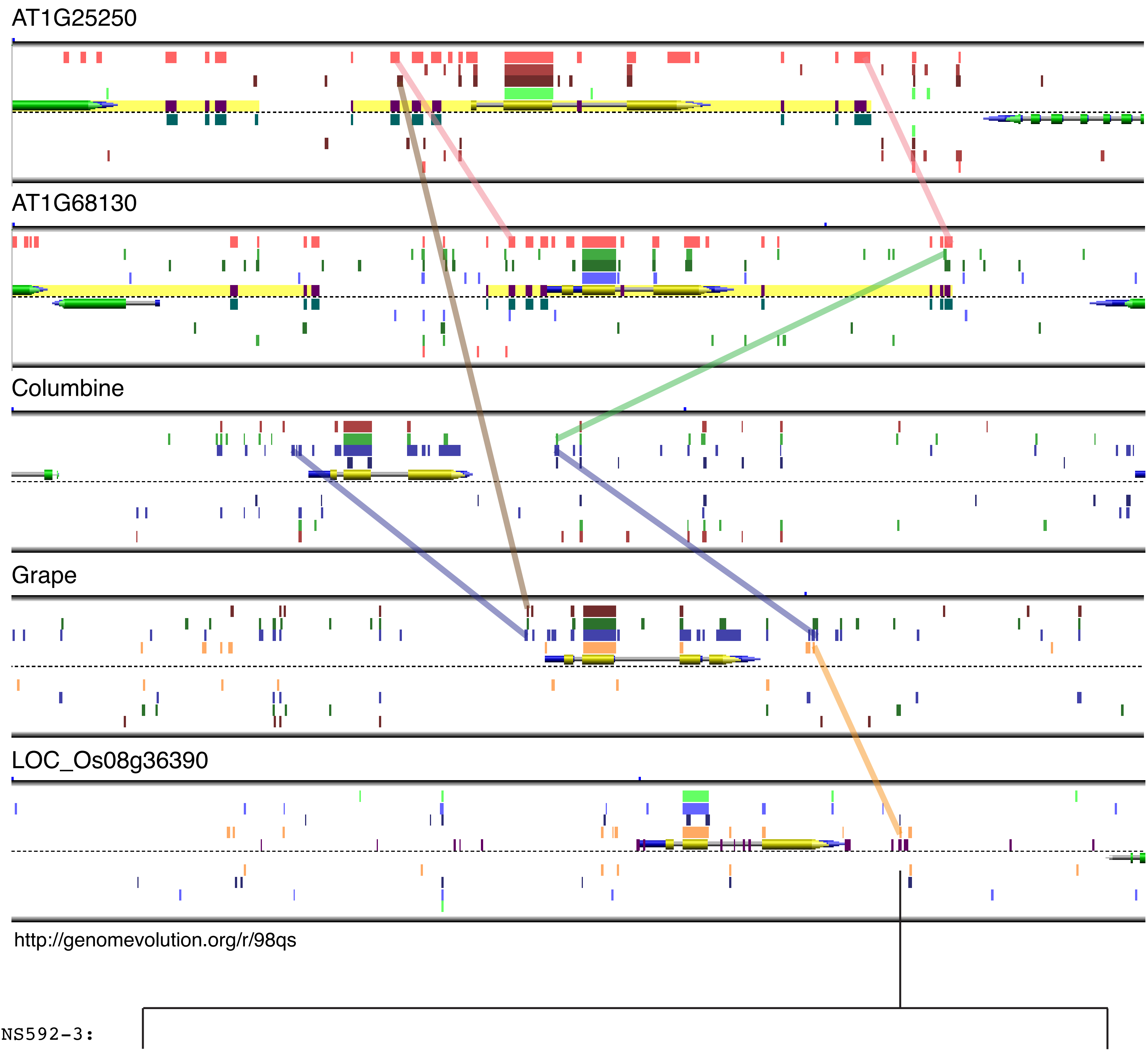
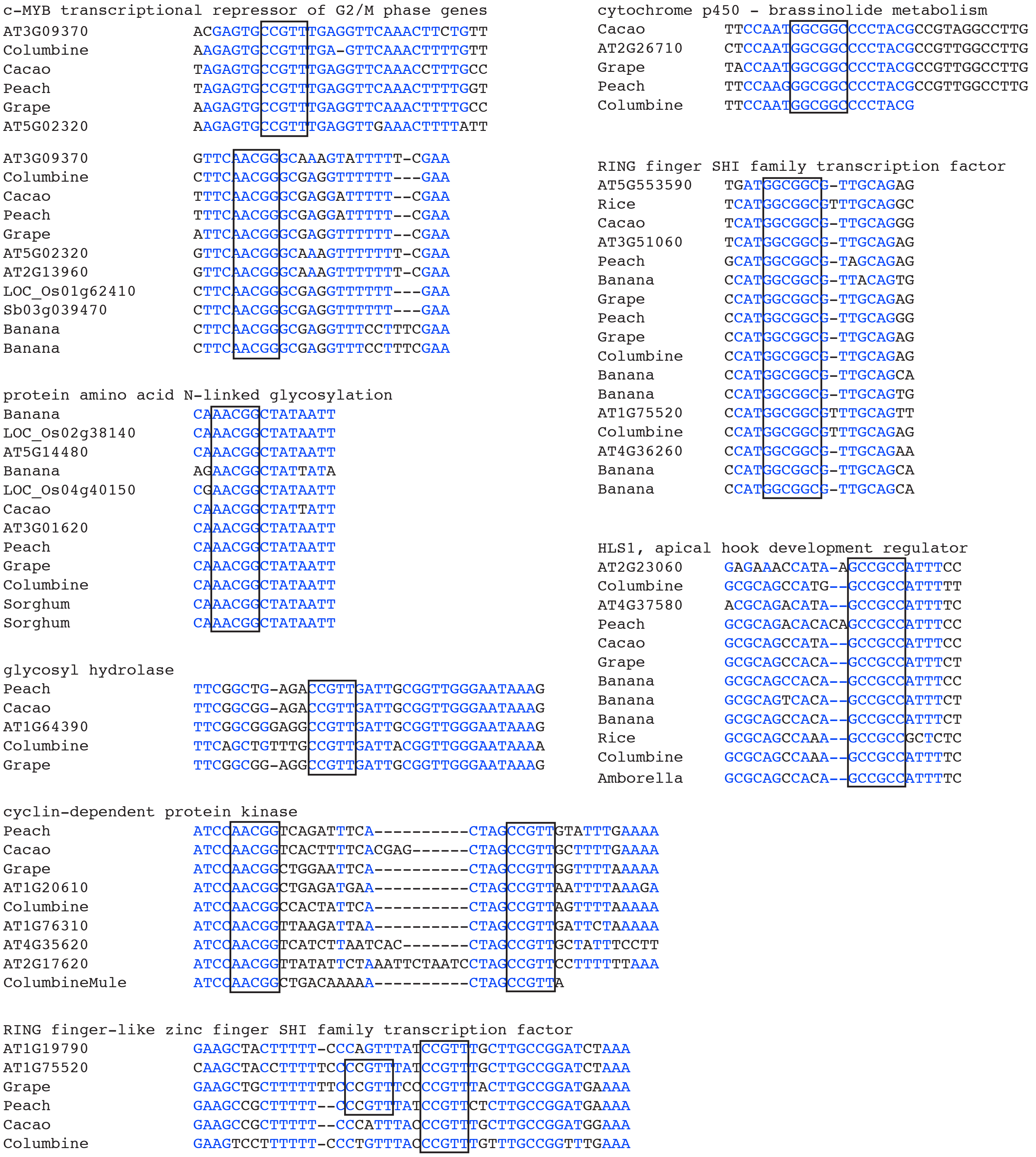


```
CNS592-3:

AT1G25250                    AACGACATGAATTTCAGTGCCAAAAAGATTGAAATGGTGGTTAGGAGTCAGGCTTCAGTCTCTCTCAC
AT1G68130                   AAACGCCATGGATTTCAGTGCCAAA--GTTTGAAA---TGGTTAGGAGTCAGGCATCAGTCTCTCTCAC
Columbine      TAAATGCGTATAAAAACACCATGGTTTTCAGTGCCATG---ATTCTAA
Grape          TAAATGTTTATAAAAACACCATGTTTTTCAGTGCCACA---GTTGAAA---TGGTTAAGA--CAGGCACTCGTCTCTCTCAA
LOC_Os08g36390                    CATGGTTTTCAGTGCCACA--G
```

