

Figure S1. Biosynthesis pathway of PA.

CHS: Chalcone synthase, CHI: Chalcone isomerase, F3H: Flavanone 3 hydroxylase, F3'H: Flavonoid 3' hydroxylase, DFR: dihydroflavonol 4-reductase, FLS: Flavonol synthase, ANS: anthocyanidin synthase, LAR: Leucoanthocyanidin reductase, ANR: anthocyanidin reductase, UGT: UDP glucosyltransferase, TT12: transparent testa 12

```

BNLGH1233 1 -----MRNPTKKDN--VGTKTTPCCSKVGLKRGPWTPPEDELLSN
MtMYB5 1 MMNLSSTDDQDKSETTNSNSNPNPKKQK--T-TTTPCCSKVGLKRGPWTPPEDEVLSE
PH4 1 -----MRTPSSSST--TSNKVTPCCSKVGLKRGPWTPPEDELLTN
VvMyb5a 1 -----MRNPASAST----SKTPCCSKVGLKRGPWTPPEDELLAN
VvMyb5b 1 -----MRNPASASAPPSSSKTPCCSKVGLKRGPWTPPEDELLAN
AtMYB5 1 -----MNSCGGKKP--VSKKTPCCSKVGLKRGPWTPPEDELLVS
DkMYB4 1 -----MGRAPCCSKVGLKRGPWTPPEDELLTK
consensus 1 .....*****.*****.*****.*****

R2
BNLGH1233 39 YINKEGEGRWRTLPKFRAGLLRCGKSCRLRWMNYLRPSVKRGOIAPDEEDLILRLHRLLLGN
MtMYB5 58 YIKKEGEGRWRTLPKFRAGLLRCGKSCRLRWMNYLRPSVKRGOIAPDEEDLILRLHRLLLGN
PH4 39 YINKEGEGRWRTLPKFRAGLLRCGKSCRLRWMNYLRPSVKRGOIAPDEEDLILRLHRLLLGN
VvMyb5a 36 YIKKEGEGRWRTLPKFRAGLLRCGKSCRLRWMNYLRPSVKRGOIAPDEEDLILRLHRLLLGN
VvMyb5b 41 YIKKEGEGRWRTLPKFRAGLLRCGKSCRLRWMNYLRPSVKRGOIAPDEEDLILRLHRLLLGN
AtMYB5 39 YIKKEGEGRWRTLPKFRAGLLRCGKSCRLRWMNYLRPSVKRGOIAPDEEDLILRLHRLLLGN
DkMYB4 28 YIQVHGGEGWRSLPKFRAGLLRCGKSCRLRWMNYLRPSVKRGOIAPDEEDLILRLHRLLLGN
consensus 61 .....*****.*****.*****.*****.*****.*****.*****.*****

R3 C1
BNLGH1233 99 RWSLIAGRIPGRTDNEIKNYWNTHLSKKLISQCIDPRTHKPLN-PQQLS-----
MtMYB5 118 RWSLIAGRIPGRTDNEIKNYWNTHLSKKLISQCIDPRTHKPLNPNPSSSIIPTNTH--
PH4 99 RWSLIAGRIPGRTDNEIKNYWNTHLSKKLISQCIDPRTHKPLNPNPSSSDDITNKLA--
VvMyb5a 96 RWSLIAGRIPGRTDNEIKNYWNTHLSKKLISQCIDPRTHKPLNPNPSPDVNAPVS--K
VvMyb5b 101 RWSLIAGRIPGRTDNEIKNYWNTHLSKKLISQCIDPRTHKPLNPNPSSVDVKASSKAK
AtMYB5 99 RWSLIAGRIPGRTDNEIKNYWNTHLSKKLISQCIDPRTHKPLDANNIHKPEEVSGG--
DkMYB4 88 RWSLIAGRIPGRTDNEIKNYWNTHLSKKLISQCIDPRTHKPLDSDSHVQEPKKRSSNKKQK
consensus 121 .....*****.*****.*****.*****.*****.*****.*****.*****

BNLGH1233 147 ---PSPFSLKPSFS-SSSSMAKPNPFPPLFVHVVNANKQNYDDGSNEDHQG--MIMN
MtMYB5 175 ---QTSTPFFAPSSSDHHHPITITNNEPDPSTQQAQCG-NLVNDVSAMDNHH--VLTN
PH4 156 ---SSPFP--SSSKANDLNPILSPTYISS-FQMEPELKGINTHPGEITSLDDQYQSNAILA
VvMyb5a 154 SIPNANPNPSSSRVGEIGSNHEVKEIESNENHKPEPNLDQYHS-PLAADSNNENWQSDGL
VvMyb5b 160 AVMNPNPNPSPSEKAAANKEAGNFKSD-----NQYQIGAAGNDGSANIQNSDGS
AtMYB5 156 ---QKYLEPFTSSHTDDTVNG-----SDGSKNSINVFVG
DkMYB4 148 NKSKENLDHTEKLVHNPKPFRIKSLASFSFSRDSSSFDWTTTTATATPSSGSSNHEGERG
consensus 181 .....*****.*****.*****.*****.*****.*****.*****.*****

C3
BNLGH1233 200 N-----DHYQQQDHDVVFSSFLNSLIN--ED-----
MtMYB5 229 NNRGDY----GDDNGNNNYGGDDVFTSFLDSLIN--DDAFAAHRSE-----
PH4 211 EYCDLNLIAVTIEEDVEMNCCTDDVFSFLNSLIN--EDMFAACNQSQ-----
VvMyb5a 213 VTGLQSTHGTSNDEDDIGFCNDDVFSFLNSLIN--EDVFGNHHHQQQQQQQQLQVVQ
VvMyb5b 211 GTGLRSSN--NEEDDDLNCCTDDVFSFLNSLIN--EDVFGQHHLQQQHHG-----
AtMYB5 190 EHGYS-----EDFGFCYDDKFSFLNSLINDVGDPEGN-----
DkMYB4 208 MLCNNGS--NGHEVGFFIGEDGSDHYHDDHMD--DSDLECCS-----
consensus 241 .....*****.*****.*****.*****.*****.*****.*****.*****

BNLGH1233 225 --DDALVSNLGLSQGWESTP-----FDQPK-----
MtMYB5 269 --NDPILSAAFPALWESPLMMTSTHNFTQNDGKAS-----
PH4 256 --TNGTFQDFDPMASSSTP-----SSDQYNF-----
VvMyb5a 271 QPSNVIAPLPHPAISVOATFSSSPRTVWEPAAALTSTSA--APLVHDQKHSMSPI--
VvMyb5b 259 --GLIAPGSDALISTSVQSFSGFTSWEAAAMTSTSVFSGIDHSKRFNDQPKRF
AtMYB5 222 ----TIFISQFLQMDCKDG-----IVGASSS-----LGHD-----
DkMYB4 247 --LEKLYEYLLQLKTEEDDQ-----GQLDSFABSL-----
consensus 301 .....*****.*****.*****.*****.*****.*****.*****.*****

