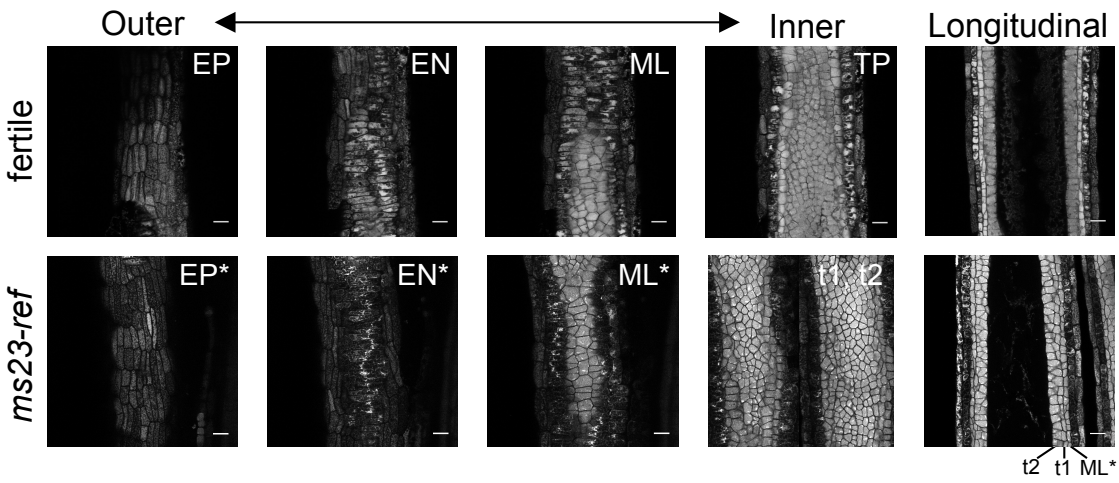
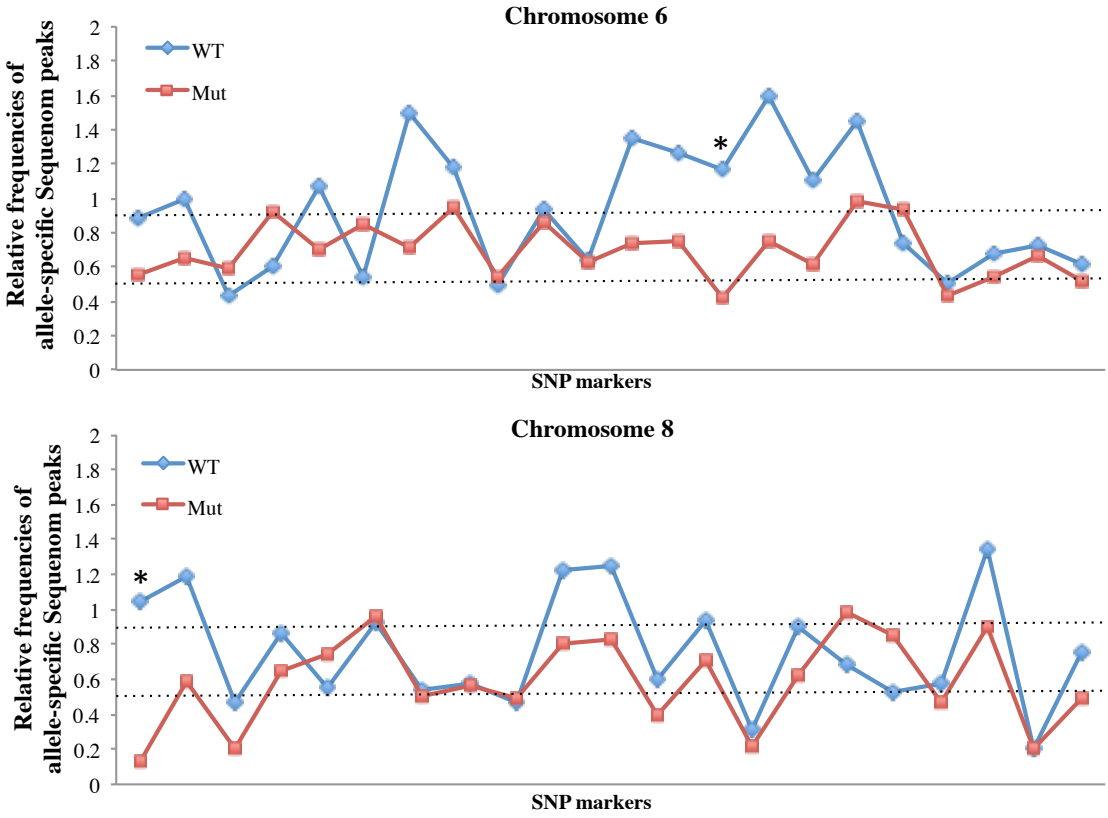


Fig. S1. Confocal images of the *ms23-ref* and its fertile sibling anthers at various depths (Z-axis) at Stage 10. The cell shape and morphology of EP versus EP*, EN versus EN, and ML versus ML* are indistinguishable at this stage. The two innermost layers (t1 and t2) in mutant anther are indistinguishable from each other under confocal from the XY-planes, indicating these two layers are sisters resulting from extra periclinal divisions while ML* is formed earlier. Judging from the central, longitudinal plane, however, t1 cells have a shorter Z-dimension, similar to ML/ML*, while t2 cells have a longer Z-dimension, resembling TP. In fertile anthers, EP – epidermis; EN – endothecium; ML – middle layer; TP – tapetum. In *ms23-ref* anthers, EP* - the presumptive epidermis; EN* - the presumptive endothecium; ML* - the presumptive middle layer; t1 – the defective fourth layer; t2 – the defective fifth layer.



Marker	WT	Mut	Chr	cM	Interval in Mb
35154W19	1.02	0.41	6	96.2	160.336 – 160.339
10013W39	1.17	0.42	6	89.8	158.443 – 158.445
27793W18	0.91	0.26	6	82.3	155.235 – 155.241
104210W9	2.75	0.35	8	19.9	4.515 - 4.517
9825W19	1.04	0.13	8	3.7	1.370 - 1.372

Fig. S2. High-throughput genetic mapping of *Ms23* via quantitative SNP-genotyping on a Sequenom chip. Segregating *ms23-ref* mutant (Mut) and wildtype (WT) genomic DNA pools were analyzed using the Maize Bulk Segregation Analysis service at the Genomic Technologies Facility at Iowa State University against a set of SNP markers (n=1,016). Using 0.9 and 0.5 cut-offs for wildtype (WT) and mutant (Mut) pools, respectively, *Ms23* was placed on chromosomes 6 and 8 at the locations indicated (*).

