bioRxiv preprint doi: https://doi.org/10.1101/140806; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cloning, molecular evolution and functional characterization of ZmbHLH16, the maize ortholog of OsTIP2 (OsbHLH142)

Yongming Liu[1], Jia Li[2], Gui Wei[1], Yonghao Sun[3], Yanli Lu[1], Hai Lan[1], Chuan Li[1], Suzhi Zhang[1], Moju Cao[1]*

[1] Maize Research Institute, Sichuan Agricultural University, Chengdu, China

[2] Tropical Crops Genetic Resources Institute, Chinese Academic of Tropical Agricultural Sciences, Danzhou, China

[3] National Key Lab of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, China

*** Correspondence:**

Moju Cao

caomj@sicau.edu.cn

Abstract Basic helix-loop-helix (bHLH) transcription factors play key roles in plant male reproduction. More than 14 bHLH proteins related to pollen development have been cloned from rice and *Arabidopsis*. However, little is known about the role of the bHLH family in maize microspore development. In this study, the bHLH transcription factor ZmbHLH16 was cloned. ZmbHLH16 shares high similarity with the OsTIP2 (OsbHLH142) protein, a master regulator of the developmental coordination of male reproduction in rice. Expression characterization analysis showed that ZmbHLH16 is preferentially expressed in male reproductive organs and is located in the nucleus. Through nucleotide variation analysis, 36 polymorphic sites in ZmbHLH16, including 23 SNPs and 13 InDels, were detected among 78 maize inbred lines. Neutrality tests and linkage disequilibrium analysis showed that ZmbHLH16 experienced no significant evolutionary pressure. A yeast one-hybrid experiment showed that the first 80 residues in the N-terminus of ZmbHLH16 had transactivation activity, whereas the full length did not. To identify potential ZmbHLH16 interactors, 395 genes that shared similar expression patterns in a genome-wide search were obtained through coexpression analysis. Among these genes, the transcription factor ZmbHLH51 had an expression pattern and subcellular localization similar to those of ZmbHLH16. The interaction between ZmbHLH51 and ZmbHLH16 was verified in yeast cells. In addition to the typical bHLH domain, other regions of ZmbHLH16 were necessary and adequate for its heterodimerization with ZmbHLH51. Our results contribute to a solid foundation for further understanding the functions and mechanisms of ZmbHLH16.

Keywords: maize, bHLH transcription factor, microspore development, coexpression analysis, molecular evolution, male sterile

1 Introduction

Maize is the most important crop in the world for its utilization in food and industrial materials. At present, there is a rising demand for maize crop yields (Ray et al., 2013). Benefitting from hybrid vigor, male sterility can be used for hybrid maize seed production to increase crop yield. Therefore, the study on male sterility is of great value in application. Till now, several maize genic male sterile (GMS) genes such as *ms8* (Wang et al., 2013), *ms9* (Albertsen et al., 2016), *ms26* (Djukanovic et al., 2013), *ms32* (Moon et al., 2013), *ms44* (Fox et al., 2017) and ms*45* (Albertsen et al., 1993), have been cloned and illuminated for their abortion mechanism. These findings not only contribute to maize heterosis utilization but also expand our understanding of the maize male reproduction. Conventionally, genic male sterile genes are mainly identified through mutant analysis. With the

bioRxiv preprint doi: https://doi.org/10.1101/140806; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

development of gene-editing technology, more male sterile genes are now from the direct editing of some key genes involving pollen development in maize (Svitashev et al., 2015; Mark Cigan et al., 2016; Qi et al., 2016; Svitashev et al., 2016). As a result, the identification of key genes in male reproduction is becoming increasingly important. More than 10,000 genes have been shown to be expressed specifically in maize male fertility development (Ma et al., 2009). Transcription factors (TFs) play key roles in regulating their spatial and temporal-specific expression. Interestingly, TFs might also be the target genes of some small RNAs in plant meiotic processes (Chen, 2004; Alonsoperal et al., 2010; Yu et al., 2013). These above findings indicate that TFs play important roles in plant reproduction. In maize, a total of 2,298 TFs have been identified, and some show tissue-specific expression (Jiang et al., 2012). Key TFs in maize meiosis have been identified using high-throughput techniques such as microarray hybridization and transcriptome sequencing (Dukowic-Schulze et al., 2014a; Dukowic-Schulze et al., 2014b; Zhang et al., 2014). However, only two pollen development-related transcription factors—*ms9* (R2R3-MYB) and *ms32* (basic helix-loop-helix (bHLH))—have been cloned in maize (Moon et al., 2013; Albertsen et al., 2016). To reveal the regulatory mechanism of maize pollen formation, it is imperative to identify additional TFs involved in maize male fertility.

The bHLH proteins compose one of the largest plant transcription factor families. In rice and maize, 178 and 276 bHLH TFs have been identified respectively (Li et al., 2006; Carretero-Paulet et al., 2010; Jiang et al., 2012). Abnormal functions of some bHLH TFs may lead to plant male sterility. In Arabidopsis, ten bHLH proteins related to pollen development have been isolated: AtAMS (Sorensen et al., 2003; Xu et al., 2014), AtDYT1 (Zhang et al., 2006; Feng et al., 2012), AtbHLH10 (Zhu et al., 2015), AtbHLH89 (Zhu et al., 2015), AtbHLH91 (Zhu et al., 2015), AtJAM1 (Nakata et al., 2013), AtJAM2 (Nakata & Ohme-Takagi, 2013), AtJAM3 (Nakata & Ohme-Takagi, 2013), AtMYC5 (Figueroa & Browse, 2015) and AtBIM1 (Xing et al., 2013). Moreover, these male sterile mutants have unique male reproduction-deficient characteristics. For example, the *ams* mutant exhibits total male sterility without any visible pollen; the *dyt1* mutant can produce few pollen grains with a low rate of self-fertility; and the single mutants of *AtbHLH10*, *AtbHLH89*, and *AtbHLH91* develop normally, with only their various double and triple combinations defective in pollen development (Sorensen et al., 2003; Li et al., 2006; Zhu et al., 2015). These differences in male sterile characteristics might result from the functional divergence of bHLH TFs. Tapetal cells provide energy and nutrition for microspore development via programmed cell death (PCD) at appropriate anther developmental stages (Zhang et al., 2008). In rice, bHLH TFs including OsUDT1 (Jung et al., 2005), OsTDR (Li et al., 2006; Zhang et al., 2008), OsEAT1/OsDTD1 (Ji et al., 2013; Niu et al., 2013), and OsTIP2 (OsbHLH142) (Fu et al., 2014; Ko et al., 2014) have been found to be essential for anther tapetal cell development. These above rice bHLH TF dysfunctions lead tapetal cells to undergo abnormal PCD, thereby causing complete male sterility. In conclusion, all above show that bHLH TFs play important roles in regulating stamen development. At present, the study about bHLH TFs related to maize male fertility is few, there is a need to characterize more bHLH proteins in maize male reproduction.

TIP2 (OsbHLH142) acts as a key regulator of tapetum development in rice (Fu et al., 2014; Ko et al., 2014). The *tip2* mutant displays complete male sterility, with three undifferentiated anther wall layers (epidermal, fibrous and middle layer) and abortion of tapetal programmed cells death(Fu et al., 2014). In this study, the transcription factor ZmbHLH16, which is homologous to OsTIP2 (OsbHLH142),

bioRxiv preprint doi: https://doi.org/10.1101/140806; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

was isolated from maize. Its structure, phylogeny, expression and subcellular localization, molecular evolution, and regulatory characteristics were investigated.

2 Materials and methods

2.1 Plant materials

Spikelets from maize inbred line A619 were collected for ZmbHLH16 (GRMZM2G021276_T02) and ZmbHLH51 (GRMZM2G139372_T07) CDS (coding sequence) cloning. Ears, main stalks, stems, and spikelets were taken from maize inbred A619 for ZmbHLH16 expression analysis. Seeds from 78 inbred lines (see Supplementary Table S1) were used to amplify the genome sequence of ZmbHLH16.

2.2 DNA and RNA extraction

Genomic DNA was extracted from seeds using a modified cetyltrimethylammonium bromide (CTAB) method (Porebski et al., 1997). Total RNAs were isolated from the above frozen samples with TRIzol reagent (Takara, China) and DNase I to eliminate any genomic DNA. One microgram of total RNA from each sample was used to synthesize cDNA via the PrimeScript™ RT Reagent Kit (Takara, China).

2.3 CDS cloning of ZmbHLH16 and phylogenetic analysis

BlastP[1] was used to identify OsTIP2 (OsbHLH142) homologous genes in the maize genome (B73 assembly v3) with the amino acid sequence of OsTIP2 (OS01G0293100). The CDS of ZmbHLH16 was amplified from cDNA templates of A619 spikelets with the following primers: 5′-ATGTATCACCCGCAGTGCGAGCT-3′ and 5′-TGTACTCGTCCACCACTTCCAT-3′. High-fidelity KOD FX (Toyobo, Japan) was used for gene cloning according to the manufacturer's instructions. The purified PCR products were inserted into the pEASY Blunt Simple cloning vector (TransGen, China) and sequenced by Tsingke Biotech with an ABI 3730XL DNA Analyzer. The ZmbHLH16 amino acid sequence was acquired based on amplifying its CDS from A619 using the online program SoftBerry FGENESH[2], and its conserved domain was predicted using the NCBI CD tool[3]. Then, 16 bHLH TFs that were reported to be involved in microspore development were retrieved from Gramene[4] to construct a phylogenetic tree using the neighbor-joining method with MEGA v5.10 (Kumar et al., 2008), and the robustness of the findings was verified via 1000 bootstrap replicates. The accession numbers of the 16 TFs are as follows: AtAMS (AT2G16910), AtDYT1 (AT4G21330), AtbHLH10 (AT2G31220), AtbHLH89 (AT1G06170), AtbHLH91 (AT2G31210), AtJAM1 (AT2G46510), AtJAM2 (AT1G01260), AtJAM3 (AT4G16430), AtMYC5 (AT5G46830), AtBIM1 (AT5G08130), OsUDT1 (OS07G0549600), OsTDR1 (OS02G0120500), OsEAT1 (OS04G0599300), OsTIP2 (OS01G0293100), ZmMS32 (GRMZM2G163233) and SlMS1035 (Solyc02g079810).