```

Figure S2. Mt MYB5 is the homolog of At MYB5. Protein sequences of BNLGH1233, PH4 (AY973324), Vv MYB5a, Vv MYB5b, At MYB5, DkMYB4 and Mt MYB5 were aligned by the Muscle program (<http://www.ebi.ac.uk/Tools/msa/muscle/>). Multiple sequence alignment image was produced by Boxshade (http://www.ch.embnet.org/software/BOX_form.html). Genbank accession numbers: At MYB5 (NP_187963.1), Dk MYB4 (BAI49721.1), Vv MYB5a (NP_001268108.1), Vv MYB5b (NP_001267854.1), Mt MYB5 (XP_003601609.1), PH4 (AAY51377.1), BNLGH1233 (AAK19611.1).

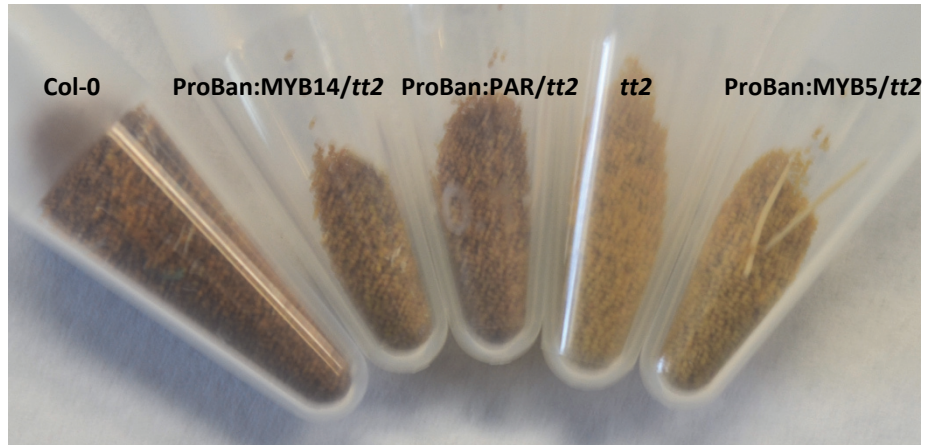


Figure S4. Genetic complementation of Arabidopsis *tt2* with Mt *Myb14*, Mt *Par* and Mt *Myb5*. Medicago *Myb14*, *Par* and *Myb5* were driven by Arabidopsis *Anr* (Banylus) promoter and used to complement Arabidopsis *tt2* mutant (Salk_005260). Fifteen to twenty independent lines were observed and representative lines were presented. Mt MYB14 and Mt PAR clearly rescue the *tt2* phenotype, while Mt MYB5 does not.

