

珙桐转录因子 *DiMYB1* 基因的生物信息学分析

戴鹏辉, 任 锐, 董旭杰, 李 萌, 曹福祥

(中南林业科技大学 生命科学与技术学院, 湖南 长沙 410004)

摘 要:利用生物信息学方法对前期通过转录组测序获得的珙桐 *DiMYB1* 基因编码蛋白序列进行分析,为后续 *DiMYB1* 基因的功能鉴定和调控研究奠定基础。结果表明:*DiMYB1* 蛋白全长 924 bp,编码 307 个氨基酸,相对分子量为 35 kDa,为亲水性蛋白。*DiMYB1* 属于 R2R3-MYB 蛋白家族,含有 1 个信号肽结构和 2 个功能结构域,亚细胞定位预测该基因位于细胞核的可能性是 91.3%。系统进化树分析表明,其与柿树 MYB2 转录因子的同源性最高,相似度为 84%。

关键词:*DiMYB1*;珙桐;生物信息学

中图分类号:S 682.36 **文献标识码:**A **文章编号:**1001-0009(2017)02-0112-05

珙桐 (*Davidia involucrata* Baill.) 是第三纪古热带孑遗物种,属于国家 I 级重点保护濒危物种^[1]。我国从 20 世纪 70 年代开始进行珙桐研究,现有的珙桐研究报道主要集中在珙桐群落、引种栽培、组培技术、次生代谢等方面,而对珙桐 MYB 转录因子的相关分子生物学方面研究极少。MYB 转录因子是参与花青素生物合成的关键转录因子之一^[2]。MYB 蛋白分为 4 类,即 R1-MYB、R2R3-MYB、R3-MYB 和 R4-MYB,其中 R2R3-MYB 主要参与类黄酮代谢^[3]。目前有很多 MYB 转录因子在其它物种中被克隆,如柑橘 *MYB15* 基因^[4]、亚洲棉 *MYB* 基因^[5]、葡萄风信子 *MYB* 基因^[6]、中国水仙 *MYB* 基因^[7]、枸杞 *LbMYB1* 基因^[8] 和柿树 *MYB4* 基因^[9]。现利用生物信息学分析方法对珙桐 *DiMYB1* 基因编码的氨基酸的理化性质、保守功能域、磷酸化位点、信号肽结构、二级结构组成等进行了分析,以期对珙桐 *DiMYB1* 基因的功能验证和调控机制提供参考依据。

第一作者简介:戴鹏辉(1991-),男,硕士研究生,研究方向为蛋白质化学与蛋白质组学。E-mail:daipenghui123@qq.com。
责任作者:曹福祥(1962-),男,博士,教授,现主要从事生物化学与分子生物学及植物学等研究工作。E-mail:csfucao@163.com。

基金项目:中南林业科技大学博士后科研基金资助项目(0490016);中南林业科技大学青年基金资助项目(QJ201512)。

收稿日期:2016-09-26

1 材料与方法

1.1 试验材料

在珙桐果实及种子转录组数据库中获得 *DiMYB1* 基因序列,基因克隆后登录号为 KR996175^[10]。使用 NCBI 中的 ORF Finder 工具找出该基因的 ORF 序列并翻译成氨基酸。

1.2 试验方法

采用生物信息学分析方法分析氨基酸序列。通过一系列在线软件,如 ExPASy-ProtParam、SMART、SignalP 4.1、MEGA 6.06 等对该蛋白质的基本理化性质、一级结构、二级结构、三级结构、保守功能域、磷酸化位点、跨膜结构域、信号肽结构、系统发育树等进行了生物信息学分析。其中系统发育树采用 NCBI 数据库中的 BLAST 将 *DiMYB1* 编码的氨基酸序列与 GenBank 中其它物种的氨基酸序列进行序列比对,使用 MEGA 6.06 软件中的邻位归并法(Neighbor Joining 法)进行了聚类分析。

2 结果与分析

2.1 *DiMYB1* 蛋白一级结构的预测分析

ExPASy-ProtParam 在线分析工具表明,*DiMYB1* 的相对分子量大小为 35 kDa,编码 307 个氨基酸,等电点(pI)为 7.02,脂肪系数(aliphatic index)为 84.50,不稳定系数(instability index)为 43.78,由参数值<40 为稳定蛋白的标准可知,*DiMYB1* 属不稳定蛋白。由图 1 可知,含量最多的氨基酸是亮氨酸(Leu),其次为天冬酰胺(Asn)和丝氨酸(Ser)。推导氨基酸序列见图 2。