2.4 Molecular evolution analysis of ZmbHLH16

[1] http://www.maizesequence.org
[2] http://www.linux1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind
[3] http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi
[4] http://www.gramene.org/

bioRxiv preprint doi: https://doi.org/10.1101/140806; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The genomic sequences of ZmbHLH16, including its 5′ and 3′ untranslated regions (UTRs), were amplified from 78 maize inbred lines (see Supplementary Table S1 for details) with the primers 5′-GGAAGGAGGAAACCAAGTCG-3′ and 5′-TGTAACGAGCAAGCGGATTTA-3′. PCR was performed according to the manufacturer's protocol using high-fidelity polymerase KOD FX (Toyobo, Japan). PCR-amplifying fragments were purified and sequenced directly using an ABI 3730XL DNA Analyzer manufactured by Tsingke Biotech. After ambiguous sequences were manually deleted, the sequence polymorphisms of ZmbHLH16 among the 78 maize inbred lines were analyzed using CodonCode Aligner 6.0.2 software (CodonCode Corporation, Dedham, MA). For molecular evolution analysis, certain parameters were calculated as follows: (1) the nucleotide diversity of common pairwise nucleotide difference per site (π) with DnaSP 5.0 (Librado & Rozas, 2009); (2) in neutrality tests, the evolutionary pressure in ZmbHLH16 via Tajima's D test (Tajima, 1989) and Fu and Li's statistics (Fu & Li, 1993); (3) the LD matrix of ZmbHLH16 was characterized by evaluating r^2 values based on SNPs and InDels (minor allele frequency (MAF) $\geq$0.05) in TASSEL 2.0 (Bradbury et al., 2007). An LD plot was obtained in Haploview 4.2 (Barrett et al., 2005), and the LD decay was assessed by averaging r^2 values with a distance of 250 bp.

2.5 Transactivation activity analysis of truncated ZmbHLH16

The ZmbHLH16 CDS contains 1098 bp encoding a protein with 365 amino acids. To investigate its transcriptional activating ability and retain its conserved bHLH domain, the ZmbHLH16 CDS sequence was divided into four parts; the first three parts each contained 240 bp (labeled A $^{1\text{-}80\ a.a.}$, B $^{81\text{-}160\ a.a.}$ and C $^{161\text{-}240\ a.a.}$), and the last part contained 375 bp (labeled D $^{241\text{-}365\ a.a.}$) (Fig 3A). Three new units—ZmbHLH16 (E) $^{1\text{-}160\ a.a.}$, (F) $^{81\text{-}240\ a.a.}$, and (G) $^{161\text{-}365\ a.a.}$—were constructed using combinations of two neighboring parts. The above seven parts, termed ZmbHLH16 (A)-(G), were artificially synthesized, and sequencing confirmation was performed by Tsingke Biotech. In total, eight fragments, including ZmbHLH16 CDS and ZmbHLH16 (A)-(G), were individually inserted into the pGBKT7 vector using the In-Fusion cloning method (Vazyme ClonExpress II One Step Cloning Kit, Vazyme Biotech Co., Ltd, China) to analyze their transcriptional activation activity (see Supplementary Table S2 for all primers used in the experiment). All recombinant pGBKT7 vectors were transformed into AH109 yeast strains (TIANDZ, China) via the lithium acetate-mediated approach. The transformants were cultivated on SD/-Trp medium for 2-3 days at 28°C. Bacterial PCR was used to identify positive clones. The positive clones were further cultured on SD/-His-Trp medium containing 50 mg/L χ-α-gal (Coolaber, China) for 2-4 days at 28°C to test their transactivation activity. The pGBKT7 and pGBKT7-GAL4 AD vectors were used as negative and positive controls, respectively.

2.6 Coexpression analysis and identification of maize male reproduction-related genes

For coexpression analysis, expression data of genome-wide maize genes in 20 tissues and 66 periods were obtained from q-teller[5], and the Pearson correlation coefficients (PCCs) between ZmbHLH16 and other genes were calculated. Cluster3.0 (de Hoon et al., 2004) was used for target gene (PCC>0.6) cluster analysis based on Euclidean distance and complete linkage. A heatmap was drawn using Java Treeview (Saldanha, 2004). Next, to gain deeper insight into the molecular mechanism underlying ZmbHLH16, all target genes (PCC>0.6) were queried with E-value<$1e^{-5}$ in the Plant

[5] http://qteller.com/maize2/

bioRxiv preprint doi: https://doi.org/10.1101/140806; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Male Reproduction Database[6], which contains 548 male-fertility-related genes in Arabidopsis. All maize gene sequences were retrieved from MaizeGDB[7]. To characterize the putative function of ZmbHLH16-coexpressed genes, GO terms for all target genes (PCC>0.6) were taken from AGRIgo[8], and GO enrichment analysis was performed using OmicShare tools[9].

2.7 Expression characteristics of ZmbHLH16 and ZmbHLH51

The primers for the semi-quantitative expression analysis of ZmbHLH16 were 5′-CCTCATGCACCTCATACC-3′ and 5′-CAGCTCCTGGATGTACTC-3′. At the same time, the primer sequences 5’-CTGGAGGTCACCAACGTCAA-3’ and 5’-AGCGAGTCCCTCAGTCTGTC-3’ were used for ZmbHLH51 expression analysis. The 18S gene in this experiment was used as the internal control, and its amplifying primers were 5‘-CTGAGAAACGGCTACCACA-3′ and 5‘-CCCAAGGTCCAACTACGAG-3′ (Hu et al., 2011).

The localization patterns of the ZmbHLH16 and ZmbHLH51 proteins were investigated using transient transformation in rice protoplasts. For ZmbHLH51 CDS cloning, we first amplified the cDNA sequence with the primers 5′-GAGCAGTGATGTGAATTGCG-3′ and 5′-TCAAGCGAGGTATTGGAGGA-3′ using high-fidelity KOD FX polymerase from A619 and inserted it into the pEASY blunt-cloning vector. The CDSs of ZmbHLH16 and ZmbHLH51 lacking the stop codons were individually fused to the N-terminus of enhanced green fluorescent protein (eGFP) in pCAMBIA2300-P_{35S} by subcloning using the In-Fusion cloning method. Two recombinants, pCAMBIA2300-P_{35S}:ZmbHLH16-eGFP and pCAMBIA2300-P_{35S}:ZmbHLH51-eGFP, were constructed to assess the localization of these proteins. The empty pCAMBIA2300-P_{35S}-eGFP vector was used as a control in this experiment. The recombinant vectors pCAMBIA2300-P_{35S}:ZmbHLH16-eGFP and pCAMBIA2300-P_{35S}:ZmbHLH51-eGFP and the control vector were transformed into rice protoplasts using polyethylene glycol (PEG), as described previously (Bart et al., 2006). The green signals (Ex=488 nm, Em=507 nm) were detected using a TCS-SP8 fluorescence microscope (Leica, Germany).

2.8 Protein–protein interactions

To confirm the interaction between ZmbHLH16 and ZmbHLH51, a yeast two-hybrid assay was conducted. The CDSs of ZmbHLH16 and ZmbHLH51 were inserted into the pGBKT7 and pGADT7 vectors, respectively. The pGBKT7-ZmbHLH16 vector without autoactivation activity was constructed as above. The pGADT7-ZmbHLH51 vector was constructed using the In-Fusion cloning method by subcloning from pCAMBIA2300-P_{35S}:ZmbHLH51. The recombinant vectors pGBKT7-ZmbHLH16 and pGADT7-ZmbHLH51 were co-transformed into AH109 yeast competent cells according to operating instructions. The transformants were cultivated on SD/-Leu-Trp medium at 28°C for 2-3 days, and positive clones were confirmed using PCR. Positive clones were further cultured on SD/-Ade-His-Leu-Trp medium containing 50 mg/L χ-α-gal at 28°C for 2–3 days. The vectors pGBKT7-T and pGBKT7-Lam were used as positive and negative controls, respectively. After confirming the interaction between ZmbHLH16 and ZmbHLH51, to investigate the interaction

[6] http://202.120.45.92/addb/
[7] ftp://ftp.ensemblgenomes.org/pub/plants/release-29/fasta/zea_mays
[8] http://bioinfo.cau.edu.cn/agriGO/analysis.php
[9] http://www.omicshare.com/tools

bioRxiv preprint doi: https://doi.org/10.1101/140806; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

196 domain in ZmbHLH16, regions of ZmbHLH16 without autoactivation activity were inserted into the
197 bait vector pGBKT7 and then co-transformed with the prey vector pGADT7-ZmbHLH51 into the
198 AH109 yeast competent cells.

199 **3 Results**

200 **3.1 ZmbHLH16 is a typical bHLH transcription factor**

201 TIP2 is a key regulator involved in rice anther development. BlastP analysis showed that the gene
202 model GRMZM2G021276 (ZmbHLH16) had the highest homology score with OsTIP2
203 (OsbHLH142). The ZmbHLH16 CDS was obtained from the maize inbred line A619. Further
204 analysis showed that the ZmbHLH16 coding sequence contained 1098 bp encoding a protein of 365
205 amino acids that had 66.85% amino acid sequence identity with OsTIP2 (OsbHLH142). Analysis
206 with NCBI CD software revealed that the amino acid sequence of ZmbHLH16 included a typical
207 bHLH DNA-binding domain (Fig 1A). The bHLH interaction and function (BIF) domain participates
208 in bHLH protein localization, transcriptional activity and dimerization (Cui et al., 2016). The BIF
209 domain was also found in the C-terminal (288-363 a.a.) region of ZmbHLH16. Furthermore,
210 phylogenetic analysis of 17 bHLH transcription factors related to microspore development illustrated
211 that these bHLH TFs were highly conserved, with most bootstrap values >90%, and that ZmbHLH16
212 was the most closely related to OsTIP2 (Fig 1B). Together, the above results indicated that
213 ZmbHLH16 protein included the typically conserved domain of bHLH TFs and might play a crucial
214 role in male reproduction in maize.