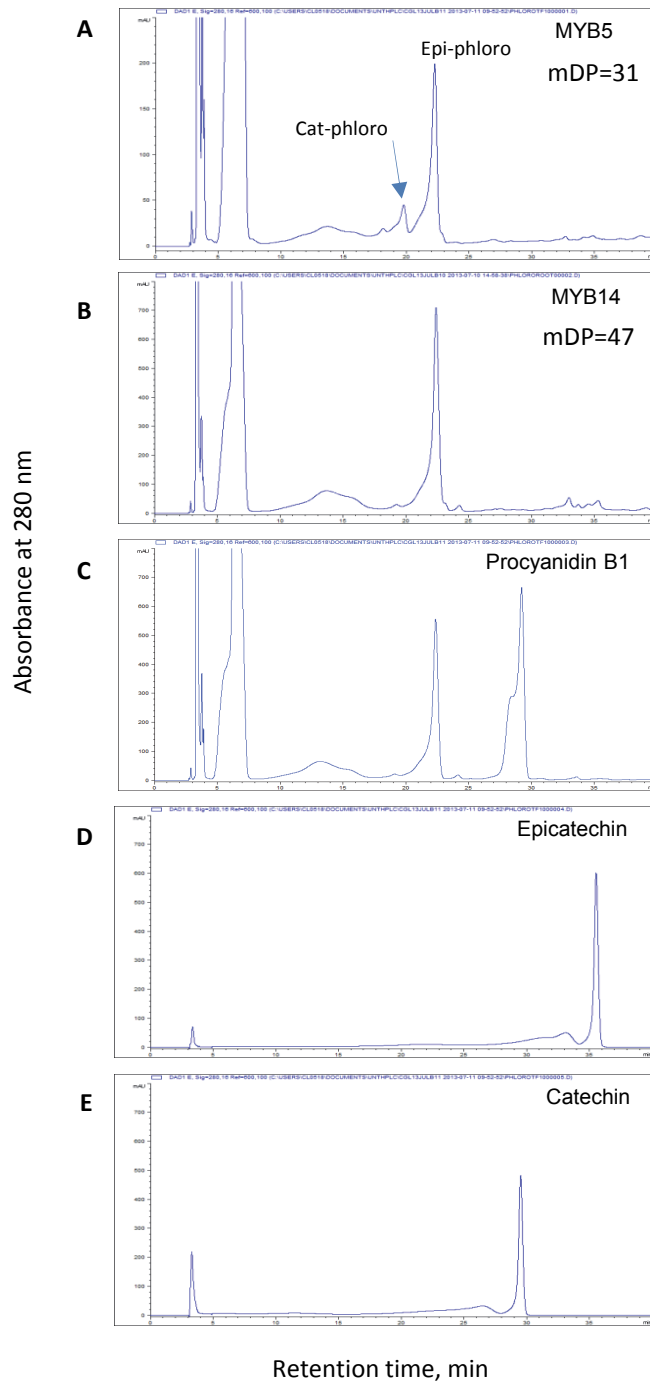


Figure S5. Phloroglucinolysis analysis of PAs in MYB5 and MYB14 over-expressing hairy roots. PAs purified by Sephadex LH20 resins were hydrolyzed by acidic phloroglucinol/mehanol solution. Hydrolyzed PA were analyzed by reverse phase HPLC. A, PAs purified from MtMYB5 over-expressing hairy roots. B, PAs purified from MtMYB14 over-expressing hairy