Plant and Cell Physiology Advance Access published June 26, 2014

Cloning and characterization of AabHLH1, a bHLH transcription factor that positively regulates artemisinin biosynthesis in *Artemisia annua*

Yunpeng Ji [a, 1], Jingwei Xiao [b, 1], Yalin Shen [a], Dongming Ma [c], Zhenqiu Li [d], Gaobin Pu [c], Xing Li [a], Lili Huang [c], Benye Liu [c], Hechun Ye [c], Hong Wang [a,*]

a University of Chinese Academy of Sciences, Beijing 100049, China

b National Institute of Occupational Health and Poison Control, Chinese Center for Disease Control and Prevention, Beijing 100050, China

c Key Laboratory of Plant Molecular Physiology, Institute of Botany, The Chinese Academy of Sciences, Beijing 100093, China

d Hebei University, Baoding 071002, China

* Corresponding author. Tel: +86-10-88256601; Fax: +86-10-88256079

E-mail address: hwang@ucas.ac.cn (H. Wang)

[1] These authors contributed equally to this article

Running title: AabHLH1 positively regulates artemisinin biosynthesis

Abstract:

Amorpha-4,11-diene synthase (ADS) and P450 monooxygenase (CYP71AV1) in *Artemisia annua* L. are two key enzymes involved in the biosynthesis of artemisinin. The promoters of *ADS* and *CYP71AV1* contain E-box elements, which are putative binding sites for bHLH transcription factors. This study successfully isolated a bHLH transcription factor from *A. annua*, designated as *AabHLH1*, from a cDNA library of the glandular secretory trichomes (GSTs) in which artemisinin is synthesized and sequestered. *AabHLH1* encodes a protein of 650 amino acids containing one putative bHLH domain. *AabHLH1* and *ADS* genes were strongly induced by ABA and the fungal elicitor, chitosan. The transient expression analysis of the AabHLH1-GFP reporter gene revealed that AabHLH1 was targeted to nuclei. Biochemical analysis demonstrated that the AabHLH1 protein was capable of binding to the E-box *cis*-elements, present in both *ADS* and *CYP71AV1* promoters, and possessed transactivation activity in yeast. In addition, transient co-transformation of *AabHLH1* and *CYP71AV1Pro*::*GUS* in *A. annua* leaves showed a significant activation of the expression of the *GUS* (*β-glucuronidase*) gene in transformed *A. annua*, but mutation of the E-boxes resulted in activation abolition, suggesting the E-box is important for the *CYP71AV1* promoter activity. Furthermore, transient expression of *AabHLH1* in *A. annua* leaves increased

transcript levels of the genes involved in artemisinin biosynthesis, such as *ADS*, *CYP71AV1* and *HMGR*. These results suggest that *AabHLH1* can positively regulate the biosynthesis of artemisinin.

Keywords: Amorpha-4,11-diene synthase (ADS) • *Artemisia annua* L. • Artemisinin biosynthesis • bHLH transcription factor • Cytochrome P450 monooxygenase (CYP71AV1).

Abbreviations: ABA, abscisic acid; ACTs, artemisinin-based combination therapies; ADS, amorpha-4,11-diene synthase; 3-AT, 3-amino-1,2,4-triazole; bHLH, basic/helix-loop-helix; BSA, bovine serum albumin; CaMV, cauliflower mosaic virus; CYP71AV1, cytochrome P450 monooxygenase; DBR2, artemisinic aldehyde Δ11(13) reductase; DMAPP, dimethylallyl diphosphate; EMSA, electrophoretic mobility shift assay; EST, expressed sequence tag; FPS, farnesyl diphosphate synthase; GFP, green fluorescent protein; GST, glandular secretory trichome; GUS, β-glucuronidase; HMGR, 3-hydroxy-3-methylglutaryl coenzyme A reductase; IPP, isopentenyl diphosphate; MEP, 2C-methyl-D-erythrito4-phosphate; NLS, nuclear localization signal; ORF, open reading frame; RACE, rapid amplification of cDNA ends; RT-PCR, reverse transcription-PCR.

Introduction

Artemisia annua, or sweet wormwood, an annual plant from the Asteraceae family, has attracted attention as the source of an alternative to quinoline drugs for the treatment of malaria. The active ingredient is artemisinin, a sesquiterpene lactone, and its semi-synthetic derivatives (Artemeter, Artesunate) have been developed as drugs by the pharmaceutical industry (Tissier 2012). Artemisinin-based combination therapies (ACTs) have now been recommended since 2001 by the World Health Organization (WHO) in order to reduce the risk of drug resistance (Mutabingwa 2005). Currently, *A. annua* is the only commercial source of artemisinin. However, artemisinin contents are low (0.01–1.00% dry weight) in the leaves and flowers of *A. annua*, but in some hybrid varieties, the content can reach 1.4%. The low content of artemisinin in *A. annua* severely limits its commercialization and has triggered numerous efforts to improve artemisinin production. An alternative microbial-based system that synthesizes an artemisinin precursor for chemical conversion has been developed (Ro et al. 2006). However, this would not replace agricultural production, which will continue to be an essential supply source (Graham et al. 2010). Owing to the importance of artemisinin in the treatment of malaria, a better understanding of its biosynthetic pathway is needed in order to increase its production.

The artemisinin biosynthetic pathway is gradually being understood as the genes involved in artemisinin biosynthesis are identified. It is believed that artemisinin biosynthesis precursor, isopentenyl diphosphate (IPP), comes from the mevalonate pathway, which is localized in the cytosol. IPP is synthesized from acetyl-CoA via mevalonic acid, a process that is catalyzed by 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR) (Argolo et al. 2000). The five-carbon building blocks of artemisinin metabolism, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) are mainly derived from the mevalonate pathway (Fig. 1), but they are also partly derived from glyceraldehyde-3-phosphate and pyruvate via the 2C-methyl-D-erythrito4-phosphate (MEP) pathway (Towler and Weathers 2007). Farnesyl diphosphate synthase (FPS) (Brodelius et al. 2002; Han et al. 2006) catalyzes condensation of two molecules of IPP and one DMAPP into farnesyl diphosphate and then the cyclization of farnesyl diphosphate is catalyzed by amorpha-4,11-diene synthase (ADS), which has been cloned by several laboratories (Chang et al. 2000; Mercke et al. 2000; Wallaart et al. 2001). Following this, armorph-4,11-diene is

hydroxylated to yield artemisinic alcohol, which is catalyzed by a cytochrome P450 dependent armorph-4,11-diene 12-hydroxylase (CYP71AV1) (Ro et al. 2006; Teoh et al. 2006). This enzyme can also covert artemisinic alcohol to artemisinic aldehyde and then into artemisinic acid. It has long been assumed that artemisinic acid is a direct precursor of artemisinin. However, recent experiments have shown that dihydroartemisinic acid is the precursor of artemisinin. Dihydroartemisinic acid is formed from artemisinic aldehyde in two steps via dihydroartemisinic aldehyde. The reduction is catalyzed by artemisinic aldehyde Δ11(13) reductase (DBR2) (Zhang et al. 2008) and oxidation to the acid is by aldehyde dehydrogenase 1 (ALDH1) (Teoh et al. 2009). The intermediate, dihydroartemisinic acid, is converted to artemisinin in a non-enzymatic reaction. An integrated artemisinin biosynthetic pathway for *A. annua* is summarized in Fig. 1.

Artemisia, like many other species from the Asteraceae, produces sesquiterpene lactones in glandular secretory trichomes (GSTs) localized on the leaves, stems and flowers. In *A. annua*, the glandular secretory trichomes are thought to be the site of biosynthesis and storage of artemisinin (Duke et al. 1994; Covello et al. 2007). In recent years, the studies on specific expression in trichomes have mainly focused on ADS and CYP71AV1. The fact that ADS is highly expressed in trichomes was confirmed (Bertea et al. 2005) and the specific expression of ADS in trichomes is shown in transgenic plants carrying ADS-promoter::GUS (Wang et al. 2011a). Very recently, the CYP71AV1 promoter has been cloned and promoter-GUS fusion exhibits a trichome-specific expression pattern (Wang et al. 2011b; Wang et al. 2013). It is probable that most of the genes involved in the artemisinin biosynthetic pathway are likely to be trichome specific (Liu et al. 2009; Olsson et al. 2009; Wang et al. 2009). These trichome specific genes are likely to be controlled by one or several transcription factors. In recent years, some studies on the effects of transcription factors on the biosynthesis of artemisinin in *A. annua* have been reported. For example, the AaWRKY1 transcription factor has been shown to positively regulate the biosynthesis of artemisinin by binding to the promoter of ADS (Ma et al. 2009) and two JA-responsive AP2/ERF subfamily transcriptional factors (AaERF1 and AaERF2) have been shown to positively regulate artemisinin biosynthesis by binding to CBF2 and RAA motifs that are present in both the ADS and the CYP71AV1 promoters in *A. annua* (Yu et al. 2012).

The basic/helix-loop-helix (bHLH) proteins are the second largest transcription factor family in the genome, behind only the MYB superfamily. Members of this family share the bHLH signature

domain, which consists of around 60 amino acids with two distinct regions: a basic stretch at the N-terminus, consisting of around 15 amino acids involved in DNA binding, and a C-terminal region of around 40 amino acids composed of two amphipathic α-helices, mainly consisting of hydrophobic residues linked by a variable loop (the "helix-loop-helix" region), which is responsible for promoting protein-protein interactions through the formation of homo- and hetero-dimeric complexes (Toledo-Ortiz et al. 2003; Osorio et al. 2012). The bHLHs are found in all three eukaryotic kingdoms and are involved in a myriad of regulatory processes. They are involved in developmental processes, such as stomata development, root hair formation and in the control of petal size (Pires and Dolan 2010). They are also reportedly involved in the metabolic regulation of various hormones, including brassinosteroids (BRs) (Friedrichsen et al. 2002), abscisic acid (ABA) (Abe et al. 2003) and auxin synthesis/homeostasis (Schlereth et al. 2010). Some bHLHs are involved in the regulation of light signaling (Castelain et al. 2012), iron and phosphate homeostasis (Yi et al. 2005; Long et al. 2010) and various abiotic stresses, including cold, drought and high salinity (Abe et al. 2003; Chinnusamy et al. 2003; Kim and Kim 2006). Interestingly, two bHLH transcription factors are reported to regulate alkaloid biosynthesis in *Catharanthus roseus* and isoquinoline alkaloid biosynthesis in *Coptis japonica* (Yamada et al. 2011; Zhang et al. 2011) . The core DNA sequence motif recognized by the bHLH proteins is a consensus hexanucleotide sequence known as the E-box (5'-CANNTG-3'). The promoters of ADS and CYP71AV1 involved in the biosynthesis of artemisinin contain E-box elements, which are putative binding sites for bHLH transcription factors. In order to further explore the transcription factors capable of regulating the biosynthesis of artemisinin, this study initiated the cloning and characterization of bHLH transcription factors from a cDNA library of glandular secretory trichomes (GSTs). The specific binding of the protein encoded by *AabHLH1* to E-box *cis*-elements was verified by EMSA (Electrophoresis Mobility Shift Analysis) and a yeast one-hybrid. Overexpression of *AabHLH1* in transiently transformed *A. annua* leaves was able to activate key enzymes related to artemisinin biosynthesis and the promoters of *ADS* and *CYP71AV1*. These results suggest that *AabHLH1* can regulate the biosynthesis of artemisinin.