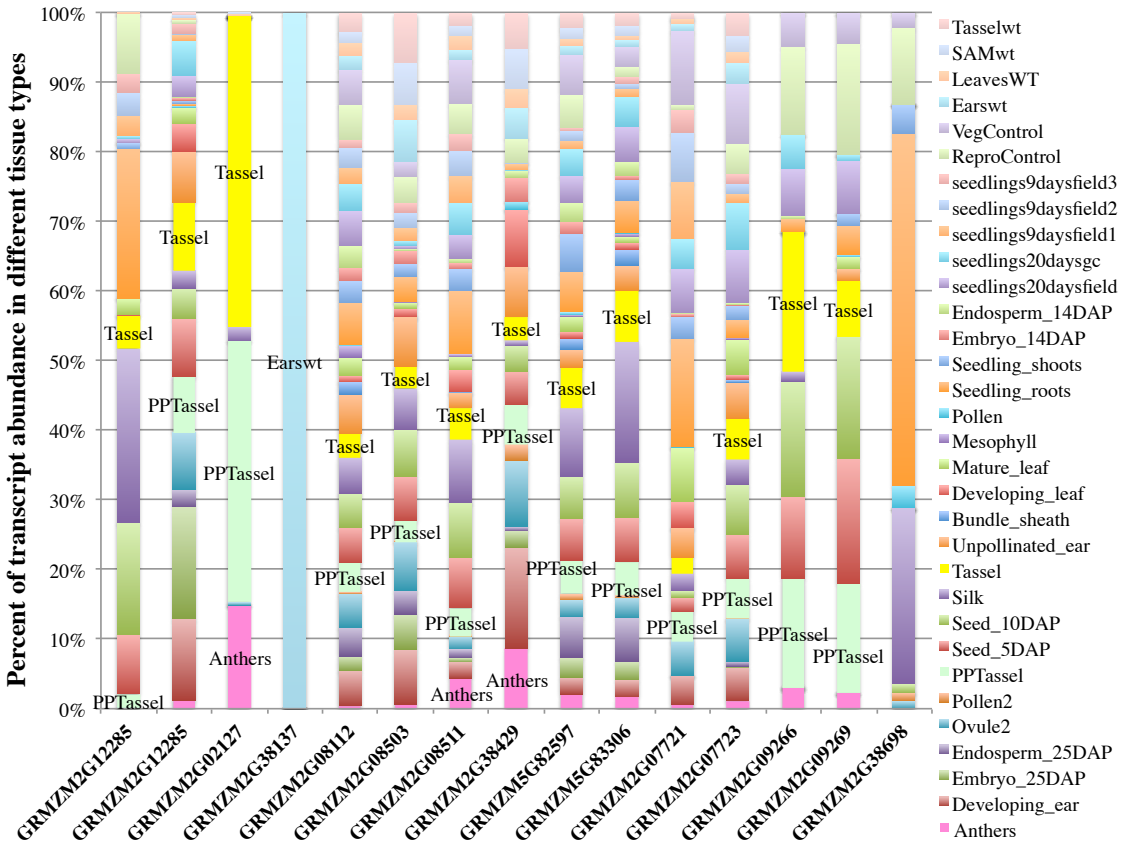
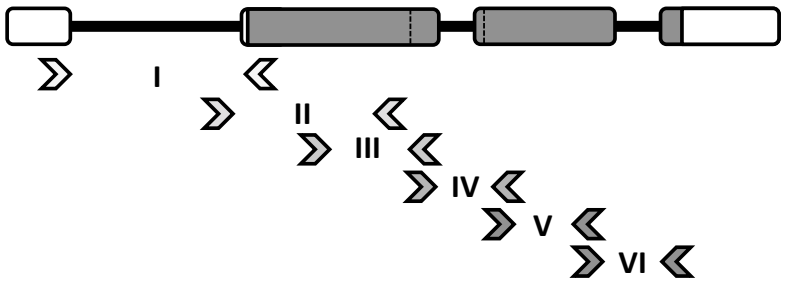


Fig. S3. Tissue-specific transcript abundances of fifteen gene models near the tip of chromosome 8 based on the data on qTeller (http://qteller.com/). Y-axis is the percentage of transcript abundance in each tissue type over the sum of abundance from all 32 tissues as 100%.

A

GTTTAATAACCTTGCACTGCCGCAGTAGCCCTTAACTGCTGCTATCTATCTCTTTTCTGAAGGAAAAAAAA
GGTTTGATACTCCTCTACCTAGCTAGTCCTGCATGCCGCTAATGTGCGTCTTGCCTGTTTATTTGTTCTTA
ATAAGGGCTGCCTATCTATTATATTTTGCACCTGTTTTGCTGTGTTCTTGGTAACTAGCTTAATTCCTTCG
CCTACAATCGTCAAATCCCCCCCATCATCAGTCAGATGAACTTTTGATCGAATTGAAGTTGTTCTTCTAAT
TCGGCCCCAGCAGCGCCCATGCATCTGGTTTTATTTGCTTTCTGTTGGGTATAATATGCAAGACCTTTTGT
TGCTAGGGCAAGGCTGCAACCACATGCGTGTACTGAACTCATGATGTAACTCATCCTTTTTGTTTGCTCAC
AGAATCACTACTCTACTGCACTTCCTTTTCATCCGATCCGCAATCTTTTTTTTCTTTTACATGCTTTAGTT
TTCTCTCTTTCTTGATTACAAACATGATTACTGGAACTTTCTTAGGCTGCCTTCCCCTTCCTTGGATCTGC
TTTAGTTTTCTTTTTTGGGCTACCGCGCGCGGCTTATTTGAGTTTATCACTTGCTGCATATACATAATATA
TATATACATGCATGCGATGGCGTTCATGTTACTCAACTACAGATCTGTTTCTGTTCGTGTGTTTCAGTTCA
GCGCGCAGTTAAGCATAGCAGGACGACCACGACGATGTATCACCCGCAGTGCGAGCTC↓CTGACGATGGCG
CACGAAACGCCGGACCTGGACGCCGGCCAGCCGCACCTAACCGTCTCCGGCGTCGCCAGCATCCCGGCAGA
GCTGAGCTTCCACCTGCTGCACTCGCTCGACGCCGCGGCGGCGGTCAATCCCGTCACGGCGCCGCCGCAGT
CCACCATCGACTACTTCCTCGGCGGCGCCGATCCCCACCAGCAGGCCATGCAGTACGAGCCGCTGCCGCCC
GCCGCGGGCGGCCACCACCAGTACACCATGGACATGTTCCGCGACTACTGCGACGGCCACTACCCCACCGC
CGAGCCGTACATCCGCGGGACAATGACTGGAGCCCTCGTGTTCGGGGCCACCGACGACGACGACTCGGCCG
CTGCCTACAT↓GCCCGGGGGGCACTTTGAGACCTCCCCGCCGCCGCCACGCGCCACCGGCCGCGGCAGGAA
GCGGGGCAGGGCGCTGGGCGGCGGCTTCCATGCTGTGCTGGCCAACGGCGTCGAGAAGAAGGAGAAGCAGC
GCCGGCTGCGGCTCACCGAGAAGTACACCGCCCTCATGCACCTCATACCCAACGTTACAAAGGTCGTACCA
AATCCTCCTCTTATGTTCGTCCATCGTTTCAAATTAAGTTAAAAAATTAATTCACGGTTCTTGTTGTTTAT
TTTTTGCGCACTGCAGACTGATAGGGCGACGGTGATCTCGGACGCGATCGAGTACATCCAGGAGCTGGGGA
GGACGGTGGAGGAGCTGACGCTGCTGGTGGAGAAGAAGCGGCGCCGGAGGGAGCTGCAGGGGGACGTCGTG
GACGCGGCGCCGGCTGCGGTGGTTGCTGCCGCCGGTGAGGCGGAGAGCTCGGAGGGCGAGGTGGCTCCTCC
GCCGCCGGCCGTGCCGCGGCAGCCGATCCGGAGCACGTACATCCAGCGGCGGAGCAAGGACACGTCCGTGG
ACGTGCGGATCGTGGAGGAGGACGTGAACATCAAGCTCACCAAGCGCCGGCGCGACGGGTGCCTCGCAGCC
GCGTCGCGCGCGCTGGATGACCTCCGCCTTGACCTCGTCCACCTCTCCGGCGGCAAGATCGGTGACTGTCA
AATCTACATGTTCAACACCAAGGTACATACGAATACGATACGTAGCCATTGATCGATCTGTAATTCTGTAG
CCTGACGATTCCGAGGTTTCTGGTGCTAAAAAATGCATCTTTTTTTCTCAGATGACAATGCTTTCTGTCTT
TGTTCACCGCAGATTCACAAGGGGTCTTCAGTGTTTGCGAGTGCAGTGGCCGGTAGGCTGATGGAAGTGGT
GGACGAGTACTAGGCTACCATGCACTTGAATTTCTAGCTAGCTCTACGTACCGCGCTGCTATGAATCTAGC
TATAGCGTTTCTTGGATGAAAGACTAGTTAGTTGTTACCTTCTATCTTTGCTTCAATTAAATCCGCTTGCT
CGTTACAGACTGAGTTTGTTTCTAAATGTCAAGGTTGTTTTGGTCAAATTGAATAAATTGGCACACTGGCC
TGTGAGGTTATTATATATATTTATGTGTTTATTACTGGTCTATTAATTTGTCTTATTATTA

B



	Forward Primer	Reverse Primer
I	TGCACCTGTTTTGCTGTGTT	TGATACATCGTCGTGGTCGT
II	CGACCACGACGATGTATCAC	CGCTGCTTCTCCTTCTTCTC
III	ACTACTGCGACGGCCACTAC	CATGAGGGCGGTGTACTTCT
IV	CGGCTCACCGAGAAGTACAC	CCGCTTCTTCTCCACCAG
V	GAGGAGCTGACGCTGCTG	CCGCTTCTTCTCCACCAG
VI	GTGGAGGAGGACGTGAACAT	ACTCGTCCACCACTTCCATC

Fig. S4. **Cloning of *Ms23*. (A)** The nucleotide sequence of GRMZM2G021276 with the four exons underlined, the translated region in bold, and the encoded bHLH domain in blue. The positions of frameshift mutations in the *ms23-6027* and *ms23-6059* alleles are indicated as orange and green arrows, respectively. **(B)** Placement and sequences of the PCR primers across the gene model of *bHLH16* (GRMZM2G021276).