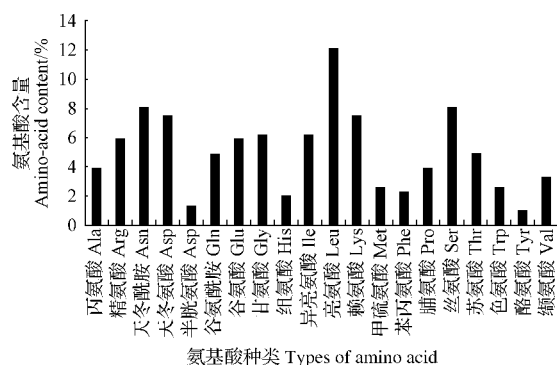


图1 氨基酸含量分布

Fig. 1 Distribution of amino acid content

2.2 DiMYB1 蛋白二、三级结构及亚细胞定位的预测

采用在线 SOPMA 对珙桐 DiMYB1 蛋白的二级结构进行了预测,由图 3 可知,DiMYB1 蛋白的二级结构由 α -螺旋、 β -折叠、 β -转角、无规则卷曲组成,其中 α -螺旋和无规则卷曲为主要构件,其次为 β -折叠和 β -转角。在线利用 ExPASy 中的 SWISS-MODEL 同源建模预测了 DiMYB1 的 3 种三级结构模型(图 4)。利用 Psort 对 DiMYB1 蛋白进行亚细胞定位,由表 1 可知,DiMYB1 蛋白 91.3%位于细胞核,位于线粒体中的可能性是 8.7%。预测了珙桐 DiMYB1 蛋白的磷酸化位点(Netphos 2.0 server),由图 5 可知,珙桐 DiMYB1 蛋白含有 11 个丝氨酸位点,3 个苏氨酸位点,1 个酪氨酸位点,在线 SignalP 4.1 Server 预测 DiMYB1 有 1 条信号肽(图 6)。

```

28  atggggaaggaaacacatgcclgccaaglgggacatgaaaaggga
M G R K P C C A K V G L K R G
73  gcatggactgctctggaagacaaattcttactagttacattaaa
A W T A L E D K I L T S Y I K
118  acccatggagaaggcaaalggagagaccltcccaaaagagctggg
T H G E G K W R D L P Q R A G
163  ctgaagcgaatgtgggaagagttgtagactccgatgggtgaactat
L K R C G K S C R L R W L N Y
208  cttagaccagatallaaaaggaglaacalltclgaggagaagaa
L R P D I K R G N I S V E E E
253  gacatcattatgaaacttcataagctacttggtaaatagggtgtcc
E L I M K L H K L L G N R W S
298  ctatcgcclggaagallacccggggcgaacagacaaatgaaatcaa
L I A G R L P G R T D N E I K
343  aaccactggaacaccaatttgagcaagagatcaatgcagggtcag
V H W N T N L S K R S M Q G Q
388  aagagtacitlaaalctgccaaacaaatcaatlaaactgaaagg
K S T L N S A K Q L N K L K G
433  aaggaatctgaaggaaattgttccctccatggcggttaattccagg
K E S E G N L S S M A L I P G
478  aatccaaatccacacagggtgagacgaacaggcclgtagagatc
N P N P H Q V I R T K A V R I
523  acaaaggttatcttccacacaattagataatcaattgtgtaag
T K V I F P Q Q L D N Q L V K
568  agaatgaagliccactlgggacclccgcgaacacacccatcc
R N E V P T W D L R D N N P S
613  tegtcaacctacttgaggatgatatacctctcttttgatggat
S S T L L E D D T S F L M D
658  tcaatallagtagactallatcagatgltcltaactcacaa
F N I S E L F I S D V L N S Q
703  actctccagatcaaaaggagatcagtaaatgataatggg
T L Q D Q K E I S N E I D N G
748  gllggtagagallacagatccatgacatgacatgacatgacat
V G G D Y S T M N M H S D F L
793  gtccctacatttgagccatgctgaacattcaaccgacgatga
V P T F E A M L N N S T D D A
838  atcatalagaaatlgaggagctagcattgatcctcttcaacaa
I I E N W R A S I D H P L Q Q
883  aatgaagatttggctccagaacacgcttgcacatcttttctaa
N E D L L L Q K Q L A S F L N
928  tcaggggagatgagatgagatgagatgagatgagatgagatgag
S G D E W I D *