**Supplemental Figure 1. Use of GEvo panels to detect deeply conserved CNSs**. This panel (http://genomevolution.org/r/98qs) shows two deep CNSs detected between orthologous *Arabidopsis* and columbine zinc finger, C2H2-type transcription factors. Orthologous genes are shown in yellow, with 5'- and 3'-UTR regions colored blue. Dark green rectangles below and purple rectangles adjacent to the *Arabidopsis* gene structure depict pre-annotated α-CNSs (Thomas et al., 2007; Turco et al., 2013). Colored rectangles above the gene structure represent blastn high scoring segment pairs (HSPs) between genomic regions, with HSPs in each panel organized from top to bottom as blastn hits to the corresponding panels. Regions shaded in yellow delineate the gene space, the CNS-associated region surrounding each *Arabidopsis* homologous gene pair. Only two of the α-CNSs shown in this figure are conserved in other eudicots. 592-1, the 5' proximal CNS, is conserved in all dicots examined including columbine. Although no direct blastn hit can be seen between *Arabidopsis* and columbine, HSPs between *Arabidopsis* and grape are visually connected to HSPs between grape and columbine. The second CNS, 592-3, is also conserved in rice and illustrates that not all CNSs are located 5' to their associated gene. The sequences retrieved for the 3' CNS from overlapping HSPs are shown below. The deep CNS sequence is in blue, which is the aligned region that is conserved between columbine and *Arabidopsis*. In this example, 15 out of 17 deep CNS positions are conserved.



