Genome-wide *in silico* identification of membrane-bound transcription factors in plant species

Shixiang Yao [1] , **Lili Deng** [1] , **Kaifang Zeng** Corresp. [1]

[1] College of Food Science, Southwest University, Chongqing, China

Corresponding Author: Kaifang Zeng
Email address: zengkaifang@hotmail.com

Membrane-bound transcription factors (MTFs) are located in cellular membranes due to their transmembrane domains. In plants, proteolytic processing is considered to be the main mechanism for MTF activation, which ensures the liberation of MTFs from membranes and further their translocation into the nucleus to regulate gene expression; this process skips both the transcriptional and translational stages, and thus it guarantees the prompt responses of plants to various stimuli. Currently, information concerning plant MTFs is limited to model organisms, including *Arabidopsis thaliana* and *Oryza sativa*, and little is known in other plant species at the genome level. In the present study, seven membrane topology predictors widely used by the research community were employed to establish a reliable workflow for MTF identification. Genome-wide *in silico* analysis of MTFs was then performed in 14 plant species spanning the chlorophytes, bryophytes, gymnosperms, monocots and eudicots. A total of 1,089 MTFs have been identified from a total of 25,850 transcription factors in these 14 plant species. These MTFs belong to 52 gene family, and the top six most abundant families are the NAC (128), SBP (77), C2H2 (70), bZIP (67), MYB-related (65) and bHLH (63) families. The MTFs have transmembrane spans ranging from one to thirteen, and 71.5% and 21.1% of the MTFs have one and two transmembrane motifs, respectively. Most of the MTFs in this study have transmembrane motifs located in either N- or C-terminal regions, indicating that proteolytic cleavage could be a conserved mechanism for MTF activation. Additionally, approximately half of the MTFs in the genome of either *Arabidopsis thaliana* or *Gossypium raimondii* could be potentially regulated by alternative splicing, indicating that alternative splicing is another conserved activation mechanism for MTFs. The present study performed systematic analyses of MTFs in plant lineages at the genome level, and provides invaluable information for the research community.

1 **Genome-wide *in silico* identification of membrane-bound transcription**

2 **factors in plant species**

3 Shixiang Yao, Lili Deng and Kaifang Zeng

4 College of Food Science, Southwest University, Chongqing 400715, China;

6 Author for correspondence:

7 Kaifang Zeng

8 Tel: +86 10 68251298

9 Email: zkaifang@swu.edu.cn; zengkaifang@hotmail.com

11 **Abstract**

12 Membrane-bound transcription factors (MTFs) are located in cellular membranes due to

13 their transmembrane domains. In plants, proteolytic processing is considered to be the

14 main mechanism for MTF activation, which ensures the liberation of MTFs from

15 membranes and further their translocation into the nucleus to regulate gene expression;

16 this process skips both the transcriptional and translational stages, and thus it guarantees

17 the prompt responses of plants to various stimuli. Currently, information concerning plant

18 MTFs is limited to model organisms, including *Arabidopsis thaliana* and *Oryza sativa*, and

19 little is known in other plant species at the genome level. In the present study, seven

20 membrane topology predictors widely used by the research community were employed to

21 establish a reliable workflow for MTF identification. Genome-wide *in silico* analysis of

22 MTFs was then performed in 14 plant species spanning the chlorophytes, bryophytes,

23 gymnosperms, monocots and eudicots. A total of 1,089 MTFs have been identified from a

24 total of 25,850 transcription factors in these 14 plant species. These MTFs belong to 52

25 gene family, and the top six most abundant families are the NAC (128), SBP (77), C2H2

26 (70), bZIP (67), MYB-related (65) and bHLH (63) families. The MTFs have

27 transmembrane spans ranging from one to thirteen, and 71.5% and 21.1% of the MTFs

28 have one and two transmembrane motifs, respectively. Most of the MTFs in this study

29 have transmembrane motifs located in either N- or C-terminal regions, indicating that

30 proteolytic cleavage could be a conserved mechanism for MTF activation. Additionally,

31 approximately half of the MTFs in the genome of either *Arabidopsis thaliana* or

32 *Gossypium raimondii* could be potentially regulated by alternative splicing, indicating that

33 alternative splicing is another conserved activation mechanism for MTFs. The present

34 study performed systematic analyses of MTFs in plant lineages at the genome level, and

35 provides invaluable information for the research community.

36

37 **Introduction**

38 Transcription factors play a primary regulatory role in gene transcription and thus ensure

39 normal growth of plants and promote their adaptation to environmental stress (Li et al.

40 2011). Transcription factors are tightly regulated at multiple levels, including the

41 transcriptional, translational, and post-translational levels (Seo et al. 2008). During the last

42 decade, a novel mechanism of post-translational regulation of transcription

43 factors—proteolytic processing dependent activation—has been extensively studied and

44 well established (Seo 2014). A small proportion of transcription factors containing

45 transmembrane (TM) motifs are translated in the cytoplasm and then rapidly anchored to

46 cellular membranes including plasma, mitochondrial, endoplasmic reticulum (ER)

47 membranes, rather than being translocated to the nucleus (Kim et al. 2010).

48 Membrane-bound transcription factors (MTFs) are stored in dormant forms and become

49 active after being released from membrane. Most MTFs with known functions have a

50 single TM domain located adjacent to their transcription factor domains located within

51 either the N-terminal or C-terminal region and they are anchored to the membrane by TM

52 motifs. Thus, in response to environmental stimuli and physiological signals, the MTFs

53 can be cleaved near the TM region by a specific protease, and the transcription factor can

54 be released from the membrane and then translocated into the nucleus to exert its

55 functions (Che et al. 2010). The proteolytic processing-dependent activation of MTFs

56 skips both transcriptional and translational steps, and it thus ensures the prompt response

57 of plants to exogenous and endogenous signals.