roots. C, Procyanin B1 standard analyzed by phloroglucinolysis. D, Epicatechin standard. E, Catechin standard. mDP: mean degree of polymerization.

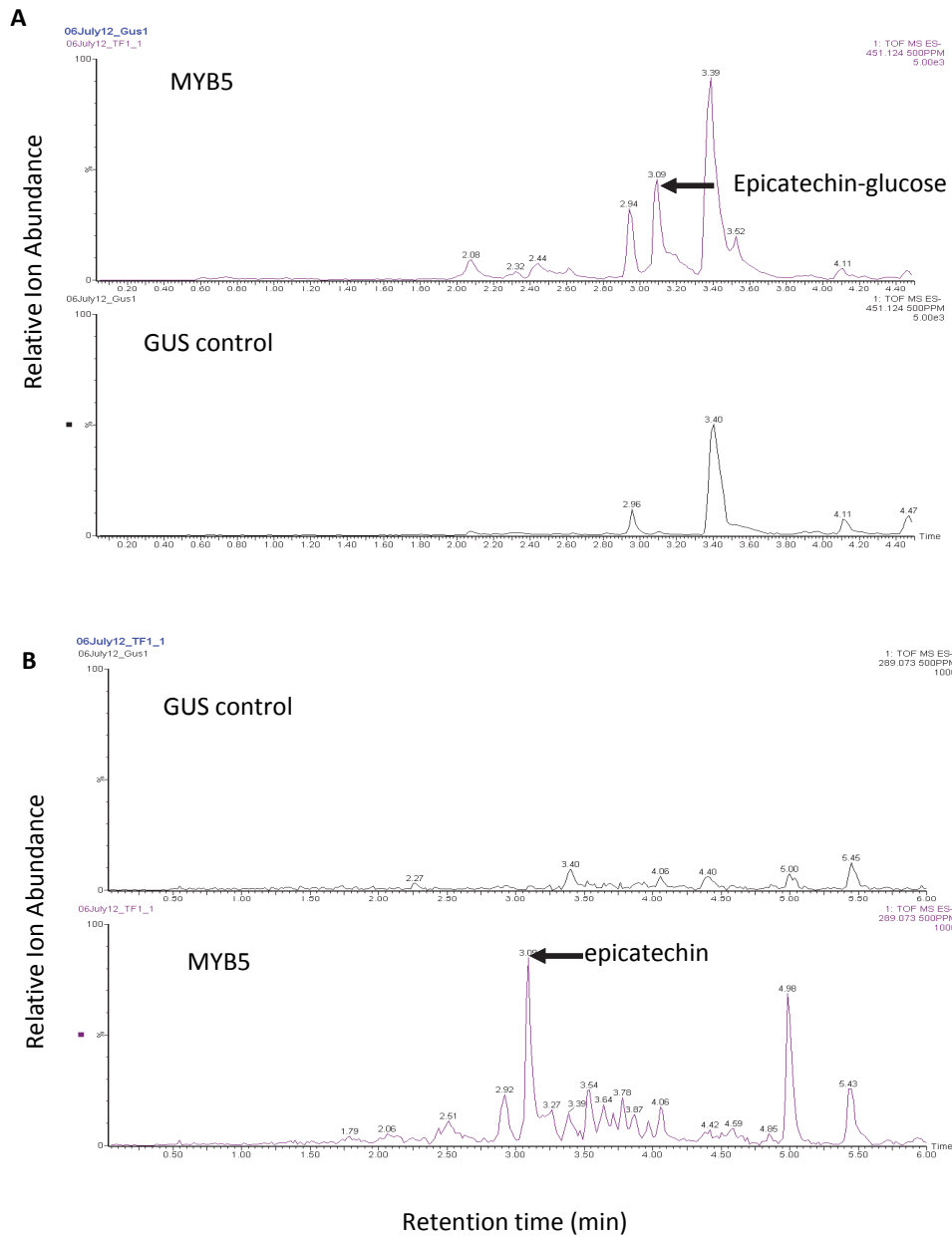


Figure S6. UPLC-MS detection of epicatechin and epicatechin-3'-O- glucoside ions in hairy roots over-expressing MYB5. A, epicatechin-3'-O- glucoside (m/z 451.124). B, epicatechin (m/z 289.073).

