHPQ
SelApo|LGDQ-2010944  --VDVREEILRKRRAKGLPGDITSVLKAWVHAHASKWPYPTEDEKARLVQETGLELQKQINNWFINQRKRWNHHH
SelApo|LGDQ-2004042  --NDVREEILRKRRAKGLPGDITSVLKTWVHAHASKWPYPSEDDKARLVQETGLELQKQINNWFINQRKRWNHNS
LycDer|PQTO-2010435  --GSLKQEFMRKKKKKGKLPRESRQQLLNWVSMHIKWPYPSEVEKASLAESTGLDQKQINNWFINQRKRHWKPS
LycDer|PQTO-2081329  --GSLKQEFMKKKKKKGKLPESRQQLLDWVSGHIKWPYPSEAEKASLAESTGLGQKQINNWFINQRKRHWKPA
LycDer|PQTO-2010278  --NDVREEILRKRRAKGLPGDITSVLKSWMFAHASKWPYPSEEEKARLVQETGLELQKQINNWFINQRKRWNHNN
LycDer|PQTO-2081967  --ADVREEILRKRRAKGLPGDITSVLKAWVQAHASKWPYPTEDEKARLVQETGLEMKQINNWFINQRKRWNHHH
LycDer|PQTO-2083351  --NDVREEILRKRRAKGLPGDITSVLKAWVHAHAKWPYPSEDEKARLVQETGLELQKQINNWFINQRKRWNHTN
SelKra|gb|AAW62517.1  --SSLKHEFLKKKKKKGKLPKDSRQQLLNWVSVHYKWPYPSESEKASLAESTGLDQKQINNWFINQRKRHWKPS
SelKra|gb|AAW62519.1  --VDVREEILRKRRAKGLPGDITSVLKAWVHAHASKWPYPTEDEKARLVQETGLELQKQINNWFINQRKRWNHHH
SelKra|ZFGK-2006217  --NDVREEILRKRRAKGLPGDITSVLKSWMHAHASKWPYPSEDDKARLVQETGLELQKQINNWFINQRKRWNHNS
SelMoe|415291  --KGLTQEYLKKKKKKGKLPKESRQQLLDWVQAQHQHWPYPNENOKSNLAOSTGLDPKQINNWFINQRKRWNHPQ
SelMoe|90744  --NDVREEILRKRRAKGLPGDITSVLKTWVHAHASKWPYPSEDDKARLVQETGLELQKQINNWFINQRKRWNHNS
IsoTeg|PKOX-2100696  --NDVREEILRKRRAKGLPGDITSVLKSWMHAHASKWPYPSEDEKARLVQETGLELQKQINNWFINQRKRWNHNS
IsoTeg|PKOX-2013942  --VDVREEILRKRRAKGLPGDITSVLKAWVHAHAKWPYPTEDEKARLVQETGLELQKQINNWFINQRKRWNHHH
IsoTeg|PKOX-2100696  --NDVREEILRKRRAKGLPGDITSVLKSWMHAHASKWPYPSEDEKARLVQETGLELQKQINNWFINQRKRWNHNS
                                     *      *      *      *      *      *      *
                                     *      *      R30      *      *      L53      *

```

- \* R (strong positive charge) in this position improves trafficking
- \* L (hydrophobic) in this position improves trafficking

Based on Xu et al. (2011), trafficking of KNOXes through PD requires the chaperonin CCT8. A comparison of the CCT8 amino acid sequences from Arabidopsis (GenBank accession Q94K05.1) and *H. selago* shoot tips shows strong conservation.