58 Since the first plant MTF, *AtbZIP60*, was elucidated at the molecular level, a major

59 focus of research on the MTFs of plants has targeted NAC and bZIP transcription factors

60 (Iwata and Koizumi 2005; Seo 2014). For instance, the function of several MTFs has been

61   well elucidated in recent years, including eight NAC and three bZIP members (Seo 2014).

62   Previous studies of plant MTFs suggested that MTFs were involved in various aspects of

63   plant growth, development and environmental responses, such as seed germination (Park

64   et al. 2011), cell division (Kim et al. 2006), root hair development (Slabaugh et al. 2011),

65   sugar signaling (Li et al. 2011), ER stress (Che et al. 2010; Iwata et al. 2008; Iwata and

66   Koizumi 2005; Yang et al. 2014), reactive oxygen species (ROS) signaling (De Clercq et

67   al. 2013; Lee et al. 2012; Ng et al. 2013), cold stress (Seo et al. 2010), heat stress (Gao et

68   al. 2008), salt stress (Che et al. 2010; Kim et al. 2008) and osmotic stress (Kim et al. 2012;

69   Yoon et al. 2008). A further genome wide analysis of MTFs in Arabidopsis (*Arabidopsis*

70   *thaliana*) and rice (*Oryza sativa*) showed that many transcription factor families contained

71   membrane-bound members, although most of them have yet to be functionally resolved

72   (Kim et al. 2010). MTFs are fast emerging as critical drivers of gene regulation in response

73   to various stimuli in plants. Because of their crucial role in the regulation of gene

74   expression, it is of great importance to identify MTFs in other plant species. However,

75   most of the MTFs that have been functionally resolved are those from Arabidopsis. This

76   may be partially because the genome-wide analyses of MTFs have been performed using

77   genomes of model organisms including Arabidopsis and rice. Thus, it is of great

78   importance that a systematic analysis of MTFs is performed in various plant lineages

79   besides Arabidopsis and rice at the genome level to not only widen the knowledge of

80   MTFs in plant species but also provide invaluable resources for the community and further

81   stimulate the functional elucidation of these transcription factors.

82   Here, we established a workflow for the identification of plant MTFs and performed

83   genome-wide analyses of MTFs in 14 plant species spanning the chlorophytes,

84   bryophytes, gymnosperms, monocots and eudicots. A total of 1089 MTFs belonging to 52

85   gene families were identified in these plant species. Most of the MTFs have one or two

86   transmembrane motifs located in either the N- or C-terminal region, indicating that

87   proteolytic cleavage could be a conserved mechanism for MTF activation. Additionally, a

88   considerable proportion of the MTFs in the Arabidopsis and cotton (*Gossypium raimondii*)

89   genome were expected to be activated by alternative splicing (ALS), indicating that ALS is

90   another conserve regulatory mechanism for the activation of MTFs.

91   ## Materials & Methods

92   **Sequence retrieval for whole membrane-bound transcription factor analysis**

93   Complete cDNA and amino acid sequences of transcription factors of the plant species

94   (Table 1) were retrieved from the recently published plant transcription factor database

95   (PlantTFDB v4.0, *http://planttfdb.cbi.pku.edu.cn/*) (Jin et al. 2017). The consequent

96   sequences were used to perform *in silico* analysis of membrane-bound proteins via the

97   following bioinformatics tools.

98   **Selection of bioinformatics methods for membrane-bound protein prediction**

99   Seven prediction methods including TMHMM 2.0 (Krogh et al. 2001), S-TMHMM (Viklund

100   and Elofsson 2004), HMMTOP (Tusnady and Simon 2001), PHOBIUS (Kall et al. 2004),

101   SCAMPI-single (Bernsel et al. 2008), MEMSAT 1.0 (Jones et al. 1994) and TOPPRED

102   (Vonheijne 1992) were chosen for TM prediction. For TMHMM 2.0, transmembrane

103   helices and the locations were predicted; the most likely locations of transmembrane

104   motifs were given, and they were included in this study if they had a probability score

105   greater than 0.90. The other six predictors were employed to predict transmembrane

106   motifs and the consequent results were filtered and/or combined by using

107   TOPCONS-single software with the default parameters (*http://single.topcons.net/*)

108   (Hennerdal and Elofsson 2011).

109   ***In silico* analysis of alternative splicing of MTFs in Arabidopsis and cotton**

110   Transcriptome sequences of *Arabidopsis thaliana* (TAIR10_cds_20101214_updated)

111   were downloaded from TAIR (The Arabidopsis Information Resource) website

112   (*https://www.arabidopsis.org/*). Transcriptome sequences of cotton (*G. raimondii*) were

113   retrieved from Phytozome (*https://phytozome.jgi.doe.gov/pz/portal.html#*).

114

Results & Discussion

Establishment of workflow for identification of MTFs

To perform genome wide analyses of the MTFs in the different plant species, seven tools that are widely used to predict transmembrane segments were selected in present study: TMHMM 2.0 (Krogh et al. 2001), HMMTOP (Tusnady and Simon 2001), PHOBIUS (Kall et al. 2004), S-TMHMM (Viklund and Elofsson 2004), TOPPRED (Vonheijne 1992), SCAMPI-single (Bernsel et al. 2008) and MEMSAT 1.0 (Jones et al. 1994). Their methods are based on different algorithms such as the hidden Markov model (HMM), the artificial neural network (ANN) and hydrophobicity (Table 2). These methods were selected because of their frequent usage in the research community and good results obtained in various studies (Schwacke et al. 2003; Zhang et al. 2015). Moreover, multiple predictors were employed to identify as many reliable membrane proteins as possible. .