```

图2 推导出的 DiMYB1 蛋白的氨基酸序列

Fig. 2 Deduced amino acid sequence of DiMYB1

酸位点,1 个酪氨酸位点,在线 SignalP 4.1 Server 预测 DiMYB1 有 1 条信号肽(图 6)。

```

          10          20          30          40          50          60          70
MGRKPCCAKVGLKRGAWTALEDKILTSYIKTHGEGKWRDLPPQAGLKRCKGKSCRLRWLNLRPDIKRGN
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SVVEEELIMKLHLKLGNRWSLIAGRLPGRTDNEIKHWNTNLSKRSMQGGKSTLNSAKQLNKLKGKESG
ehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
NLSMMLIPGNPNPQVIRTKAVRITKVIFPQQLDNQLVKRNEVPTWDLRDNDPSSSTLEDDDTSLMD
chhhheeeettcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
FVISELFISDVLSQTLQDQKEISNEIDNGVGGDYSTMHSDFLVPTFEAMLANSTDDAIIENWRASID
cehhheehhhheeecccccccccccccccccccccccccccccccccccccccccccccccccccccccc
HPLQQNEDLLKQLASFLNSGDEWID
cccccthhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

Sequence length:307

SOPMA:

Alpha helix	(Hh):	125is	40.72%
3 ₁₀ helix	(Gg):	0is	0.00%
Pi helix	(Ii):	0is	0.00%
Beta bridge	(Bb):	0is	0.00%
Extended strand	(Ec):	43is	14.01%
Beta turn	(Tt):	29is	9.45%
Bend region	(Ss):	0is	0.00%
Random coil	(Cc):	110is	35.83%
Ambiguous states	(?):	0is	0.00%
Other states	:	0is	0.00%

注:Hh. α -螺旋;Ee. β -折叠;Tt. β -转角;Cc. 无规则卷曲。Note: Hh. α -helix; Ee. β -sheet; Tt. β -turn; Cc. Random coil.

图3 二级结构预测

Fig. 3 Secondary structure prediction

2.3 DiMYB1 蛋白跨膜结构域预测和分析

利用 TMHMM server v 2.0 在线预测珙桐

DiMYB1 氨基酸序列的跨膜结构域(图 7),发现其不存在跨膜结构。由此推测 DiMYB1 蛋白不经过跨膜

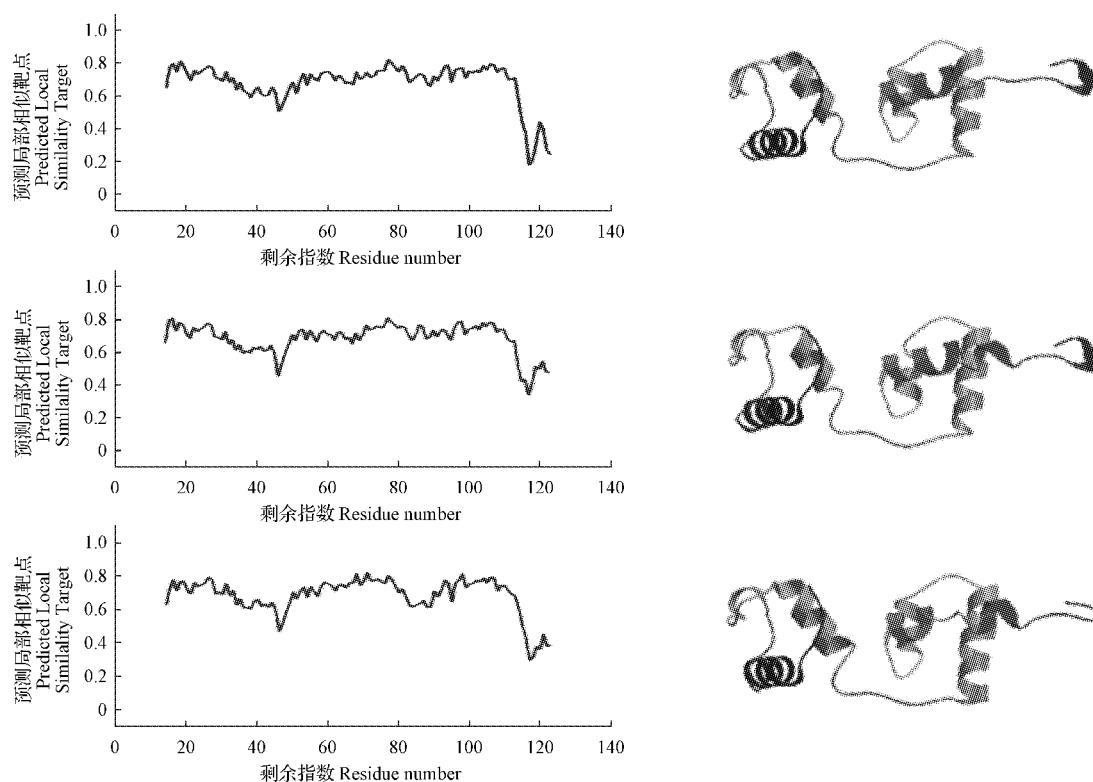


图 4 三级结构预测模型

Fig. 4 Three-dimensional constitution predicted by Swiss-model

表 1 亚细胞定位预测

Table 1 Subcellular localization prediction

亚细胞定位 Localization	可能性 Probability/%
细胞核 Nuclear	91.3