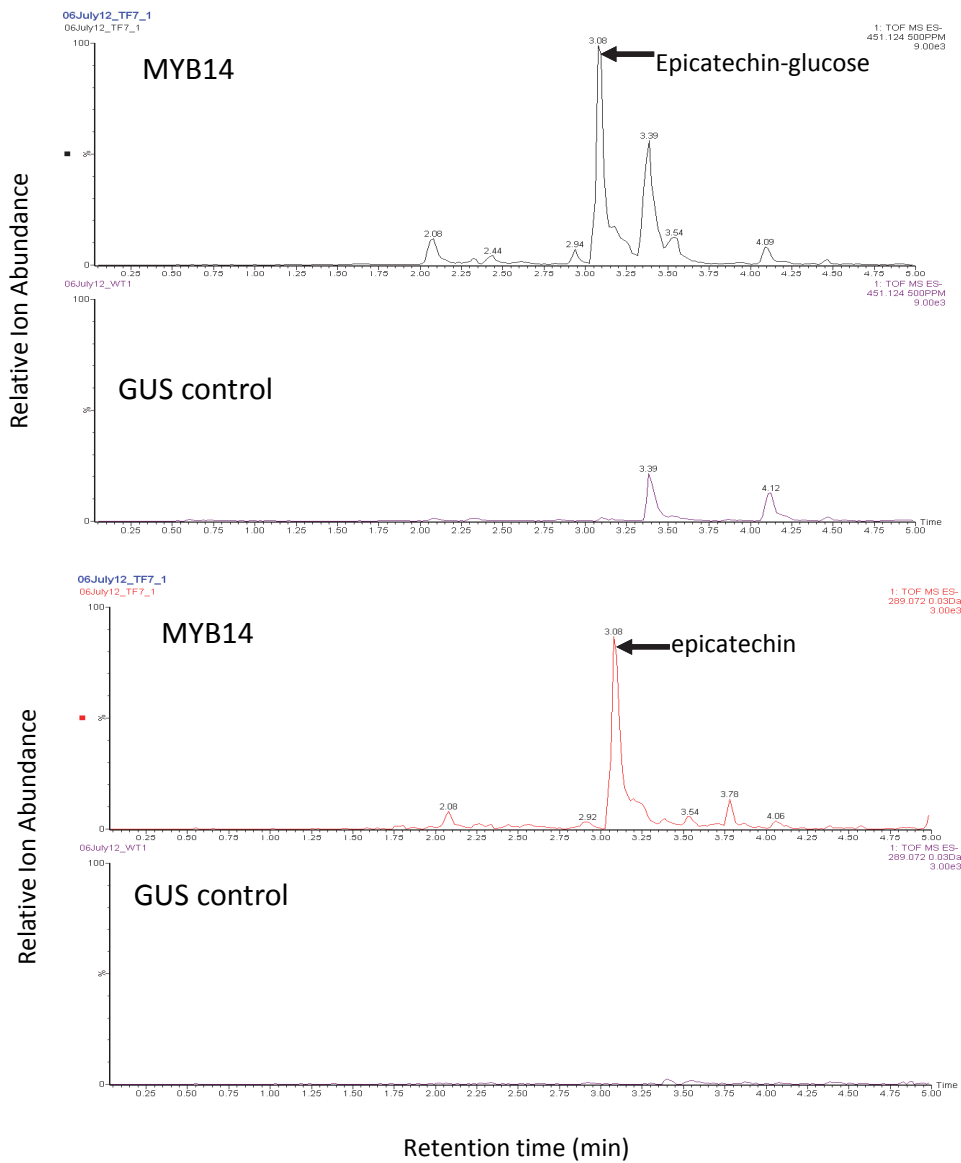


Figure S7. UPLC-MS detection of epicatechin and epicatechin-3'-O- glucoside ions in hairy roots over-expressing MYB14. A, epicatechin-3'-O- glucoside (m/z 451.124). B, epicatechin (m/z 289.073).