Sequences of Arabidopsis transcription factors were extracted from the recently published Plant Transcription Factor Database (PlantTFDB 4.0) (Jin et al. 2017), and they were then subjected to transmembrane region screening using the above predictors. In *A. thaliana*, for example, the genome was predicted to contain 37 to 60 membrane-bound transcription factors (MTFs) depending on the methods used (Fig. 1a). Regardless of the prediction method, most of the resulting MTFs have a single TM span, ranging from 63% to 93% (Fig. 1a). The numbers of both predicted MTFs and predicted TM motifs per protein varied among the different methods. This is because the algorithms of the methods are different; some methods may overlook membrane regions that are recognized by other methods (Schwacke et al. 2003). Therefore, a consensus decision was made for topology predictions based on different prediction methods, which could generate data with high accuracy and confidence (Hennerdal and Elofsson 2011). The workflow for MTF identification in the present study is shown in Fig. 1b; plant transcription factor databases were subjected to analysis using the seven bioinformatics tools, including TMHMM 2.0, HMMTOP, PHOBIUS, S-TMHMM, TOPPRED, SCAMPI-single and MEMSAT 1.0. The TOPCONS-single algorithm (Hennerdal and Elofsson 2011) was

143   employed to obtain consensus prediction data for membrane proteins predicted by all of

144   the methods except TMHMM 2.0, and the results were then combined with the data

145   predicted by TMHMM 2.0 to constitute the membrane transcription factor database with

146   high confidence (Fig. 1b). The results generated by TMHMM 2.0 were separately

147   analyzed because TMMHM 2.0 is the most accurate software currently available for

148   membrane protein prediction, as it is characterized by a low false-positive rate but a high

149   false-negative rate (Schwacke et al. 2003).

150   We employed this established workflow to identify membrane-bound members from

151   the Arabidopsis transcription factors. The Arabidopsis genome was found to contain 64

152   MTFs (Fig. 1c), among which the number of proteins containing one, two and three

153   membrane segments was 52, 11 and one, respectively. These proteins are classified as

154   belonging to 24 gene families, including 17 NAC members and 5 bZIP members (Fig. 1d).

155   The total number of MTFs identified in the present study was less than that in the

156   previously predicted MTF database (85) in Arabidopsis (Kim et al. 2010), and 43 members

157   overlapped between the two studies, which accounted for 67% of the MTFs in the present

158   study (Fig. 1e). In contrast to the *in silico* analysis of Arabidopsis MTFs, this study

159   identified 85 MTFs in the rice (*Oryza sativa*) genome database (Table 1), which is much

160   more than the 45 members identified in previous research (Kim et al. 2010). The

161   difference between the two studies may be due to the different algorithms used, which is

162   common in the research on membrane protein prediction (Schwacke et al. 2003).

163   However, 85 MTFs were identified in rice in the present study, while only 45 MTFs were

164   identified in rice by Kim et al. 2010 (Table 1). Both the previous (Kim et al. 2010) and the

165   present study have tried to identify all possible MTFs for the scientific community. One

166   major advantage of the workflow in the present study is that these methods are freely

167   available to the scientific community and are applicable to more plant species. The

168   Arabidopsis MTFs identified in this study are highly reliable, because all of the Arabidopsis

169   MTFs functionally elucidated in previous research (Table S1) were present in this study

170   except for a PHD transcription factor (AT5G35210) that was not deposited in PlantTFDB

171   4.0 (Sun et al. 2011). These results further indicated that the workflow for MTFs

172   identification established in this study can identify MTF with considerable reliability.

Genome-wide *in silico* identification of MTFs in plant species

174   Genome-wide scale screening of MTFs has been performed in only two plant species:

175   *Arabidopsis thaliana* and rice (*Oryza sativa*) (Kim et al. 2010; Seo 2014). To considerably

176   broaden this dataset and cover other plant species, we performed extensive

177   genome-wide *in silico* analyses of various plant species as follows (Table 1)*. First, the

178   sequences of plant transcription factors ranging from 230 to 3714 loci were retrieved for

179   different species from PlantTFDB 4.0 (Jin et al. 2017). Second, the transcription factor

180   databases were screened for transmembrane motifs. Finally, a highly reliable database

181   has been established that consists of 1089 MTFs (Table 1). The number of MTFs in each

182   species varied from 37 to 117 and accounted for approximately 3% to 16% of the total

183   transcription factors. Relevant information about the MTFs, including the gene loci,

184   transmembrane region and gene family, are shown in Table S2.

Gene family classification of MTFs in plant species

186   The 1089 MTFs identified in 14 plant species in this study belong to 52 families (Table 1).

187   Twenty-eight additional gene families of MTFs were identified from plant species besides

188   *A. thaliana* (At) (Table 3). A total of 26 families containing MTFs were identified from more

189   than seven plant species including the AP2, B3, bHLH, bZIP, C2H2, C3H, Dof, ERF,

190   FAR1, G2-like, GATA, GRAS, HB-other, HD-ZIP, LBD, MIKC, M-type, MYB, MYB-related,

191   NAC, NF-X1, Nin-like, S1Fa-like, SBP, Trihelix and WRKY families (Table 3). Moreover,

192   nine gene families of MTFs, including bHLH, C2H2, G2-like, NAC, C3H, ERF, HB-other,

193   M-type and NF-X1, were identified from more than 12 plant species. Notably, the bZIP,

194   MYB-related, S1Fa-like and SBP families of MTFs were identified from all 14 of the plant

195   species. In contrast, several gene families of MTFs were identified from only a few plant

196   species (no more than two), namely, the BBR-BPC, EIL, GeBP, GRF, RAV, SAP and

197   Whirly families. We further summarized the total number of MTFs in the 14 species, and

198   the top six families with the highest numbers of MTFs   were as follows: NAC (128), SBP

199   (77), C2H2 (70), MYB-related (67) and bHLH (65) (Fig. S1). These results showed that

200   MTFs are highly diverse in various plant species. Considering that only a limited number

201   MTFs (Seo 2014), *e.g.*, eight NAC and three bZIP transcription factors in Arabidopsis,

202   have been functionally studied, the other MTFs deserve much more attention in further

203   research due to the important role of MTFs in gene expression.