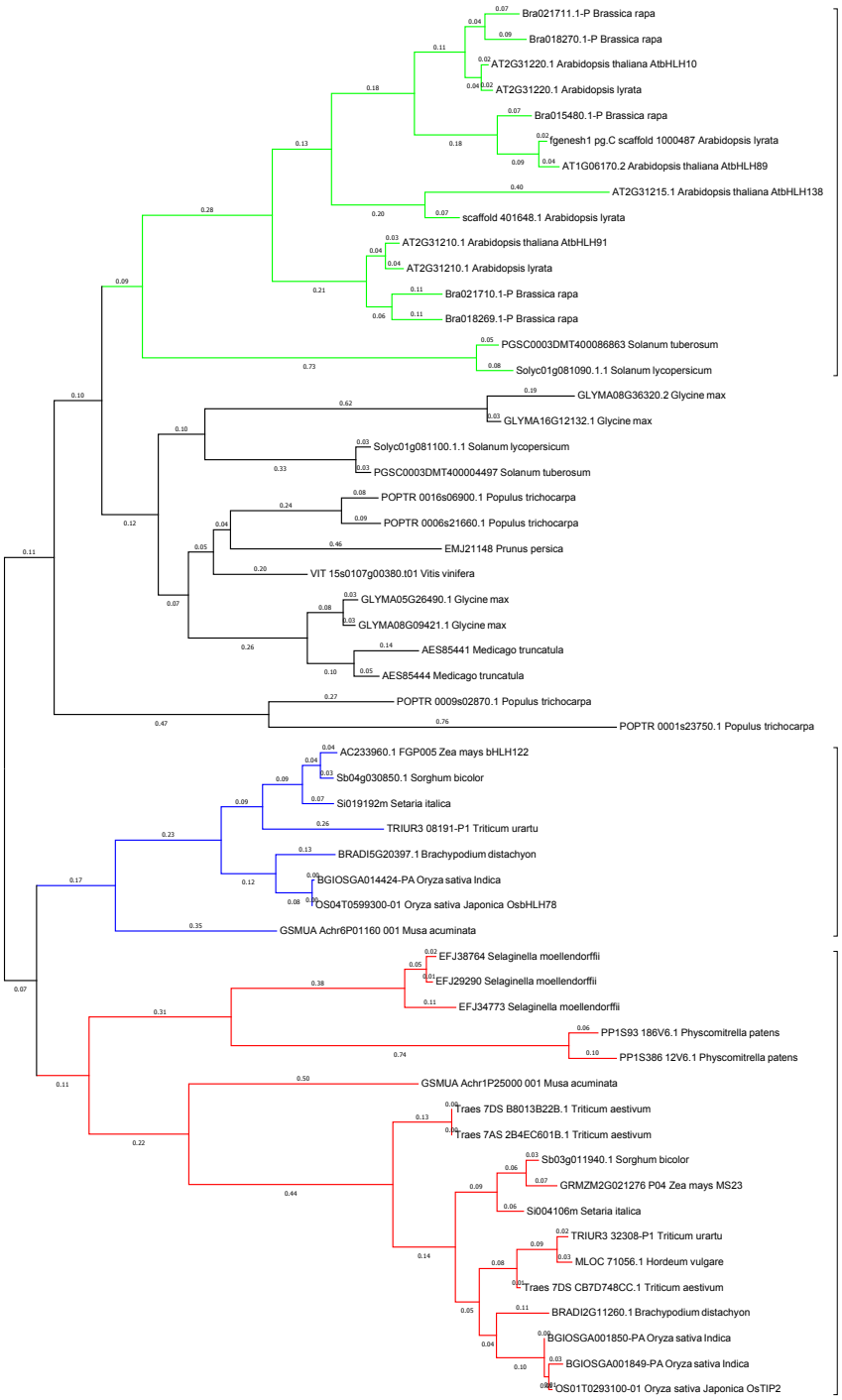


Fig. S5. Phylogenetic analysis of the orthologs of both MS23 and its paralog, bHLH122, (http://www.gramene.org/). The tree was generated using MEGA 6.06 (Tamura et al., 2013) by the maximum likelihood method based on the Poisson model; the tree with the highest log likelihood (-29072.8504) is shown. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 55 amino acid sequences. There were a total of 1,146 positions in the final dataset.

Gene	Species	[BASIC][HELIX]◡LOOP◡[HELIX]	No.of amino acids
MS23	*Zea mays*	ANGVEKKEKQRRLRLTEKYTALMHLIPNVTKTDRATVISDAIEYIQELGRTVEELTLLV	59
TIP2/bHLH142	*Oryza sativa*	ANGVEKKEKQRRLRLTEKYNALMLLIPNRTKEDRATVISDAIEYIQELGRTVEELTLLV	59
bHLH89	*Arabidopsis thaliana*	KRKIFPTERERRVHFKDRFGDLKNLIPNPTKNDRASIVGEAIDYIKELLRTIDEFKLLV	59
bHLH10	*Arabidopsis_thaliana*	KSRTSPTERERRVHFNDRFFDLKNLIPNPTKIDRASIVGEAIDYIKELLRTIEEFKMLV	59
bHLH91	*Arabidopsis thaliana*	KNKPFTTERERRCHLNERYEALKLLIPSPSKGDRASILQDGIDYINELRRRVSELKYLV	59
bHLH122	*Zea_mays*	KAN-FATERERREQFNVKYGALRSLFPNPTKNDRASIVGDAIEYINELNRTVKELKILL	58
EAT1/DTD	*Oryza_sativa*	KAN-FATERERREQLNVKFRTLRMLFPNPTKNDRASIVGDAIEYIDELNRTVKELKILV	58
bHLH51	*Zea mays*	QCKNLEAERKRRKKLNERLYKLRSLVPNISKMDRAAILGDAIDYIVGLQNQVKALQDEL	59
TDR	*Oryza sativa*	QCKNLEAERKRRKKLNGHLYKLRSLVPNITKMDRASILGDAIDYIVGLQKQVKELQDEL	59
AMS	*Arabidopsis thaliana*	QAKNLMAERRRRKKLNDRLYALRSLVPRITKLDRASILGDAINYVKELQNEAKELQDEL	59
MS32	*Zea mays*	KSKNLEAERKRRGKLNRNILALRAVVPNITKMSKESTLSDAIDLIKRLQNQVLELQRQL	59
UDT1	*Oryza sativa*	KSKNLEAERRRRGRLNGNIFALRAVVPKITKMSKEATLSDAIEHIKNLQNEVLELQRQL	59
DYT1	*Arabidopsis thaliana*	KSPNLEAERRRREKLHCRLMALRSHVPIVTNMTKASIVEDAITYIGELQNNVKNLLETF	59
		*:.** :: * .* :: : . :.: : * : .	

Fig. S6. Multiple sequence alignment of the bHLH domains of fourteen tapetum-associated bHLH proteins from maize, rice, and Arabidopsis. A partially duplicated pseudogene, *bHLH138* from Arabidopsis is excluded. An asterisk (*) indicates positions that have an invariant residue; a column (:) indicates conservation between groups of strongly similar properties with scoring > 0.5 in the Gonnet PAM 250 matrix; a period (.) indicates conservation between groups of weakly similar properties with scoring ≤ 0.5.