线粒体 Mitochondrial	8.7

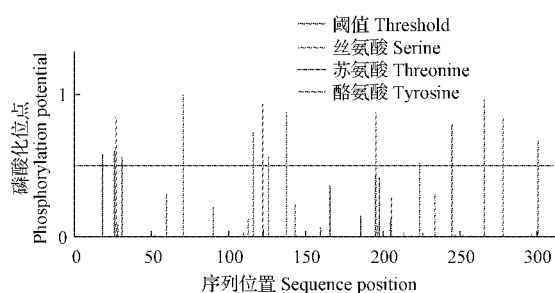


图 5 磷酸化位点预测

Fig. 5 Phosphorylation site prediction

转运,在细胞质的基质中合成。

2.4 DiMYB1 蛋白保守结构域分析

运用 SMART 软件分析 *DiMYB1* 基因的结构域,由图 8 可知,*DiMYB1* 蛋白含有 2 个 SANT 结构功能域。利用 NCBI 中的 CDD(Conserved Domain Database)对珙桐 *DiMYB1* 基因所编码的氨基酸序列进行分析,珙桐 *DiMYB1* 蛋白的 2 个典型 MYB 保

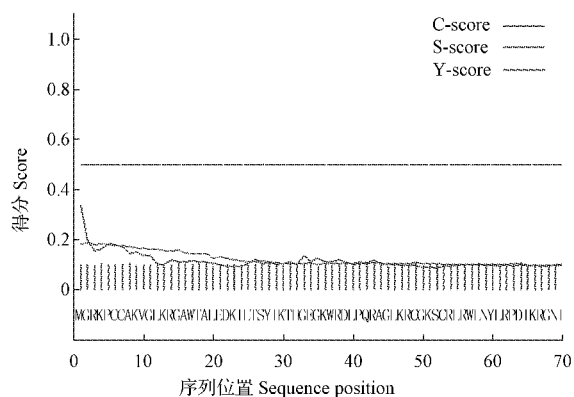


图 6 信号肽位点预测

Fig. 6 Protein signal peptide site prediction

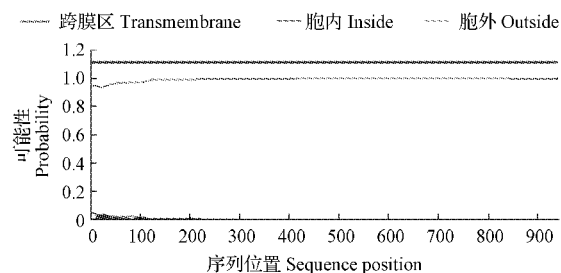


图 7 跨膜区域预测分析

Fig. 7 Transmembrane domain prediction

守结构域处于第 17~61 位和第 71~111 位的氨基酸之间,符合 MYB 基因家族保守结构域的特点,预测珙桐 DiMYB1 转录因子属于植物 R2R3-MYB 转录因子家族之一。

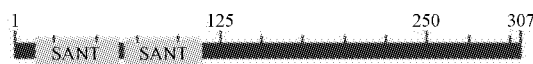


图 8 保守结构域预测

Fig. 8 Conserved domain prediction

1	DiMYB	MGR	KPCCAKVGLK	KGAWTALEDK	ILTSYIKTHG	EGKWRDLQPR	AGLKRCGKSC	RLRWLNLYRP	D1-KRGNISV	EEEEILMKLH	KLLGNRWSLI	AGRLPGRTDN	110
	EsMYB10	MGR	SPCCSKIEGLA	KGAWTSTEDK	ILKDYIRVIG	EGKWRSLPPR	AGLKRCGKSC	RLRWLNLYRP	D1-KRGNITI	DEEELIRLH	KLLGNRWSLI	AGRLPGRTDN	
	GrMYB6	MGR	SPCCSKIEGLA	KGAWTALEDK	ILTSYIHVIG	EGKWRSLPPR	AGLKRCGKSC	RLRWLNLYRP	D1-KRGNISI	DEEELIRLH	KLLGNRWSLI	AGRLPGRTDN	
	GhMYB36	MGR	SPCCSKIEGLA	KGAWTALEDK	ILASYIHVIG	EGKWRSLPPR	AGLKRCGKSC	RLRWLNLYRP	D1-KRGNISI	DEEELIRLH	KLLGNRWSLI	AGRLPGRTDN	
	AtMYB28	MSR	KPKKVGEGGLK	KGAWTTEEDK	KLISYIHDHG	EGGWRDIPQK	AGLKRCGKSC	RLRWLNLYRP	E1-KRGEFSS	EEEQIIMLH	ASRGNKWSVI	ARHLPRRTDN	
	DkMYB2	MGR	KPCCAKVGLK	KGAWTALEDK	ILANYIKLHG	EGKWRDIPQK	AGLKRCGKSC	RLRWLNLYRP	D1-KRGNISE	EEEEILRLH	KLLGNRWSLI	ARHLPGRTDN	
	EgMYB114	MYVCAKRLS	SSSSAAAALN	KGAWTALEDK	KLAAQCIEMHG	PTRWKFLALK	AGLKRCGKSC	RLRWLNLYRP	D1-KRGNISD	AEEDLILRLH	RLGNRWSLI	AGRLPGRTDN	
	PpR2R3MYB	MGR	KLPKDKDVVK	KGAWCPKEDK	VLADYIKTHG	EGKWSIAVKK	TGLNRSCKSC	RLRWLNLYRP	DV-KRGNISP	EEEDLIRLH	KLLGNRWSLI	AGRLPGRTDN	