204   **Analysis of transmembrane motifs of MTFs in plant species**

205   Among the 1089 MTFs identified from 14 plant species, the number of TM motifs varied

206   from one to 13. Most MTFs in the plant species, except for those in *C. reinhardtii*, have

207   one or two TM motifs (Fig. 2a). We summarized the total number of TM regions of the

208   MTFs in all plant species; proteins containing one and two TM motifs accounted for 71.5%,

209   and 21.1% of the MTFs, respectively.

210   The location patterns of TM motifs within protein sequences were analyzed, and the

211   major types of TM motif locations were shown in Fig. 2b. Among the MTFs containing one

212   TM region, a total of 311 (39.9%), 360 (46.2%), and 108 (13.9%) proteins had their TM

213   motifs in the N-terminal, C-terminal and central region of their sequences, respectively.

214   This was slightly different from a previous study on the MTFs of *A. thaliana*, in which more

215   than 70% members had TM motifs in their C-terminal regions (Kim et al. 2010). For the

216   230 MTF members containing two TM motifs, a total of 51 (22.2%) and 46 (20%)

217   members had their TM regions in the N-terminal and C-terminal regions, respectively; an

218   additional 31 MTF members (13.5%) had one TM motif located in the N-terminal region

219   and the other motif located in the C-terminal region (Fig. 2b). Taken together, most MTFs

220   in plant species have one or two TM domains, which are located in either their N- or

221   C-terminal region. These results indicated that proteolytic cleavage is a conserved

222   mechanism for MTFs in various plant species, as it is likely to release an intact functional

223   transcription factor since the MTF was cleaved by protease near the site of the

224   transmembrane domain. Such proteolytic process-dependent activation has been well

225   established in Arabidopsis MTFs (Seo 2014). Furthermore, the variance in TM motifs in

226 MTFs may lead to slight differences in the details of the mechanism of MTF activation by

227 proteolytic processing.

Possible regulation of MTFs by alternative splicing

229 MTFs were once considered to be activated solely by proteolytic processing. Interestingly,

230 recent research concerning *ZmbZIP60* and *OsbZIP74* suggested that alternative splicing

231 (ALS) is also responsible for the activation of MTFs (Li et al. 2012; Lu et al. 2012). To test

232 whether ALS-assisted activation of MTFs was also employed by other plants' MTFs, MTF

233 databases of Arabidopsis and cotton (*G. raimondii*) were selected, and *in silico* analyses

234 were performed. The results are shown in Fig. 3. In an Arabidopsis MTF member

235 (AT5G1050.3), for example, two splice variants (AT5G1050.1 and AT5G1050.2) encode

236 proteins that maintain the nuclear transcription factor domain while deleting the TM region

237 (Fig. 3a and b). Thus, ALS was proposed to be another mechanism for the activation of

238 MTFs, in which the splice variant without the sequence encoding a TM motif would be

239 transcribed under specific physiological conditions. In this case, the nuclear transcription

240 factors translocate into the nucleus and become active immediately since they were

241 translated in the cytoplasm. The regulation of MTFs by ALS may be important in the late

242 stage of plants in response to various stimuli. Interestingly, approximately 30% (19/64)

243 and 53% (59/111) of genes encoding MTFs in Arabidopsis and cotton, respectively, were

244 also shown to generate alternative splicing variants coding for proteins that only maintain

245 the transcription factor domain with nuclear localization signal while deleting the TM motif

246 (Fig. 3c & d and Table S3). Hence, these MTFs could be regulated by ALS in Arabidopsis

247 and cotton. Moreover, ALS seems to be involved in the activation of a wide range of MTF

248 families, such as the NAC, bHLH and MYB-related families, in Arabidopsis and cotton

249 (Table S3). Thus, an ALS-dependent activation mechanism may also be employed by the

250 MTFs of other plant species. However, a further "wet experiment" is needed to verify this

251 hypothesis.

252

253 **Conclusion**

254   The present study performed a systematic analysis of MTFs in plant lineages at the

255   genome level. A total of 1089 MTFs belonging to 52 gene families were identified in 14

256   plant species. Most of the MTFs have TM domains located in either the N- or C-terminal

257   regions, indicating proteolytic cleavage could be a conserved mechanism of MTF

258   activation in plant species. Additionally, approximately half of the MTFs in the genome of

259   either Arabidopsis or cotton can potentially be regulated by ALS, indicating that ALS may

260   be another conserved activation mechanism for MTFs. The present study provides

261   invaluable information for the research community.

262

263   **References:**

264   Bernsel A, Viklund H, Falk J, Lindahl E, von Heijne G, Elofsson A (2008) Prediction of

265   membrane-protein topology from first principles. P NATL ACAD SCI USA 105: 7177-7181

266   Che P, Bussell JD, Zhou W, Estavillo GM, Pogson BJ, Smith SM (2010) Signaling from the

267   endoplasmic reticulum activates brassinosteroid signaling and promotes acclimation to stress

268   in Arabidopsis. SCI SIGNAL 3: ra69.