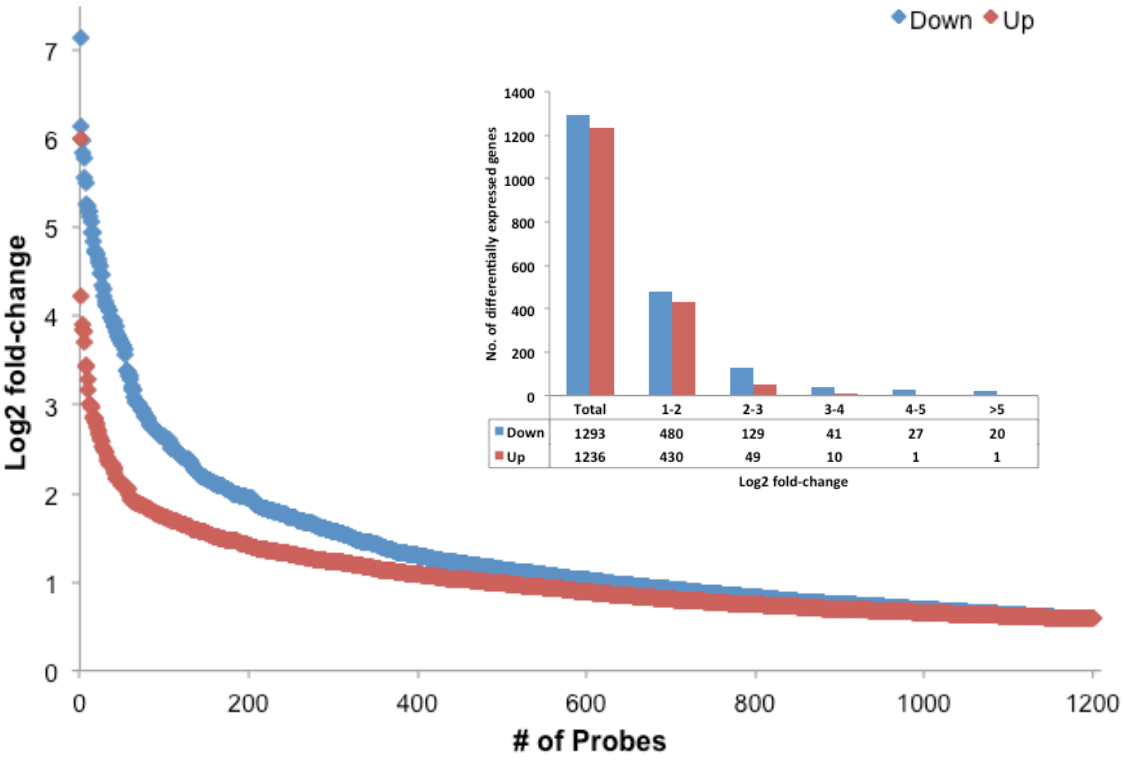


Fig. S7. Distribution of differentially expressed genes in *ms23-ref* mutant on the array. Up- and down-regulated genes comparing *ms23-ref* to its fertile sibling at Stage 9 were plotted based on their $\log_2$ fold change values. Inset: Counts of genes at different ranges of fold changes.

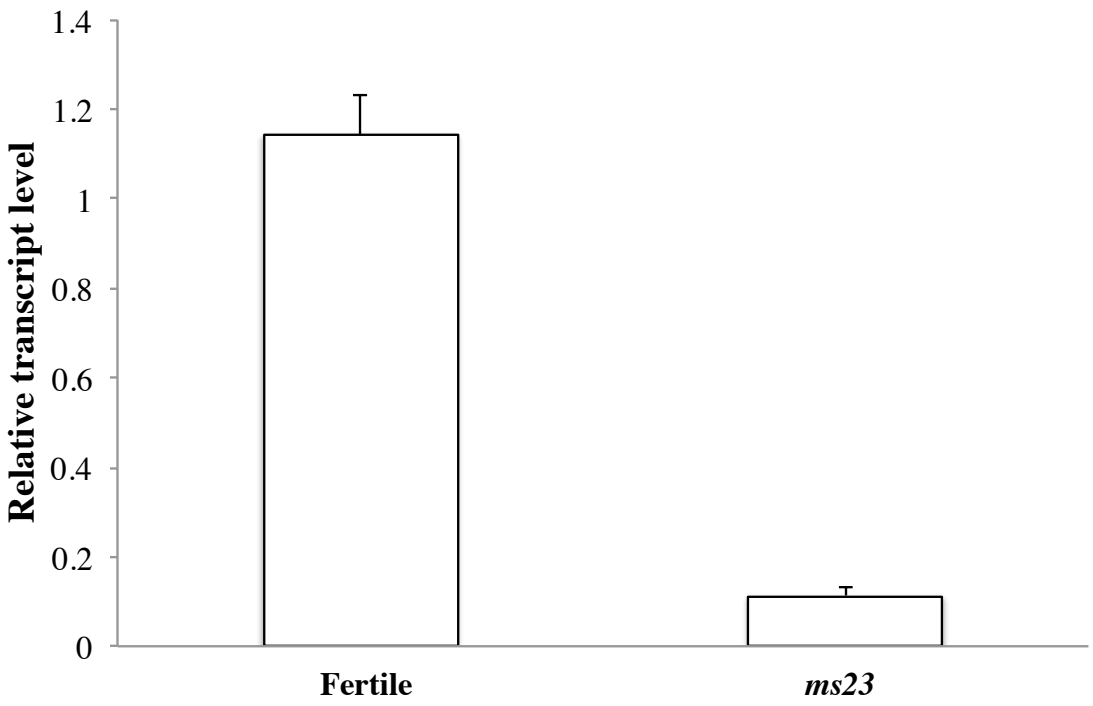


Fig. S8. Quantitative RT-PCR assays of *Dcl5* expression. *Dcl5* transcript levels in fertile and *ms23-ref* anthers at Stage 8. Expression values were normalized to the cyanase gene.

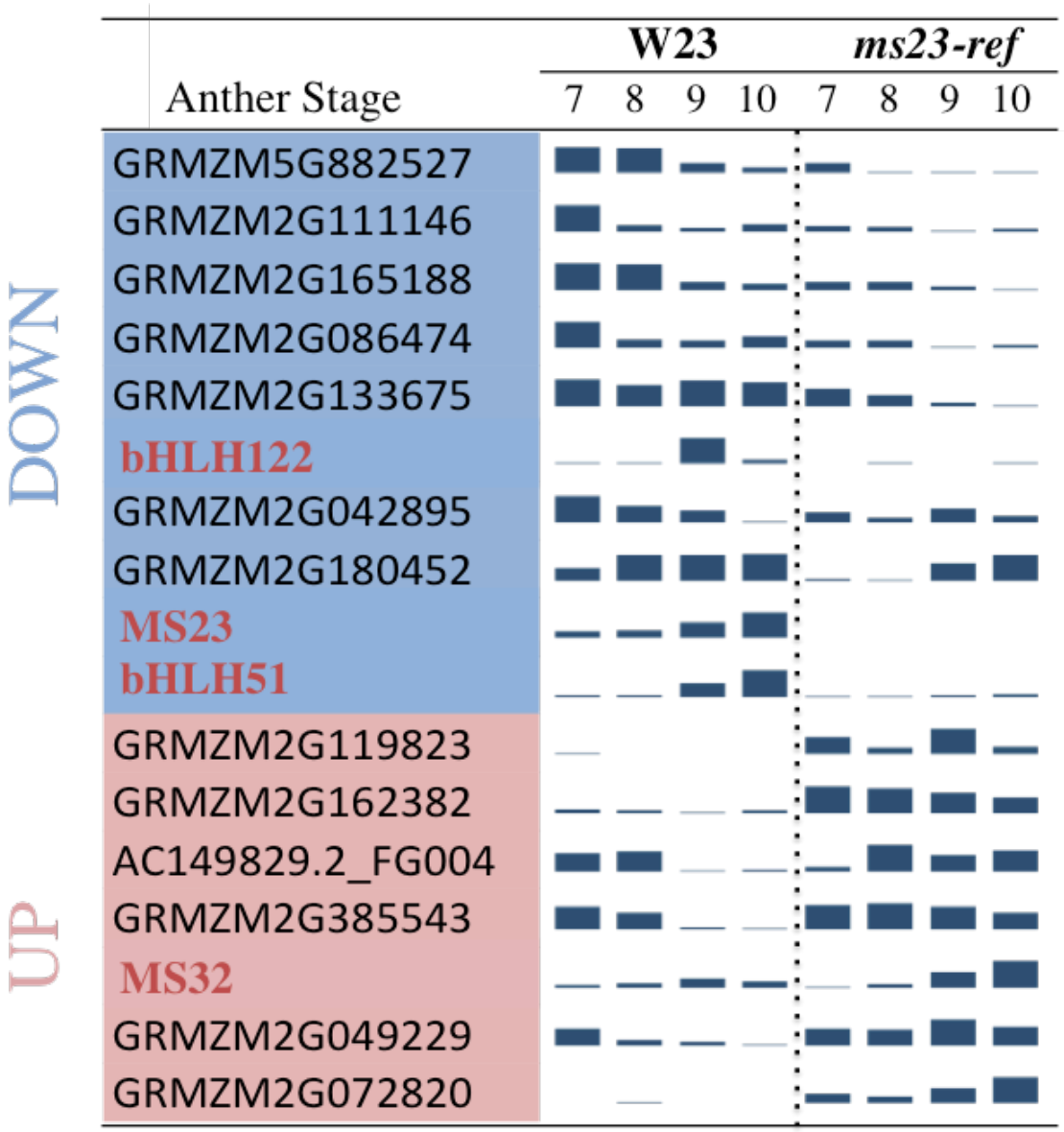


Fig. S9. Role of *Ms23* in regulating expression of the bHLHs with highest transcript abundance in fertile anthers. Maize bHLH genes with the sum of abundance in the top quartile (n=75) and with 4-fold or more difference (n=18) between *ms23-ref* mutant and W23 fertile anthers. Data are from RNA-seq analysis.