215 **3.2 ZmbHLH16 is highly evolutionarily conserved in maize germplasm**

216 To analyze its molecular evolution, the DNA sequences of ZmbHLH16 were amplified from 78
217 maize inbred lines. The ZmbHLH16 genomic sequence is divided into seven regions with a length of
218 2,514 bp (Table 1). Based on the MAF≥0.05, 36 polymorphism sites within ZmbHLH16 were
219 identified, including 23 SNPs and 13 InDels, with one SNP/InDel per 109/193 bp (Table 1). Among
220 23 SNPs, 13 (57%) and 10 (43%) involved transitions and transversions, respectively. Further
221 analysis showed that there were 4 amino acid variations in ZmbHLH16 among 78 inbred lines, with
222 3 mutations in the first exon and the fourth mutation in the second exon. Comparison analysis
223 showed that nucleotide variations were not evenly distributed in ZmbHLH16, and introns had higher
224 sequence diversity (3.1 polymorphisms/100 bp) than UTRs (1.24 polymorphisms/100 bp) and exons
225 (1.18 polymorphisms/100 bp). Moreover, a nucleotide polymorphism test in DnaSP showed the
226 highest nucleotide diversity ratio ($\pi=6.15\times10^{-3}$) in the first intron but no significant nucleotide
227 variation in the 3rd intron or 3'-UTR (Table 2).

228 To investigate whether ZmbHLH16 has experienced selection pressure, various regions of
229 ZmbHLH16 were assessed (Table 2). In Tajima's D test and Fu and Li's test, no significant
230 difference was obtained across all regions of ZmbHLH16. This result indicated that ZmbHLH16
231 experienced no significant selective pressure and underwent neutral selection. Elevated LD is usually
232 expected for genes under selection (Bomblies & Doebley, 2005). Thus, to further confirm whether
233 ZmbHLH16 underwent directional selection, its LD patterns and LD decay were also calculated. In
234 the LD matrix, no obvious LD block was found in the ZmbHLH16 genome sequence (Fig 2A). A
235 schematic diagram of LD decay represented by plots of r^2 showed that the LD level dropped to 0.1 at

bioRxiv preprint doi: https://doi.org/10.1101/140806; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

approximately 1,300 bp (Fig 2B), indicating a more rapid decay rate than the average 1.5 kb of several genes under selection in maize (Remington et al., 2001). Therefore, our above results also reflected the conserved evolution of ZmbHLH16 in the maize germplasm.

3.3 Only the N-terminal first 80 residues of ZmbHLH16 have transactivation activity

To identify the activating function of ZmbHLH16, eight fragments of ZmbHLH16 were analyzed in yeast (Fig 3A). Yeast cells with pGBKT7-ZmbHLH16(A) $^{1\text{-}80\ a.a.}$ or pGBKT7-ZmbHLH16(E) $^{1\text{-}160a.a.}$ grew normally on both SD/-Trp and SD/-His-Trp selective media and turned the indicator blue. However, the other six yeast transformants with ZmbHLH16 could only live on the SD/-Trp medium (Fig 3B). In comparison with the living conditions of the transformants containing ZmbHLH16 (A) $^{1\text{-}80a.a.}$, (B) $^{81\text{-}160a.a.}$ and (E) $^{1\text{-}160a.a.}$, this finding suggested that the N-terminal first eighty amino acids of ZmbHLH16 possessed transcriptional activation activity. Based on a comparison with the whole coding region (1-365 a.a.) and region (E) $^{1\text{-}160\ a.a.}$, it was inferred that domain (G) $^{161\text{-}365\ a.a.}$ of ZmbHLH16 might inhibit its transcriptional activation activity. However, the deficiency of transactivation activity of the full-length ZmbHLH16 might also be due to its poor expression in yeast. In conclusion, the above results indicated that the first 80 amino acids in the N-terminus of bHLH16 could activate transcription in yeast, whereas the full-length version did not.

3.4 ZmbHLH16 coexpresses with many male reproduction-related genes

Functionally associated genes are more likely to share similar expression patterns (Fu & Xue, 2010). Coexpression analysis was therefore conducted to identify potential ZmbHLH16 cooperators. A total of 395 maize genes with calculated PCC values above 0.6 were found (Fig 4A and Supplementary Table S3). Among them were three male sterile genes, including *ms8* (GRMZM2G119265), *ms26* (GRMZM2G091822) and *ms44* (AC225127.3_FG003), which shared expression PCC values of 0.9937, 0.8375 and 0.9964 with ZmbHLH16, respectively. In a search of the PMRD database, in maize whole genome we identified 2405 (6.0%) genes showing homologous to Arabidopsis male fertility genes (Supplementary Table S4). In comparison, a higher ratio of 8.8% (35/395) at a significant level (χ^2=4.92, p=0.0265) was got when ZmbHLH16 coexpressed genes were queried (Table 3). The similar expression pattern to a number of plant reproduction-related genes indicated that ZmbHLH16 might be closely associated with maize male fertility.

Next, the 395 coexpressed genes were subjected to GO term analysis (Fig 4B and Supplementary Table S3). In the cellular component, 155 GO terms were enriched and most of these genes were categorized under cells, cell parts, membranes and organelles. For the molecular function category, binding and catalytic activity were the most abundant subcategories. Similarly, previous reports have confirmed that the binding activity and catalytic activity functions are essential for alterations in male fertility (Qu et al., 2015; Zhu et al., 2015; Mei et al., 2016). For biological processes, there were 780 enriched GO terms and cellular processes, metabolic processes, and single-organism processes were the most abundant clusters. Through hypergeometric test at the 0.05 significance level, it was found that some enriched Go terms such as pollen wall assembly, pollen exine formation, pollen development, gametophyte development reached to significant levels compared with the maize background (Supplementary Table S3). These above results supported that ZmbHLH16 might participate in maize pollen formation. Moreover, in the reproduction GO term (GO:0000003), a

bioRxiv preprint doi: https://doi.org/10.1101/140806; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bHLH transcription factor family member, ZmbHLH51 (GRMZM2G139372), was found, which shared a PCC score of 0.8990 with ZmbHLH16. ZmbHLH51 was homologous to the male sterile gene OsTDR. Accordingly, ZmbHLH51 might be an important factor in maize pollen development. Some studies have indicated that the interactions among bHLH TFs are important for pollen development (Niu et al., 2013; Zhu et al., 2015). Therefore, we next aimed to analyze the interaction between ZmbHLH16 and ZmbHLH51.

3.5 ZmbHLH16 and ZmbHLH51 have similar expression characteristics

The expression patterns of ZmbHLH16 and ZmbHLH51 were simultaneously analyzed using semi-quantitative PCR for reproductive and vegetative organs. Both ZmbHLH16 and ZmbHLH51 showed a higher expression level in spikelets than organs (Fig 5A). This finding indicated that ZmbHLH16 and ZmbHLH51 might be closely associated with maize male fertility.

Based on the above results, ZmbHLH51 is homologous to the male sterile gene *OsTDR* and might interact with ZmbHLH16. Therefore, the subcellular localization of both ZmbHLH16 and ZmbHLH51 was investigated. As depicted in Fig 5B, the recombinant fusion proteins ZmbHLH16-eGFP and ZmbHLH51-eGFP were only located in the nucleus, whereas the control eGFP was localized to both the cytoplasm and the nucleus. The similar expression profiles and protein localization patterns between ZmbHLH16 and ZmbHLH51 supported their interaction.

3.6 ZmbHLH51 interacts with ZmbHLH16

Because the aforementioned results indicated that the two bHLH TFs ZmbHLH16 and ZmbHLH51 had similar expression characteristics and subcellular localization patterns, a yeast two-hybrid assay was used to verify the interaction between the two proteins. As shown in Fig 6A, yeast cells containing pGBKT7-ZmbHLH16 and pGADT7-ZmbHLH51 were able to grow on both SD/-Leu-Trp and SD/-Ade-His-Leu-Trp media, similar to the positive control. The above results confirmed the interaction between ZmbHLH16 and ZmbHLH51.

To map the domains involved in the ZmbHLH16-ZmbHLH51 interaction, fragments without transcriptional activation activity, including ZmbHLH16 (B) $^{81\text{-}160\ a.a.}$, (C) $^{161\text{-}240\ a.a.}$, (D) $^{241\text{-}365\ a.a.}$, (F) $^{81\text{-}240\ a.a.}$ and (G) $^{161\text{-}365\ a.a.}$, were further analyzed using yeast two-hybrid assays. The conserved bHLH domain is reported to participate in protein homo- or heterodimerization (Pires & Dolan, 2010). As expected, the regions containing the bHLH domain, i.e., (C) $^{161\text{-}240\ a.a.}$, (F) $^{81\text{-}240\ a.a.}$ and (G) $^{161\text{-}365\ a.a.}$, could grow normally on SD/-Ade-His-Leu-Trp media and turned the media blue (Fig 6B). Interestingly, transformants (B) and (D), which lacked the bHLH domain, also survived on the SD/-Ade-His-Leu-Trp synthetic dropout medium. These results indicated that not only the bHLH domain but also other regions in ZmbHLH16 were sufficient and necessary for its heterodimerization with ZmbHLH51.

4 Discussion

Male reproduction is a complicated process in plants that involves thousands of genes and many biological processes (Dukowic-Schulze & Chen, 2014; Zhou & Pawlowski, 2014; Rutley & Twell, 2015). Some genes involved in plant anther and pollen development are conserved in plants (Gómez

bioRxiv preprint doi: https://doi.org/10.1101/140806; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

et al., 2015). This phenomenon provides the possibility of elucidating key genes in other species based on homology analysis. Thus, here, we isolated ZmbHLH16 based on homology cloning from OsTIP2, which has been reported to be a master regulator of pollen formation.