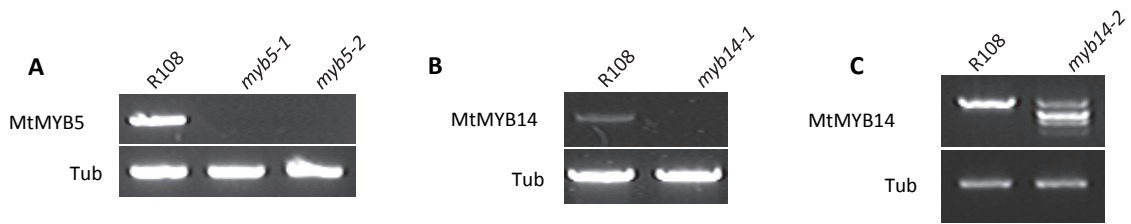


Figure S8. RT-PCR to detect MYB5 and MYB14 transcripts in *myb5* and *myb14* mutant seeds.

A, *myb5*. B, *myb14-1*. C, *myb14-2*.

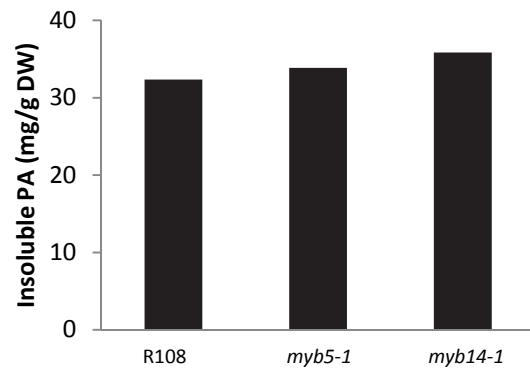


Figure S9. Butanol-HCl analysis of insoluble PA-like material in R108, *myb5* and *myb14* mutant seeds.

Insoluble PAs were hydrolyzed in butanol-HCl solution and levels expressed as procyanidin B1 equivalents.

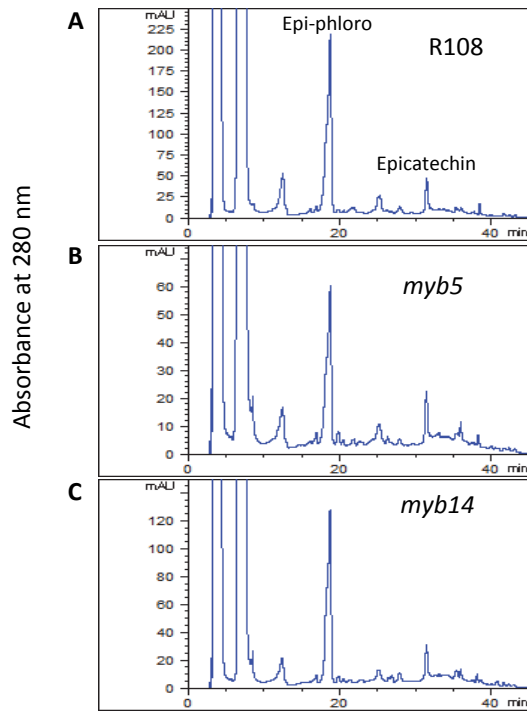


Figure S10. Phloroglucinolysis analysis of soluble PAs in the seeds of wild-type (R108), *myb5-1* and *myb14-1*.

PAs purified with Sephadex LH20 resin were hydrolyzed by acidic phloroglucinol/methanol solution. Hydrolyzed PAs were analyzed by reverse phase HPLC. A, PAs purified from R108 seeds. B, PAs purified from *mtmyb5* seeds. C, PAs purified from *myb14* seeds.