	ZmMYB64	MSTKGSISRR	SFAAAHATA	TVARQSTDS	TATGRTIRTH	AASERIESVRR	RRLCRAGG	RIICEVPLRA	GAHQGRVDQ	GGGPAPIRLI	Q-GPRRGLL	EVATKSRRAA	
	VvMYB				MTAYVILQHG	UGNWRALPKQ	AGLKRCGKSC	RLRWLNLYRP	D1-KRGNISR	EEEDTIELH	EKLGKWSAI	AASLPGRTDN	
	Consensusm.r.....l..gawt...odk..l..yi..Hg	eg.wr.lp..agL.RcGksc	RlrwlnlyRp	di.krGn.s.	ccc.liir\$H	lIlgnrwslI	agrlpgRtdn					
111	DiMYB	ETKNYWNITL	SKRSMQGGKS	TLNSAKQLAK	L-----KGKE	SEGNLSSMAL	IPGNPPIQVQ	IRTKAVRIT-	KVILPQQLDN	QLVKRNEVPT	WDLRDNDPSS	STILLEDDTS	220
	EsMYB10	EIKNYWNITL	AKR-VGGQSS	TCTEKQLTHQ	T-----KRRK	VLPKIAQPPQ	IPDTKSS-VV	VRTKATKCT-	KSFFHLQKSE	KDSYGMCSN	YPKPNRPDS	DNIMTESSSR	
	GrMYB6	EIKNYWNITL	GKR-----AKA	QASIEAKTIP	T-----ESRL	NPKSKSSTKI	-----EV	IRTKAIRCSS	KVMVPLQPPA	THQHGQHHT	NNNEEMGGGI	STIEAHNGIQ	
	GhMYB36	EIKNYWNITL	GKR-----AKA	QASIDAKTIP	T-----ESRL	NPKSKSSTKI	-----EV	IRTKAIRCSS	KVMVPLQPPA	THQHGQHCT	NNNEEMGGGI	ATIEAHNGIQ	
	AtMYB28	EIKNYWNITL	KKRLMEQGID	PVTHIKPLASS	SNPTVIDENL	SPNASSSDIKQ	YSRSSSMPIFL	SRPPISSCNM	VSKVSELSN	DGTPIQGSSL	SKKRIKKSS	STSRILNKVA	
	DkMYB2	EIKNYWNSL	SKRI-----	-----NIIGD	NYSKNSRKQ	RSPIRQKHTV	IRTKAVRCTK	VVILPQELNVT	AMIEKREAPA	ASQPDYLLD	DAAGDLVTD		
	EgMYB114	EIKNYWNAHL	KKKA-----	-----KLME	KLPAPSTAT	ILQDNNEIDR	EKTVGSSSCT	GAVEILKEITQ	ESKORDEYWS	SDDDDCGGD	VERIDSSRRK		
	PpR2R3MYB	EIKNYWNTL	SKRVPPEARPR	RI-----	-----KINA	KNDGADGGN	CTAKDDQVQV	IQSTILNCTK	AAAMIPSHH	LHDGQDR			
	ZmMYB64	EVREELPAAV	DQLP-----	-----AFGP	QARQLQRGGG	RADHQVPRV	RQQVVAHRR	AAGQDQGRK	KLEHAHQAE	AARPRHGPAD	PPAAGRSASR		
	VvMYB	EIKNFHW											
	Consensus	e!k#.wn..l..kr.....v.....c.....										
221	DiMYB	FLMDFNISEL	FISDVLSQT	LQDQKEISNE	IDN-GVGGDY	STNMHSDFL	VPTFEAMLN	S--TDDAITE	NWRASIDHPL	QQNEDLLQK	QLASFLNSGD	-EWID	330