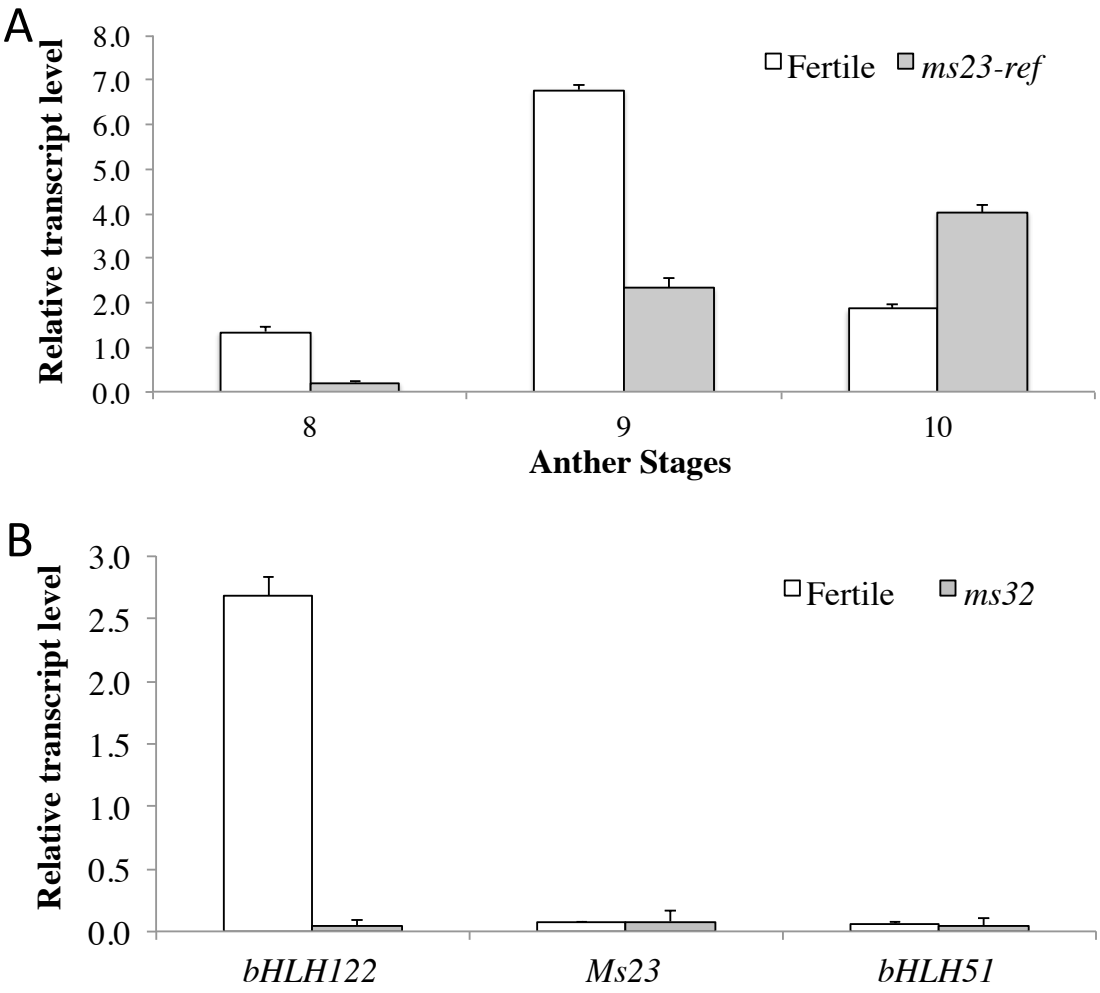


Fig. S10. Quantitative RT-PCR assays of gene expression. **(A)** *Ms32* transcript levels in fertile and *ms23-ref* anthers at various anther stages. **(B)** *Ms23*, *bHLH122*, and *bHLH51* transcript levels in fertile and *ms32* anthers at Stage 9. Expression values were normalized to the cyanase gene.

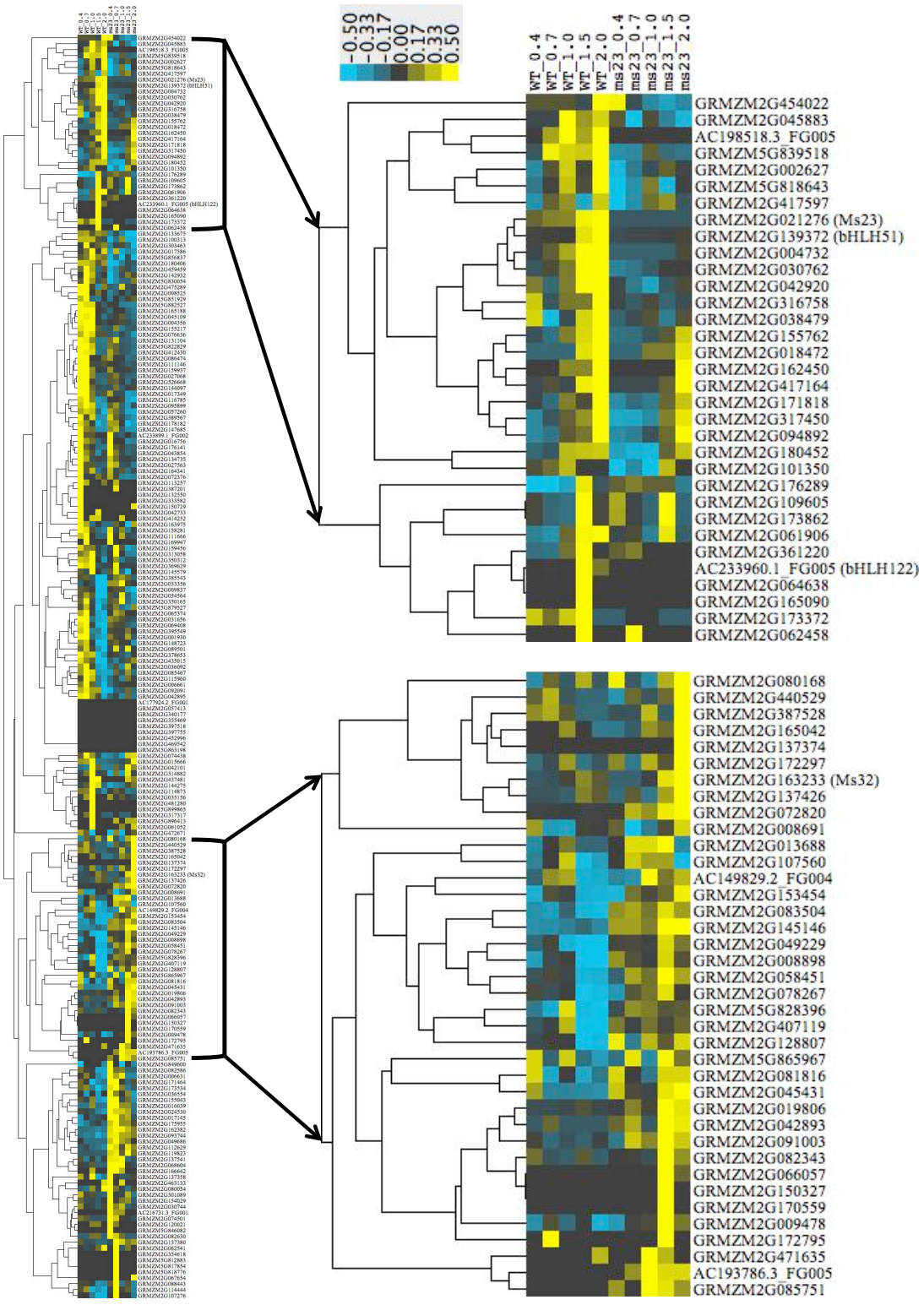


Fig. S11. K-median clustering of 213 maize bHLHs based on the expression patterns in both W23 fertile and *ms23-ref* anthers at stages 4, 7, 8, 9, and 10. The heat map was generated using Cluster 3.0 program based on Euclidean similarity and average hierarchical linkage.