Results

Isolation and sequence analysis of *AabHLH1* cDNA

In previous studies, a cDNA library of *A. annua*, was constructed using glandular secretory trichomes (GSTs) as the source of mRNA and three EST fragments of the candidate bHLH transcription factor were found by alignment through the NCBI in the cDNA library (Ma et al. 2009). Since amorpha-4,11-diene synthase (ADS) and P450 monooxygenase (CYP71AV1) in *A. annua* are key enzymes involved in the biosynthesis of artemisinin and the promoters of *ADS* and *CYP71AV1* contain E-box elements, which are putative binding sites for bHLH transcription factors, RACE-PCR was used in this study in order to try and get the whole cDNA sequences of the three EST fragments in the cDNA library. However, only the cDNA sequences of two bHLH transcription factors, designated as *AabHLH1* and *AabHLH2*, were obtained. In addition, AabHLH1 was found to bind to the E-box *cis*-elements (Figs. 5 and 6) while AabHLH2 failed to bind to the E-box (data not shown). As a result, this study only concentrated on AabHLH1. The full-length cDNA of *AabHLH1* was found to be a 1,950 bp long open reading frame (ORF). The ORF encoded a protein of 650 amino acids with a calculated mol. wt. of 71.83 kDa and an isoelectric point of 5.75. AabHLH1 contained one basic helix-loop-helix conservative domain at the N-teminus and AabHLH1 belonged to Group B of the bHLH family, based on the Atchley theory (Atchley et al. 2000). In general, the basic region in the bHLH domain determines the DNA binding activity of the protein. Toledo-Ortiz et al. (2003) identified 147 proteins as members of the bHLH family in *Arabidopsis* and, depending on the presence or absence of basic residues in the first 17 positions of the bHLH domain, two major categories were defined: (1) DNA binding bHLHs and (2) non-DNA binding bHLHs. The DNA binding bHLH category can be further subdivided into two subcategories based on the predicted DNA binding sequence: (1) the E-box binders and (2) the non-E-box binders. This subdivision is based on the presence or absence of two specific residues: Glu (E) and Arg (R), in the basic region (Toledo-Ortiz et al. 2003). AabHLH1 belongs to the E-box binders because it contains an average of six basic residues and has Glu (E) and Arg (R) in the first 17 positions of basic region (Table 2). In previous studies, bHLH transcription factors were also found in other plants besides *Arabidopsis*. The alignment of AabHLH1 with some of the bHLH proteins from other plants is shown in Fig. 2. AabHLH1 exhibited 48.43% sequence identity with CrMYC2 protein, 44.37% sequence identity with LjbHLH22 protein, 47.64% identity with NbbHLH1 protein, 47.91% sequence identity with StbMYC1 protein, 48.82% sequence identity with NtMYC1a protein and 44.63% identity with

PvPG1 protein (Fig. 2).

Expression pattern of *AabHLH1*

AabHLH1 was highly expressed in the flowers and moderately expressed in the leaves and stems. A relatively weak expression was observed in the roots (Fig. 3A). Chitosan has been shown to be an effective elicitor of artemisinin biosynthesis, both in hairy root cultures and in the *A. annua* plant (Putalun et al. 2007). Abscisic acid (ABA) at different concentrations (1, 10 and 100 μmol l^{-1}) was tested by treating *A. annua* plants and the results showed that the artemisinin content in ABA-treated plants significantly increased (Jing et al. 2009). In *Arabidopsis*, bHLH transcription factor AtMYC2 could be induced by drought and ABA treatment (Abe et al. 2003). Since the ADS promoter contains E-box elements, the induced expression of the *AabHLH1* and *ADS* under separate ABA and chitosan treatments was investigated using semi-quantitative RT-PCR. The results showed that the expression of AabHLH1 was induced rapidly and strongly following treatment with 100 μmol l^{-1} ABA or 150 mg l^{-1} chitosan. In ABA-treated leaves, the transcript levels of AabHLH1 increased rapidly and peaked within 30 min. This activation persisted for up to 1 h post-treatment and then declined to their original levels by 1.5 h post-treatment (Fig. 3B). ADS showed a slower rate of induction than that of AabHLH1, with the transcript level increasing after 1 h post-treatment. This is consistent with the assumption that AabHLH1 transcription factor regulates the expression of ADS. In chitosan-treated leaves, the expression level of ADS showed similar kinetics to that of ABA induction, whereas the level of AabHLH1 increased relatively slowly compared to that of ABA induction, which showed strong induction after 1 h post-treatment (Fig. 3C).

AabHLH1 was targeted to the nucleus

Analysis of the AabHLH1 protein sequence using the PSORT program (Nakai and Kanehisa 1992) revealed the presence of four putative nuclear localization signals: KKKR between positions 393 and 396 of the amino acid sequence, KKPR between positions 451 and 454 of the amino acid sequence, KPRK between positions 452 and 455 of the amino acid sequence and PRKR between positions 453 and 456 of the amino acid sequence. To examine the cellular localization of AabHLH1, AabHLH1 was fused with green fluorescent protein (GFP) and the chimeric gene was

under the control of the *CaMV35S* promoter. The resulting plasmid was then bombarded into onion epidermal cells. As shown in Fig. 4, the AabHLH1-GFP fusion protein was localized exclusively in the nucleus, whereas the control GFP (*CaMV35S::GFP*) was distributed in both the cytoplasm and the nucleus. These results indicate that AabHLH1 is a nuclear protein, which is consistent with its function as a transcriptional regulator.

Electrophoretic mobility shift assay (EMSA)

The promoter of the *ADS* gene was isolated from *A. annua* (001) according to DQ448294 in the NCBI (Kim et al. 2008) and it was found to contain fourteen E-box elements, $_{-2108}$CACATG$_{-2103}$, $_{-1999}$CATATG$_{-1994}$, $_{-1877}$CATTTG$_{-1872}$, $_{-1743}$CATCTG$_{-1738}$, $_{-1533}$CAGTTG$_{-1528}$, $_{-1393}$CAAATG$_{-1388}$, $_{-1289}$CAAATG$_{-1284}$, $_{-1144}$CACATG$_{-1139}$, $_{-917}$CAAATG$_{-912}$, $_{-659}$CATTTG$_{-654}$, $_{-625}$CATATG$_{-620}$, $_{-390}$CATTTG$_{-385}$, $_{-158}$CAATTG$_{-153}$ and $_{-90}$CAATTG$_{-85}$. In addition, the *CYP71AV1* gene promoter was also obtained from *A. annua* (001) according to Wang et al. (2011b), and six E-boxes ($_{-1840}$CATATG$_{-1835}$, $_{-1823}$CAATTG$_{-1818}$, $_{-1749}$CAAGTG$_{-1744}$, $_{-1218}$CAAATG$_{-1213}$, $_{-587}$CATCTG$_{-582}$ and $_{-32}$CATTTG$_{-27}$) were found in the *CYP71AV1* gene promoter.

It is known that the E-box is a putative recognition site for bHLH transcription factors. In order to test the AabHLH1 protein-E-box DNA interactions, an electrophoretic mobility shift assay (EMSA) was performed. Recombinant AabHLH1 was overexpressed in *Escherichia coli* and purified using immobilized metal ion affinity chromatography. The molecular weight of the purified protein was about 73 kDa (Supplementary Fig. S1). The DNA fragments containing the 3 x E-box were used as probes. A shift band was observed with the recombinant AabHLH1 protein incubated with the 3 x E-box probe (Fig. 5). No binding signal was detected in the negative control, where the 3 x E-box was incubated with 2 μg BSA protein (Fig. 5). Therefore, the observed probe retardation was assigned to the DNA binding activity of AabHLH1. To verify AabHLH1 binding capacity, the same amounts of E-box DNA were incubated with 34 ng, 54 ng, 81 ng and 102 ng AabHLH1 proteins in lanes 2, 3, 4 and 5, respectively. The amount of shift protein-DNA complexes showed a gradually increasing tendency; whereas, free DNA levels decreased accordingly. The results indicate that AabHLH1 is able to recognize and interact with the E-box elements *in vitro*.

Yeast one-hybrid assay

To determine whether AabHLH1 binds to the E-box *in vivo*, a yeast one-hybrid assay was performed. The 3 x E-box fragment was fused with a HIS reporter gene to generate a pHIS-3 x E-box. Similarly, a pHIS-3 x m_1E-box and a pHIS-3 x m_2E-box were obtained by fusing a HIS reporter gene with two mutants of the E-box. All the above three plasmids were introduced into yeast strain Y187. The ORF of AabHLH1 was inserted into the pGADT7::AD vector in order to generate pGADT7::AabHLH1. The pGADT7::AabHLH1 was introduced into a yeast strain Y187, which contained a plasmid harboring the pHIS-3 x E-box, the pHIS-3 x m_1E-box or the pHIS-3 x m_2E-box. At the same time, the empty pGADT7::AD plasmid was introduced into a yeast strain Y187, which contained a plasmid harboring the pHIS-3 x E-box as a control. On SD/-Trp/-Leu/-His selective medium containing 30 mM 3-amino-1,2,4-triazole (3-AT), only the yeast harboring pGADT7::AabHLH1 and the pHIS-3 x E-box could grow (Fig. 6). When the core sequence of the E-box was mutated, as shown in the 3 x m_1E-box and the 3 x m_2E-box, AabHLH1 could no longer bind to the E-box. This result indicates that AabHLH1 could bind to the E-box and activate transcription specifically *in vivo*.

AabHLH1 activated ADS, CYP71AV1 promoters, as well as the genes involved in artemisinin biosynthesis

In order to study the interactions of AabHLH1 with ADS and CYP71AV1 promoters, this study cloned an ADS promoter according to DQ448294 in the NCBI (Kim et al. 2008) and a CYP71AV1 promoter according to Wang et al. (2011b). Then two constructs were generated using the amplified ADS and CYP71AV1 promoters: pCAMBIA-*ADSpro* and pCAMBIA-*CYP71AV1pro*. The interactions between AabHLH1 and the ADS and CYP71AV1 promoters were tested in an *Agrobacterium*-mediated transient expression system in *A. annua* leaves. When only the reporter plasmid, pCAMBIA-*ADSpro* or pCAMBIA-*CYP71AV1pro*, was introduced into *A. annua*, the GUS expression was about three or two times lower than that of *35S*::*AabHLH1*/*ADSpro*::*GUS* or *35S*::*AabHLH1*/*CYP71AV1pro*::*GUS* co-transformation, respectively (Fig. 7A).

Since AabHLH1 can bind to the E-box *in vitro* and *in vivo*, and there are several E-boxes in *ADS* and *CYP71AV1* promoters, it is of interest to know whether AabHLH1 can activate *ADS* and

CYP71AV1 in the artemisinin biosynthetic pathway. Hence, the pBI-AabHLH1 construct and the empty vector pBI121 were introduced into the leaves of *A. annua* by *Agrobacterium* co-cultivation with the help of a vacuum. When the pBI-AabHLH1 construct was introduced into *A. annua* leaves, the transcript levels of AabHLH1 were much higher than those of the control after 36 h co-cultivation. Semi-quantitative RT-PCR results showed that the expression levels of *ADS*, *CYP71AV1* and *HMGR* had significantly increased with the overexpression of *AabHLH1*; the expression level of *DBR2* had increased mildly and the expression levels of *FPS* and *ALDH1* were almost unchanged. However, the expression level of *RED1*, a gene that has a putatively negative role in artemisinin biosynthesis, decreased with the overexpression of *AabHLH1* (Fig. 7B). Thus, overexpression of *AabHLH1* could strongly activate the expression of *ADS* and *CYP71AV1*, members of the artemisinin biosynthetic genes.