	EsMYB10	TSISECKGGL	W-SIEPIGMVS	FSPSREGDYS	LDHFGIDLIV	GELCLITEFL	CSDPQRLTEE	SLKVNDDNS	SLPLCLVNLQ	LIRDDQYDTK	WHEEQCEP		
	GrMYB6	MLES-----L	YSDGSGNLS	FE-----INE	LLKSHDGGEF	EENPMQGHITP	LGEAMIKDWS	TCPCLDNAGA	TDLESIAFLI	DTDEWP			
	GhMYB36	MLES-----L	YSDGSGNLS	FE-----INE	LLKSHDGGEF	EENPMQGHITP	LGEAMIKDWS	TCPCLDNAGA	TDLESIAFLI	DTDEWP			
	AtMYB28	AKATSTKDL	SASMEGSLA	TTISHAFEF	GFTEQIRNEE	DSNSTSLNT	LAEFDSLHTP	SLYPEHEINA	T--SDLNMDQ	DYDFSQFFFK	FGGDNHNEEN	-SMNDLLMSD	
	DkMYB2	LLDLEIPPEP	NENDGGG-G	DGDGGDFSGD	VCGFPPSEAW	WSLNDLSDEI	ILEADENWRA	STPEPPQHGL	G				
	EgMYB114	TKGLNWKTF	LELEEEES										
	PpR2R3MYB												
	ZmMYB64	RAAAAALPLP	DGATEAGGGG	GPIFGRRCTIQ	LAGGQQPQQR	RRRGAGAGAT	AAAAAAPPRH	RPEIPVIQPRG	LPAAGGNRQR	IRATDGDDEA	GGNGGGGV	LQQPVRVPCR	
	VvMYB												
	Consensus											

注:方框表示 Myb-like DNA 保守结合域、SANT 结构域、亲脂性区域、Rap1_C 区、Pox_H7 区;DiMYB1. 珙桐 (*Davidia involucreata* Baill., KR996175);DkMYB2. 柿树 (*Diospyros kaki*, AB503699);GrMYB6. 雷蒙德氏棉 (*Gossypium raimondii*, XM_012632600);GhMYB36. 陆地棉 (*Gossypium hirsutum*, AF336284);PpR2R3MYB. 碧桃 (*Prunus persica*, KT893408);EsMYB10. 淫羊藿 (*Epimedium sagittatum*, JN426957);VvMYB. 葡萄 (*Vitis vinifera*, DQ277669);AtMYB28. 拟南芥 (*Arabidopsis thaliana*, NP_200950);ZmMYB64. 玉米 (*Zea mays*, NM_001152681)。

Note:Box means Myb-like DNA conserved domain, SANT domain, amphipathic domain, Rap1_C domain, Pox_H7 domain;DiMYB1. *Davidia involucreata* Baill., KR996175;DkMYB2. *Diospyros kaki*, AB503699;GrMYB6. *Gossypium raimondii*, XM_012632600;GhMYB36. *Gossypium hirsutum*, AF336284;PpR2R3MYB. *Prunus persica*, KT893408;EsMYB10. *Epimedium sagittatum*, JN426957;VvMYB. *Vitis vinifera*, DQ277669;AtMYB28. *Arabidopsis thaliana*, NP_200950;ZmMYB64. *Zea mays*, NM_001152681.