269   De Clercq I, Vermeirssen V, Van Aken O, Vandepoele K, Murcha MW, Law SR, Inze A, Ng S,

270   Ivanova A, Rombaut D, van de Cotte B, Jaspers P, Van de Peer Y, Kangasjarvi J, Whelan J,

271   Van Breusegem F (2013) The membrane-bound NAC transcription factor ANAC013 functions

272   in mitochondrial retrograde regulation of the oxidative stress response in Arabidopsis. PLANT

273   CELL 25: 3472-3490

274   Gao H, Brandizzi F, Benning C, Larkin RM (2008) A membrane-tethered transcription factor

275   defines a branch of the heat stress response in Arabidopsis thaliana. P NATL ACAD SCI USA

276   105: 16398-16403

277   Hennerdal A, Elofsson A (2011) Rapid membrane protein topology prediction.

278   BIOINFORMATICS 27: 1322-1323

279   Iwata Y, Fedoroff NV, Koizumi N (2008) Arabidopsis bZIP60 is a proteolysis-activated

280   transcription factor involved in the endoplasmic reticulum stress response. PLANT CELL 20:

281   3107-3121

282   Iwata Y, Koizumi N (2005) An Arabidopsis transcription factor, AtbZIP60, regulates the

283   endoplasmic reticulum stress response in a manner unique to plants. P NATL ACAD SCI USA

284   102: 5280-5285

285   Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, Gao G (2016) PlantTFDB 4.0: toward a

286   central hub for transcription factors and regulatory interactions in plants. NUCLEIC ACIDS

287   RES

288   Jones DT, Taylor WR, Thorton JM (1994) A model recognition approach to the prediction of

289   all-helical membrane-protein structure and topology. BIOCHEMISTRY-US 33: 3038-3049

290   Kall L, Krogh A, Sonnhammer E (2004) A combined transmembrane topology and signal

291   peptide prediction method. J MOL BIOL 338: 1027-1036

292 Kim MJ, Park M, Seo PJ, Song J, Kim H, Park C (2012) Controlled nuclear import of the
293 transcription factor NTL6 reveals a cytoplasmic role of SnRK2.8 in the drought-stress
294 response. BIOCHEM J 448: 353-363

295 Kim S, Lee A, Yoon H, Park C (2008) A membrane-bound NAC transcription factor NTL8
296 regulates gibberellic acid-mediated salt signaling in Arabidopsis seed germination. PLANT J
297 55: 77-88

298 Kim S, Lee S, Seo PJ, Kim S, Kim J, Park C (2010) Genome-scale screening and molecular
299 characterization of membrane-bound transcription factors in Arabidopsis and rice. GENOMICS
300 95: 56-65

301 Kim Y, Kim S, Park J, Park H, Lim M, Chua N, Park C (2006) A membrane-bound NAC
302 transcription factor regulates cell division in Arabidopsis. PLANT CELL 18: 3132-3144

303 Krogh A, Larsson B, von Heijne G, Sonnhammer E (2001) Predicting transmembrane protein
304 topology with a hidden Markov model: Application to complete genomes. J MOL BIOL 305:
305 567-580

306 Lee S, Seo PJ, Lee H, Park C (2012) A NAC transcription factor NTL4 promotes reactive
307 oxygen species production during drought-induced leaf senescence in Arabidopsis. PLANT J
308 70: 831-844

309 Li P, Wind JJ, Shi X, Zhang H, Hanson J, Smeekens SC, Teng S (2011) Fructose sensitivity is
310 suppressed in Arabidopsis by the transcription factor ANAC089 lacking the membrane-bound
311 domain. P NATL ACAD SCI USA 108: 3436-3441

312 Li Y, Humbert S, Howell SH (2012) ZmbZIP60 mRNA is spliced in maize in response to ER
313 stress. BMC research notes 5: 144

314 Lu S, Yang Z, Sun L, Sun L, Song Z, Liu J (2012) Conservation of IRE1-regulated bZIP74
315 mRNA unconventional splicing in rice (*Oryza sativa* L.) involved in ER stress responses. MOL
316 PLANT 5: 504-514

317 Ng S, Ivanova A, Duncan O, Law SR, Van Aken O, De Clercq I, Wang Y, Carrie C, Xu L,
318 Kmiec B, Walker H, Van Breusegem F, Whelan J, Giraud E (2013) A membrane-bound NAC
319 transcription factor, ANAC017, mediates mitochondrial retrograde signaling in Arabidopsis.
320 PLANT CELL 25: 3450-3471

321 Park J, Kim Y, Kim S, Jung J, Woo J, Park C (2011) Integration of auxin and salt signals by the
322 NAC transcription factor NTM2 during seed germination in Arabidopsis. PLANT PHYSIOL 156:
323 537-549

324 Schwacke R, Schneider A, van der Graaff E, Fischer K, Catoni E, Desimone M, Frommer WB,
325 Flugge UI, Kunze R (2003) ARAMEMNON, a novel database for Arabidopsis integral
326 membrane proteins. PLANT PHYSIOL 131: 16-26

327 Seo PJ (2014) Recent advances in plant membrane-bound transcription factor research:
328 Emphasis on intracellular movement. J INTEGR PLANT BIOL 56: 334-342

329 Seo PJ, Kim MJ, Park J, Kim S, Jeon J, Lee Y, Kim J, Park C (2010) Cold activation of a
330 plasma membrane-tethered NAC transcription factor induces a pathogen resistance response
331 in Arabidopsis. PLANT J 61: 661-671

332 Seo PJ, Kim S, Park C (2008) Membrane-bound transcription factors in plants. TRENDS
333 PLANT SCI 13: 550-556

334   Slabaugh E, Held M, Brandizzi F (2011) Control of root hair development in *Arabidopsis*

335   *thaliana* by an endoplasmic reticulum anchored member of the R2R3-MYB transcription factor

336   family. PLANT J 67: 395-405

337   Sun X, Feng P, Xu X, Guo H, Ma J, Chi W, Lin R, Lu C, Zhang L (2011) A chloroplast

338   envelope-bound PHD transcription factor mediates chloroplast signals to the nucleus. NAT