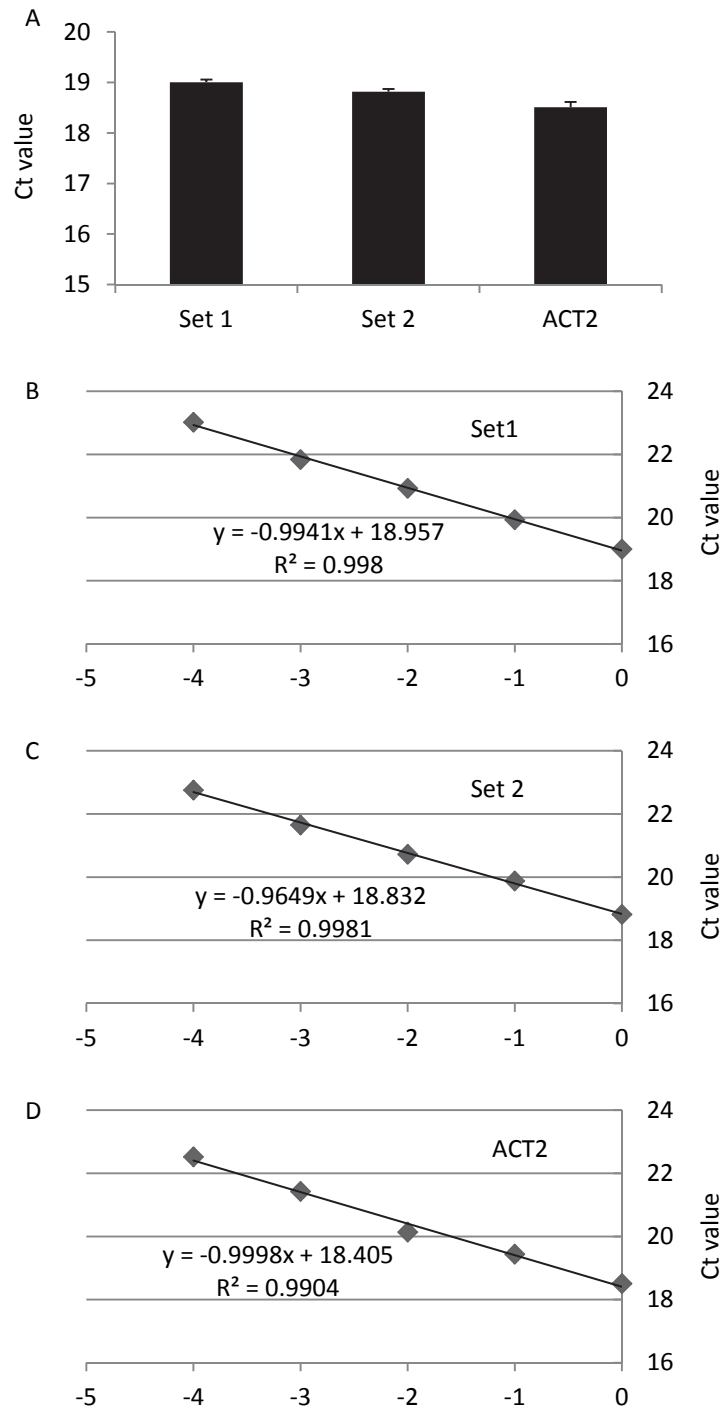


Figure S11. Measurement of MYB14 copy number in the *Medicago truncatula* genome by qPCR.

A, The Ct values of MYB14 determined by two sets of primers (Set 1 and Set 2, Supplemental Table 4) and reference ACT2 genes.

B, The amplification efficiency of primer set 1 determined by serial dilution of genomic DNA. Y-axis, Ct value. X-axis, log₂ value of dilution factors.

C, The amplification efficiency of primer set 2 determined by serial dilution of genomic DNA. Y-axis, Ct value. X-axis, log₂ value of dilution factors.

D, The amplification efficiency of primer pairs of ACT2 determined by serial dilution of genomic DNA. Y-axis, Ct value. X-axis, log₂ value of dilution factors.

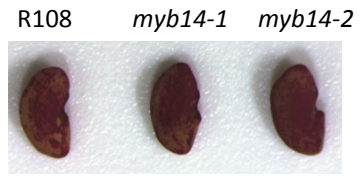


Figure S12. Ruthenium red staining showing that mucilage levels in *myb14* are similar to those in wild-type R108.