A

```
c-MYB transcriptional repressor of G2/M phase genes
AT3G09370            ACGAGTGCCGTTTGAGGTTCAAACTTCTGTT
Columbine            AAGAGTGCCGTTTGA-GTTCAAACTTTTGTT
Cacao                TAGAGTGCCGTTTGAGGTTCAAACCTTTGCC
Peach                TAGAGTGCCGTTTGAGGTTCAAACTTTTGGT
Grape                AAGAGTGCCGTTTGAGGTTCAAACTTTTGCC
AT5G02320            AAGAGTGCCGTTTGAGGTTGAAACTTTTATT

AT3G09370            GTTCAACGGGCAAAGTATTTTT-CGAA
Columbine            CTTCAACGGGCGAGGTTTTTT---GAA
Cacao                TTTCAACGGGCGAGGATTTTT--CGAA
Peach                TTTCAACGGGCGAGGATTTTT--CGAA
Grape                ATTCAACGGGCGAGGTTTTTT--CGAA
AT5G02320            GTTCAACGGGCAAAGTTTTTTT-CGAA
AT2G13960            GTTCAACGGGCAAAGTTTTTTT-CGAA
LOC_Os01g62410       CTTCAACGGGCGAGGTTTTTT---GAA
Sb03g039470          CTTCAACGGGCGAGGTTTTTT---GAA
Banana               CTTCAACGGGCGAGGTTTCCTTTCGAA
Banana               CTTCAACGGGCGAGGTTTCCTTTCGAA

protein amino acid N-linked glycosylation
Banana               CAAACGGCTATAATT
LOC_Os02g38140       CAAACGGCTATAATT
AT5G14480            CAAACGGCTATAATT
Banana               AGAACGGCTATTATA
LOC_Os04g40150       CGAACGGCTATAATT
Cacao                CAAACGGCTATTATT
AT3G01620            CAAACGGCTATAATT
Peach                CAAACGGCTATAATT
Grape                CAAACGGCTATAATT
Columbine            CAAACGGCTATAATT
Sorghum              CAAACGGCTATAATT
Sorghum              CAAACGGCTATAATT

glycosyl hydrolase
Peach                TTCGGCTG-AGACCGTTGATTGCGGTTGGGAATAAAG
Cacao                TTCGGCGG-AGACCGTTGATTGCGGTTGGGAATAAAG
AT1G64390            TTCGGCGGGAGGCCGTTGATTGCGGTTGGGAATAAAG
Columbine            TTCAGCTGTTTGCCGTTGATTACGGTTGGGAATAAAA
Grape                TTCGGCGG-AGGCCGTTGATTGCGGTTGGGAATAAAG

cyclin-dependent protein kinase
Peach                ATCCAACGGTCAGATTTCA----------CTAGCCGTTGTATTTGAAAA
Cacao                ATCCAACGGTCACTTTTCACGAG------CTAGCCGTTGCTTTTGAAAA
Grape                ATCCAACGGCTGGAATTCA----------CTAGCCGTTGGTTTTAAAAA
AT1G20610            ATCCAACGGCTGAGATGAA----------CTAGCCGTTAATTTTAAAGA
Columbine            ATCCAACGGCCACTATTCA----------CTAGCCGTTAGTTTTAAAAA
AT1G76310            ATCCAACGGTTAAGATTAA----------CTAGCCGTTGATTCTAAAAA
AT4G35620            ATCCAACGGTCATCTTAATCAC-------CTAGCCGTTGCTATTTCCTT
AT2G17620            ATCCAACGGTTATATTCTAAATTCTAATCCTAGCCGTTCCTTTTTTAAA
ColumbineMule        ATCCAACGGCTGACAAAAA----------CTAGCCGTTA

RING finger-like zinc finger SHI family transcription factor
AT1G19790            GAAGCTACTTTTT-CCCAGTTTATCCGTTTGCTTGCCGGATCTAAA
AT1G75520            CAAGCTACCTTTTTCCCCGTTTATCCGTTTGCTTGCCGGATCTAAA
Grape                GAAGCTGCTTTTTTTCCCGTTTCCCCGTTTACTTGCCGGATGAAAA
Peach                GAAGCCGCTTTTT--CCCGTTTATCCGTTCTCTTGCCGGATGAAAA
Cacao                GAAGCCGCTTTTT--CCCATTTACCCGTTTGCTTGCCGGATGGAAA
Columbine            GAAGTCCTTTTTT-CCCTGTTTACCCGTTTGTTTGCCGGTTTGAAA
```