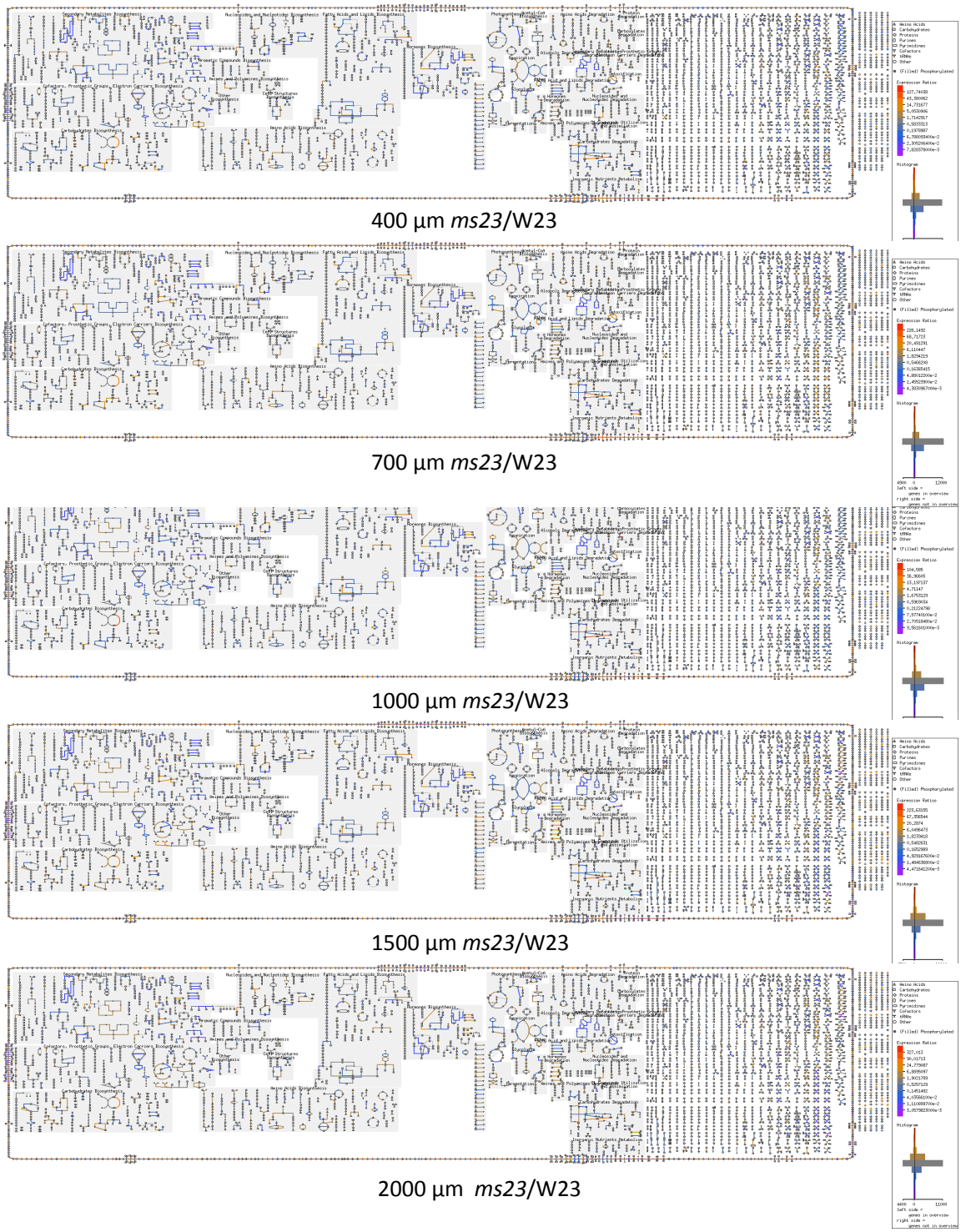


Fig. S12. Pathway mapping using the transcriptomic analysis tool (http://pathway.gramene.org/maizecyc.html) comparing ms23-ref and the W23 fertile inbred at the 400-, 700-, 1,000-, 1,500-, and 2,000-µm stages based on RNA-seq. Data are from the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession nos. GSE52290 described in Zhai et al., 2015).

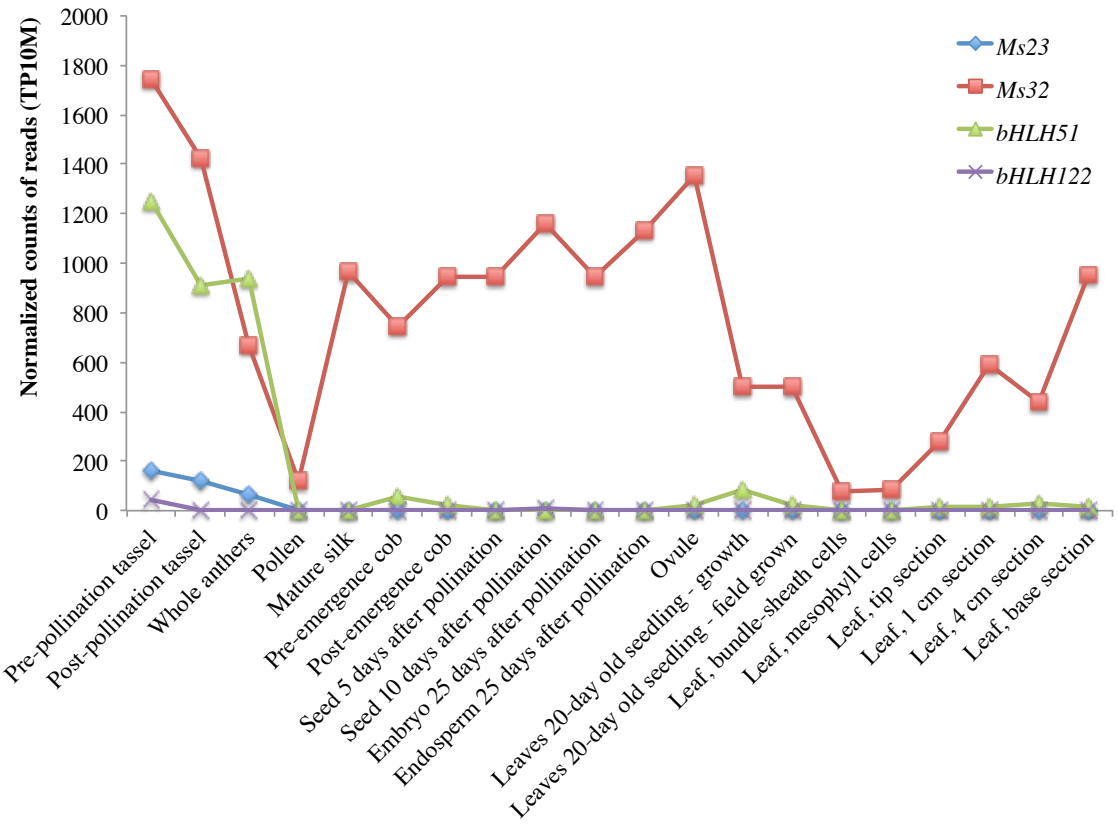


Fig. S13. Expression patterns of *Ms23*, *Ms32*, *bHLH51*, and *bHLH122* in various organs. Normalized sum of abundances from twenty vegetative and reproductive tissues across a wide range of developmental stages were compiled from published RNA-seq data. The total counts of RNA-seq reads in each gene model are normalized to 10 million reads (TP10M) (https://mpss.danforthcenter.org/dbs/index.php?SITE=maize_RNAseq).