```

HsCCT8 1  MSSFHLQPEHGLQSMLEKGGHRLLSGLDEAILKNIIDACKELSNIITRTSLGPNGMNKMIIINHL 60
AtCCT8 1  MVGMSMOPYGIQSMLEKGYRRLLSGLDEAVIKNIIDACKELSNIITRTSLGPNGMNKMVINHL 60

HsCCT8 61  DRLFVTS DAGTIVNELELVQHPAAKLLVLAASKAQQEEIGDGANLVTISFAGELLQCAEELIR 120
AtCCT8 61  DKLFVTNDAAITIVNELEIQHPAAKLLVLAASKAQQEEIGDGANLVTISFAGELLQCAEELIR 120

HsCCT8 121 TGLHPSEIISGYTKATKALTLEELVEPSSETMDVRSKEQVIQRMKSVVASKQYQEDI 180
AtCCT8 121 MGLHPSEIISGYTKAVSKAVLELLEQLVETGSETMDVRNKDEVIQSRMRAAVASKQFGQEEI 180

HsCCT8 181  ITPLIAEACIQVCPKNPANFNVDNVRVAKILGGGIHDC HVVIRGMVLKVEALGSIKHKVKN 240
AtCCT8 181  ICSLVTDACIQVCPKNPTNFNVDNVRVSKLLGGGLHNSCIVIRGMVLKSDAVGSIKRMKA 240

HsCCT8 241  KIAVFGGGVDTAATETKGTVLIKTAEQLENYAKTEEAKIEELIKAVASSGANVIVSGGAV 300
AtCCT8 241  KVAVFAGGVDTTATETKGTVLIHSAEQLENYAKTEEAKVEELIKAVAESGAKVIVSGGSI 300

HsCCT8 301  GEMALHFCERYKLMVLKISSKFELRRFCRTTGATSLVKLGHPGPELGHADSISVDEIGG 360
AtCCT8 301  GEMALHFCERYKIMVLKISSKFELRRFCRTAGAVAHLLKLSRPSPELGLGYVDSISVDEIGG 360

HsCCT8 361  TRVTIVKNEESGNLVSSVIIRGSTDSVLDDVERAVDDGVNVYKAMCKDSRIIPGAGASEV 420
AtCCT8 361  VTVTIARNEEGGNSISTVVLRGSTDSILDLLERAVDDGVNTYKAMCRDSRIVPGAATEI 420

HsCCT8 421  ELAKKINEFGDKETGLDQYAIKFAQSLEVPRTLAENAGLNSTDVISTLYAAHAGNVK 480
AtCCT8 421  ELAQRLEKYANAETGLDKYAITKYAESFEFVPKTLADNAGLNAMETIAALYTGHGSGNTK 480

HsCCT8 481  AGVDVEQGGFKDMTMENVVDLEFVTKYWAIKLAARAVCTVLQVDQIIMAKQAGGPNKKDRP 540
AtCCT8 481  LGIDLEEGACKDVSSETKVVWDLFATKLFALKYASDAACTVLRVDQIIMAKPAGGPRRDAAQ 540