339   COMMUN 2:477.

340   Tusnady GE, Simon I (2001) The HMMTOP transmembrane topology prediction server.

341   BIOINFORMATICS 17: 849-850

342   Viklund H, Elofsson A (2004) Best alpha-helical transmembrane protein topology predictions

343   are achieved using hidden Markov models and evolutionary information. PROTEIN SCI 13:

344   1908-1917

345   Vonheijne G (1992) Membrane-protein structure prediction-hydrophobicity analysis and the

346   positive-inside rule. J MOL BIOL 225: 487-494

347   Yang Z, Lu S, Wang M, Bi D, Sun L, Zhou S, Song Z, Liu J (2014) A plasma

348   membrane-tethered transcription factor, NAC062/ANAC062/NTL6, mediates the unfolded

349   protein response in Arabidopsis. PLANT J 79: 1033-1043

350   Yoon H, Kim S, Kim S, Park C (2008) Regulation of leaf senescence by NTL9-mediated

351   osmotic stress signaling in Arabidopsis. MOL CELLS 25: 438-445

352   Zhang Y, Li Y, Zhang Y, Wang Z, Zhao M, Su N, Zhang T, Chen L, Wei W, Luo J, Zhou Y, Xu

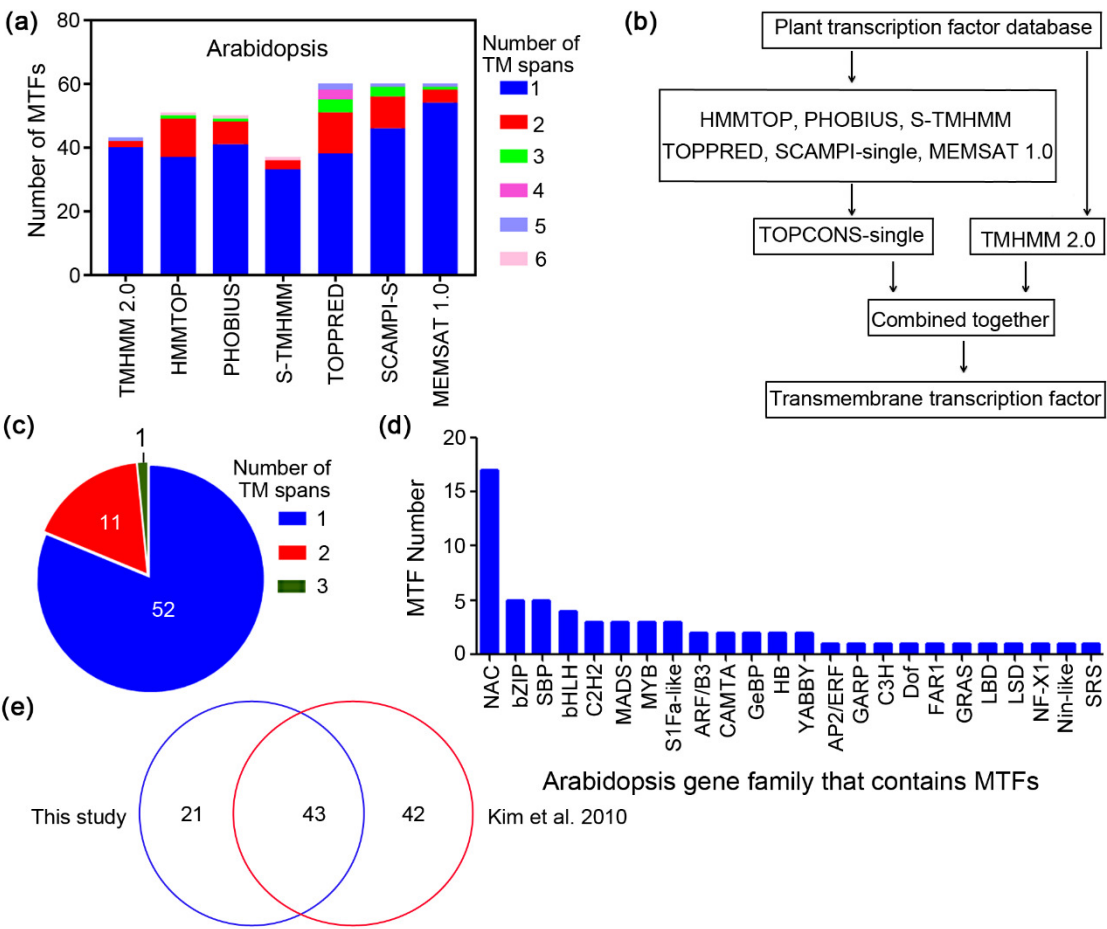
353   Y, Xu Y, Xu P, Li W, Tao Y (2015) Quantitative proteomics reveals membrane

354   protein-mediated

355   hypersaline sensitivity and adaptation in halophilic *Nocardiopsis xinjiangensis*. J PROTEOME

356   RES 15: 68-85

357

358

360 **Fig. 1**. Workflow for identification of membrane-bound transcription factors in plant
361 species. (a) Prediction of transmembrane (TM) spans of Arabidopsis by different methods;
362 (b) workflow for identification of MTFs in plant species; (c) Arabidopsis MTFs identified in
363 present study; (d) gene family of Arabidopsis MTFs; (e) Comparison of the Arabidopsis
364 MTFs identified in this study with the previous research.