Abolished activation of AabHLH1 by mutated E-boxes of *CYP71AV1* promoter

As mentioned above, *ADS* gene promoter contained 14 E-boxes, *CYP71AV1* gene promoter contained 6 E-boxes, so we selected 750 bp-long *CYP71AV1* gene promoter containing 2 E-boxes for the mutated experiment. Transient co-transformation of *AabHLH1* and *CYP71AV1Pro*::*GUS* in *A. annua* leaves significantly activated the expression of the *GUS* (*β-glucuronidase*) gene in transformed *A. annua*, but mutation of the E-boxes showed a clear reduction of GUS activity (Fig. 8, $P<0.05$**),** this suggested that the E-box was important for the *CYP71AV1* gene promoter activity. The results indicated that the E-boxes themselves are actually required for the activity of *CYP71AV1* gene promoter**.**

Discussion

Using isolated glandular secretory trichomes as a source of mRNA, Teoh et al. constructed a cDNA library and CYP71AV1, which catalyzes multiple oxidations of the sesquiterpene intermediate amorpha-4,11-diene to artemisinic acid, was cloned using expressed sequence tag (ESTs) methods (Teoh et al. 2006). Similarly, an artemisinic aldehyde Δ11(13) reductase (DBR2) was isolated from the cDNA library of GSTs (Zhang et al. 2008). In a previous report, AaWRKY1 transcription factor, which can regulate ADS gene expression, was also cloned from the cDNA

library of GSTs (Ma et al. 2009). So, this approach is fundamental in identifying genes and transcription factors that could play a role in artemisinin biosynthesis. In the present study, the evaluation of expressed sequence tags (ESTs) from the cDNA library of GSTs revealed a bHLH cDNA, designated as AabHLH1. Amorpha-4,11-diene synthase (ADS), a sesquiterpene synthase, and CYP71AV1, a cytochrome P450 monooxygenase, are two key enzymes in the biosynthesis pathway of artemisinin. The promoters of *ADS* and *CYP71AV1* contain several E-box elements, which are putative recognition sites for bHLH transcription factors. A total of fourteen E-box elements were found in the promoter of the *ADS* gene (Supplementary Fig. S2, Kim et al. 2008). Very recently, two *CYP71AV1* gene promoters were isolated from different chemtypes of *A. annua*. One was isolated by (Wang et al. 2011b) contained six E-boxes (Supplementary Fig. S3) and the other by (Wang et al. 2013) and contained five E-boxes ($_{-8}$CATTTG$_{-3}$, $_{-74}$CAAATG$_{-69}$, $_{-689}$CACTTG$_{-684}$, $_{-763}$CAACTG$_{-758}$ and $_{-1021}$CATTTG$_{-1016}$). Alignment of the above two sequences (Wang et al. 2011b; Wang et al. 2013) with the other published *CYP71AV1* gene promoter sequences (GenBank entries HM48927 and EF015297) showed that the nucleotide sequences were different. Notably, the number of E-boxes was different. No matter how many E-boxes *CYP71AV1* promoters have, the consensus hexanucleotide sequence is consistent with E-box (5'-CANNTG-3'). Yeast one-hybrid and electrophoretic mobility shift assay (EMSA) showed that AabHLH1 was able to bind to E-box motifs present in both *ADS* and *CYP71AV1* promoters. Furthermore, Yeast one-hybrid and transient transformation of *A. annua* with mutated E-boxes indicated that the E-boxes themselves are actually needed for the activity of *CYP71AV1* gene promoters (Figs 5, 6 and 8). Activation of AabHLH1 was abolished when two E-boxes were mutated. In a general sense, E-box elements (CANNTG) serve as a binding site of bHLH transcription factors, when the CANNTG motif is mutated, the binding activity is lost (Abe et al. 1997; Hong et al. 2012).

ABA plays an important role in plant response to different biotic and abiotic stresses and the content of ABA often increases in response to drought stress. When soil-grown *A. annua* plants were sprayed with 1–100 μmol l^{-1} ABA, there was a significant increase in artemisinin content compared to the controls (Jing et al. 2009). Chitosan was shown to be an effective elicitor of artemisinin biosynthesis in hairy root cultures as well as in the *A. annua* plant (Putalun et al. 2007;

Lei et al. 2011). In this experiment, *AabHLH1* expression was induced rapidly by 100 μmol l^{-1} ABA after 30 min of treatment. *ADS* expression also increased after treatment with 100 μmol l^{-1} ABA for 1 h. In chitosan-treated leaves, the transcript levels of *AabHLH1* and *ADS* showed similar kinetics to those of ABA induction (Fig. 3). For *CYP71AV1* expression, the transcript level increased in ABA-treated plants after 8 h treatment (Jing et al. 2009). In the *A. annua* plant with foliar application of chitosan, transcript levels increased after 4 h treatment (Lei et al. 2011). Thus, as the transcript level of *AabHLH1* increased rapidly and transiently upon elicitation, it could be a transcriptional regulator, which was substantiated by the nucleus-targeting experiment, the yeast one-hybrid and the EMSA experiments (Figs. 4, 5 and 6).

Drought can affect artemisinin levels in plants. Pot-grown plants deprived of water for either 38 or 62 h showed increased artemisinin levels in leaves (Marchese et al. 2010). Yang et al. substantiated this response by showing increased *ADS* in shoots when plant roots were allowed to dry for 6 h (Yang et al. 2010). Yu et al. (2008) showed that chilling could affect both artemisinin content and the transcript levels of *ADS* and *CYP71AV1* in *in vitro* cultures. Yang et al. (2010) also showed that marked increases in *ADS* and *CYP71AV1* occurred in response to chilling (Yin et al. 2008; Yang et al. 2010; Nguyen et al. 2011). Overall, the above results indicate that *ADS* and *CYP71AV1* are induced in response to drought and cold in *Artemisia annua*. Abscisic acid (ABA) is involved in many physiological processes, including seed germination, dormancy, seedling growth and stomatal aperture activity, as well as in the responses to environmental stresses, such as drought, salt and cold (Finkelstein et al. 2002). In our experiment, the expression of both *AabHLH1* and *ADS* was induced by extracellular ABA treatment and overexpression of *AabHLH1* in transiently transformed *A. annua* was able to regulate the transcripts of *ADS* and *CYP71AV1* (Fig. 3 and Fig. 7). Two bHLH type transcription factors: AtAIB and AtMYC2, have been reported to be induced by ABA and acted as transcriptional activators in the ABA signaling pathway (Abe et al. 2003; Li et al. 2007). Hence it is likely that AabHLH1 acts as a transcriptional activator by binding to the E-boxes of *ADS* and *CYP71AV1* promoters in an ABA-dependent manner.

In recent years, a great number of studies have highlighted the importance of the transcriptional

regulation of target genes through transcription factors in secondary metabolites found in plant cells. Transcription factors have been isolated and characterized for several plant secondary metabolic pathways, including those leading to the biosynthesis of flavonoids, terpenoid indole alkaloids, benzylisoquinoline alkaloids and sesquiterpenes (Borevitz et al. 2000; van der Fits and Memelink 2000; Xu et al. 2004; Kato et al. 2007; Ma et al. 2009). Transcription factors can often activate several genes in the same pathway, e.g. overexpression of the MYB transcription factor PAP1 from *Arabidopsis* resulted in strongly enhanced expression of phenylpropanoid biosynthesis genes in anthocyanin biosynthesis (Borevitz et al. 2000). In the plant species, *Catharanthus roseus*, a jasmonate-responsive transcriptional regulator of both primary and secondary metabolism, ORCA3, was shown to regulate the jasmonate-responsive activation of several terpenoid indole alkaloid biosynthesis genes (van der Fits and Memelink 2000). A novel basic helix-loop-helix (bHLH) protein, CjbHLH1, was reported to positively regulate isoquinoline alkaloids biosynthesis in IQA-producing *Coptis japonica* (Yamada et al. 2011). In the present study, AabHLH1 can positively regulate the transcription of *ADS* and *CYP71AV1* by binding to the E-boxes in the promoters and transient expression of the AabHLH1 transcription factor in leaves of *A. annua* resulted in increased levels of *HMGR*, *ADS* and *CYP71AV1*. Very recently, two JA-responsive AP2/ERF subfamily transcriptional factors (AaERF1 and AaERF2) were reported to positively regulate artemisinin biosynthesis by binding to CBF2 and RAA motifs present in both the *ADS* and *CYP71AV1* promoters in *A. annua* (Yu et al. 2012). In *Catharanthus roseus*, the AP2/ERF-domain transcription factor, ORCA3, controls the jasmonate-responsive expression of several genes encoding enzymes involved in terpenoid indole alkaloid biosynthesis. A basic helix-loop-helix (bHLH) transcription factor, CrMYC2, acts as the major activator of MeJA-responsive ORCA3 gene expression. CrMYC2 binds to the qualitative sequence in the ORCA3 JRE *in vitro*. These results show that MeJA-responsive expression of alkaloid biosynthesis genes in *C. roseus* is controlled by a transcription factor cascade consisting of the bHLH protein, CrMYC2, which regulates the AP2/ERF-domain transcription factor ORCA gene expression, and the AP2/ERF-domain transcription factor, ORCA3. It should be interesting to investigate in the future whether there is a transcription factor cascade consisting of AabHLHs regulating AP2/ERF-domain transcription factors (AaERF1 and AaERF2) in artemisinin biosynthesis in *A. annua*.

The effects of biotic elicitors (e.g. chitosan) involving AaWRKY1 transcription factor on artemisinin biosynthesis in plants have been reported previously (Ma et al. 2009). Foliar application of chitosan on *A. annua* plant resulted in an increase of dihydroartemisinic acid and artemisinin by 72% and 53%, respectively, and semi-quantitative PCR showed increased levels of *ADS*, *CYP71AV1* and *DBR2* after the application of chitosan (Lei et al. 2011). Transient expression of the AaWRKY1 transcription factor in leaves of *A. annua* resulted in increased expression levels of *HMGR*, *ADS*, *CYP71AV1* and *DBR2* (Ma et al. 2009). In this study, AabHLH1 could also be induced by chitosan and transient expression of the AabHLH1 transcription factor in the leaves of *A. annua* could result in increased levels of *HMGR*, *ADS* and *CYP71AV1*. In fact, more than one transcription factor can coordinate and regulate the expression of pathway genes, for example, in maize. The majority of the structural genes encoding enzymes committed to anthocyanin biosynthesis were coordinated and regulated by the bHLH protein-encoding gene, R and the MYB gene, C1, in the aleurone layer of maize kernels (Mol et al. 1998). Since *AaWRKY1*, *AabHLH1*, *AaERF1* and *AaERF2* transcript levels are increased by abiotic stress, such as ABA and JA, or by biotic stress, it is of interest to know whether stress stimulation coordinates transcriptional control of biosynthetic genes. Further elucidation of the interaction among transcription factors will help to elucidate the regulation of artemisinin biosynthesis.

Materials and methods

Plant materials

The high artemisinin-yielding strain, 001, of *Artemisia annua* from Sichuan Province, China (Wang et al. 2007) was used in this study. *A. annua* seedlings were micropropagated as described previously (Zhang et al. 2006). The rooted *in vitro* plantlets were transplanted into pots and grown in a greenhouse with a 16 h/25°C day and an 8 h/20°C night.

AabHLH1 cDNA isolation

As reported previously (Ma et al. 2009), the glandular secretory trichomes were isolated, a cDNA library was constructed and 1985 ESTs were sequenced. The candidate bHLH EST fragments, by

alignment from the NCBI, were subjected to RACE-PCR. For *AabHLH1* cDNA isolation, RACE was undertaken according to the protocol of the 5' RACE System for Rapid Amplification of cDNA Ends (Invitrogen). After first-strand cDNA was synthesized, the first round of PCR was undertaken using a specific GSP1 primer (5'-TCTGGCAATGAGATAGCAGAGGAAGACG-3') and a universal AAP primer (Invitrogen protocol). Then the product of the first PCR reaction was diluted and the second round of PCR was carried out using a specific GSP2 primer (5'-TCCTCTTTATCACATTGGGTGGTGTCG-3') and a universal UAUP primer (Invitrogen protocol).

Subcellular localization of *AabHLH1*

The ORF of AabHLH1 was amplified using forward primer, 5'-GATTCTAGAATGACGGAGTACCGCATGAATC-3' (with the *Xba* I site underlined) and reverse primer, 5'-GTACCCGGGTGGTTCGGCTTCGTATACG-3' (with the *Sma* I site underlined). The PCR product cut by the *Xba* I and *Sma* I enzymes was ligated into pBI221 digested with *Xba* I and *Sma* I in order to generate the CaMV35S::*AabHLH1*-GFP construct, which was verified by sequencing. The CaMV35S::GFP construct was used as a control. Onion inner epidermal cells were stretched out inside-up on MS medium and bombarded with DNA-coated gold particles using the PDS-1000/He system (Bio-Rad, Hercules, CA, USA) at 1100 p.s.i. The transformed cells were cultured on MS medium at 25°C for 24 h and then observed under a confocal microscope (Leica TCS SP2, Germany). Green fluorescent protein (GFP) was excited at 488 nm with an argon laser and fluorescence was detected at 500–540 nm and imaged with a 40.0 x /1.25 HCX PL APO objective lens.