HsCCT8 541  MDG---DDE 546
AtCCT8 541  AAGAGAEED 549

```

The KNOX homeobox domain is necessary and sufficient for trafficking of class I KNOX proteins through PD (Kim et al., 2005), a process which requires the binding of a chaperonin (Xu et al., 2011). Chen et al. (2014) used targeted mutagenesis to find out which amino acids were required for trafficking, and found that a positively charged amino acid at the end of the alpha-1 domain (R30) and a hydrophobic residue at the beginning of the alpha-3 domain (L53) played important roles. Comparison with the lycophyte KNOX protein sequences used in Figure 4 showed that none of the lycophyte KNOXes contained the R30. Furthermore, all class I KNOXes of lycophytes did not contain the L53, while most class II KNOXes did (the only exception, from *Huperzia deuterodensum*, contained a methionine in that position). The analyses of Chen et al. (2014) were based on trafficking of maize KN1 through secondary PD separating clonally

non-related epidermis and mesophyll cells in Arabidopsis leaves. It is possible that trafficking through primary PD has different structural requirements; in this case, class I KNOX proteins might act mostly cell-autonomously in SAMs of Lycopodiales but not of Selaginellales. It is also possible that the requirements do not differ between both types of PD and class I KNOX proteins act mostly cell-autonomously in all lycophytes.

## References

- Chen H, Jackson D, Kim JY. Identification of evolutionarily conserved amino acid residues in homeodomain of KNOX proteins for intercellular trafficking. *Plant Signal Behav.* 2014;9:e28355.
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl Acids Res.* 2004;32:1792-1797.
- Kim J-Y, Rim Y, Wang J, Jackson D. A novel cell-to-cell trafficking assay indicates that the KNOX homeodomain is necessary and sufficient for intercellular protein and mRNA trafficking. *Genes Dev.* 2005;19:788–93.
- Nicholas KB, Nicholas HB Jr, Deerfield DW II. GeneDoc: Analysis and visualization of genetic variation. *EMBNEW.NEWS.* 1997;4:14.
- Xu C, Jiao C, Zheng Y, Sun H, Liu W, Cai X, Wang X, Liu S, Xu Y, Mou B, Dai S, Fei Z, Wang Q. *De novo* and comparative transcriptome analysis of cultivated and wild spinach. *Sci Rep.* 2015;5:17706.

# ARP

Sequence comparison between

**(A)** the ARP proteins of *Arabidopsis* (AtAS1, GenBank accession O80931.1) and *Selaginella kraussiana* (SkARP, GenBank accession AAW62520.1) and **(B)** the corresponding cDNA sequences (AtAS1, GenBank accession AF175996.1; SkARP, GenBank accession AY667452.1). The N-terminal part shows strong conservation.