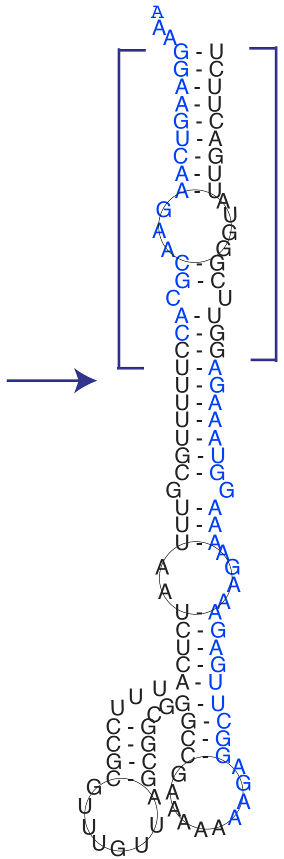
B

```
cytochrome p450 - brassinolide metabolism
Cacao          TTCCAATGGCGGCCCCTACGCCGTAGGCCTTG
AT2G26710      CTCCAATGGCGGCCCCTACGCCGTTGGCCTTG
Grape          TACCAATGGCGGCCCCTACGCCGTTGGCCTTG
Peach          TTCCAAGGGCGGCCCCTACGCCGTTGGCCTTG
Columbine      TTCCAATGGCGGCCCCTACG

RING finger SHI family transcription factor
AT5G553590     TGATGGCGGCG-TTGCAGAG
Rice           TCATGGCGGCGTTTGCAGGC
Cacao          TCATGGCGGCG-TTGCAGGG
AT3G51060      TCATGGCGGCG-TTGCAGAG
Peach          GCATGGCGGCG-TAGCAGAG
Banana         CCATGGCGGCG-TTACAGTG
Grape          CCATGGCGGCG-TTGCAGAG
Peach          CCATGGCGGCG-TTGCAGGG
Grape          CCATGGCGGCG-TTGCAGAG
Columbine      CCATGGCGGCG-TTGCAGAG
Banana         CCATGGCGGCG-TTGCAGCA
Banana         CCATGGCGGCG-TTGCAGTG
AT1G75520      CCATGGCGGCGTTTGCAGTT
Columbine      CCATGGCGGCGTTTGCAGAG
AT4G36260      CCATGGCGGCG-TTGCAGAA
Banana         CCATGGCGGCG-TTGCAGCA
Banana         CCATGGCGGCG-TTGCAGCA

HLS1, apical hook development regulator
AT2G23060      GAGAAACCATA-AGCCGCCATTTCC
Columbine      GCGCAGCCATG--GCCGCCATTTTT
AT4G37580      ACGCAGACATA--GCCGCCATTTTC
Peach          GCGCAGACACACAGCCGCCATTTCC
Cacao          GCGCAGCCATA--GCCGCCATTTCC
Grape          GCGCAGCCACA--GCCGCCATTTCT
Banana         GCGCAGCCACA--GCCGCCATTTCC
Banana         GCGCAGTCACA--GCCGCCATTTCT
Banana         GCGCAGCCACA--GCCGCCATTTCT
Rice           GCGCAGCCAAA--GCCGCCGCTCTC
Columbine      GCGCAGCCAAA--GCCGCCATTTTC
Amborella      GCGCAGCCACA--GCCGCCATTTTC
```