In this study, the molecular evolution of ZmbHLH16 was investigated. In the analysis of selective pressure, no significant signal was found in ZmbHLH16 according to Tajima's D and Fu and Li's tests. Moreover, a lower nucleotide diversity ratio ($\pi=2.58\times10^{-3}$) was observed in all regions of ZmbHLH16 than in the average ($\pi=6.3\times10^{-3}$) of 18 maize genes in previous reports (Ching et al., 2002). This finding implied weak or no natural selection pressure on ZmbHLH16 and provided evidence that ZmbHLH16 is highly evolutionarily conserved in maize. The target gene sequence polymorphism also reflects evolutionary pressure during maize improvement (Wang et al., 2005). Previous studies found one polymorphic site per 60.8 bp in maize (Ching et al., 2002). In the present experiment, a lower frequency was obtained for ZmbHLH16 in 78 maize inbred lines (one SNP or InDel every 69.8 bp). The global LD decay of ZmbHLH16 investigated in our study ($r^2<0.1$ within 1,300 bp) was also less than the average intragenic level ($r^2<0.1$ within 1,500 bp) (Remington et al., 2001). The above nucleotide polymorphism testing results confirmed the conserved evolution of ZmbHLH16. The conserved molecular evolution of ZmbHLH16 further hinted at its crucial function in maize male reproduction.

Most bHLH proteins consist of a classical helix-loop-helix (HLH) domain to form homo- or heterodimers with other HLH-proteins to regulate downstream target genes (Murre et al., 1989). bHLH-bHLH or bHLH-MYB complexes have been reported to be involved in plant fertility (Niu et al., 2013; Qi et al., 2015; Chen et al., 2016). Our experiments showed that ZmbHLH16 lacks the ability of transcriptional activation. Thus, we speculate that ZmbHLH16 might regulate target gene expression by interacting with other proteins. One of its interacting factors, ZmbHLH51, was identified and confirmed using genome-wide coexpression and yeast two-hybrid analyses. In rice, the BIF domain is necessary for DYT1-bHLH protein dimerization (Cui et al., 2016). The present study showed that the BIF domain is also present in ZmbHLH16(D) $^{241\text{-}365\ a.a.}$ and participates in the interaction between ZmbHLH16 and ZmbHLH51. Previous studies related to bHLH proteins mainly focused on their conserved domains. In the investigation of the interaction between ZmbHLH16 and ZmbHLH51, we noticed that not only the conserved bHLH and BIF domains but also the ZmbHLH16(B) $^{81\text{-}160\ a.a.}$ region could form heterodimers with ZmbHLH51. Moreover, the ZmbHLH16(G) $^{161\text{-}365\ a.a.}$ fragment may have a negative effect on activation, leading to a reduced transcription activation capacity of the full-length ZmbHLH16 protein. Taken together, our findings provide new evidence that in bHLH proteins, other regions are of importance in molecular function in addition to the typical bHLH and BIF domains.

Coincidently, it was recently reported that ZmbHLH16 was a candidate gene for the ms23 mutant (Nan et al., 2017). The tapetal layer of the ms23 mutant undergoes abnormal periclinal division instead of tapetal differentiation (Chaubal et al., 2000). These results strongly support the hypothesis that ZmbHLH16 is involved in tapetal specification and maturation. In Nan's paper (Nan et al., 2017), the researchers mainly focused on the abortion mechanism in the *ms23* mutant, combining RNA-seq with proteomics data. These authors also confirmed the interaction between ZmbHLH16 and ZmbHLH51, similar to our result. In contrast, we paid more attention to the ZmbHLH16 nucleotide polymorphisms, molecular evolution, expression features, subcellular location and regulatory

bioRxiv preprint doi: https://doi.org/10.1101/140806; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mechanisms. Through coexpression analysis, a group of genes potentially involved in maize male reproduction were also revealed in this study. Our results might help uncover the mechanism of ZmbHLH16 regulating the pollen abortion in the ms23 mutant.

Funding

This work was supported by grants from the National Key Research and Development Program of China (2016YFD0101206 and 2016YFD0102104) and the Platform for Mutation Breeding by Radiation in Sichuan (2016NZ0106).

Acknowledgments

We thank the Chinese Maize Industry Technology System for providing maize inbred lines in this experiment with the aid of Prof. Guangtang Pan and Prof. Lujiang Li. We sincerely thank Prof. Shibin Gao for providing the genomic DNA of several maize inbred lines. We also thank Prof. Yufeng Hu for providing the original vectors. We are very grateful to Dr. Yibing Yuan and Jingtao Qu for assisting in data analysis.

References:

Albertsen, M.C., Fox, T., Leonard, A., LI, B., Loveland, B., and Trimnell, M. 2016. CLONING AND USE OF THE MS9 GENE FROM MAIZE. US Patent 20,160,024,520.

Albertsen, M.C., Trimnell, M.R., and Fox, T.W. 1993. Tagging, cloning and characterizing a male fertility gene in maize. *American Journal of Botany* 80.

Alonsoperal, M.M., Li, J., Li, Y., Allen, R.S., Schnippenkoetter, W., Ohms, S., White, R.G., and Millar, A.A. 2010. The MicroRNA159-Regulated GAMYB-like genes inhibit growth and promote programmed Cell Death in Arabidopsis. *Plant Physiology* 154:757-771.

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263-265.

Bart, R., Chern, M., Park, C.J., Bartley, L., and Ronald, P.C. 2006. A Novel System For Gene Silencing Using Sirnas In Rice Leaf And Stem-Derived Protoplasts. *Plant Methods* 2:292-307.

Bomblies, K., and Doebley, J.F. 2005. Molecular evolution of FLORICAULA/LEAFY orthologs in the Andropogoneae (Poaceae). *Molecular Biology and Evolution* 22:1082-1094.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., and Buckler, E.S. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633-2635.

Carretero-Paulet, L., Galstyan, A., Roig-Villanova, I., Martinez-Garcia, J.F., Bilbao-Castro, J.R., and Robertson, D.L. 2010. Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in Arabidopsis, poplar, rice, moss, and algae. *Plant Physiology* 153:1398-1412.

Chaubal, R., Zanella, C., Trimnell, M.R., Fox, T.W., Albertsen, M.C., and Bedinger, P. 2000. Two male-sterile mutants of Zea Mays (Poaceae) with an extra cell division in the anther wall. *American Journal of Botany* 87:1193-1201.

Chen, X. 2004. A MicroRNA as a Translational Repressor of APETALA2 in Arabidopsis Flower Development. *Science* 303:2022-2025.

Chen, X., Huang, H., Qi, T., Liu, B., and Song, S. 2016. New perspective of the bHLH-MYB complex in jasmonate-regulated plant fertility in arabidopsis. *Plant signaling & behavior* 11:e1135280.

Ching, A., Caldwell, K.S., Jung, M., Dolan, M., Smith, O.S., Tingey, S., Morgante, M., and Rafalski, A.J. 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genetics* 3:19.

Cui, J., You, C., Zhu, E., Huang, Q., Ma, H., and Chang, F. 2016. Feedback Regulation of DYT1 by Interactions with

bioRxiv preprint doi: https://doi.org/10.1101/140806; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Downstream bHLH Factors Promotes DYT1 Nuclear Localization and Anther Development. *The Plant Cell* 28.

de Hoon, M.J., Imoto, S., Nolan, J., and Miyano, S. 2004. Open source clustering software. *Bioinformatics* 20:1453-1454.

Djukanovic, V., Smith, J., Lowe, K., Yang, M., Gao, H., Jones, S., Nicholson, M.G., West, A., Lape, J., and Bidney, D. 2013. Male-sterile maize plants produced by targeted mutagenesis of the cytochrome P450-like gene (MS26) using a re-designed I–CreI homing endonuclease. *The Plant Journal* 76:888-899.

Dukowic-Schulze, S., Harris, A., Li, J., Sundararajan, A., Mudge, J., Retzel, E.F., Pawlowski, W.P., and Chen, C. 2014. Comparative transcriptomics of early meiosis in Arabidopsis and maize. *Journal of Genetics and Genomics* 41:139-152.

Dukowic-Schulze, S., Sundararajan, A., Mudge, J., Ramaraj, T., Farmer, A.D., Wang, M., Sun, Q., Pillardy, J., Kianian, S., and Retzel, E.F. 2014. The transcriptome landscape of early maize meiosis. *BMC Plant Biology* 14:1-15.

Dukowic-Schulze, S., and Chen, C. 2014. The meiotic transcriptome architecture of plants. *Frontiers in Plant Science* 5:220.

Feng, B., Lu, D., Ma, X., Peng, Y., Sun, Y., Ning, G., and Ma, H. 2012. Regulation of the *Arabidopsis* anther transcriptome by DYT1 for pollen development. *The Plant Journal* 72:612-624.

Figueroa, P., and Browse, J. 2015. Male sterility in Arabidopsis induced by overexpression of a MYC5-SRDX chimeric repressor. *The Plant Journal* 81:849-860.

Fox, T., DeBruin, J., Haug Collet, K., Trimnell, M., Clapp, J., Leonard, A., Li, B., Scolaro, E., Collinson, S., and Glassman, K. 2017. A single point mutation in Ms44 results in dominant male sterility and improves nitrogen use efficiency in maize. *Plant Biotechnology Journal.* 10.1111/pbi.12689

Fu, F.F., and Xue, H.W. 2010. Coexpression analysis identifies Rice Starch Regulator1, a rice AP2/EREBP family transcription factor, as a novel rice starch biosynthesis regulator. *Plant Physiology* 154:927-938.

Fu, Y., and Li, W. 1993. Statistical tests of neutrality of mutations. *Genetics* 133:693-709.