**(C)** Sequence alignment of ARP proteins from angiosperms, *Selaginella kraussiana* and *Pinus taeda*. *Amborella*: GenBank XP\_006855819.1 from *Amborella trichopoda*; *Pisum*: GenBank AF299140\_1 from *Pisum sativum*; *Arabidopsis*: GenBank NP\_181299.1 from *Arabidopsis thaliana*; *Zea mays*: GenBank AF126489\_1; *Selaginella*: XP\_002962315.1 from *Selaginella moellendorffii*; *Pinus taeda*: Congenie PITA\_000017758).

Sequence alignments were performed using MUSCLE (multiple sequence comparison by log-expectation; Edgar, 2004) at the EMBL-EBI website, and displayed using GeneDoc (Nicholas et al. 1997).

## (A)

<b>AtAS1</b>	1	MKERQRWSGEEDALLRAYVRFGRWHLVSERMKNPLNRDAKSLERWKNYLKPGIKKG	60
<b>SkARP</b>	1	MKDQRWQPEEDALLCAYVKQYGFNDNVLVSERMAPLDRDPKSCHERWKNYLKPGIKRG	60

\* \* \* \* \*

<b>AtAS1</b>	61	SLTEEEQRLVIRLQEKHGKWKKIAAEVPGRTAKRLGKWWEVHEKQOREEKESNKRVPEP	120
<b>SkARP</b>	61	FLSEEEQNLVIRLQEKYGNKWKRIA AEVPGRTAKRLGKWWEVHEKERROKEAIQRHORIQ-	119

\* \* \* \* \*

<b>AtAS1</b>	121	IDESKYDRILESFAEKLVKERSNVVPA AAAAAATVVMANSNGGFLHSEQQVQ-PPNPVIPP	179
<b>SkARP</b>	120	--TGVHTSHLSMRYGQTV---APFIPPAQSFS TCAEVVS SSSSASEGESQCRNEPRMNLPA	174

\* \* \* \* \*

<b>AtAS1</b>	180	WLATSNNGNVVARPPSVTL--TLPSTVAAAAPQPFPI-----FWLQQQPERAEN	228
<b>SkARP</b>	175	AFPPTSSPEVTLTGPTVLDLPAWKPAAPRAASTSELPSLMAPEAIMKPNLSLSLDSGAES	234

\* \* \* \* \*

<b>AtAS1</b>	229	GP-----GGLVVLGSMMPS----CSGSSES VFLSELVE---CCRELEEGHRAWADHKK	273
<b>SkARP</b>	235	CDTDTGTHFNKNKVVSTIIPKDFECNEINS DTPGELIPLLGLVKELEENKESWNVQKK	294

\* \* \* \* \*

<b>AtAS1</b>	274	EAAWRRLRRLLEIQLSEKTCRQREKMEIEIAKMKALREEQKNAMEKIEGEYREQLVGILRRD	333
<b>SkARP</b>	295	NAASTLRELKQOLECERIEK RKQKMLEVESKIQALRKEEKLYLDKLELDYAELVAKLDRD	354

\* \* \* \* \*

<b>AtAS1</b>	334	AEAKDQKLADQWTSRHIRLTKFLEQQMG-----CRLDRP-	367
<b>SkARP</b>	355	AELKEEKLIVESWSLKYKLVLMFEQTMORYSSFHGPIFOAIQMRGMSPA	404

## (B)

<b>AtAS1</b>	1	-----ATGAAAGA	8
<b>SkARP</b>	661	CTTTGCTCATTCTACACAAAGTATTTGATCAGAAAAGTTGTATAATAAAGCGATGAAGGA	720

\* \* \* \* \*

<b>AtAS1</b>	9	GAGACAACTGGAGTGGTGAAGAAGATGCATGTGTACGTGCTTACGTTAGACAGTTCTGG	68
<b>SkARP</b>	721	CAAGCAGCGTTGGCAGCCTGAAGAGGACGCACTCCTCTGTGCTACGTGAAGCAATACGG	780

\* \* \* \* \*

<b>AtAS1</b>	69	TCCGAGAGAACTGGCATCTTGCTCTGAGCGTATGAACAAACCTTTGAACCGTGACGCCAA	128
<b>SkARP</b>	781	CCCAAATGACTGGAACCTTAGTCTCTGAAAAGGATGGCAACGCCTTTGGATCGTGATCCAAA	840

\* \* \* \* \*

<b>AtAS1</b>	129	GTCTTGTTTGAGAGATGGAAGAATATCTTAAGCCAGGGATCAAGAAAGGGTCTTTGAC	188
<b>SkARP</b>	841	GTCTTGCCATGAACGCTGGAAGAACTATCTTAAGCCGGGACTCAAGAGGGGGCCATTGTC	900

<i>AtAS1</i>	189	AGAGGAAGAGCAGAGGCTTGTGATCCGTCTTCAAGAGAAACACGGCAACAAGTGGAAAGAA	248
<i>SkARP</i>	901	TGAAGAGGGAGCAGAACCTCTGTGATACGTCTTCAAGAAAAATACGGTAATAAATGGAAGCG	960
<i>AtAS1</i>	249	GATTGCTGCTGAGGTTCCCGGAGGACGGCAAAACGGTTAGGSAAGTGGTGGGAACTGTT	308
<i>SkARP</i>	961	TATAGCGGCTGAAGTTCCCGGCCGGACGGCCAAACGTCTGGGAAAATGGTGGGAGCTCCA	1020
<i>AtAS1</i>	309	TAAGGAGAAGCAACAGAGAGAAGAGAAAGAGACTAACAAAGAGACTTGAGCCTATTGACGA	368
<i>SkARP</i>	1021	CAAAGAACGCAGGCAGAAAGAAGCGATTCCAGCGCATCAGCGGATCCA-----AACGG	1073