图 9 同源序列比对

Fig. 9 Homologous alignments of DiMYB1 protein in *Davidia involucreata* Baill.

Lys 取代。系统发育树表明(图 10),DiMYB1 与柿树的 MYB2 的亲缘关系最近,相似度为 84%。

2.6 蛋白质相互作用分析

借助与拟南芥中相关 MYB 蛋白的相互作用辅助了解 DiMYB1 蛋白的功能和调节机制。采用在线 STRING 交互式数据库搜索与 ATMYB4 蛋白相互作用的蛋白构建蛋白互作网络(图 11),发现其与肉

2.5 DiMYB1 氨基酸序列同源性及聚类分析

将克隆到的 DiMYB1 氨基酸序列与 GenBank 中已经注册的不同物种的氨基酸序列在 Multalin version 5.4.1 上进行同源性在线比对,由图 9 可知,珙桐 DiMYB1 转录因子的氨基酸序列含 H-T-H 跨膜螺旋区、Pox_H7 区、Rap1_C 区、Myb-like DNA 结合区、SANT 结构域,与其它物种保守序列不同的是:DiMYB1 蛋白跨膜螺旋区及高度保守结构域中第 14 位的赖氨酸(Lys)和第 107 位的酪氨酸(Tyr)分别被精氨酸(Arg)和组氨酸(His)取代,第 80 位的 Arg 被

桂酸羟化酶(C4H)的功能相近,C4H 是参与花青素合成途径中的一种关键酶,预测 DiMYB1 在原花青素合成过程中也可能起到了很重要的调控作用。

3 讨论

该研究在珙桐果实及种子转录组数据库的基础上应用生物信息学分析的方法和工具对 DiMYB1 基因编码蛋白的理化性质、信号肽、亲水性/疏水性、二

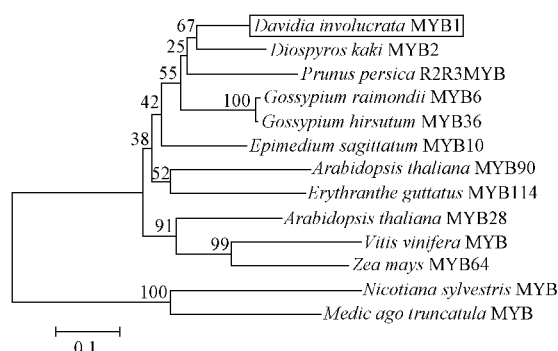


图 10 系统进化分析

Fig. 10 Phylogenetic analysis

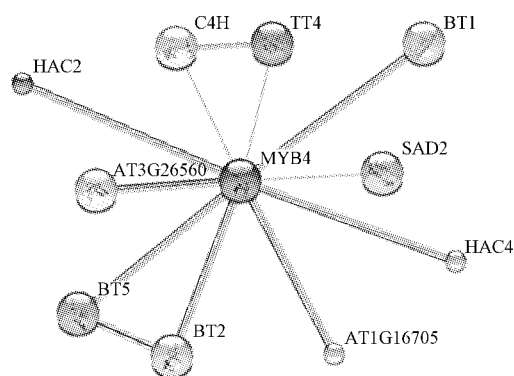


图 11 蛋白质相互作用分析

Fig. 11 Protein interaction analysis

级结构、三级结构、结构域功能等进行了预测,对 *DiMYB1* 氨基酸同源性及蛋白质的相互作用进行了分析。结果表明, *DiMYB1* 基因全长 924 bp, 编码 307 个氨基酸, 理论等电点为 7.02, 含有 15 个磷酸化

位点, 1 个信号肽, 其亲水性很强。 *DiMYB1* 蛋白的二级结构以 α -螺旋和无规则卷曲为主要构件, 三级结构预测结果比较可靠。该蛋白的 MYB-like DNA 保守结合域、2 个 SANT 结构域, 符合 MYB 基因家族保守结构域的特点, 属于 R2R3-MYB 基因家族。该研究对珙桐 *DiMYB1* 基因进行的生物信息学分析为后续验证 *DiMYB1* 基因的功能提供了参考。

参考文献

- [1] 中国植被编委会. 中国植被 [M]. 北京: 科学出版社, 1980: 298-299.
- [2] ALLAN A C, HELLENS R P, LAING W A. MYB transcription factors that colour our fruit[J]. Trends in Plant Science, 2008, 13(3): 99-102.
- [3] DUBOS C, STRACKE R, GROTEWOLD E, et al. MYB transcription factors in *Arabidopsis* [J]. Trends in Plant Science, 2010, 15(10): 573-581.