Fu, Z., Yu, J., Cheng, X., Zong, X., Xu, J., Chen, M., Li, Z., Zhang, D., and Liang, W. 2014. The rice basic helix-loop-helix transcription factor TDR INTERACTING PROTEIN2 is a central switch in early anther development. *The Plant Cell* 26:1512-1524.

Gómez, J.F., Talle, B., and Wilson, Z.A. 2015. Anther and pollen development: A conserved developmental pathway. *Journal of Integrative Plant Biology* 57:876-891.

Hu, Y., Li, Y., Zhang, J., Liu, H., Chen, Z., and Huang, Y. 2011. PzsS3a, a novel endosperm specific promoter from maize (Zea mays L.) induced by ABA. *Biotechnology Letters* 33:1465-1471.

Ji, C., Li, H., Chen, L., Xie, M., Wang, F., Chen, Y., and Liu, Y.G. 2013. A Novel Rice bHLH Transcription Factor, DTD, Acts Coordinately with TDR in Controlling Tapetum Function and Pollen Development. *Molecular Plant* 6:1715-1718.

Jiang, Y., Zeng, B., Zhao, H., Zhang, M., Xie, S., and Lai, J. 2012. Genome-wide transcription factor gene prediction and their expressional tissue-specificities in Maize. *Journal of Integrative Plant Biology* 54:616-630.

Jung, K., Han, M., Lee, Y., Kim, Y., Hwang, I., Kim, M., Kim, Y., Nahm, B.H., and An, G. 2005. Rice Undeveloped Tapetum1 is a major regulator of early tapetum development. *The Plant Cell* 17:2705-2722.

Ko, S., Li, M., Ku, M.S., Ho, Y., Lin, Y., Chuang, M., Hsing, H., Lien, Y., Yang, H., and Chang, H. 2014. The bHLH142 transcription factor coordinates with TDR1 to modulate the expression of EAT1 and regulate pollen development in rice. *The Plant Cell* 26:2486-2504.

Kumar, S., Nei, M., Dudley, J., and Tamura, K. 2008. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics* 9:299-306.

Li, N., Zhang, D., Liu, H., Yin, C., Li, X., Liang, W., Yuan, Z., Xu, B., Chu, H., and Wang, J. 2006. The rice tapetum

bioRxiv preprint doi: https://doi.org/10.1101/140806; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

degeneration retardation gene is required for tapetum degradation and anther development. *The Plant Cell* 18:2999-3014.

Li, X., Duan, X., Jiang, H., Sun, Y., Tang, Y., Yuan, Z., Guo, J., Liang, W., Chen, L., and Yin, J. 2006. Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and Arabidopsis. *Plant Physiology* 141:1167-1184.

Librado, P., and Rozas, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451-1452.

Ma, J., Skibbe, D.S., Fernandes, J., and Walbot, V. 2009. Male reproductive development: gene expression profiling of maize anther and pollen ontogeny. *Genome Biology* 9:R181. 10.1186/gb-2008-9-12-r181

Mark Cigan, A., Singh, M., Benn, G., Feigenbutz, L., Kumar, M., Cho, M.J., Svitashev, S., and Young, J. 2016. Targeted mutagenesis of a conserved anther-expressed P450 gene confers male sterility in monocots. *Plant Biotechnology Journal*. 10.1111/pbi.12633

Mei, S., Liu, T., and Wang, Z. 2016. Comparative Transcriptome Profile of the Cytoplasmic Male Sterile and Fertile Floral Buds of Radish (Raphanus sativus L.). *International Journal of Molecular Sciences* 17:42. 10.3390/ijms17010042

Moon, J., Skibbe, D., Timofejeva, L., Wang, C.R., Kelliher, T., Kremling, K., Walbot, V., and Cande, W.Z. 2013. Regulation of cell divisions and differentiation by MALE STERILITY32 is required for anther development in maize. *The Plant Journal* 76:592-602.

Murre, C., McCaw, P.S., Vaessin, H., Caudy, M., Jan, L.Y., Jan, Y.N., Cabrera, C.V., Buskin, J.N., Hauschka, S.D., and Lassar, A.B. 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* 58:537-544.

Nakata, M., Mitsuda, N., Herde, M., Koo, A.J., Moreno, J.E., Suzuki, K., Howe, G.A., and Ohme-Takagi, M. 2013. A bHLH-type transcription factor, ABA-INDUCIBLE BHLH-TYPE TRANSCRIPTION FACTOR/JA-ASSOCIATED MYC2-LIKE1, acts as a repressor to negatively regulate jasmonate signaling in Arabidopsis. *The Plant Cell* 25:1641-1656.

Nakata, M., and Ohme-Takagi, M. 2013. Two bHLH-type transcription factors, JA-ASSOCIATED MYC2-LIKE2 and JAM3, are transcriptional repressors and affect male fertility. *Plant signaling & behavior* 8:e26473.

Nan, G., Zhai, J., Arikit, S., Morrow, D., Fernandes, J., Mai, L., Nguyen, N., Meyers, B.C., and Walbot, V. 2017. MS23, a master basic helix-loop-helix factor, regulates the specification and development of tapetum in maize. *Development* 144:163-172.

Niu, N., Liang, W., Yang, X., Jin, W., Wilson, Z.A., Hu, J., and Zhang, D. 2013. EAT1 promotes tapetal cell death by regulating aspartic proteases during male reproductive development in rice. *Nature Communications* 4:1445.

Pires, N., and Dolan, L. 2010. Origin and diversification of basic-helix-loop-helix proteins in plants. *Molecular Biology and Evolution* 27:862-874.

Porebski, S., Bailey, L.G., and Baum, B.R. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter* 15:8-15.

Qi, T., Huang, H., Song, S., and Xie, D. 2015. Regulation of jasmonate-mediated stamen development and seed production by a bHLH-MYB complex in Arabidopsis. *The Plant Cell* 27:1620-1633.

Qi, W., Zhu, T., Tian, Z., Li, C., Zhang, W., and Song, R. 2016. High-efficiency CRISPR/Cas9 multiplex gene editing using the glycine tRNA-processing system-based strategy in maize. *BMC Biotechnology* 16:58.

Qu, C., Fu, F., Liu, M., Zhao, H., Liu, C., Li, J., Tang, Z., Xu, X., Qiu, X., Wang, R., and Lu, K. 2015. Comparative Transcriptome Analysis of Recessive Male Sterility (RGMS) in Sterile and Fertile Brassica napus Lines. *PLoS One* 10:e144118. 10.1371/journal.pone.0144118

Ray, D.K., Mueller, N.D., West, P.C., and Foley, J.A. 2013. Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLoS One* 8:e66428.

Remington, D.L., Thornsberry, J.M., Matsuoka, Y., Wilson, L.M., Whitt, S.R., Doebley, J., Kresovich, S., Goodman,

bioRxiv preprint doi: https://doi.org/10.1101/140806; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

M.M., and Buckler, E.S. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc Natl Acad Sci U S A* 98:11479-11484.

Rutley, N., and Twell, D. 2015. A decade of pollen transcriptomics. *Plant Reproduction* 28:73-89.

Saldanha, A.J. 2004. Java Treeview-extensible visualization of microarray data. *Bioinformatics* 20:3246-3248.

Sorensen, A., Krober, S., Unte, U.S., Huijser, P., Dekker, K., and Saedler, H. 2003. The Arabidopsis ABORTED MICROSPORES (AMS) gene encodes a MYC class transcription factor. *The Plant Journal* 33:413-423.

Svitashev, S., Schwartz, C., Lenderts, B., Young, J.K., and Cigan, A.M. 2016. Genome editing in maize directed by CRISPR–Cas9 ribonucleoprotein complexes. *Nature Communications* 7:13274.

Svitashev, S., Young, J.K., Schwartz, C., Gao, H., Falco, S.C., and Cigan, A.M. 2015. Targeted mutagenesis, precise gene editing, and site-specific gene insertion in maize using Cas9 and guide RNA. *Plant Physiology* 169:931-945.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585-595.

Wang, D., Skibbe, D.S., and Walbot, V. 2013. Maize Male sterile 8 (Ms8), a putative β-1, 3-galactosyltransferase, modulates cell division, expansion, and differentiation during early maize anther development. *Plant Reproduction* 26:329-338.

Wang, H., Nussbaum-Wagler, T., Li, B., Zhao, Q., Vigouroux, Y., Faller, M., Bomblies, K., Lukens, L., and Doebley, J.F. 2005. The origin of the naked grains of maize. *Nature* 436:714-719.

Xing, S., Quodt, V., Chandler, J., Hohmann, S., Berndtgen, R., and Huijser, P. 2013. SPL8 acts together with the brassinosteroid-signaling component BIM1 in controlling *Arabidopsis thaliana* male fertility. *Plants* 2:416-428.

Xu, J., Ding, Z., Vizcay-Barrena, G., Shi, J., Liang, W., Yuan, Z., Werck-Reichhart, D., Schreiber, L., Wilson, Z.A., and Zhang, D. 2014. ABORTED MICROSPORES acts as a master regulator of pollen wall formation in *Arabidopsis*. *The Plant Cell* 26:1544-1556.

Yu, J., Zhao, Y.X., Qin, Y.T., Yue, B., Zheng, Y.L., and Xiao, H.L. 2013. Discovery of MicroRNAs Associated with the S Type Cytoplasmic Male Sterility in Maize. *Journal of Integrative Agriculture* 12:229-238.

Zhang, D., Liang, W., Yuan, Z., Li, N., Shi, J., Wang, J., Liu, Y., Yu, W., and Zhang, D. 2008. Tapetum degeneration retardation is critical for aliphatic metabolism and gene regulation during rice pollen development. *Molecular Plant* 1:599-610.

Zhang, H., Egger, R.L., Kelliher, T., Morrow, D., Fernandes, J., Nan, G., and Walbot, V. 2014. Transcriptomes and proteomes define gene expression progression in pre-meiotic maize anthers. *G3: Genes| Genomes| Genetics* 4:993-1010.