<i>AtAS1</i>	369	GAGTAAGTACGATCGGATCTCTCGAGAGTTTCGCTGAGAAGCTTGCAAAAGACCGGTCTAA	428
<i>SkARP</i>	1074	GAGTTCATACCTCTCACCTCTCGATGTTCT-----ACGGCCAAA	1112
<i>AtAS1</i>	429	C-GTTGTCCTCTGCTGCTGCCGCTGCTGCAACCGTT-----GTGATGGCTAATTCG----	477
<i>SkARP</i>	1113	CTGTTGCTCTCTTCATTCCTCCAGCT-CAAAGCTTCTCTACGTG-TGCCGAAGTCGTTTC	1170
<i>AtAS1</i>	478	-----AATGGAGGGTTTTTACATCTGAAACAACAAGTTCAGCCTC-----C	518
<i>SkARP</i>	1171	ATCATCGAGTGCATCGGAAGGTG-----AGTCTCAATGTCCGGAACGAGCCTCGGATGAAC	1225
<i>AtAS1</i>	519	TAACC--CAGTGCATCCCGCC--TTGGTT--AGCTACTTCTAACAAATGGGAACAATGTTG	571
<i>SkARP</i>	1226	TTACCGGCAGCGTTTCCACCAGCTAGTTCTGAGCCGGTTCTTACATGGGTCCGACAGTG	1285
<i>AtAS1</i>	572	TTGCAAGGCCCTCCCTCGGTAACCTTGACATT---ATCGCCTTCCACAGTGGCTGCAGCTG	628
<i>SkARP</i>	1286	TTG---GACCTTCTTC-----CGGCATGGAAACCGGCACCGAGAGCAGCCTCGACT-	1333
<i>AtAS1</i>	629	CGCTCAACCGCCAATCCCG-TGGCTGCAGCAGCAAC-----AGCCTGA-----	671
<i>SkARP</i>	1334	--TCTGAACTGCCAAGCCTGATGGCTCCAGAGGGGATCATGAGCCGAACCTTTCGTTAT	1391
<i>AtAS1</i>	672	-----GAGAGCAGAGAACGG---TCCAG---GGGACTTGTGTT-----	704
<i>SkARP</i>	1392	CTCTGGACTCGGCTGCTGAGTCCGGAGATACAGATACAGGTACTCATTTCACAATAATA	1451
<i>AtAS1</i>	705	AGGGAGTAT-----GATGCC--TCTTGTAGTGGAGTAGCGAGAGTG	745
<i>SkARP</i>	1452	AGAAGGTATCTACGATAATACCTAAGGACGACCGAATTTTGCAATGAAATCAACTCGGATA	1511
<i>AtAS1</i>	746	TGTTCTTGTGAGAGCTTGTG-----GAGTGTGTAGAGACTTGGAGGAAGGCACC	796
<i>SkARP</i>	1512	TATCACCTGAGAGCTGATACCTCTACTTGGGCTTGTCAAGGAGCTCGAGGAGAACAAGG	1571
<i>AtAS1</i>	797	GAGCTTGGGCAGACCATAAGAAAGAGGCTGCATGGAGGCTAAGAAGCTGGAGCTGCAGC	856
<i>SkARP</i>	1572	AAAGTTGGAATGTGCAGAAGAAGAAATCGGCATCCACACTGAGGAACTGAAGCAGCAGC	1631
<i>AtAS1</i>	857	TAGAGTCAGAGAAGACGTGTAGACAAAGGGAGAAGATGAGGAGATTGAGGCAAAGATGA	916
<i>SkARP</i>	1632	TGGAGTGTGAGAGAATTGAAAAGAGGAAGCAGAAGATGTTAGAGTGGAGTCAAAGATTC	1691
<i>AtAS1</i>	917	AAGCTCTTAGGAAAGAGCAGAAAGAACGCAATGGAGAAGATCGAAGGAGAGTACAGAGAAC	976
<i>SkARP</i>	1692	AAGCTCTAAGAAAGGAGGAAAAGTTGTATTGGACAAGCTTGAAGCTGATTATGCTGAAC	1751
<i>AtAS1</i>	977	AGCTCGTTGGTTTGAGGCAGACGCAGAGGCCAAAAGACCAGAACTGGCTGATCAATGGA	1036
<i>SkARP</i>	1762	TTGTTGCAAAGCTGGACAGAGATGCTGAACTTAAGGAAGAAAAGCTTGTGGAAAGCTGGA	1811
<i>AtAS1</i>	1037	CCTCTAGGCATATCAGACTCACCAAGTTCTTGAACAACAATG-----	1080
<i>SkARP</i>	1812	GCTTGAAGTACAACAACACTGGTTCTTATGTTTGAAGCAGACAATCCAACGGTATTCGTCT	1871



(C)

*Selaginella* 1 MSIATKDRQRWQPEEDAILCAYVTQYGADDWNLISERMGEFLDRDPKSCHERWKNYLKPG 60  
*Amborella* 1 ----MKERQRWQPEEDALLRAYVKQYGPKEWNLVSQRMGRFLHRDAKSCLERWKNYLKPG 56  
*Pinus taeda* 1 ----MKERQRWQPEEDALLRAYVKQYGPKEWNLVSQRMGKTLDRDAKSCLERWKNYLKPG 56  
*Zea mays* 1 ----MKERQRWRPEEDAVLRAYVRYQYGPKEWHLVSQRMNVALDRDAKSCLERWKNYLKPG 56  
*Pisum* 1 MSLEMKDRQRWRAEEDALLRAYVKQYGPKEWNLVSQRMNTFLNRDAKSCLERWKNYLKPG 60  
*Arabidopsis* 1 ----MKERQRWQSGEEDALLRAYVRFQYGPKEWHLVSERMNKEFLNRDAKSCLERWKNYLKPG 56

*Selaginella* 61 IKKGPLTDEEQQLVIRLQTKYGNKWKRIAAEVPGRGTAKRLLGKWWVEVYKERLTKDKKLLS 120  