**Supplemental Figure 2. Motifs over-represented in deep CNSs are phylogenetically conserved.** A. MSA box (AACGG). B. GCC box (GGCGGC CNS sequences are associated with the following genes. MSA box - C-MYB transcriptional repressor of G2/M phase genes: AT3G09370, AcoGoldSmith_v1.006219m, Tc05_g004560, PAC:17659278, GSVIVG01034171001, AT5G02320, AT2G13960, LOC_Os01g62410, Sb03g039470, GSMUA_Achr7G10520_001, GSMUA_Achr10G26610_001; protein amino acid N-linked glycosylation: GSMUA_Achr8G28540_001, LOC_Os02g38140, AT5G14480, GSMUA_Achr10G16790_001, LOC_Os04g40150, Tc04_g001940, AT3G01620, PAC:17662084, GSVIVG01032386001, AcoGoldSmith_v1.006484m, Sb04g024790, Sb06g020110; glycosyl hydrolase: PAC_17645685, Tc03_g028870, AT1G64390, AcoGoldSmith_v1.002101m, GSVIVG01019523001; cyclin-dependent protein kinase: PAC:17653262, Tc01_g000600, GSVIVG01009895001, AT1G20610, AcoGoldSmith_v1.005274m, AT1G76310, AT4G35620, AT2G17620, columbine Mule; RING finger-like zinc finger SHI family transcription factor: AT1G19790, AT1G75520, GSVIVG01009812001, PAC:17653900, Tc08_g002910, AcoGoldSmith_v1.008533m. GCC-box - cytochrome p450: Tc09_g010780, AT2G26710, GSVIVG01015485001, PAC:17643034, AcoGoldSmith_v1.003329m; SHI family transcription factor: AT5G66350, LOC_Os06g49830, Tc08_g002910, AT3G51060, PAC:17656898, GSMUA_Achr5G14410_001, GSVIVG01018845001, PAC:17653900, GSVIVG01009812001, AcoGoldSmith_v1.008533m, GSMUA_Achr3G12550_001, GSMUA_Achr2G12310_001, AT1G75520, AcoGoldSmith_v1.023667m, AT4G36260, GSMUA_Achr7G05730_001, GSMUA_Achr1G18650_001; HLS1: AT2G23060, AcoGoldSmith_v1.023967m, AT4G37580, PAC:17669007, Tc01_g006370, GSVIVG01000131001, GSMUA_Achr5G01700_001, GSMUA_Achr10G11370_001, GSMUA_Achr4G23250_001, LOC_Os07g02540, AcoGoldSmith_v1.006140m, evm.TU.AmTr_v1.0_scaffold00137.5.

```
Arabidopsis  ATAAATGC-ATTGCACCCTCTCTATGCGCTC-ATCAATAACTCC      GGAAGTCAAGAACGCACCT.........GGTTCGGGTATTGACTTCT
Arabidopsis  ATAAATGCAAT-GCACCCTCTCTATGCATTT-CGCAATAACTCC      GGAAGTCAAGAACGCACCT.........GGTTCGGGTATTGACTTCT
Cacao        TTAAATGCAATTGCACCCTCTCACTACGCTC-TGCAATAACTCT      AGAAGTCAAGAGCGCACCT.........GGTTCGGGAATTGACTTCC
Peach        ATAAATGCTATCGCACCCTCTCAATACGCTC-TGCATTAACTCC      AGAAGTCAAGAACGCACCT.........GGTTCGGGTATTGACTTTC
Grape        ATAAATGCTATTGCACCCTCTCAATACGCTC-TGCAATAACTCC      AGAAGTCAAGAACGCACCT.........GGTTCGGGTATTGACTTTG
Banana       ATAAATGCAATTGCACCCTCTCTAGAGGCTC-TGCAATAACTCA      GGAAGTCAAGAACGCACCT.........GGTTCGGGCATTGACTTCC
Rice         ATAAATGCGATTGCACCCTCTCAACACTCCGCGGCAATAACTCC      GGAAGTCAAGAACGCACCT.........GGTTCGGGTATTGACTTCC
Amborella    TTAAATGCAATTGCACCCCCTC--TACGCTC-TGCAATTACTCC      GGAAGTCAAGAACGCACCG.........GGTTCGGGTATTGACTTTC
```