Zhang, H., Liang, W., and Zhang, D. 2008. Research progress on tapetum programmed cell death. *Journal of Shanghai Jiaotong University (Agr Sci)* 26:86-90.

Zhang, W., Sun, Y., Timofejeva, L., Chen, C., Grossniklaus, U., and Ma, H. 2006. Regulation of Arabidopsis tapetum development and function by DYSFUNCTIONAL TAPETUM1 (DYT1) encoding a putative bHLH transcription factor. *Development* 133:3085-3095.

Zhou, A., and Pawlowski, W.P. 2014. Regulation of meiotic gene expression in plants. *Frontiers in Plant Science* 5:413.

Zhu, E., You, C., Wang, S., Cui, J., Niu, B., Wang, Y., Qi, J., Ma, H., and Chang, F. 2015. The DYT1 interacting proteins bHLH010, bHLH089 and bHLH091 are redundantly required for Arabidopsis anther development and transcriptome. *The Plant Journal* 83:976-990.

Zhu, Q., Song, Y., Zhang, G., Ju, L., Zhang, J., Yu, Y., Niu, N., Wang, J., and Ma, S. 2015. De Novo Assembly and Transcriptome Analysis of Wheat with Male Sterility Induced by the Chemical Hybridizing Agent SQ-1. *PLoS One* 10:e123556. 10.1371/journal.pone.0123556

Table 1 Single-nucleotide polymorphisms (SNPs) and insertion-deletion polymorphisms (InDels)

bioRxiv preprint doi: https://doi.org/10.1101/140806; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

524 **of ZmbHLH16 among 78 maize inbred lines**. Underlined letters represent minor alleles.

Locus name	Position in gene (bp)/region	SNP or insertion sequence	Amino acid change	Minor allele frequency
S230	230/5'-UTR	$(C)_n$		0.06
S236	236/5'-UTR	C/A		0.12
S238	238/5'-UTR	C/A		0.12
S244	244/5'-UTR	C/T		0.12
S246	246/5'-UTR	C/T		0.12
S331	331/5'-UTR	TATA		0.10
S381	381/5'-UTR	C/G		0.12
S409	409/5'-UTR	G/T		0.12
S504	504/5'-UTR	C/T		0.40
S753	753/5'-UTR	ATCCG		0.10
S882	882/5'-UTR	T/G		0.12
S929	929/5'-UTR	C/T		0.10
S933	933/5'-UTR	ATAT		0.10
S1349	1349/EXON1	C/A	His/Asn	0.17
S1436	1436/EXON1	GCC	Ala	0.14
S1510	1510/EXON1	G/A		0.09
S1540	1540/EXON1	T/G		0.08
S1602	1602/EXON1	AC/TT	Tyr/Phe	0.31
S1603	1603/EXON1			
S1606	1606/EXON1	C/G		0.31
S1636	1636/EXON1	A/G		0.31
S1645	1645/INTRON1	ACGGCGTACTTGCGCGCGG		0.24
S1684	1684/INTRON1	TCG		0.24
S1698	1698/INTRON1	T/C		0.40
S1701	1701/INTRON1	AAATTAAGTTAAA		0.31
S1731	1731/INTRON1	TTG		0.31
S1859	1859/EXON2	C/T		0.09
S1963	1963/EXON2	C/T	Pro/Leu	0.08
S1988	1988/EXON2	G/C		0.10
S2006	2006/EXON2	G/A		0.09
S2117	2117/EXON2	T/C		0.26
S2250	2250/INTRON2	TTTCATGCATTACTT		0.36
S2278	2278/INTRON2	GTGCTATACTACCTAATCTA		0.36
S2314	2314/INTRON2	T/C		0.15
S2323	2323/INTRON2	C		0.21
S2509	2509/3'-UTR	T		0.06

525 **Table 2 ZmbHLH16 nucleotide diversity and neutrality test**. Due to a lack of significant
526 polymorphism sites (MAF≥0.05), values for neutrality tests in Exon 3 and the 3'-UTR are not given.

Region	5'-UTR	Exon 1	Intron 1	Exon 2	Intron 2	Exon 3	3'-UTR	Overall
$\pi(\times 10^{-3})$	1.85	4.13	6.15	2.48	2.39	0	0	2.58
Tajima's D	-0.65	0.48	0.42	-0.28	-0.91	/	/	-0.26

bioRxiv preprint doi: https://doi.org/10.1101/140806; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fu andLi's D	-0.92	-0.64	-1.01	0.22	-1.90	/	/	-1.21
Fu andLi's F	-0.98	-0.30	-0.67	0.06	-1.86	/	/	-1.01
Length (bp)	1042	597	121	432	168	72	82	2514

527 **Table 3 ZmbHLH16 coexpressed genes homologous to Arabidopsis male-sterility**
528 **(MS)/reproduction (MR) genes.**

Query gene ID	PCC-value	Subject ID	E-value	Score	Symbol	Status	Description
GRMZM2G014651	0.9972	AT1G62940	1.65E-35	147	ACOS5	MS gene (cloned)	acyl-CoA synthetase 5 (ACOS5
GRMZM2G145179	0.9972	AT1G62940	1.65E-35	147	ACOS5	MS gene (cloned)	acyl-CoA synthetase 5 (ACOS5)
GRMZM2G108894	0.9959	AT1G02050	2.06E-26	116	LAP6	MS gene (cloned)	LESS ADHESIVE POLLEN 6 (LAP6)
GRMZM2G119265	0.9937	AT1G33430	1.90E-22	104		MR gene (GO evidence)	Galactosyltransferase family protein
GRMZM2G498736	0.9800	AT4G13640	1.57E-13	73.4	UNE16	MR gene (GO evidence)	unfertilized embryo sac 16 (UNE16)
GRMZM2G073982	0.9799	AT1G16060	9.11E-13	71.6	ADAP	MR gene (mutant evidence)	ARIA-interacting double AP2 domain protein (ADAP)
GRMZM2G380650	0.9326	AT1G02050	2.93E-66	250	LAP6	MS gene (cloned)	LESS ADHESIVE POLLEN 6 (LAP6)
GRMZM2G136353	0.9323	AT3G49670	1.87E-11	68	BAM2	MR gene (GO evidence)	BARELY ANY MERISTEM 2 (BAM2)
GRMZM2G700011	0.8913	AT5G56110	2.69E-09	59	AtMYB103	MR gene (GO evidence)	myb domain protein 103 (MYB103)
GRMZM2G177928	0.8662	AT3G59530	1.95E-06	51.8	LAP3	MR gene (GO evidence)	Calcium-dependent phosphotriesterase superfamily protein
GRMZM2G040905	0.8613	AT4G13640	1.19E-12	73.4	UNE16	MR gene (GO evidence)	unfertilized embryo sac 16 (UNE16)
GRMZM2G091822	0.8375	AT1G69500	1.37E-17	87.8	CYP704B1	MR gene (GO evidence)	cytochrome P450, family 704, subfamily B, polypeptide 1 (CYP704B1)
GRMZM2G435993	0.8225	AT4G03550	4.84E-55	214	ATGSL05	MR gene (GO evidence)	glucan synthase-like 5 (GSL05)
GRMZM2G093408	0.7987	AT3G04620	6.70E-12	68		MR gene (GO evidence)	Alba DNA/RNA-binding protein
GRMZM2G368610	0.7867	AT2G44810	2.58E-06	50	DAD1	MR gene (GO evidence)	DEFECTIVE ANTHER DEHISCENCE 1 (DAD1)
GRMZM2G055279	0.7760	AT5G59030	1.67E-11	66.2	COPT1	MR gene (GO evidence)	copper transporter 1 (COPT1)
GRMZM2G155699	0.7598	AT2G35270	6.55E-07	51.8	GIK	MR gene (GO evidence)	GIANT KILLER (GIK)
GRMZM2G027522	0.7481	AT4G24972	2.85E-08	57.2	TPD1	MR gene (GO evidence)	TAPETUM DETERMINANT 1 (TPD1)
GRMZM2G060513	0.7017	AT3G04620	2.61E-14	77		MR gene (GO evidence)	Alba DNA/RNA-binding protein
GRMZM2G161680	0.6947	AT1G54830	6.31E-45	179	NF-YC3	MR gene (mutant evidence)	nuclear factor Y, subunit C3 (NF-YC3)
GRMZM2G166855	0.6895	AT4G21150	3.32E-20	98.7	HAP6	MS gene (cloned)	HAPLESS 6 (HAP6)
GRMZM2G007146	0.6734	AT4G28395	4.05E-06	51.8	ATA7	MR gene (GO evidence)	ANTHER 7 (A7)
GRMZM2G083111	0.6656	AT1G25540	8.61E-06	48.2	PFT1	MS gene (cloned)	PHYTOCHROME AND FLOWERING TIME 1 (PFT1)
GRMZM2G113959	0.6534	AT4G13890	8.50E-06	48.2	EDA36	MR gene (GO evidence)	EMBRYO SAC DEVELOPMENT ARREST 37 (EDA36)
GRMZM2G168228	0.6487	AT1G64110	3.25E-16	84.2		MR gene (GO evidence)	P-loop containing nucleoside triphosphate hydrolases superfamily protein
GRMZM2G123776	0.6425	AT1G60420	5.19E-25	113		MR gene (GO evidence)	DC1 domain-containing protein
GRMZM2G045287	0.6314	AT3G02230	1.32E-108	390	ATRGP1	MS gene (cloned)	reversibly glycosylated polypeptide 1 (RGP1)
GRMZM2G056039	0.6300	AT1G09080	3.38E-47	187	BIP3	MR gene (GO evidence)	BIP3
GRMZM2G062154	0.6254	AT1G53320	6.62E-38	156	AtTLP7	MR gene (mutant evidence)	tubby like protein 7 (TLP7)
GRMZM2G317051	0.6200	AT1G65470	1.96E-09	62.6	FAS1	MS gene (cloned)	FASCIATA 1 (FAS1)
GRMZM5G857992	0.6133	AT5G22420	8.33E-06	50	FAR7	MR gene (GO evidence)	fatty acid reductase 7 (FAR7)
GRMZM2G128771	0.6130	AT5G57880	1.65E-08	60.8	ATPRD2	MR gene (GO evidence)	MULTIPOLAR SPINDLE 1 (MPS1)
GRMZM2G453832	0.6123	AT4G28580	2.17E-10	64.4	AtMGT5	MS gene (cloned)	magnesium transport 5 (MGT5)
GRMZM2G036916	0.6058	AT2G35060	4.60E-21	100	KUP11	MR gene (GO evidence)	K+ uptake permease 11 (KUP11)
GRMZM2G013607	0.6044	AT4G05450	3.79E-21	100	ATMFDX1	MR gene (GO evidence)	mitochondrial ferredoxin 1 (MFDX1)