365

366

367

368

369

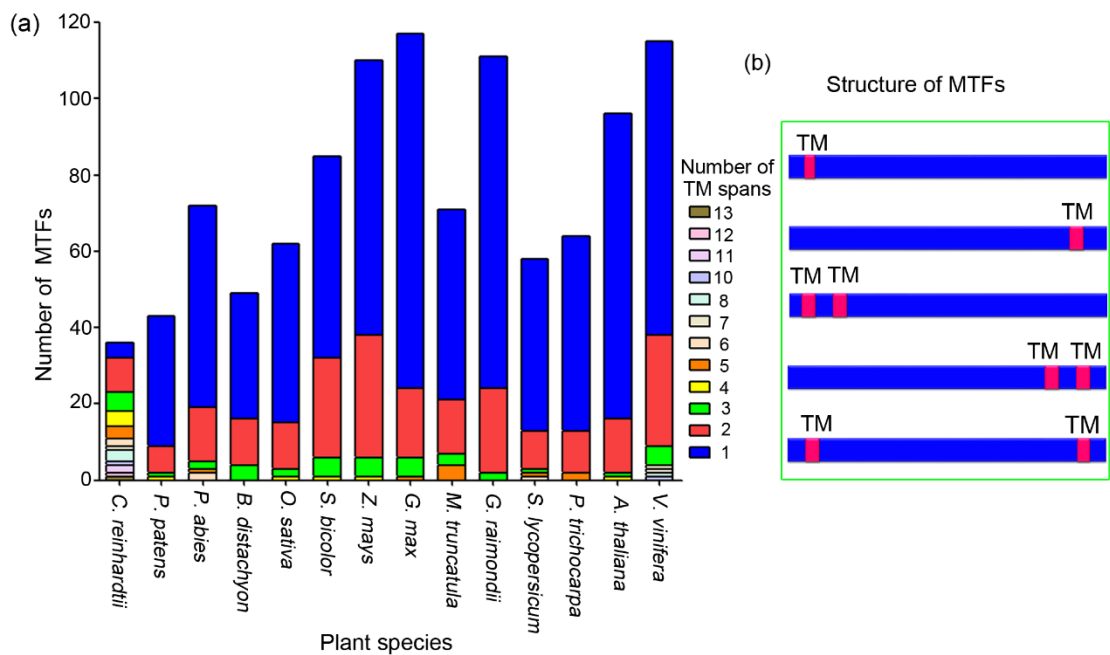



Fig. 2. Analysis of transmembrane motifs of membrane-bound transcription factors in plant species. (a) Number of TM spans in MTFs; (b) typical locations of TM spans in MTFs. TM, transmembrane; MTF, membrane-bound transcription factor.

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

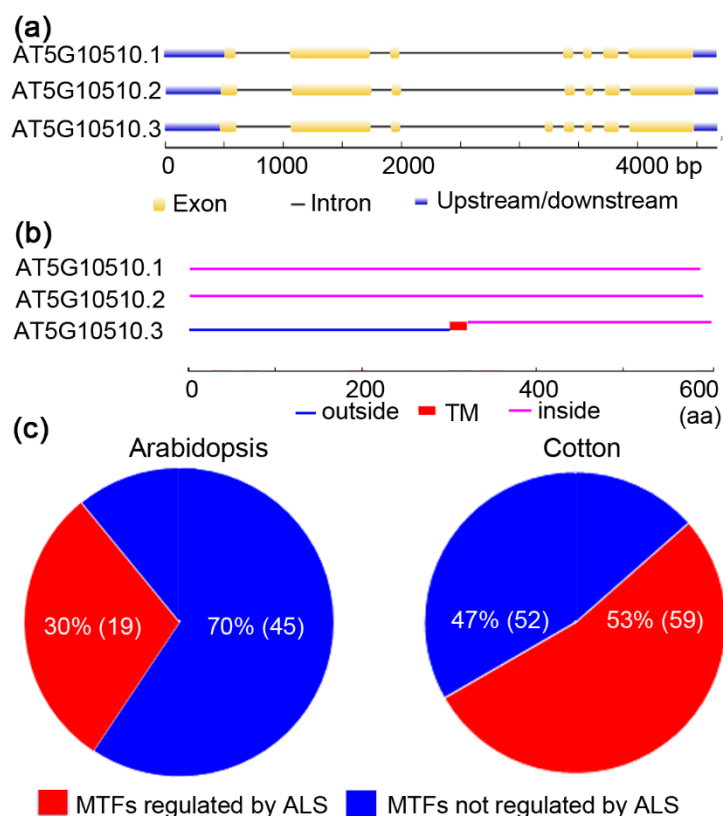



Fig. 3. Alternative splicing of membrane-bound transcription factors in Arabidopsis
(*Arabidopsis thaliana*) and cotton (*Gossypium raimondii*). (a) Arabidopsis
membrane-bound transcription factor (AT5G10510) undergoes alternative spicing (ALS);
(b) Alternatively spliced forms of Arabidopsis AT5G10510 lack a TM domain; (c) Potential
ALS dependent activation of membrane-bound transcription factors in genome of
Arabidopsis and cotton.

Table 1. Genome information and sizes of the MTFs of analyzed plant species.

Species	Common name	Genome size (Mbp)	Chromosome (1N)	TF loci	MTF loci
Chlorophyta					
Chlamydomonas reinhardtii	Green algae	112	17	230	37
Bryophyta					
Physcomitrella patens	Moss	480	27	1079	49
Gymnosperm					
Picea abies	Norway spruce	19.6 Gb	12	1851	71
Monocot					
Brachypodium distachyon	Brachypodium	272	5	1557	50
Oryza sativa	Japanese rice	372	12	1859	85
Sorghum bicolor	Sorghum	730	10	1826	62
Zea mays	Maize	3,000	10	2231	110
Eudicot					
Glycine max	Soybean	979	20	3714	117
Medicago truncatula	Barrel medic	258	8	1577	71
Gossypium raimondii	Cotton	880	13	2634	111
Solanum lycopersicum	Tomato	900	12	1845	58
Populus trichocarpa	Western balsam poplar	423	19	2455	89
Arabidopsis thaliana	Thale cress	135	5	1716	64
Vitis vinifera	Wine grape	487	19	1276	115

Topology predictor	Algorithm	Reference
TMHMM 2.0	HMM	(Krogh et al., 2001)
HMMTOP	HMM	(Tusnady & Simon, 2001)
PHOBIUS	HMM	(Kall et al., 2004)
S-TMHMM	HMM	(Viklund and Elofsson, 2004)
TOPPRED	Hydrophobicity profiles	(von Heijne, 1992)
SCAMPI-single	Hydrophobicity+Model	(Bernsel et al., 2008)
MEMSAT 1.0	ANN+Grammer	(Jones et al., 1994)

Table 2. The membrane topology prediction methods used in the current study.