- [4] 郭文芳, 刘德春, 杨莉, 等. 柑橘 MYB15 基因的克隆与表达分析 [J]. 植物科学学报, 2015(6): 808-818.
- [5] 杨召恩, 杨作仁, 刘坤, 等. 一个亚洲棉 MYB 家族新基因克隆及特征分析 [J]. 中国农业科学, 2013(1): 195-204.
- [6] 姜玲. 葡萄风信子 MYB 转录因子基因的克隆及花色基因的表达分析 [D]. 杨凌: 西北农林科技大学, 2014.
- [7] 杨晔. 中国水仙 MYB 转录因子基因的克隆与表达分析 [D]. 福州: 福建农林大学, 2013.
- [8] 郑蕊, 岳思君, 王丽娟. 枸杞 *LbMYB1* 基因的克隆与表达分析 [J]. 江西农业大学学报, 2016(1): 174-179.
- [9] AKAGI T, IKEGAMI A, TSUJIMOTO T, et al. DkMYB4 is a MYB transcription factor involved in proanthocyanidin biosynthesis in persimmon fruit [J]. Plant Physiology, 2009, 151(4): 2028-2045.
- [10] 戴鹏辉, 任锐, 曹福祥, 等. 珙桐 MYB 转录因子 *DiMYB1* 基因的克隆及表达分析 [J]. 植物生理学报, 2016(8): 1255-1262.

Bioinformatics Analysis of *DiMYB1* Gene in *Davidia involucreta* Baill.

DAI Penghui, REN Rui, DONG Xujie, LI Meng, CAO Fuxiang

(College of Life Science and Technology, Central South University of Forestry and Technology, Changsha, Hunan 410004)

Abstract: The paper was to lay the foundation for functional identification and regulation of gene *DiMYB1* from *Davidia involucreta* Baill. transcriptome data. The results showed that the fragment of *DiMYB1* included an ORF (open reading frame) of 924 bp, which encoded for a protein product of 307 amino acids, with molecular weight 35 kDa. *DiMYB1* belonged to R2R3-MYB protein family, containing a signal peptide and the two functional domains. Subcellular localization to predict the likelihood of the gene in the nucleus was 91.3%. Phylogenetic tree showed with *Diospyros kaki* MYB2 transcription factor had the highest homology, the similarity was 84%.

Keywords: *DiMYB1*; *Davidia involucreta* Baill.; bioinformatics analysis