529

bioRxiv preprint doi: https://doi.org/10.1101/140806; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure

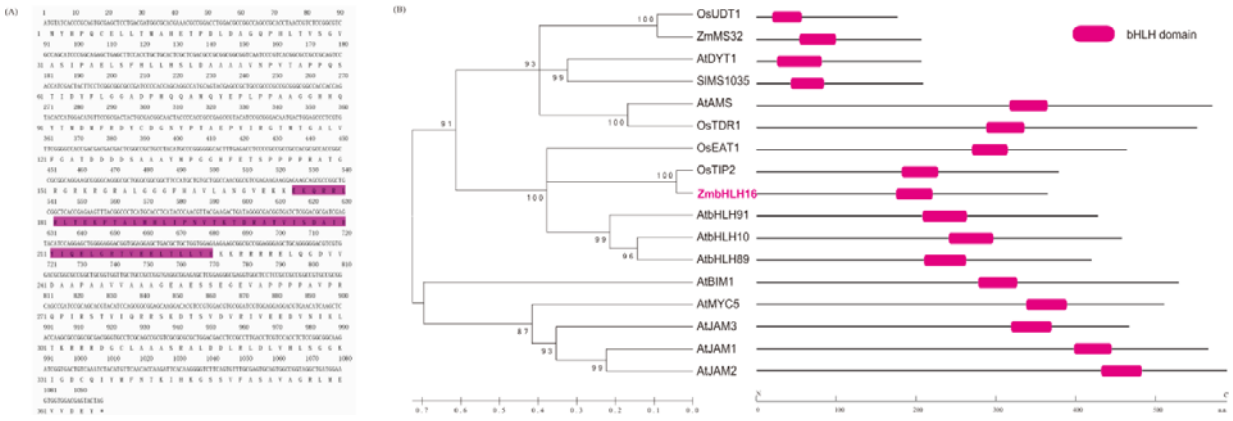


Fig 1 Structure and phylogenetic analysis of ZmbHLH16. (A) Nucleotide and deduced amino acid sequences of ZmbHLH16 CDS. Shaded regions are the conserved DNA-binding domains of the bHLH protein. Bold letters show conserved tryptophan residues in the bHLH domain. **(B)** Phylogenetic relationship of ZmbHLH16 and other bHLH proteins related to microspore development. The two scale bars indicate branch length and amino acid length.

bioRxiv preprint doi: https://doi.org/10.1101/140806; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

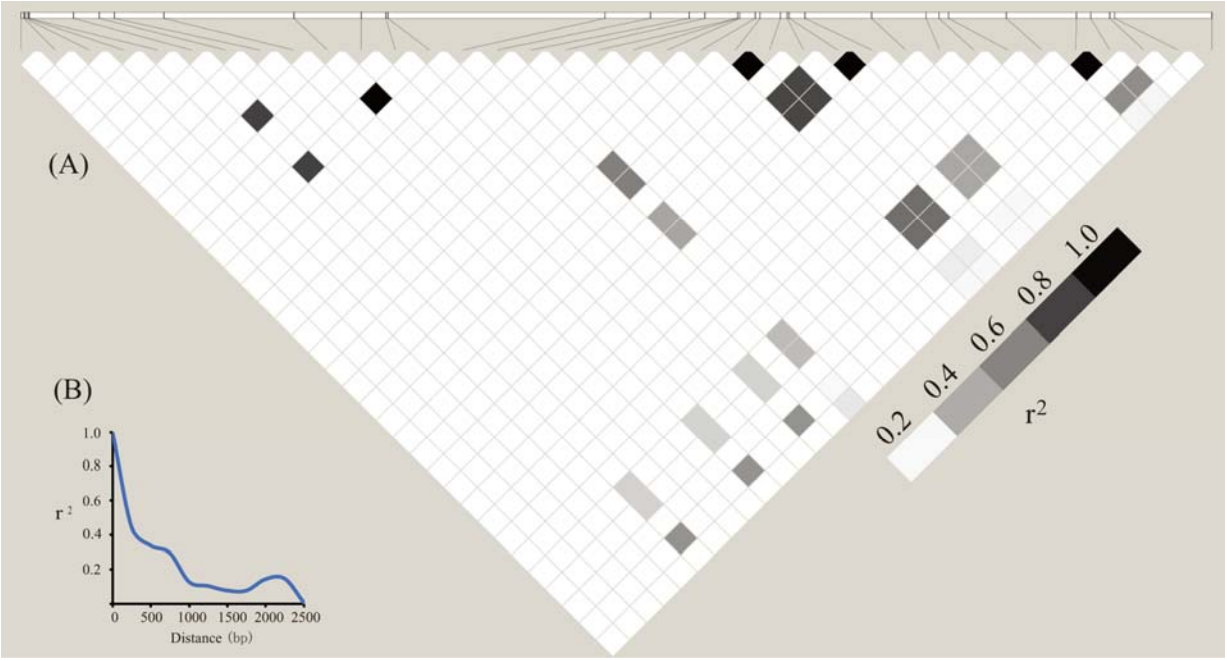


Fig 2 LD block and LD decay pattern of ZmbHLH16. (A) Matrix of pairwise LD of DNA polymorphisms (MAF≥0.05) in ZmbHLH16. The shaded boxes indicate the LD standard (r^2). **(B)** LD decay in the DNA sequence of ZmbHLH16 in maize. The *x* axis represents the distance between polymorphic sites, and the *y* axis represents the average r^2 value for each distance category (250 bp).

bioRxiv preprint doi: https://doi.org/10.1101/140806; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

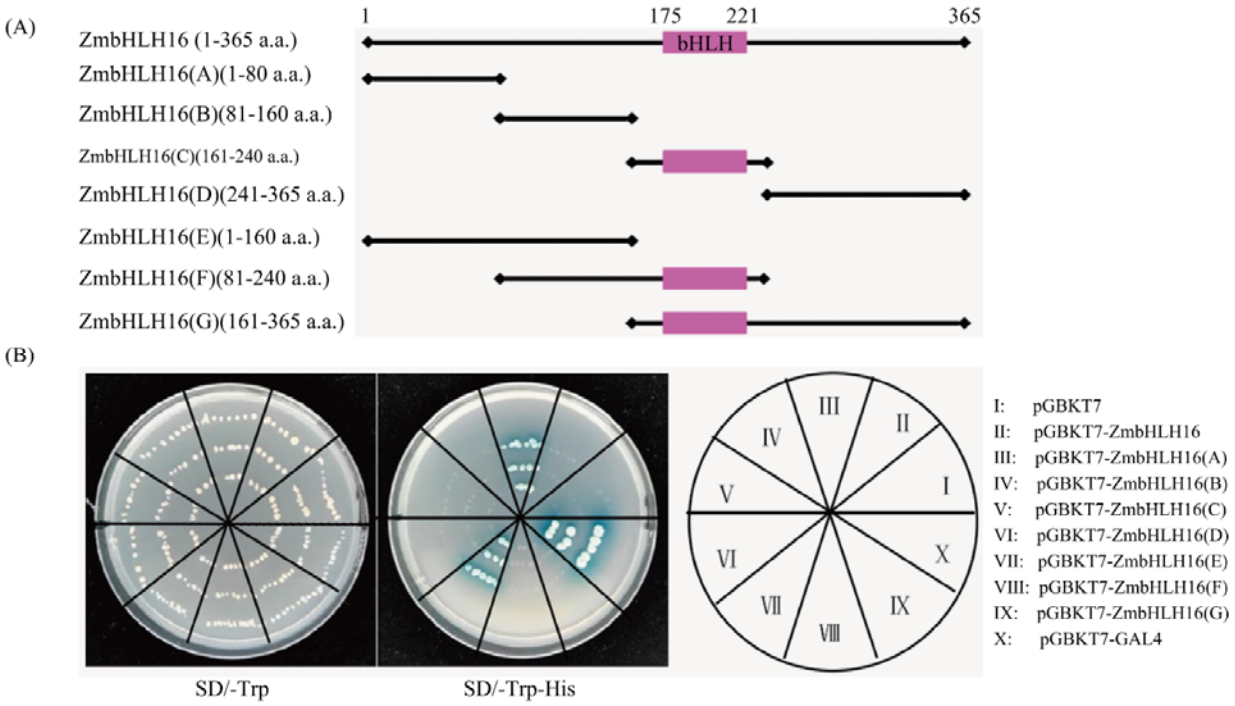


Fig 3 Transactivation activity assays of ZmbHLH16 in yeast. (A) Diagram of ZmbHLH16 activation domain. **(B)** Growth of yeast containing various fragments of ZmbHLH16 on selective media (SD/-Trp and SD/-Trp-His).

bioRxiv preprint doi: https://doi.org/10.1101/140806; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