L1 box



**Supplemental Fig 3. ATML1 and PDF2 have a 5' proximal experimentally-validated L1 box (Abe et al. 2001; Abe et al. 2003; Takada and Jurgens 2007) and a 3' UTR CNS with a potential RNA-fold.** GEvo panel showing blastn hits (colored rectangles) between *Arabidopsis* PDF2 and (from top to bottom) orthologous genes from *Arabidopsis* (*ATML1*), cacao, peach, grape, banana, rice, and amborella (http://genomevolution.org/r/97su). The sequences of the three deep CNSs associated with this gene family are shown below the corresponding blastn hits. Nucleotides conserved in at least 6 of the 8 sequences are colored blue. The 5' proximal CNS has an experimentally-verified epidermis-specific transcription factor binding site, the L1 box, which is boxed in yellow. The two CNSs present in the 3' UTR have the potential to form RNA secondary structure. The RNAfold-predicted optimal secondary structure for *ATML1*, with a minimal free energy of -42.90 kcal/mol, is shown to the right of the sequences, with bases colored blue corresponding to dsRNA reads. The bracketed region is deeply conserved throughout angiosperms and includes nucleotides present in looped-out regions. HD-ZIPIV homeodomain genes: *Arabidopsis* AT4G04890 (PDF2, *Arabidopsis* AT4G21750 (ATML1), cacao Tc02_g17643801, peach PAC:17643801, grape GSVIVG01012643001, banana GSMUA_Achr8G07310_001, rice LOC_Os04g53540 and amborella evm.TU.AmTr_v1.0_scaffold00004.843.

**Supplemental Table 1.** Genes associated with deep CNS sequences described in Figure 3.

| | |
|---|---|
| NGA1-4 | AcoGoldSmith_v1.017991m, AT2G46870, AT3G61970, PAC:17653345, Tc03_g022470, GSMUA_Achr1G03600_001, GSMUA_Achr4G24970_001, GSMUA_Achr2G10230_001, LOC_Os03g02900, Sb01g049150, GSVIVG01027463001, GSMUA_Achr4G30170_001, LOC_Os04g49230, Sb06g026370, LOC_Os02g45850, PAC:17648747, Tc00_g048490, AT1G01030 and AT4G01500, evm.TU.amTr_v1.0_scaffold00018.22 |
| PAR2 | AcoGoldSmith_v1.016845m, grape, PAC_17664766, Tc01_g014560, AT3G58850 |
| polygalacturonase | AT1G60590, AcoGoldSmith_v1.002735m, AT1G10640, Tc02_g023950, PAC:17665277, GSVIVG01013772001, AcoGoldSmith_v1.003611m |
| bZIP | GSMUA_Achr3G31390_001, GSMUA_Achr3G03550_001, GSMUA_Achr7G25310_001, LOC_Os01g11350, Sb03g002050, LOC_Os02g14910, Sb04g008840, GSVIVG01020658001, AcoGoldSmith_v1.026203m, AT3G58120, PAC:17652691, Tc00_g087150 |
| AT-hook motif | AcoGoldSmith_v1.009502m, PAC:scaffold_1_4872201_4872989, GSVIVG01036210001, AT4G14465, Tc04_g025020 |

**Supplemental Table 2.** Genes associated with deep CNS sequences used to derive the RNAz-predicted consensus RNA structures depicted in Figure 4.