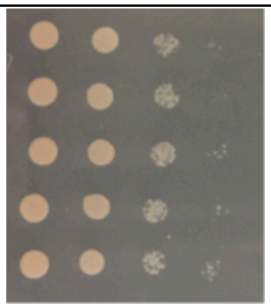
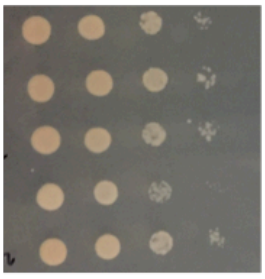
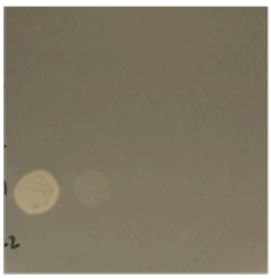
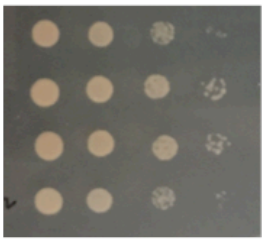
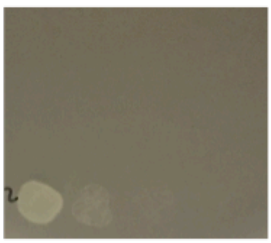
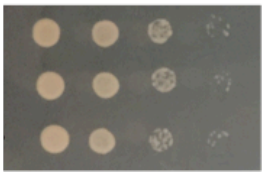
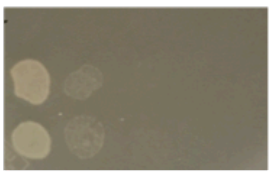
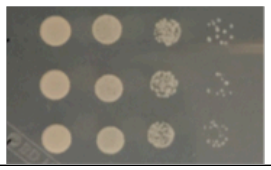
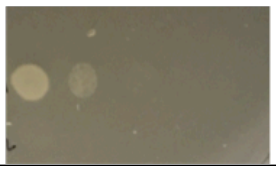
Bait	Prey	Low Stringency	High Stringency	Interaction
Empty	Empty			No
Empty	MS23			No
Empty	MS32			No
Empty	bHLH51			No
Empty	bHLH122			No
MS23	Empty			No
MS23	MS23			No
MS23	MS32			No
MS23	bHLH51			Yes
MS23	bHLH122			No
MS32	Empty			No
MS32	MS32			No
MS32	bHLH51			No
MS32	bHLH122			Yes
MS51	Empty			No
bHLH51	bHLH51			Yes
bHLH51	bHLH122			Yes
bHLH122	Empty			No
bHLH122	bHLH51			Yes
bHLH122	bHLH122			No

Fig. S14. Protein interactions determined based on yeast two-hybrid assays. Serial dilutions (1, 10^{-1}, 10^{-2}, 10^{-3}) of yeast two-hybird clones under low (-Leu, -Thr) and high (-Leu,-Thr,-His, -Ala) stringency conditions. Each bait and prey vector contains a bHLH domain from MS23, MS32, bHlH51, or bHLH122. Empty bait (pGBK) and empty prey (pGAD) were used as negative controls.

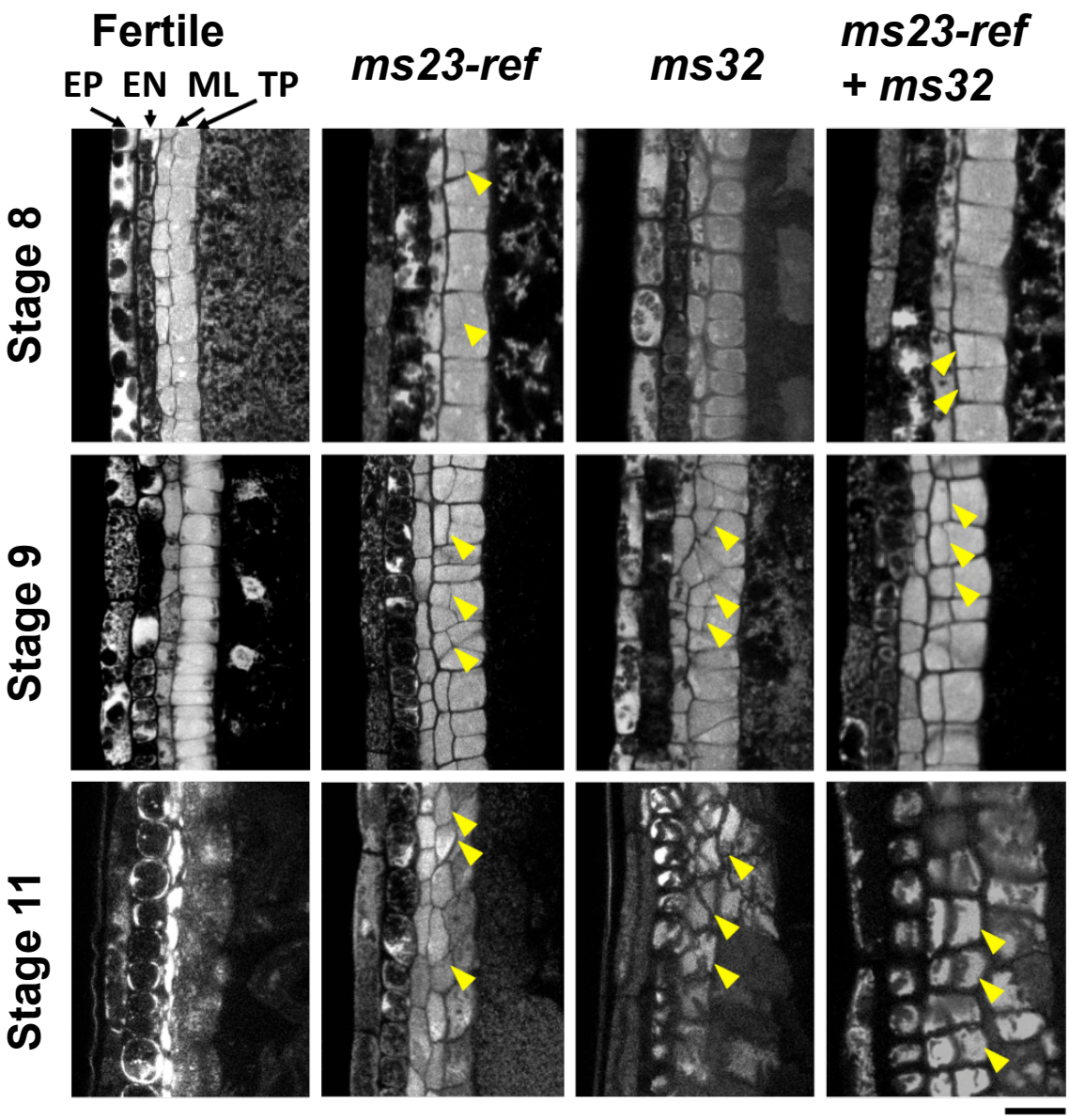


Fig. S15. Longitudinal confocal images of *ms23-ref*, *ms32*, and the *ms23-ref, ms32* double mutant at three developmental stages. Yellow triangles point to cells produced by extra periclinal divisions. Scale bar: 20 µm