Heterologous expression and purification of *AabHLH1*

The ORF of *AabHLH1* was amplified with forward primer, 5'-GATAAGCTTGCATGACGGAGTACCGCATGAATC-3' (with the *Hin*d III site underlined) and the reverse primer, 5'-GTAGCGGCCGCCTACAGCTTCGCTAATTGCCTTAAC-3' (with the *Not* I site underlined) and then fused to the *Hin*d III-*Not* I sites in the pET-30a (+) vector (Novagen,

Madison, WI, USA). After sequence confirmation, the pET-AabHLH1 plasmid was transformed into *E. coli* strain BL21 (DE3) competent cells for protein expression. Overexpression and purification of the protein in *E. coli* strain BL21 (DE3) was carried out as described previously (Ma et al. 2009). The purification efficiency was monitored by SDS-PAGE. Protein concentration was determined by the Bradford method with bovine serum albumin (BSA) as the standard (Bradford 1976).

Electrophoretic mobility shift assay

The DNA binding ability of AabHLH1 was analyzed by EMSA. Two oligonucleotides were synthesized, namely the sense strand sequence of the E-box (5' -AATGCAAATGTTGGGGAGCACGTGTTGGGGAGCACATGGGGA-3') and the corresponding antisense strand sequence (5' -TCCCCATGTGCTCCCCAACACGTGCTCCCCAACATTTGCATT-3'). The two oligonucleotides were incubated for 5 min at 95°C, and then 5 min at 65°C in order to renature the products. The EMSA was performed using the Electrophoretic Mobility Shift Assay Kit (Invitrogen, USA) following the manufacturer's instructions. The DNA-protein binding reaction was performed by incubating double-stranded oligonucleotides with purified protein at room temperature for 30 min in a total volume of 15 µl. The reaction system contained 20 mmol Tris-HCl (pH 7.6), 30 mmol KCl, 0.2% (w/v) Tween-20, 1 mmol DTT and 10 mmol $(NH_4)_2SO_4$. The binding mixture was resolved on a 6% non-denaturing polyacrylamide gel in 0.5 x TBE buffer. The gel was first stained with SYBR Green EMSA stain so that the DNA could be visualized and photographed. Then the same gel was stained with SYPRO Ruby EMSA stain in order to monitor and photograph the protein.

Binding assay in a yeast one-hybrid system

The yeast reporter plasmid was constructed by inserting annealed complementary oligonucleotides, containing the triple tandem copies of the E-box fragment (annealed to yield *Eco*R I−and *Sac* I−compatible ends), into the pHIS-3 x E-box. The sequence of the 3 x E-box fragment was 3 x E-box F (5' -AATTCAATTGTAATCATTTGTAATCACATGAGCT- 3')

and 3 x E-box R (5' -CATATGATTACATGTGATTACAAATG- 3'). In the same way, the pHIS-3 x m_1E-Box was constructed using the 3 x m_1E-Box F (5' -AATTACATTGTAATACTTTGATTA ACCATGAGCT- 3'), and the 3 x m_1E-Box R (5' -CATAGT ATTACATGGTATTACAAAGT- 3'). Similarly, the 3 x m_2E-Box F (5' -AATTCATAGTTAATCATTGTATTACACAGTAGCT- 3'), and 3 x m_2E-Box R (5' -ACTGGTATTAACAATGATTAACTATG- 3') were used as the primers for the pHIS-3 x m_2E-Box. The ORF of *AabHLH1* was amplified using forward primer: 5' -GATCATATGTATGACGGAGTACCGCATGAATC- 3' (with the *Nde* I site underlined) and reverse primer: 5' -GTACCCGGGCTACAGCTTCGCTAATTGCCTTAAC- 3' (with the *Sma* I site underlined). The PCR product, cut by *Nde* I and *Sma* I enzymes, was ligated into the pGADT7-AD vector in order to generate pGADT7::AabHLH1 using the restriction sites: *Nde* I (forward) and *Sma* I (reverse). Then group 1 (pGADT7::AabHLH1 and the pHIS-3 x E-box), group 2 (pGADT7::AabHLH1 and the pHIS-3 x m_1E-Box), group 3 (pGADT7::AabHLH1 and the pHIS-3 x m_2E-Box), and group 4 (pGADT7-AD and the pHIS-3 x E-box) were co-transformed into yeast Y187 using the LiAc method (Clontech protocol, PT3024-1). Positive clones were screened on SD/-Trp/-Leu medium and then the positive strains were grown on SD/-Trp/-His/-Leu selective medium containing different concentrations of 3-AT at 30°C for 3 d.

The transient transformation of *A. annua*

For the reporter construct, the ADS promoter (Kim et al. 2008) was amplified from *A. annua* 001 genome by PCR with the primer pairs: ADSpro F: 5' -CGTACTGCAGGCATAAGAACATACAAAGCA-3' (with the Kpn I site underlined) and ADSpro R: 5'-GCTACCATGGGATTTTCAAAACTTTGAATATATG-3' (with the Nco I site underlined). The 2.4 kb long ADS promoter was inserted into the pCAMBIA1301 vector, generating pCAMBIA-ADSpro::GUS. In the same way, the 2 kb long *CYP71AV1* promoter was amplified according to Wang et al. (2011b), with the primer pairs: CYP71AV1pro F: 5'-CGTACTGCAGTAGGCTAGGCACATGGTGATC-3' (with the *Kpn* I site underlined) and CYP71AV1pro R: 5'-GCTACCATGGTGCTTTCAGTATACTCTTTATGGTC-3' (with the *Nco* I site underlined), and then inserted into the pCAMBIA1301 vector, generating pCAMBIA-*CYP71AV1pro::GUS*. Double mutated E_{ab}*CYP71AV1pro* was carried out in the same way

as Ma et al. (2009) by using E_a-box F (5'-ACCCATGCATGCATGACGTCTGCTTATTTA-3') and E_a-box R (5'-TAAATAAGCAGACGTCATGCATGCATGGGT-3'); followed by E_b-box F (5'-TATAAACAATCTCATCATTCGCGACCATAA-3') and E_b-box R (5'-GTATGTTGTTATATTTGTTAGAGTAGTAA-3'). The resulting 750 bp-long promoter included two mutated sites (Ea: CATCTG into CGTCTG) and (Eb: CATTTG into CATTCG). For the effector construct, the coding region of *AabHLH1* was inserted into the reconstructed pBI121 vector (PCR-based mutagenesis was applied in order to create an *Xho* I site immediately after the GUS TGA site) with *Xba* I and *Sac* I restriction sites under the control of the cauliflower mosaic virus (CaMV) 35S promoter, generating pBI-*AabHLH1*. Then pCAMBIA-*ADSpro::GUS*, pCAMBIA-*CYP71AV1pro::GUS*, pCAMBIA-*35S::GUS*, pBI-*AabHLH1*/pCAMBIA-*ADSpro::GUS*, and pBI-*AabHLH1*/pCAMBIA-*CYP71AV1pro::GUS*, pCAMBIA-E_{ab}*CYP71AV1pro::GUS*, pBI-*AabHLH1*/pCAMBIA-E_{ab}*CYP71AV1pro::GUS* were introduced into the *Agrobacterium tumefaciens* EHA105 strain respectively as described by Daley et al. (1998). After this, the above constructs were introduced into *A. annua* leaves by the transformation method described in (Ma et al. 2009) with a modification to the co-cultivation at 25°C for 36 h.

GUS activity assay

For the GUS expression measurements in *A. annua*, 4 independent experiments were carried out by vacuum-assisted transformation as described before (Ma et al, 2009). In each treatment, three leaves were collected for GUS measurement. GUS activity was determined by an F-4500 Fluorescence Spectrophotometer (Hitachi, Tokyo, Japan) using 4-umbelliferyl- D-glucuronide as a substrate and 4-methyl-umbelliferone for the fluorometer calibration. Protein content was determined according to the Bradford protein assay using BSA as the standard (Bradford 1976). Relative GUS activity was determined using *CYP71AV1pro::GUS* as a control. The measurement data were expressed as mean± SD. Multiple comparison tests (Least Significant Differences, LSD) were done using IBM SPSS Statistics software for Windows, Version 22. A P-value of <0.05 was considered statistically significant and represented by different letters.

Gene expression analysis by semi-quantitative RT-PCR

For tissue expression analysis, total RNA was isolated from the roots, leaves, stems and flowers of *A. annua* with an RNeasy Mini Kit (Qiagen). For the elicitor treatments, 4-week-old plants were sprayed with either 100 μmol l^{-1} ABA or 150 mg l^{-1} chitosan. In order to investigate whether AabHLH1 can activate the genes involved in the artemisinin biosynthetic pathway, pBI-AabHLH1 or the empty vector, pBI121 (control), were transiently transformed into *A. annua.* Following the transformation, *A. annua* leaves were washed twice in aseptic distilled water and total RNA was isolated from explants using a RNeasy Mini Kit (Qiagen). The conserved actin gene was used as an internal control. RT-PCR was undertaken as described previously (Ma et al. 2009). The gene specific primers used in the RT-PCR analyses are listed in Table 1. The PCR products were electrophoresed on a 1% agarose gel and visualized after ethidium bromide staining. Pictures of gels were taken using a GelDoc EQ imager (Bio-Rad).

Funding

This work was supported by the National Natural Science Foundation of China (Nos. 61173098, 61379081) and the Knowledge Innovation Program of the Chinese Academy of Sciences (KSCX2-EW-J-29).

Acknowledgements

The authors would like to thank Professor Shihua Shen (Institute of Botany, the Chinese Academy of Sciences) for his help with the yeast one-hybrid assay.

References

Abe, H., Urao, T., Ito, T., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2003). *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *Plant Cell* 15: 63-78.

Abe, H., Yamaguchi-Shinozaki, K., Urao, T., Iwasaki, T., Hosokawa, D. and Shinozaki, K. (1997). Role of arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. *Plant Cell* 9: 1859-1868.

Argolo, A. C., Charlwood, B. V. and Pletsch, M. (2000). The regulation of solasodine production by *Agrobacterium* rhizogenes-transformed roots of Solanum aviculare. *Planta Med* 66: 448-451.

Atchley, W., Wollenberg, K., Fitch, W., Terhalle , W. and Dress, A. (2000). Correlations among amino acid sites in bHLH protein domains: An information theoretic analysis. *Molecular Biology and Evolution* 17: 164-178.

Bertea, C. M., Freije, J. R., van der Woude, H., Verstappen, F. W., Perk, L., Marquez, V., De Kraker, J. W., Posthumus, M. A., Jansen, B. J., de Groot, A., Franssen, M. C. and Bouwmeester, H. J. (2005). Identification of intermediates and enzymes involved in the early steps of artemisinin biosynthesis in *Artemisia annua*. *Planta Med* 71: 40-47.

Borevitz , J. O., Xia , Y., Blount , J., Dixon , R. A. and Lamb , C. (2000). Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell* 12: 2383 – 2394.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72: 248-254.

Brodelius, M., Lundgren, A., Mercke, P. and Brodelius, P. E. (2002). Fusion of farnesyldiphosphate synthase and epi-aristolochene synthase, a sesquiterpene cyclase involved in capsidiol biosynthesis in *Nicotiana tabacum*. *Eur J Biochem* 269: 3570-3577.

Castelain, M., Le Hir, R. and Bellini, C. (2012). The non-DNA-binding bHLH transcription factor PRE3/bHLH135/ATBS1/TMO7 is involved in the regulation of light signaling pathway in *Arabidopsis*. *Physiol Plant* 145: 450-460.