*Amborella* 57 IKKGSLSQEEQALVVVALQAKYGNKWKKIAAEVPGRGTAKRLLGKWWVEVFKKQAKDKQRRLLQ 116  
*Pinus taeda* 57 IKKGSLSLDEEQSLVIRLQAKYGNKWKKIAAEVPGRGTAKRLLGKWWVEVFKKQKLERK--E 114  
*Zea mays* 57 IKKGSLSLDEEQQLVIRLQAKYGNKWKKIAAEVPGRGTAKRLLGKWWVEVFKKQKQRELRD--S 114  
*Pisum* 61 IKKGSLSLDEEQHLVIRLQATHGNKWKKIAAQVPGRTAKRLLGKWWVEVFKKQKQRETKG--- 117  
*Arabidopsis* 57 IKKGSLSLDEEQQLVIRLQEKYGNKWKKIAAEVPGRGTAKRLLGKWWVEVFKKQKQREEKE--- 113

*Selaginella* 121 TH-----AATGNCDMSVMEAMHLQALAPGFSRPFSSST----- 153  
*Amborella* 117 QNHSHHHHHHQHSSSISSSQPPVGLSSDSSVPSGRYDHILETFAEKYAQQAQKSCCLPP 176  
*Pinus taeda* 115 RH-----NIDSANCREGGKYDHILETFAEKYVQKKICPTSLP 151  
*Zea mays* 114 RR-----PPEPSPDERGRYEWLLENFAEKLWGER----- 144  
*Pisum* 118 -----INKTVDPINDSKYEHILETFAEKLWKER----- 145  
*Arabidopsis* 113 -----SNKRVEPIDESKYDRILETFAEKLWKER----- 141

*Selaginella* 154 P--DLCVNGAPAFSTSTPDAND-----ICGPPT--VC----- 181  
*Amborella* 177 PPPPILLSDPPTLLSLNSAGTTTQRAPAATETPALLPS----- 214  
*Pinus taeda* 152 PILMPCPSSSPLLTLSVPTGTYADNICLANIKTTPSVLSGPIVSQMANLPGSMGFTLPV 211  
*Zea mays* 145 P--QAAAAAPSPLLMAA-----PV----- 161  
*Pisum* 146 P-----SPSFVMAASNSS-YLHTDAQAATPG---L----- 171  
*Arabidopsis* 142 SNVVPAAAAAATVVMANSNGG-FLHSEQVQPPN--PV----- 176

*Selaginella* 181 ----- 181  
*Amborella* 214 ----- 214  
*Pinus taeda* 212 IKGKMPAREIEPRLTLESIGSTRPMLEPELALQCSSSAMDLDLTPTYLGTSSSAVDATS 271  
*Zea mays* 161 ----- 161  
*Pisum* 171 ----- 171  
*Arabidopsis* 176 ----- 176

*Selaginella* 182 -QDHFIGN-----NASGVTTTEALS-----GDETLTLTLPSTTF 213  
*Amborella* 215 -LPPWLS-----AAGGLKNFGAVPSVSPPPPPPP-----PPSVLSLSPTS-- 255  
*Pinus taeda* 172 SLPSWMSNVPTAQVPSGGSDVKTTSVAFILEKDTPLSSKSIHLNSSPSVLSLSPSASD 331  
*Zea mays* 162 -LPPWLS-----SNAGPAAAAAAVAHPPPRPP-----SPSVTLASAA-- 200  
*Pisum* 172 -LPSWLS-----NSNNTA-----PVRPN-----SPSVTLSPST-- 200  
*Arabidopsis* 177 -LPPWLAT-----SNNNNVVAR-----PPSVTLTSPST-- 205

*Selaginella* 214 QKQFPMEVAVATSGVTMK-WI----PKKLEIQIASTLTSSLSLSSSRTADEGL----- 262  
*Amborella* 256 -----VAT----- 258  
*Pinus taeda* 332 PVSSSGASAPASEVIAFRSWLQDTMRHHENINGVAGGVAKKVNFPWCALQKQVGTSSF 391  
*Zea mays* 201 -----VAP-GPPAPAPWM-----PDRAAADAAPYGFPSPSQH----- 231  
*Pisum* 201 -----VAA-----EPPWM-----QPVRGPDNA-----P--LVL----- 221  
*Arabidopsis* 206 -----VAAAAPQPPIPWLQQQPERAENG-----PGGLVL----- 235

*Selaginella* 263 -DCLDFASESSDNNEGSNLS-MFELFKELREORENWIQQKKGISSKLEKELKQOLECEKAE 320  
*Amborella* 259 -ECTMQAAASGV-----VQQLAQWCREVEEGRQAWVQKKEAAWRVRRVEQOLESEKSR 311  
*Pinus taeda* 392 DPNLASSKQAMDSLIMQQLPTFLQYCKDLEEGRQSWFMHKKKATWRLSRLEQOLESEKAR 451  
*Zea mays* 232 -GGAAPPMAVVDGQA--LAELAECRELEEGRRWAAHRRAAWRLKRVEQOLEMEREM 288  
*Pisum* 222 -GNVAPHGAVLSYGENMVMSELIDCKELEEGHHAALAAHKKKAAWRLSRVLEQOLESEKAS 280  
*Arabidopsis* 236 -GSMMPSCSGSSESVF--LSELVECCRELEEGHRAWADHKKAAWRLRLELEQOLESEKTC 292