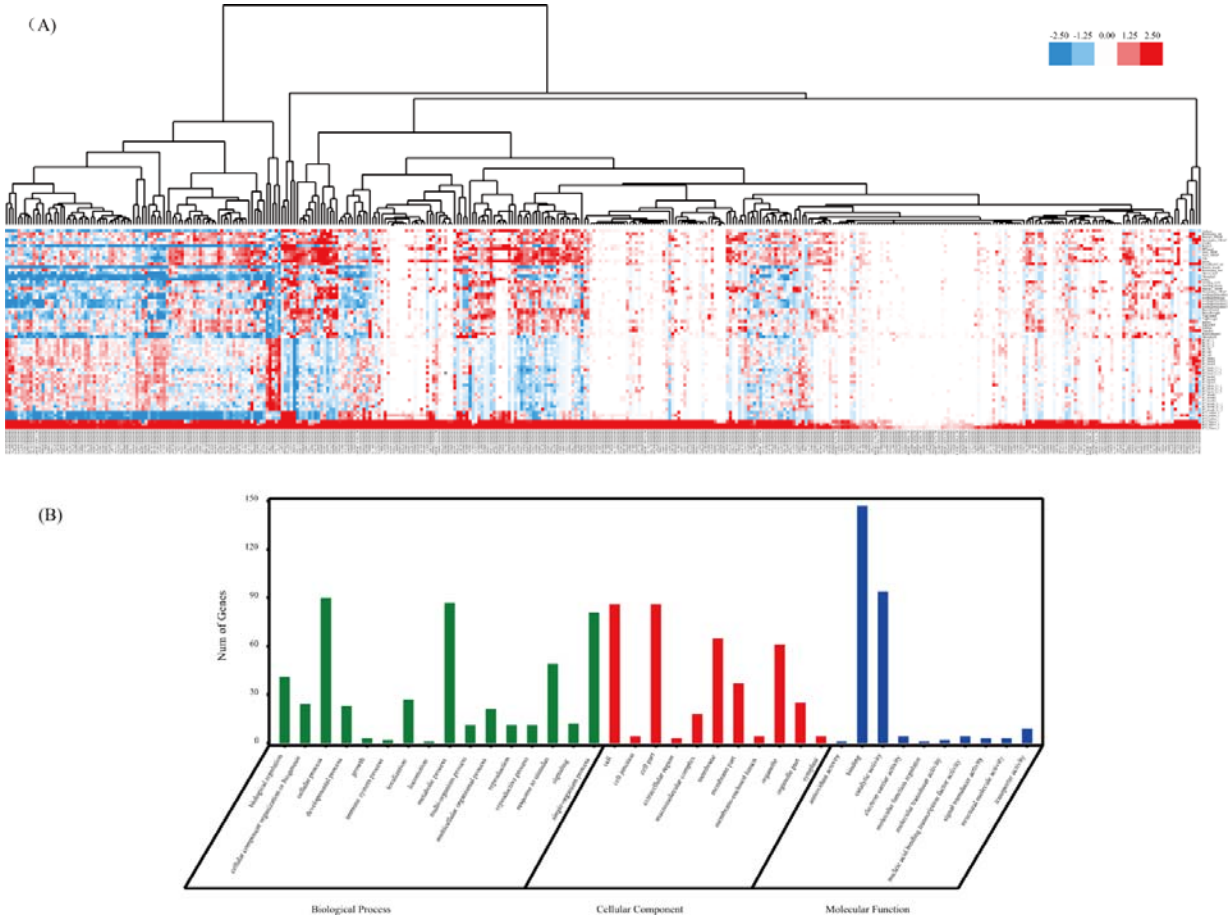
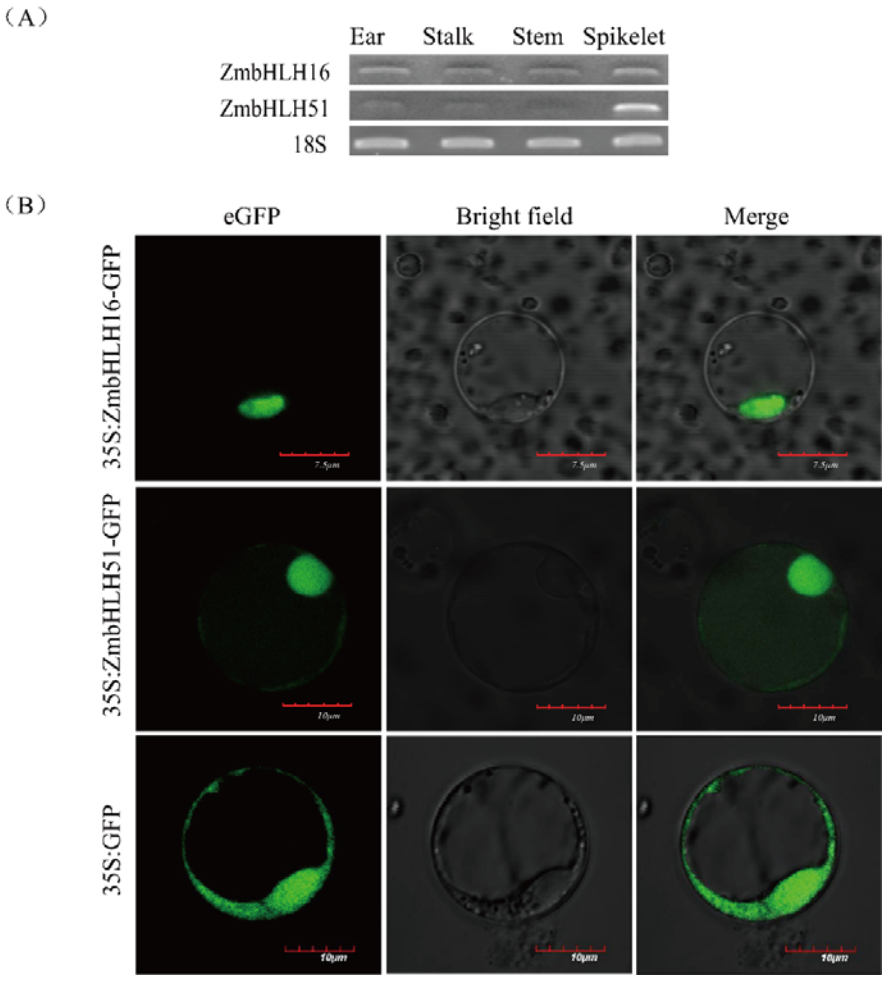


Fig 4 Expression patterns and GO annotations of ZmbHLH16 coexpressed genes. (A) A cluster of 395 coexpressed genes based on expression characteristics. Gene expression data were downloaded from q-teller (http://www.qteller.com/qteller3/index.php); the bar indicates the relative gene expression level, which was $\log_2$-normalized (original data+1). **(B)** GO analysis of 395 coexpressed genes.

bioRxiv preprint doi: https://doi.org/10.1101/140806; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



573

574 **Fig 5 Expression characteristics of ZmbHLH16 and ZmbHLH51. (A)** Semi-quantitative analysis
575 of ZmbHLH16 and ZmbHLH51 in various organs. The expression of 18S was taken as the reference.
576 **(B)** Subcellular localization analysis of ZmbHLH16 and ZmbHLH51 in rice protoplasts.

bioRxiv preprint doi: https://doi.org/10.1101/140806; this version posted May 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

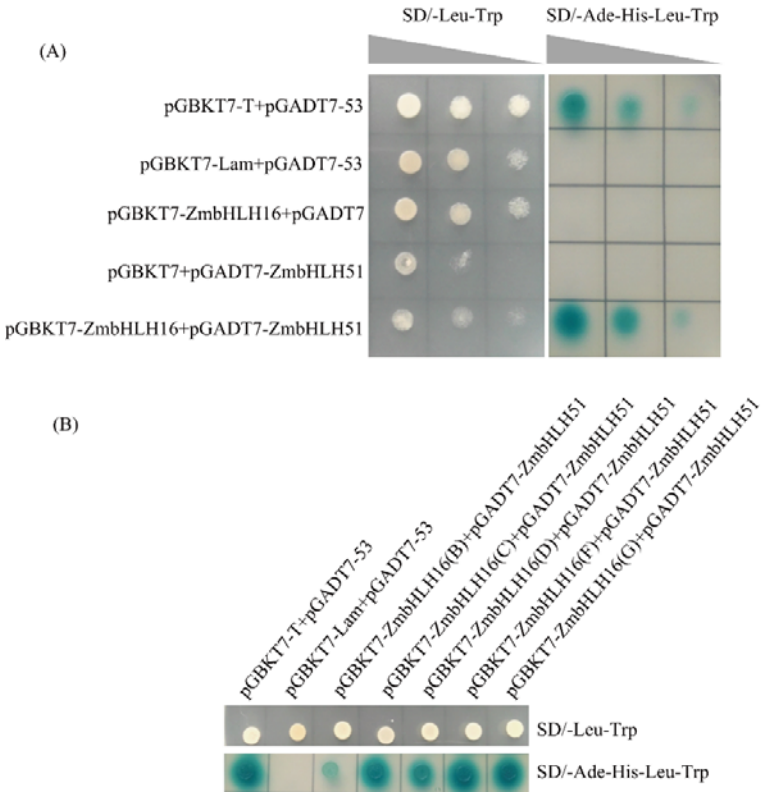


Fig 6 Determination of the interaction between ZmbHLH16 and ZmbHLH51 using yeast two-hybrid analysis. (A) Yeast two-hybrid analysis of the interaction between ZmbHLH16 and ZmbHLH51 proteins. Each transformant was stained in media with three relative concentrations (1, 0.1, 0.01) from left to right. **(B)** Yeast two-hybrid mapping of domains involved in the ZmbHLH16-ZmbHLH51 interaction. Regions without transcriptional activation activity, including ZmbHLH16(B) $^{81\text{-}160\ a.a.}$, (C) $^{161\text{-}240\ a.a.}$, (D) $^{241\text{-}365\ a.a.}$, (F) $^{81\text{-}240\ a.a.}$ and (G) $^{161\text{-}365\ a.a.}$, were chosen for analysis.