| | | |
|---|---|---|
| (B) bHLH | *Arabidopsis* | AT4G29100 |
| | peach | PAC:17643560 |
| | grape | GSVIVG01015353001 |
| | columbine | AcoGoldSmith_v1.007621m |
| | banana | GSMUA_Achr2G15350 |
| | *Amborella* | evm.TU.AmTr_v1.0_scaffold00109.1 |
| (C) DEAD box RNA helicase | *Arabidopsis* | AT3G01540, AT5G14610 |
| | rice | LOC_Os01g36860 |
| | banana | GSMUA_Achr8G31890 |
| | *Amborella* | evm.TU.AmTr_v1.0_scaffold00096.5 |

**Supplemental Table 3**. CNSs with predicted RNA-folds

| *Arabidopsis* **genes** | **Rice genes** | **Gene description** | **Position** | **dsRNA reads present[a]** |
|---|---|---|---|---|
| **AT4G04890** | LOC_Os04g53540 | | | |
| **AT4G21750[b]** | LOC_Os08g08820[c] | HD-ZIPIV transcription factor | 3' UTR | Yes |
| **AT2G20100** | | | | |
| **AT4G29100** | LOC_Os02g17680 | bHLH | 3' UTR | Yes |
| **AT2G29630** | LOC_Os03g47610 | THIC | 3' UTR | Yes |
| | LOC_Os02g10700 | | | |
| **AT5G25350** | LOC_Os06g40360 | F-box and LRR | 3' UTR | Yes |
| **AT3G09050** | | | | |
| **AT5G01175** | LOC_Os01g63410 | thylakoid membrane protein | 3' UTR | Yes |
| **AT1G20880** | | RNA recognition motif | | |
| **AT1G76460** | | (RRM)-containing protein | 3' UTR | |
| **AT3G48390** | | | | |
| **AT5G63190** | | eIF-4 gamma | 5' UTR | Yes |
| **AT1G57680** | LOC_Os01g66190 | G-protein coupled receptor 1 | 5' UTR | |
| **AT3G01540** | | | 5' UTR/ | |
| **AT5G14610** | LOC_Os01g36860 | DEAD-box RNA helicase | intron | Yes |
| **AT1G71692** | LOC_Os08g02070 | MADS box/MIKCc type box | 5' UTR | |
| **AT4G26850** | LOC_Os12g08810 | VTC2, rate-limiting step in | | |
| **AT5G55120** | LOC_Os12g41800 | ascorbate biosynthesis | 5' UTR | Yes |
| | | adenosylmethionine | | |
| **AT5G18930** | | decarboxylase | 5' UTR | |
| **AT4G27000** | LOC_Os04g53440 | | | |
| **AT5G54900** | LOC_Os08g09100 | RNA binding family protein | intron | |
| | | pre-mRNA processing factor | | |
| **AT1G04080** | LOC_Os03g11200 | 39 | intron | |

*Arabidopsis* and rice genes with conserved potential RNA-fold structures.

[a] Zheng et al., 2010

[b] Also AT1G17920, AT1G73360, AT4G17710, AT5G46880, AT1G05230, AT2G32370, AT4G00730, and AT3G61150

[c] Also LOC_Os08g04190, LOC_Os06g10600, LOC_Os10g42490, LOC_Os04g48070, LOC_Os02g45250, and LOC_Os09g35760