Chang, Y. J., Song, S. H., Park, S. H. and Kim, S. U. (2000). Amorpha-4,11-diene synthase of *Artemisia annua*: cDNA isolation and bacterial expression of a terpene synthase involved in artemisinin biosynthesis. *Arch Biochem Biophys* 383: 178-184.

Chinnusamy, V., Ohta, M., Kanrar, S., Lee, B. H., Hong, X., Agarwal, M. and Zhu, J. K. (2003). ICE1: a regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. *Genes Dev* 17: 1043-1054.

Covello, P. S., Teoh, K. H., Polichuk, D. R., Reed, D. W. and Nowak, G. (2007). Functional genomics and the biosynthesis of artemisinin. *Phytochemistry* 68: 1864-1871.

Daley , M., Knauf, V. C., Summerfelt, K. R. and Turner, J. C. (1998). Co-transformation with one *Agrobacterium tumefaciens* strain containing two binary plasmids as a method for producing marker-free transgenic plants. *Plant Cell Rep* 17: 489-496.

Duke, M. V., Paul, R. N., Elsohly, H. N., Sturtz, G. and Duke, S. O. (1994). Localization of artemisinin and artemisitene in foliar tissues of glanded and glandless biotypes of *Artemisia annua*. *Int. J. Plant Sci.* 155: 365-372.

Finkelstein, R. R., Gampala, S. S. and Rock, C. D. (2002). Abscisic acid signaling in seeds and seedlings. *Plant Cell* 14 Suppl: S15-45.

Friedrichsen, D. M., Nemhauser, J., Muramitsu, T., Maloof, J. N., Alonso, J., Ecker, J. R., Furuya, M. and Chory, J. (2002). Three redundant brassinosteroid early response genes encode putative bHLH

transcription factors required for normal growth. *Genetics* 162: 1445-1456.

Graham, I. A., Besser, K., Blumer, S., Branigan, C. A., Czechowski, T., Elias, L., Guterman, I., Harvey, D., Isaac, P. G., Khan, A. M., Larson, T. R., Li, Y., Pawson, T., Penfield, T., Rae, A. M., Rathbone, D. A., Reid, S., Ross, J., Smallwood, M. F., Segura, V., Townsend, T., Vyas, D., Winzer, T. and Bowles, D. (2010). The genetic map of *Artemisia annua* L. identifies loci affecting yield of the antimalarial drug artemisinin. *Science* 327: 328-331.

Han, J. L., B. Liu and Ye, H. C. (2006). Effects of overexpression of the endogenous farnesyl diphosphate synthase on the artemisinin content in *Artemisia annua* L. *J Integrat Plant Biol* 48: 482-487.

Hong, G. J., Xue, X. Y., Mao, Y. B., Wang, L. J. and Chen, X. Y. (2012). *Arabidopsis* MYC2 interacts with DELLA proteins in regulating sesquiterpene synthase gene expression. *Plant Cell* 24: 2635-2648.

Jing , F., Zhang, L., Li , M., Tang, Y., Wang , Y., Wang , Y., Wang, Q., Pan , Q., Wang , G. and Tang , K. (2009). Abscisic acid (ABA) treatment increases artemisinin content in *Artemisia annua* by enhancing the expression of genes in artemisinin biosynthetic pathway. *Biologia* 64: 319-323.

Kato, N., Dubouzet, E., Kokabu, Y., Yoshida, S., Taniguchi, Y., Dubouzet, J. G., Yazaki, K. and Sato, F. (2007). Identification of a WRKY protein as a transcriptional regulator of benzylisoquinoline alkaloid biosynthesis in *Coptis japonica*. *Plant Cell Physiol* 48: 8-18.

Kim, J. and Kim, H. Y. (2006). Functional analysis of a calcium-binding transcription factor involved in plant salt stress signaling. *FEBS Lett* 580: 5251-5256.

Kim, S. H., Chang, Y. J. and Kim, S. U. (2008). Tissue specificity and developmental pattern of amorpha-4,11-diene synthase (ADS) proved by ADS promoter-driven GUS expression in the heterologous plant, *Arabidopsis thaliana*. *Planta Med* 74: 188-193.

Lei, C., Ma, D., Pu, G., Qiu, X., Du, Z., Wang, H., Li, G., Ye, H. and Liu, B. (2011). Foliar application of chitosan activates artemisinin biosynthesis in *Artemisia annua* L. *Ind Crop Prod* 33: 176–182.

Li, H., Sun, J., Xu, Y., Jiang, H., Wu, X. and Li, C. (2007). The bHLH-type transcription factor AtAIB positively regulates ABA response in *Arabidopsis*. *Plant Mol Biol* 65: 655-665.

Liu, S., Tian, N., Li, J., Huang, J. and Liu, Z. (2009). Isolation and identification of novel genes involved in artemisinin production from flowers of *Artemisia annua* using suppression subtractive hybridization and metabolite analysis. *Planta Med* 75: 1542-1547.

Long, T. A., Tsukagoshi, H., Busch, W., Lahner, B., Salt, D. E. and Benfey, P. N. (2010). The bHLH transcription factor POPEYE regulates response to iron deficiency in *Arabidopsis* roots. *Plant Cell* 22: 2219-2236.

Ma, D., Pu, G., Lei, C., Ma, L., Wang, H., Guo, Y., Chen, J., Du, Z., Li, G., Ye, H. and Liu, B. (2009). Isolation and characterization of AaWRKY1, an *Artemisia annua* transcription factor that regulates the amorpha-4,11-diene synthase gene, a key gene of artemisinin biosynthesis. *Plant Cell Physiol* 50: 2146-2161.

Marchese, J., Ferreira, J., Rehder, V. and Rodrigues, O. (2010). Water deficit on the accumulation of biomass and artemisinin in annual wormwood (*Artemisia annua* L., Asteraceae). *Congr. Braz. Soc. Plant Pathol* 22: 1-9.

Mercke, P., Bengtsson, M., Bouwmeester, H. J., Posthumus, M. A. and Brodelius, P. E. (2000). Molecular cloning, expression, and characterization of amorpha-4,11-diene synthase, a key enzyme of artemisinin biosynthesis in *Artemisia annua* L. *Arch Biochem Biophys* 381: 173-180.

Mol , J., Grotewold , E. and Koes , R. (1998). How genes paint flowers and seeds. *Trends Plant Sci* 3 212 -217.

Mutabingwa, T. K. (2005). Artemisinin-based combination therapies (ACTs): best hope for malaria treatment but inaccessible to the needy! *Acta Trop* 95: 305-315.

Nakai, K. and Kanehisa, M. (1992). A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics* 14: 897-911.

Nguyen, K. T., Arsenault, P. R. and Weathers, P. J. (2011). Trichomes + roots + ROS = artemisinin: regulating artemisinin biosynthesis in *Artemisia annua* L. *In Vitro Cell Dev Biol Plant* 47: 329-338.

Olsson, M. E., Olofsson, L. M., Lindahl, A. L., Lundgren, A., Brodelius, M. and Brodelius, P. E. (2009). Localization of enzymes of artemisinin biosynthesis to the apical cells of glandular secretory trichomes of *Artemisia annua* L. *Phytochemistry* 70: 1123-1128.

Osorio, M. B., Bucker-Neto, L., Castilhos, G., Turchetto-Zolet, A. C., Wiebke-Strohm, B., Bodanese-Zanettini, M. H. and Margis-Pinheiro, M. (2012). Identification and in silico characterization of soybean trihelix-GT and bHLH transcription factors involved in stress responses. *Genet Mol Biol* 35: 233-246.

Pires, N. and Dolan, L. (2010). Origin and diversification of basic-helix-loop-helix proteins in plants. *Mol Biol Evol* 27: 862-874.

Putalun, W., Luealon, W., De-Eknamkul, W., Tanaka, H. and Shoyama, Y. (2007). Improvement of artemisinin production by chitosan in hairy root cultures of *Artemisia annua* L. *Biotechnol Lett* 29: 1143-1146.

Ro, D. K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., Chang, M. C., Withers, S. T., Shiba, Y., Sarpong, R. and Keasling, J. D. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 440: 940-943.

Schlereth, A., Moller, B., Liu, W., Kientz, M., Flipse, J., Rademacher, E. H., Schmid, M., Jurgens, G. and Weijers, D. (2010). MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. *Nature* 464: 913-916.

Teoh, K. H., Polichuk, D. R., Reed, D. W. and Covello, P. S. (2009). Molecular cloning of an aldehyde dehydrogenase implicated in artemisinin biosynthesis in *Artemisia annua*. *Botany* 87: 635-642.

Teoh, K. H., Polichuk, D. R., Reed, D. W., Nowak, G. and Covello, P. S. (2006). *Artemisia annua* L. (Asteraceae) trichome-specific cDNAs reveal CYP71AV1, a cytochrome P450 with a key role in the biosynthesis of the antimalarial sesquiterpene lactone artemisinin. *FEBS Lett* 580: 1411-1416.

Tissier, A. (2012). Trichome Specific Expression: Promoters and Their Applications, Transgenic Plants -Advances and Limitations. *PhD. Yelda Ozden Çiftçi (Ed.)* ISBN: 978-953-51-0181-9, InTech.

Toledo-Ortiz, G., Huq, E. and Quail, P. H. (2003). The *Arabidopsis* basic/helix-loop-helix transcription factor family. *Plant Cell* 15: 1749-1770.

Towler, M. J. and Weathers, P. J. (2007). Evidence of artemisinin production from IPP stemming from both the mevalonate and the nonmevalonate pathways. *Plant Cell Rep* 26: 2129-2136.

van der Fits, L. and Memelink, J. (2000). ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism. *Science* 289: 295-297.

Wallaart, T. E., Bouwmeester, H. J., Hille, J., Poppinga, L. and Maijers, N., C. (2001). Amorpha-4,11-diene synthase: cloning and functional expression of a key enzyme in the biosynthetic pathway of the novel antimalarial drug artemisinin. *Planta* 212: 460-465.

Wang, H., Han, J., Kanagarajan, S., Lundgren, A. and Brodelius, P. E. (2013). Trichome-specific expression of the amorpha-4,11-diene 12-hydroxylase (cyp71av1) gene, encoding a key enzyme of artemisinin biosynthesis in *Artemisia annua*, as reported by a promoter-GUS fusion. *Plant Mol Biol* 81:

119-138

Wang, H., Liu, Y., Chong, K., Liu, B. Y., Ye, H. C., Li, Z. Q., Yan, F. and Li, G. F. (2007). Earlier flowering induced by over-expression of *CO* gene does not accompany increase of artemisinin biosynthesis in *Artemisia annua*. *Plant Biol (Stuttg)* 9: 442-446.

Wang, H., Olofsson, L., Lundgren, A. and Brodelius, P. (2011a). Trichome-specificexpression of amorpha-4,11-diene synthase, a key enzyme of artemisinin biosynthesis in *Artemisia annua*, as reported by a promoter-GUS fusion. *Am J Plant Sci* 2: 619–628.

Wang, W., Wang, Y., Zhang, Q., Qi, Y. and Guo, D. (2009). Global characterization of *Artemisia annua* glandular trichome transcriptome using 454 pyrosequencing. *BMC Genomics* 10: 465.

Wang, Y., Yang, K., Jing, F., Li, M., Deng, T., Huang, R., Wang, B., Wang, G., Sun, X. and Tang, K. X. (2011b). Cloning and characterization of trichome-specific promoter of cpr71av1 gene involved in artemisinin biosynthesis in *Artemisia annua* L. *Mol Biol* 45: 751-758.

Xu, Y. H., Wang, J. W., Wang, S., Wang, J. Y. and Chen, X. Y. (2004). Characterization of GaWRKY1, a cotton transcription factor that regulates the sesquiterpene synthase gene (+)-delta-cadinene synthase-A. *Plant Physiol* 135: 507-515.