```

          *           *           *           *           *           *
Selaginella 321 KQRQKIQEVDAARVKALKKEKKQFLQKVEQDYSSELVSNLEERDAEMKEKKLTFEAWTLKYDKL 380
Amborella    312 KRREKMEEVEAKIRRLREEEAFAVDRLBADCREQLAAVQRDAEMKEVKLMEQWAAKHLKL 371
Pinus        452 KRREKIEEVGSKIRALREEEITYLDKLETECREQLSSLQRDAEMKEAKMELWATKHLQL 511
Zea          289 RRREVWEEFEAKMRTMRLEQAAAAERVERDHRREKVAELRRDAQVKEEKMAEQWAAKHARV 348
Pisum        281 RRREKMEEIEAKIKALREEQAVALDRIEGEYREQLAGLRRDAEAKQKLAEQWAAKHLRL 340
Arabidopsis  293 RQREKMEEIEAKMKALREEQKNAMEKIEGEYREQLVGLRRDAEAKDQKLADQWTSRHIRL 352

```

```

          *           *
Selaginella 381 VHTYERYTLSGPFVFAE----- 397
Amborella    372 ATFVEHLSSQHPSSSCFLKDSL 394
Pinus taeda  512 TKFVEQMLYQFPDAQRLFSKEMH 534
Zea mays     349 AKFVEQMGGCSRSWSSATDMNC- 370
Pisum        341 TKFLE-QVGCRSRHAEQNGR--- 359
Arabidopsis  353 TKFLEQQMGCRLLDRP----- 367

```

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792-1797.

Nicholas KB, Nicholas HB Jr, Deerfield DW II (1997) GeneDoc: Analysis and visualization of genetic variation. *EMBNW.NEWS* 4:14.

# YABBY

**(A) Intron exon structure of YABBY genomic sequence.** The zinc finger domain is indicated in green, the myb domain in red. The size bar denotes 100 bp. **(B) Gene tree of YABBY proteins.** As there is no known ortholog in more early diverging plants, *H. selago* YABBY is used as an outgroup. Bootstrap support is only shown for values >60.

**A**