Table 3. Membrane-bound transcription factors (MTFs) in gene families of plant species.

Family													
Chlorophyte	Bryo-phyte	Gymno-sperm	Monocot				Eudicot						
Cr	Pp	Pa	Bd	Os	Sb	Zm	Gm	Mt	Gr	Sl	Pt	At	Vv

Header note: "Plant species (MTF number)" spans the species columns.

Family	Cr	Pp	Pa	Bd	Os	Sb	Zm	Gm	Mt	Gr	Sl	Pt	At	Vv
AP2	4	0	0	1	1	0	0	1	1	0	0	0	1	2
ARF	0	0	3	1	0	0	2	2	0	2	0	1	0	0
ARR-B	0	0	0	0	0	0	0	1	0	1	0	1	0	1
B3	0	4	0	3	2	1	0	1	2	1	0	0	2	2
BBR-BPC	0	0	1	0	0	0	0	0	0	0	0	1	0	0
BES1	0	1	1	0	0	0	1	1	0	1	0	1	0	0
bHLH	4	2	3	3	6	5	8	13	3	4	0	4	4	4
bZIP	3	6	5	3	4	4	6	5	3	5	5	9	5	4
C2H2	0	3	4	4	5	6	5	9	6	11	5	3	3	6
C3H	3	2	7	3	3	0	2	2	3	2	0	5	1	5
CAMTA	0	0	0	0	0	0	2	2	2	5	0	0	2	0
CO-like	1	0	0	0	0	0	1	0	0	0	0	2	0	0
CPP	1	0	0	0	0	0	0	0	0	0	1	1	0	1
DBB	0	0	0	0	0	0	1	2	0	0	0	2	0	0
Dof	0	0	2	1	1	0	1	1	0	1	0	0	1	4
E2F/DP	0	0	1	0	0	0	0	2	0	1	0	1	0	0
EIL	0	0	0	0	1	0	0	0	1	0	0	0	0	0
ERF	0	3	3	1	3	3	7	1	1	3	2	1	0	10
FAR1	0	0	0	3	11	2	3	0	10	3	0	2	1	3
G2-like	1	2	1	2	4	0	3	2	1	4	3	1	1	2
GATA	2	1	0	1	2	1	1	2	2	0	0	1	0	2
GeBP	0	0	0	0	0	1	0	0	0	0	0	0	2	0
GRAS	0	0	3	1	4	11	6	0	1	0	2	0	1	4
GRF	0	0	0	0	0	0	0	1	0	1	0	0	0	0
HB-other	1	0	3	1	2	0	4	3	1	4	3	5	2	4
HD-ZIP	0	1	0	1	0	1	6	3	0	6	0	1	0	3
HSF	0	0	1	2	0	1	1	0	0	0	0	0	0	1
LBD	0	2	1	0	0	0	0	0	1	1	1	0	1	1
LSD	0	0	0	0	2	0	0	1	0	0	0	1	1	1
MIKC	0	0	0	2	1	2	3	2	0	4	0	1	1	2
M-type	0	1	3	1	3	1	2	4	2	3	7	3	2	0
MYB	1	1	0	0	1	2	1	2	0	4	0	4	1	11
MYB-related	4	1	6	2	4	2	10	7	6	6	5	8	2	2

Continued-Table 3. Membrane-bound transcription factors (MTFs) in gene families of plant species.

Family	Plant species (MTF number)													
	Chlorophyte	Bryop-hyte	Gymno-sperm	Monocot				Eudicot						
	Cr	Pp	Pa	Bd	Os	Sb	Zm	Gm	Mt	Gr	Sl	Pt	At	Vv
NAC	0	6	7	7	6	7	11	14	5	15	11	12	17	10
NF-X1	1	1	0	0	1	1	1	2	1	1	1	1	1	2
NF-YA	0	0	1	0	1	0	1	1	1	0	0	0	0	0
NF-YB	0	0	1	0	0	0	0	3	0	1	2	0	0	1
NF-YC	0	1	1	1	1	0	0	0	1	0	1	0	0	0
Nin-like	1	2	0	1	1	0	0	2	1	1	0	0	1	2
RAV	0	0	0	0	0	0	0	0	0	0	0	2	0	0
S1Fa-like	1	2	2	1	2	1	2	4	3	4	1	2	3	2
SAP	0	0	0	0	0	0	0	0	1	1	0	0	0	0
SBP	9	4	2	4	5	5	7	9	3	9	3	8	5	4
SRS	0	0	0	0	0	0	0	1	2	1	0	1	1	1
TALE	0	0	0	0	0	1	0	0	0	0	1	0	0	1
TCP	0	0	2	0	0	0	3	0	0	0	0	1	0	5
Trihelix	0	2	2	0	0	0	3	3	0	4	2	1	0	5
Whirly	0	0	0	0	0	0	0	0	0	0	0	0	0	1
WOX	0	0	0	0	0	0	1	0	0	0	1	0	0	1
WRKY	0	1	5	0	6	3	3	4	5	1	1	0	0	2
YABBY	0	0	0	0	0	1	2	3	0	0	0	2	2	0
ZF-HD	0	0	0	0	2	0	0	1	2	0	0	0	0	3