Yamada, Y., Kokabu, Y., Chaki, K., Yoshimoto, T., Ohgaki, M., Yoshida, S., Kato, N., Koyama, T. and Sato, F. (2011). Isoquinoline alkaloid biosynthesis is regulated by a unique bHLH-type transcription factor in *Coptis japonica*. *Plant Cell Physiol* 52: 1131-1141.

Yang, R. Y., Zeng, X. M., Lu, Y. Y., Lu, W. J., Feng, L. L., Yang, X. Q. and Zeng, Q. P. (2010). Senescent leaves of *Artemisia annua* are one of the most active organs for overexpression of artemisinin biosynthesis responsible genes upon burst of singlet oxygen. *Planta Med* 76: 734-742.

Yi, K., Wu, Z., Zhou, J., Du, L., Guo, L., Wu, Y. and Wu, P. (2005). OsPTF1, a novel transcription factor involved in tolerance to phosphate starvation in rice. *Plant Physiol* 138: 2087-2096.

Yin, L., Zhao, C., Huang, Y., Yang, R. and Zeng, Q. (2008). Abiotic stress-induced expression of artemisinin biosynthesis genes in *Artemisia annua* L. *Chin. J Appl. Environ. Biol.* 14: 1-5.

Yu, Z. X., Li, J. X., Yang, C. Q., Hu, W. L., Wang, L. J. and Chen, X. Y. (2012). The jasmonate-responsive AP2/ERF transcription factors AaERF1 and AaERF2 positively regulate artemisinin biosynthesis in *Artemisia annua* L. *Mol Plant* 5: 353-365.

Zhang, H., Hedhili, S., Montiel, G., Zhang, Y., Chatel, G., Pre, M., Gantet, P. and Memelink, J. (2011). The basic helix-loop-helix transcription factor CrMYC2 controls the jasmonate-responsive expression of the ORCA genes that regulate alkaloid biosynthesis in *Catharanthus roseus*. *Plant J* 67: 61-71.

Zhang , L., Ye , H. C. and Li , G. F. (2006). Effect of development stage on the artemisinin content and the sequence characterized amplified region (SCAR) marker of high-artemisinin yielding strains of *Artemisia annua* L. *J. Integr. Plant Biol.* 48: 1054 – 1062.

Zhang, Y., Teoh, K. H., Reed, D. W., Maes, L., Goossens, A., Olson, D. J., Ross, A. R. and Covello, P. S. (2008). The molecular cloning of artemisinic aldehyde Delta11(13) reductase and its role in glandular trichome-dependent biosynthesis of artemisinin in *Artemisia annua*. *J Biol Chem* 283: 21501-21508.

Figure legends

Fig. 1 Biosynthetic pathway for artemisinin in *Artemisia annua*. Below are enzyme abbreviations. HMGR: 3-hydroxy-3-methyl-glutaryl coenzyme A reductase; FPS: farnesyl diphosphate synthase; ADS: amorpha-4,11-diene synthase; CYP71AV1: cytochrome P450 monooxygenase; DBR2: artemisinic aldehyde Δ11(13) reductase; ALDH1: aldehyde dehydrogenase 1; RED1: dihydroartemisinic aldehyde reductase.

Fig. 2 Alignment of AabHLH1 with related bHLH proteins from *Catharanthus roseus*, *Lotus japonicus*, *Nicotiana benthamiana*, *Nicotiana tabacum*, *Phaseolus vulgaris* and *Solanum tuberosum*. The deduced amino acid sequence of AabHLH1 was aligned with CrMYC2 (AAQ14332), LjbHLH22 (ACN21638), NbbHLH1 (ADH04262), NtMYC1a (ADH04267), PvPG1 (AAB00686) and StMYC1 (CAF74710) using DNAMAN software. The amino acid residues shaded with blue, pink, blue-green and those without shadings represent respectively the high, intermediate, low and very low sequence similarity of AabHLH1 with respect to the other species.

Fig. 3 Expression patterns of *AabHLH1*. A, Ethidium bromide-stained agarose gels showing RT-PCR results for roots (R), leaves (L), stems (S) and flowers (F). B and C, *AabHLH1* and *ADS* expression in response to ABA and chitosan, respectively. Leaves were sprayed with ABA (100 μmol l^{-1}) or chitosan (150 mg l^{-1}). Leaves were harvested at the indicated times after treatment for preparation of total RNA. Accumulation of *AabHLH1* and *ADS* transcripts was monitored by semiquantitative RT-PCR. Actin amplification was used as constitutive control.

Fig. 4 Nuclear localization of the AabHLH1 protein. A–C, The control GFP protein. D–F, The AabHLH1–GFP fusion protein. Onion epidermal cells were bombarded with plasmids harboring the GFP coding region 35S::GFP (A–C), or 35S::AabHLH1–GFP fusion construct (D–F). GFP accumulated in the cytosolic and nuclear compartments. AabHLH1–GFP fusion protein accumulated in the nucleus. Detection of fluorescence was performed 24 h after bombardment under a confocal laser scanning microscope (Leica TCS SP2, Germany). In the control group (A-C), the whole epidermal cells of the onion transformed by 35S::GFP showed the green

fluorescence. However the green fluorescence was only observed in the nuclei of the epidermal cells of the onion transformed by the fusion expression vector 35S::AabHLH1–GFP (D-F).

Fig. 5 The binding ability of the AabHLH1 protein to the E-box *cis*-acting elements, as analyzed by EMSA. A, titration of 3 x E-box DNA with AabHLH1 protein stained with SYBR Green EMSA for visualizing DNA; B, the same gel in "A" stained with SYPRO Ruby EMSA for visualizing protein. Increasing amounts of AabHLH1 protein were added to 3 x E-box DNA in a final 15 μl volume. M: DNA maker (DL 2000). Lane 1: *cis*-acting elements DNA only. Lanes 2-5: *cis*-acting elements DNA with increasing amounts of AabHLH1 protein (34 ng, 54 ng, 81 ng and 102 ng). Lane 6: AabHLH1 protein only (102 ng). Lane 7: *cis*-acting elements DNA with 2 μg BSA.

Fig. 6 AabHLH1 interact with E-box in yeast cell. A, Distribution of yeast cells Y187 carrying pGAD::AabHLH1/pHIS-3xE-box, pGAD::AabHLH1/pHIS-3xm_1E-box, pGAD::AabHLH1/ pHIS-$m_2$3xE-box or pHIS3xE-box. B, The transformants containing four constructs are grown for 3 d at 30 ℃ in SD/-Trp/-His/-Leu. C, The transformants only carrying pGAD::AabHLH1/ pHIS-3xE-box were grown in SD/-Trp/-His/-Leu plus 30 mM 3-amino-1,2,4- triazole (3-AT), while others' growth were prohibited due to failing to produce His.

Fig. 7 AabHLH1 activated *ADS*, *CYP71AV1* promoters, as well as the genes involved in artemisinin biosynthesis by *Agrobacterium*-mediated transient expression system in *A. annua* leaves. A, Transient assay of GUS activities was done in *A. annua* leaves 36 h after co-cultivation with *A. tumefaciens* cells harboring (1) *35S::GUS*, (2) *35S::AabHLH1/CYP71AV1pro::GUS*, (3) *35S::AabHLH1/ADSpro::GUS*, (4) *CYP71AV1pro::GU*S, (5) *ADSpro::GUS*. GUS activity was normalized against *CYP71AV1pro::GU*S. B, RT-PCR analysis of genes expression levels in the reconstructed pBI121empty vector (PCR-based mutagenesis was applied to create a *Xho* I site immediately after the GUS TGA site) and pBI::AabHLH1 transiently transformed *A. annua*. The genes transcript levels of the empty vector-transiently transformed *A. annua* were set to CK and compared with the transcript levels in pBI::AabHLH1 transiently transformed *A. annua*.

Fig. 8 AabHLH1 activated *CYP71AV1* promoter by *Agrobacterium*-mediated transient expression system in *A. annua* leaves, and mutated E-boxes abolished the activation of AabHLH1. Transient assay of GUS activities was carried out in *A. annua* leaves 36 h after co-cultivation with *A. tumefaciens* cells harboring (1) *CYP71AV1pro::GUS*; (2) *EabCYP71AV1pro::GUS*; (3) *35S::AabHLH1*/*CYP71AV1pro::GUS*; (4) *35S::AabHLH1/EabCYP71AV1pro::GUS*. GUS activity was normalized against construct (1) *CYP71AV1pro::GU*S. Multiple comparison test was done by SPSS software, a P-value of <0.05 was considered statistically significant shown in letter "a" and "b".

Table 1 Nucleotide sequence of primers used

Name	Primer sequence
actin F	5'-AACTGGGATGACATGGAGAAGATAT-3'
actin R	5'-TCACACTTCATGATGGAGTTGTAGG-3'
AabHLH1 F	5'-ATGGAAAGTGTTTGTGTTTAT-3'
AabHLH1 R	5'-AAATTTGAAATCAAGGTCTAA-3'
ADS F	5'-ATGTCACTTACAGAAGAAAAA-3'
ADS R	5'-TATACTCATAGGATAAACGAG-3'
FPS F	5'-ATGGGTAGCATCGATCTGAAAT-3'
FPS R	5'-TTTGCCTCTTGTAGATTTTACC-3'
HMGR F	5'-ATGGATCTCCGTCGTAAACTGCCA-3'
HMGR R	5'-GAGCGTTTGAGCCTGGTGATTCTA-3'
ALDH1 F	5'-TTAAAGCCACGGGGATAT-3'
ALDH1 R	5'-CACCATGTCTGAAAAACCAACCTTG-3'
DBR2 F	5'-CACCATGTCTGAAAAACCAACCTTG-3'
DBR2 R	5'-GCTCATAAGATGCACCTTAATAAG-3'
CYP71AV1 F	5'-CTAGGATCCATGGCACTCTCACTGACCACT-3'
CYP71AV1 R	5'-TCATATACTCATAGGATAAAC-3'
RED1 F	5'-CGAAATATGTCGTCAGTTGGCCTC-3'
RED1 R	5'-GACCATCATCGGGCAACAAAGC-3'