**Supplemental Table 4**. Experimentally-validated deep CNSs

| Family | Locus ID | Biological role | CNS role | Reference |
|---|---|---|---|---|
| **ATML1/PDF2 HD ZIPIV homeodomain transcription factor** | AT4G04890<br>AT4G21750 | Regulating epidermal cell fate | L1 box specifying epidermal expression | Abe et al. (2001); Abe et al. (2003); Takada and Jürgens (2007) |
| **AGL15 MADS-domain transcription factor** | AT5G13790 | Regulating embryogenesis | CArG3 box; negative autoregulation | Zhu and Perry (2005) |
| **SHI/STY RING-like zinc finger transcription factors** | AT3G51060<br>AT5G66350 | Transcriptional activators of auxin biosynthetic genes | GCC box required for expression in aerial tissues | Eklund et al. (2011) |
| **YABBY zinc finger transcription factors** | AT2G45190<br>AT4G00180 | Abaxial cell type specification | cis-element suppressing adaxial expression | Watanabe and Okada (2003) |
| **MIR165a/b** | AT1G01183<br>AT4G00885 | Xylem specification, floral stem termination, leaf polarity | *cis*-element suppressing adaxial expression | Yao et al. (2009) |
| **SPL transcription factors** | AT1G53160<br>AT3G15270 | Mediate flowering time via activation of floral meristem identity genes | miRNA responsive element subject to translational repression by miR156 | Gandikota et al. (2007) |
| **MYB44/MYB77 transcription factors** | AT3G50060<br>AT5G67300 | Induction of stress genes | VRE box bound by VIP1 | Pitzschke et al. (2009) |
| **CUC2 NAC transcription factor** | AT5G53950 | Boundary formation | Restricts expression to intercotyledonary region | Larsson et al. (2012) |
| **BAS1** | AT2G26710 | Steroid hydroxylase inactivating brassinosteroid | LBD motif bound by LOB | Bell et al. (2012) |
| **CESA4 cellulose synthase** | AT5G44030 | Secondary wall cellulose synthesis | cis-element for MYB44 | Kim et al. (2013a); Kim et al. (2013b) |
| **glyceraldehyde-3-phosphate dehydrogenase** | AT1G12900<br>AT3G26650 | Photosynthetic carbon fixation | GATA box bound by GATA-1 | Jeong and Shih (2003) |
| **THIC** | AT2G29630 | Thiamine biosynthesis | Binds TPP aptamer which prevents 3' UTR intron splicing thus lowering transcript levels | Wachter et al. (2007) |

## Supplemental References:

**Abe, M., Katsumata, H., Komeda, Y., and Takahashi, T. (2003). Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in Arabidopsis. Development 130, 635-643.**

**Bell, E.M., Lin, W.C., Husbands, A.Y., Yu, L., Jaganatha, V., Jablonska, B., Mangeon, A., Neff, M.M., Girke, T., and Springer, P.S. (2012). Arabidopsis lateral organ boundaries negatively regulates brassinosteroid accumulation to limit growth in organ boundaries. Proc. Natl. Acad. Sci. USA 109, 21146-21151.**

**Hanada, K., Zhang, X., Borevitz, J.O., Li, W.H., and Shiu, S.H. (2007). A large number of novel coding small open reading frames in the intergenic regions of the Arabidopsis thaliana genome are transcribed and/or under purifying selection. Genome Res. 17, 632-640.**

**Kim, W.-C., Kim, J.-Y., Ko, J.-H., Kim, J., and Han, K.-H. (2013a). Transcription factor MYB46 is an obligate component of the transcriptional regulatory complex for functional expression of secondary wall-associated cellulose synthases in *Arabidopsis thaliana*. J. Plant Physiol. 170, 1374-1378.**

**Kim, W.C., Ko, J.H., Kim, J.Y., Kim, J.M., Bae, H.J., and Han, K.H. (2013b). MYB46 directly regulates the gene expression of secondary wall-associated cellulose synthases in Arabidopsis. Plant J. 73, 26-36.**

**Pérez-Rodríguez, P., Riaño-Pachón, D.M., Corrêa, L.G., Rensing, S.A., Kersten, B., and Mueller-Roeber, B. (2010). PlnTFDB: updated content and new features of the plant transcription factor database. Nucleic Acids Res. 38, D822-827.**

**Pitzschke, A., Djamei, A., Teige, M., and Hirt, H. (2009). VIP1 response elements mediate mitogen-activated protein kinase 3-induced stress gene expression. Proc. Natl. Acad. Sci. USA 106, 18414-18419.**

**Takada, S., and Jürgens, G. (2007). Transcriptional regulation of epidermal cell fate in the Arabidopsis embryo. Development 134, 1141-1150.**

**Watanabe, K., and Okada, K. (2003). Two discrete cis elements control the Abaxial side-specific expression of the FILAMENTOUS FLOWER gene in Arabidopsis. Plant Cell 15, 2592-2602.**

**Yao, X., Wang, H., Li, H., Yuan, Z., Li, F., Yang, L., and Huang, H. (2009). Two types of cis-acting elements control the abaxial epidermis-specific transcription of the MIR165a and MIR166a genes. FEBS Lett. 583, 3711-3717.**