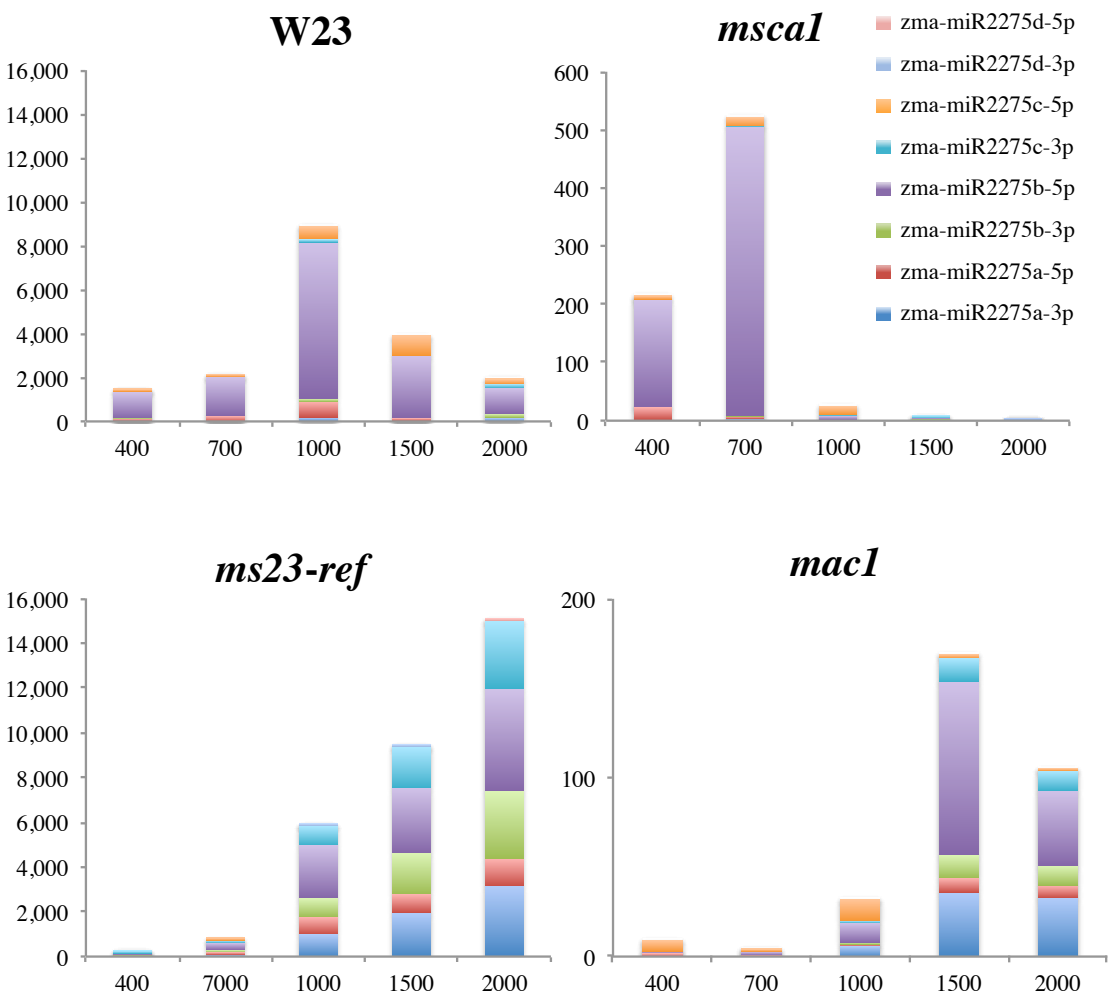


Fig. S16. miR2275 abundances in W23, *ms23-ref, msca1,* and *mac1* anthers based on small RNA-seq data. (A) The miR2275 profiles of 0.4-, 0.7-, 1.0-, 1.5-, and 2.0-mm anthers of homozygous *ms23-ref*, *msca1*, and *mac1* mutants comparing to the W23 fertile line. The total counts of RNA-seq reads are normalized to 10 million reads (TP10M) (Zhai et al., 2015).

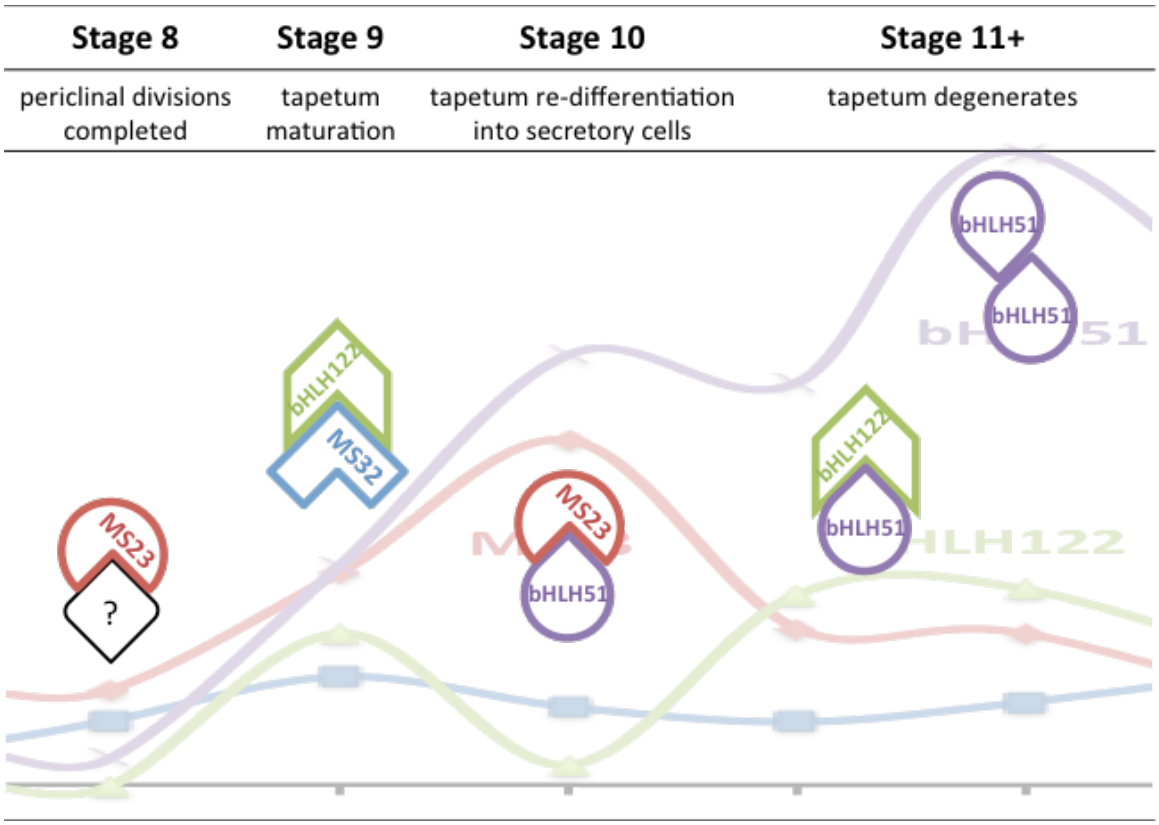


Fig. S17. A proposed "partner exchange" model of sequential steps of tapetal differentiation in maize. We propose that MS23 prevents periclinal division, likely as a heterodimer with an unknown factor to regulate the last periclinal division in the SPC and the pre-tapetal cells during their anticlinal proliferation in Stage 8. The MS23:unknown factor complex is required for tapetal cell fate. The MS32:bHLH122 complex acts to prevent excessive, aberrant divisions beyond Stage 9 after most tapetal cells are formed. The absence of bHLH122 at Stage 10 facilitates another partner exchange in which newly synthesized bHLH51 can interact with MS23, critical for the re-differentiation of the TP to serve its secretory roles at Stage 10. With bHLH122 reappears and MS23 levels off at Stage 11, bHLH51 is released to form new complexes (bHLH51:bHLH122 and bHLH51:bHLH51) for the post-meiotic, late development of the TP, e.g., tetrad callose dissolution, exine deposition onto pollen, maturation of somatic walls to permit pollen attachment, and ultimately programmed cell death of the TP.

Table S1. Comparison of the amino acid identity to MS23 with twelve other bHLH proteins important for tapetal functions in maize, rice, and Arabidopsis anthers.
MS23 protein sequence was used as the basis for the comparison.

Protein Name (species)	Total length (amino acids)	Amino acid identity (%)	bHLH domain identity (%)
MS23 (*Zea mays*)	365	100.00	100.00
TIP2 (*Oryza sativa)*	379	72.68	91.11
bHLH122 (*Zea mays*)	491	27.38	52.27
EAT1/DTD (*Oryza sativa)*	464	26.36	50.00
bHLH91 (*Arabidopsis thaliana*)	428	25.43	46.67
bHLH89 (*Arabidopsis thaliana*)	420	26.43	42.22
bHLH10 (*Arabidopsis thaliana*)	458	26.21	42.22
bHLH51 (*Zea mays*)	625	13.93	42.22
TDR (*Oryza sativa)*	551	14.12	42.22
AMS (*Arabidopsis thaliana*)	571	12.75	42.22
DYT1 (*Arabidopsis thaliana*)	207	19.25	37.78
MS32 (*Zea mays*)	219	15.46	35.56
UDT1 (*Oryza sativa)*	177	20.12	37.78

Table S2

Click here to Download Table S2

Table S3

Click here to Download Table S3

Table S4

Click here to Download Table S4

Table S5

Click here to Download Table S5

Table S6. Anther stages in maize, rice, and Arabidopsis. Based on the developmental events previously defined (Kelliher et al., 2014; Sanders et al., 1999; Zhang and Wilson, 2009), anther stages in maize, rice, and Arabidopsis are compared and listed. AR – archesporial; EN – endothecium; SPC – secondary parietal cell; PMC – pollen mother cells.

Events	Anther stages		
	Maize	Rice	Arabidopsis
Pluripotent primordium	1	1	1
AR cells forming	2	2	2
EN and SPC forming	3	3	3
Three-wall-layer stage with EN and SPC complete	4	4	3-4
SPC periclinal divisions starts	5	4	4
Anticlinal divisions to add anther girth	6	4	4
Final SPC periclinal divisions	7	4	4-5
AR cells mature to PMC	8	5	5
Meiosis starts	9	6	5-6
Entering Meiosis II	10	7	6-7
Meiosis complete	11	8	7

Tables S7 – S9

Click here to Download Tables S7 - S9