Table 2 AabHLH1 and AtbHLH proteins predicted to have DNA-binding capacity

bHLH	PID	BASIC		HELIX	LOOP	HELIX	
Number	Number	10	**E R** 20	30	40	50	60
A:							
AtbHLH67	CAB71902	EIENQRINHI	AV**E**RN**R**RRQM	NEHINSLRAL	LPPSYIQ.RG	DQASIVGGAI	NYVKVLEQII
AtbHLH57	CAB77716	EVENQRMTHI	AV**E**RN**R**RRQM	NEHLNSLRSL	MPPSFLQ.RG	DQASIVGGAI	DFIKELEQLL
AtbHLH95	AAG13058	EESPDHEIHI	WT**E**RE**R**RKKM	RDMFSKLHAL	LPQ.LPP.KA	DKSTIVDEAV	SSIKSLEQTL
AtbHLH92	BAB11628	PEKERSRRHM	LK**E**RT**R**REKQ	KQSYLALHSL	LPFA...TKN	DKNSIVEKAV	DEAIAKLQRL
AtbHLH10	AAD20667	GRGSRKSRTS	PT**E**RE**R**RVHF	NDRFFDLKNL	IPNP...TKI	DRASIVGEAI	DYIKELLRTI
AtbHLH89	AAF80214	GRGSKKRKIF	PT**E**RE**R**RVHF	KDRFGDLKNL	IPNP...TKN	DRASIVGEAI	DYIKELLRTI
AtbHLH19	AAC63587	RSPVLAKEHV	LA**E**RK**R**REKL	SEKFIALSAL	LPGL...KKA	DKVTILDDAI	SRMKQLQEQL
AtbHLH20	AAC63588	REPHLLKEHV	LA**E**RK**R**RQKL	NERLIALSAL	LPGL...KKT	DKATVLEDAI	KHLKQLQERV
AabHLH1		NGREEPLNHV	EA**E**RQ**R**REKL	NQRFYALRAV	VPNV...SKM	DKASLLGDAI	LYIKELKSKV
AtbHLH14	AAB62853	KHHPAVLSHV	EA**E**KQ**R**REKL	NHRFYALRAI	VPKV...SRM	DKASLLSDAV	SYIESLKSKI
AtbHLH3	CAB78685	NGREEALNHV	EA**E**RQ**R**REKL	NQRFYALRAV	VPNI...SKM	DKASLLADAI	TYITDMQKKI
B:							
AtbHLH135	AAF15922	RRSRSRQSSG	TS**E**RISEDQI	NDLIIKLQQL	LPRS...DKV	SAARVLQDTC	NYIRNLHREV
AtbHLH26	AAK15282	REVPSVTRKG	SK**R**RR**R**DEKM	SNKMRKLQQL	VPNC..HK.T	DKVSVLDKTI	EYMKNLQLQL
AtbHLH132	BAC10690	KRKRNAEAYN	SP**E**RNQRNDI	NKKMRTLQNL	LPNS..HK.D	DNESMLDEAI	NYMTNLQLQV
AtbHLH145	BAB10287	RISFLKRSKL	SSNKIGEEKI	FETVSLLRSV	VPGE...ELV	DPILVIDRAI	DYLKSLKMEA
AtbHLH108	NM102341	KSSDKSDHDT	LLKKK**R**RERI	RRQLETLKEI	TPNC...PQS	DINAILDCVI	EYTNNLRLAH
C:							
AtbHLH83	AAG27834	PTTSPKDPQS	LAAKN**R**RERI	SERLKILQEL	VPNG..TK.V	DLVTMLEKAI	SYVKFLQVQV
AtbHLH86	BAB10359	ATTSPKDPQS	LAAKN**R**RERI	SERLKVLQEL	VPNG..TK.V	DLVTMLEKAI	GYVKFLQVQV
AtbHLH140	CAB81914	TSTLSTDPQS	VAARD**R**RHRI	SDRFKILQSM	VPGG..AK.M	DTVSMLDEAI	SYVKFLKAQI
AtbHLH142	BAB09865	STKEDTGSGL	SN**E**QSSKDKI	RTALKILESV	VPGA...KGN	EALLLLDEAI	DYLKLLKRDL

A, Putative E-box binders, which have conserved residues in positions E13 and R16 (highlighted in bold). AabHLH1 is marked with red box. B, Non E-box binders, proteins with a basic region, but not predicted to have E-box binding capacity, do not contain both of the amino acids E13 and R16, considered to be required to recognize E-Box. C, Non-DNA-binders, lack capacity to bind DNA, based on the absence of amino acid E13 and/or R16, and the presence of few basic residues in the first 17 positions. Proteins are listed according to their PID Number. Positions within the bHLH domain are numbered at the top of the table.

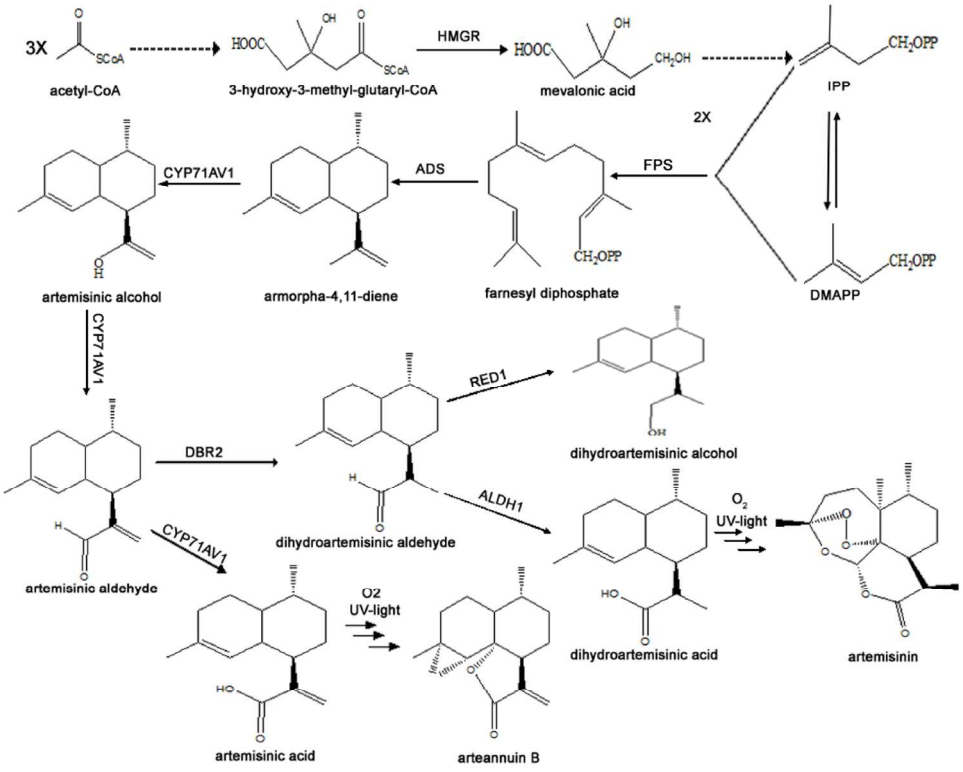


Fig. 1 Biosynthetic pathway for artemisinin in Artemisia annua. Below are enzyme abbreviations. HMGR: 3-hydroxy-3-methyl-glutaryl coenzyme A reductase; FPS: farnesyl diphosphate synthase; ADS: amorpha-4,11-diene synthase; CYP71AV1: cytochrome P450 monooxygenase; DBR2: artemisinic aldehyde Δ11(13) reductase; ALDH1: aldehyde dehydrogenase 1; RED1: dihydroartemisinic aldehyde reductase.
119x96mm (300 x 300 DPI)

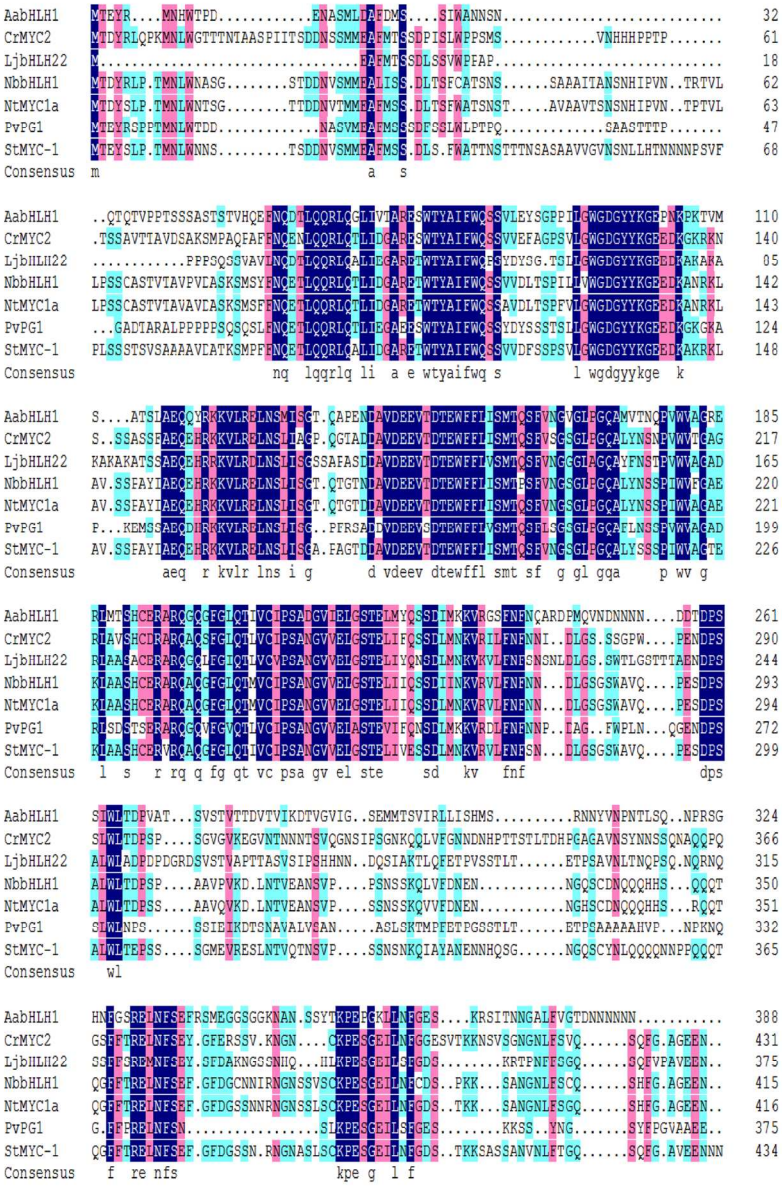


Fig. 2 Alignment of AabHLH1 with related bHLH proteins from Catharanthus roseus, Lotus japonicus, Nicotiana benthamiana, Nicotiana tabacum, Phaseolus vulgaris and Solanum tuberosum. The deduced amino acid sequence of AabHLH1 was aligned with CrMYC2 (AAQ14332), LjbHLH22 (ACN21638), NbbHLH1 (ADH04262), NtMYC1a (ADH04267), PvPG1 (AAB00686) and StMYC1 (CAF74710) using DNAMAN software. The amino acid residues shaded with blue, pink, blue-green and those without shadings represent respectively the high, intermediate, low and very low sequence similarity of AabHLH1 with respect to the other species.
119x180mm (300 x 300 DPI)

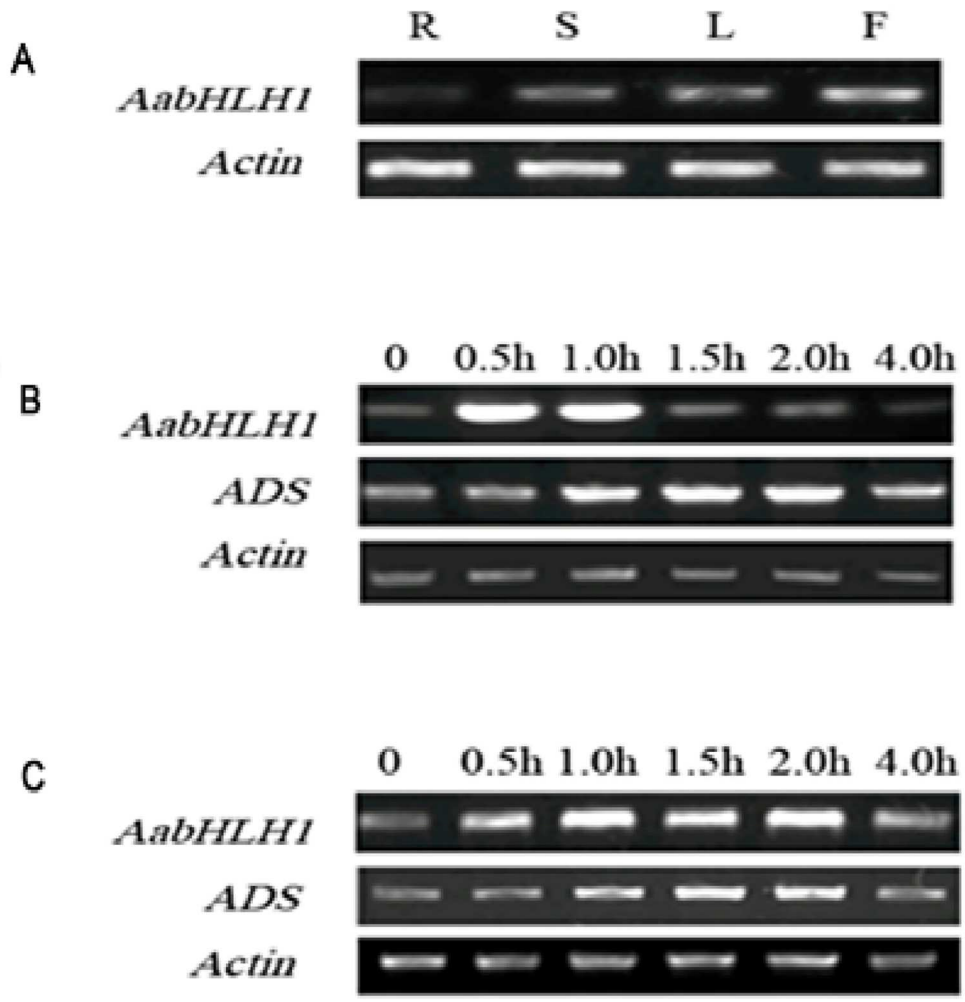


Fig. 3 Expression patterns of AabHLH1. A, Ethidium bromide-stained agarose gels showing RT-PCR results for roots (R), leaves (L), stems (S) and flowers (F). B and C, AabHLH1 and ADS expression in response to ABA and chitosan, respectively. Leaves were sprayed with ABA (100 μmol l-1) or chitosan (150 mg l-1). Leaves were harvested at the indicated times after treatment for preparation of total RNA. Accumulation of AabHLH1 and ADS transcripts was monitored by semiquantitative RT-PCR. Actin amplification was used as constitutive control.
119x125mm (300 x 300 DPI)

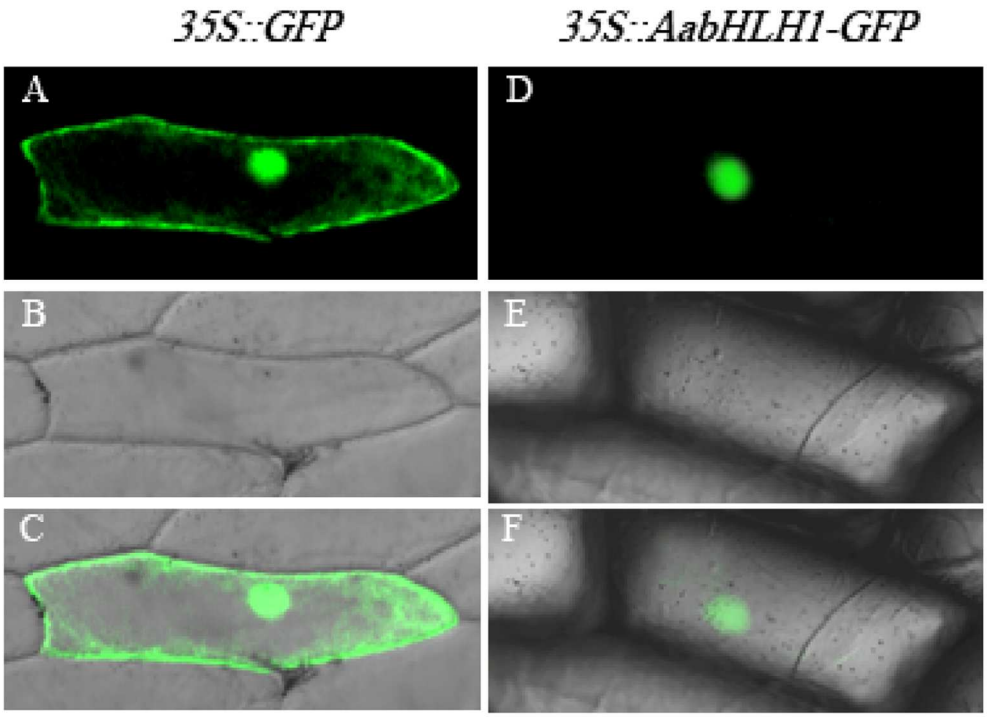


Fig. 4 Nuclear localization of the AabHLH1 protein. A–C, The control GFP protein. D–F, The AabHLH1–GFP fusion protein. Onion epidermal cells were bombarded with plasmids harboring the GFP coding region 35S::GFP (A–C), or 35S::AabHLH1–GFP fusion construct (D–F). GFP accumulated in the cytosolic and nuclear compartments. AabHLH1–GFP fusion protein accumulated in the nucleus. Detection of fluorescence was performed 24 h after bombardment under a confocal laser scanning microscope (Leica TCS SP2, Germany). In the control group (A-C), the whole epidermal cells of the onion transformed by 35S::GFP showed the green fluorescence. However the green fluorescence was only observed in the nuclei of the epidermal cells of the onion transformed by the fusion expression vector 35S::AabHLH1–GFP (D-F).
119x90mm (300 x 300 DPI)

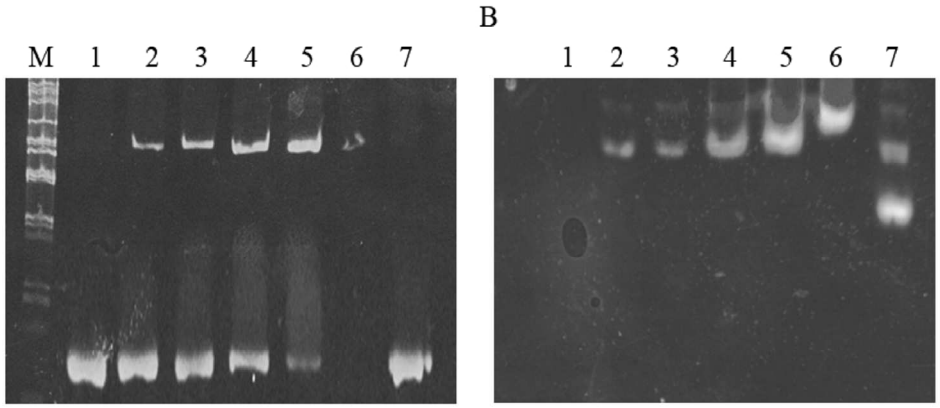


Fig. 5 The binding ability of the AabHLH1 protein to the E-box cis-acting elements, as analyzed by EMSA. A, titration of 3 x E-box DNA with AabHLH1 protein stained with SYBR Green EMSA for visualizing DNA; B, the same gel in "A" stained with SYPRO Ruby EMSA for visualizing protein. Increasing amounts of AabHLH1 protein were added to 3 x E-box DNA in a final 15 µl volume. M: DNA maker (DL 2000). Lane 1: cis-acting elements DNA only. Lanes 2-5: cis-acting elements DNA with increasing amounts of AabHLH1 protein (34 ng, 54 ng, 81 ng and 102 ng). Lane 6: AabHLH1 protein only (102 ng). Lane 7: cis-acting elements DNA with 2 µg BSA.
119x55mm (300 x 300 DPI)

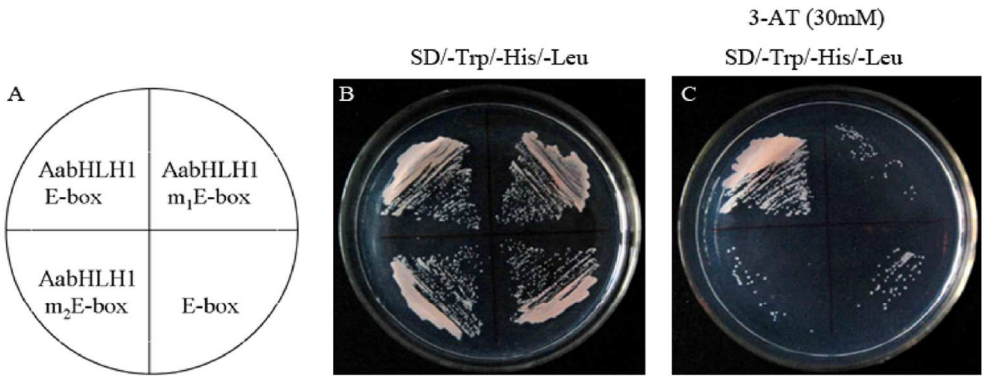
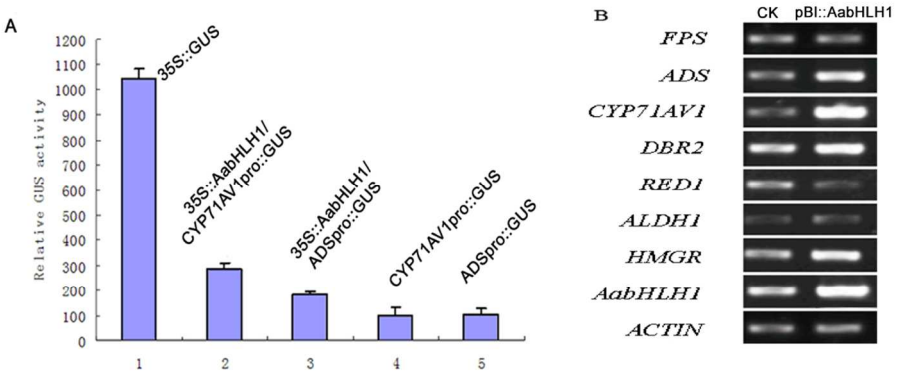


Fig. 6 AabHLH1 interact with E-box in yeast cell. A, Distribution of yeast cells Y187 carrying pGAD::AabHLH1/pHIS-3xE-box, pGAD::AabHLH1/pHIS-3xm1E-box, pGAD::AabHLH1/ pHIS-m23xE-box or pHIS3xE-box. B, The transformants containing four constructs are grown for 3 d at 30 °C in SD/-Trp/-His/-Leu. C, The transformants only carrying pGAD::AabHLH1/ pHIS-3xE-box were grown in SD/-Trp/-His/-Leu plus 30 mM 3-amino-1,2,4- triazole (3-AT), while others' growth were prohibited due to failing to produce His
119x50mm (300 x 300 DPI)



119x48mm (300 x 300 DPI)

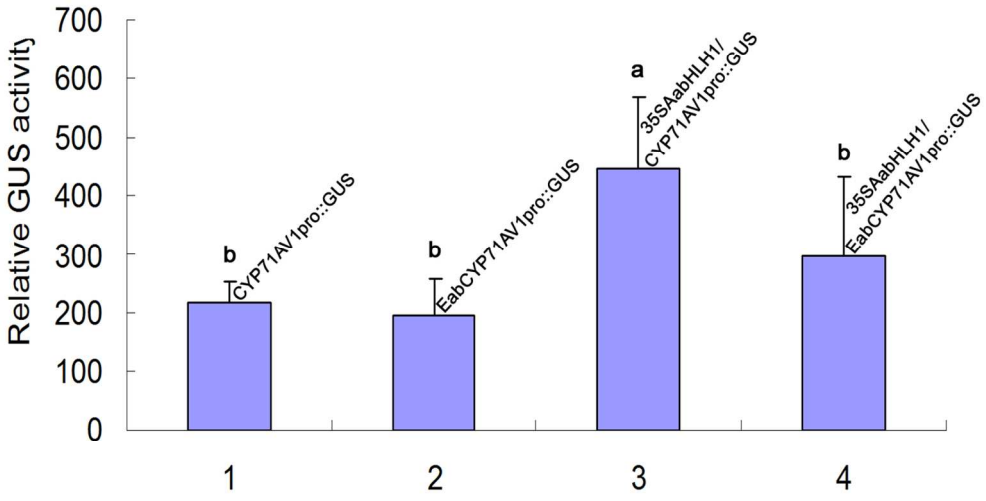


Fig. 8 AabHLH1 activated CYP71AV1 promoter by Agrobacterium-mediated transient expression system in A. annua leaves, and mutated E-boxes abolished the activation of AabHLH1. Transient assay of GUS activities was carried out in A. annua leaves 36 h after co-cultivation with A. tumefaciens cells harboring (1) CYP71AV1pro::GUS; (2) EabCYP71AV1pro::GUS; (3) 35S::AabHLH1/CYP71AV1pro::GUS; (4) 35S::AabHLH1/EabCYP71AV1pro::GUS. GUS activity was normalized against construct (1) CYP71AV1pro::GUS. Multiple comparison tests were done by SPSS software, a P-value of <0.05 was considered statistically significant and shown in letter "a" and "b".
119x69mm (300 x 300 DPI)