

蛇根木WRKY转录因子的鉴定及特征分析

杨致荣^{1,*}, 陈钊¹, 樊娟², 李润植^{3,*}, 王兴春²

山西农业大学¹文理学院, ²生命科学学院, ³分子农业与生物能源研究所, 山西太谷030801

摘要: WRKY是一个转录因子大家族, 在植物次生代谢过程起着重要的调控作用。本文对蛇根木RsWRKY转录因子的数目、类型、保守基序及其表达模式等进行了分析。从蛇根木的34 109个蛋白中鉴定出48个RsWRKY转录因子; 这48个转录因子根据其WRKY域的特点可分为1、2和3三类, 第1类RsWRKY包括13个成员, 其中8个含有两个WRKY域; 第2类有29个成员, 可分为2-a、2-b、2-c、2-d和2-e五个亚类; 第3类RsWRKY包括6个成员, 全部属于3-a类。这48个WRKY基因的表达具有组织器官特异性, 按照其表达谱可分为五大类群。第I、II和V类群的RsWRKY在根、茎和叶中的表达量相对较高, 表达模式分别与第II、III和I类阿吗灵生物合成关键基因的表达模式相似, 是未来研究RsWRKY介导的次生代谢调控网络的候选靶基因。

关键词: WRKY转录因子; 次生代谢; 表达谱; 蛇根木

Identification and Characterization of WRKY Transcription Factors in *Rauvolfia serpentina*

YANG Zhi-Rong^{1,*}, CHEN Zhao¹, FAN Juan², LI Run-Zhi^{3,*}, WANG Xing-Chun²

¹College of Arts and Science, ²College of Life Sciences, ³Institute of Molecular Agriculture & Bioenergy, Shanxi Agricultural University, Taigu, Shanxi 030801, China

Abstract: WRKY is one of the largest transcription factor families involved in plant secondary metabolism regulation. In this study, the gene number, classification, conserved motif and expression pattern of the WRKY transcriptional factors in *Rauvolfia serpentina* were analyzed. A total of 48 RsWRKYs were identified from all of the 34 109 proteins in *R. serpentina*. According to the structural feature of the WRKY domain, the 48 RsWRKYs were classified into 3 groups. The group 1 included 13 WRKY members, among which only 8 members had two WRKY domains. The group 2 included 29 RsWRKYs, and they were classified into 5 subgroups (2-a, 2-b, 2-c, 2-d and 2-e). The group 3 had 6 members and all of them belonged to group 3-a. These 48 RsWRKY genes displayed different expression patterns, and could be divided into 5 groups. Among these five groups, the genes of group I, II and V were highly expressed in roots, stems and leaves, and their expression patterns were similar with the group II, III, and I genes involved ajmaline biosynthesis respectively. Thus, these WRKY genes are the putative targets for second metabolite engineering mediated by RsWRKYs in *R. serpentina* in the future.

Key words: WRKY transcription factor; secondary metabolism; expression pattern; *Rauvolfia serpentina*

WRKY是植物中特有的一个转录因子大家族, 由于该家族蛋白N端都含有保守的“WRKYGQK”基序而得名。除了WRKY域外, 该家族蛋白还包含其它的保守基序如“WRKYGKK”和“WRKYGRN”等(Wang等2014)。根据WRKY域的数目及其锌指结构的类型可将WRKY蛋白分为三类: 第1类含有2个WRKY结构域(N末端WRKY域和C末端WRKY域), 第2和3类只有1个WRKY域。第1和2类蛋白锌指结构为C₂H₂型, 而第3类为C₂HC型(Brand等2013; Wen等2014)。第2类WRKY转录因子又包括a、b、c、d和e五个亚类(Dou等2014)。

WRKY转录因子不仅参与植物生长和发育的调控、生物和非生物胁迫的响应(Li等2012; 孙利军等2014), 同时还形成了一个庞大的信号网络参与次生代谢的转录调控中, 是药用植物有效成分合成途径的关键调控因子(Suttipanta等2011; Yang等2013; 杨致荣等2013)。WRKY转录因子通

收稿 2014-12-14 修定 2015-03-04

资助 山西省自然科学基金(2013011028-1)、山西农业大学引进人才博士科研启动资金项目(2013YJ42)和山西省留学人员科技活动择优资助项目(晋人社厅函[2014]779号)。

* 共同通讯作者(E-mail: zryangsx@163.com, Tel: 0354-6288341; E-mail: rli2001@hotmail.com, Tel: 0354-6288063)。

过结合靶基因启动子区域的W-box等元件调控靶基因的表达(Yang等2013; Wang等2014)。抗癌药用植物长春花(*Catharanthus roseus*)中有47个WRKY基因,其中超表达CrWRKY1基因能提高长春花中阿吗碱和蛇根碱等药用成分含量,表明基于WRKY转录因子的次生代谢基因工程在植物源药物研发和生产上具有广阔的应用前景(Suttipanta等2011; Yang等2013; 杨致荣等2013)。

蛇根木又称印度萝芙木,和长春花同属于夹竹桃科植物,富含阿吗灵(ajmaline)和利血平(reserpine)等生物碱,具有降血压、抗癫、消炎、抗心律失常和抗癌等功效,是一种重要的药用植物(Srivastava等2006; Yu和Chen2014)。其中,阿吗灵是蛇根木根中的主要有效成份,其生物合成主要涉及色氨酸脱羧酶(tryptophan decarboxylase, TDC)、异胡豆苷合酶(strictosidine synthase, STR)(Kutchan等1988; Bracher和Kutchan 1992)、异胡豆苷糖苷酶(strictosidine-O-beta-D-glucosidase, SG)、聚精液素醛酯酶(polyneuridine aldehyde esterase, PNAE)(Dogru等2000)、维诺任碱合酶(vinorine synthase, VS)(Bayer等2004)、霹雳萝芙木碱还原酶(perakine reductase, PR)(Sun等2008)、萝卡辛糖苷酶(raucaffricine-O-beta-D-glucosidase, RG)(Warzecha等2000)和乙酰阿吗碱酯酶(acetylajmalan esterase, AAE)等。但由于次生代谢途径极其复杂,通过对个别酶的改造或遗传操作很难从根本上解决药用有效成分含量低的问题。因此,在全面、深入了解药用有效成分合成代谢的基础上,发现并鉴定对代谢途径起调控作用的转录因子,揭示次生代谢产物的转录调控机制,可以实现对药用植物的进一步开发和利用。本研究通过对药用植物数据库中蛇根木蛋白序列的生物信息学分析,鉴定出48个蛇根木RsWRKY转录因子,并根据氨基酸序列将其分为三大类;对不同类RsWRKY蛋白的WRKY域中的WRKY核心序列及锌指结构进行比对及保守性分析,并通过已有的表达数据对部分鉴定的WRKY蛋白进行表达分析,为今后蛇根木RsWRKY基因的功能研究奠定了基础。

材料与方法

1 蛋白序列的收集及整理

蛇根木(*Rauvolfia serpentina* Benth. Ex Kurz)

和拟南芥(*Arabidopsis thaliana* L.)的蛋白序列分别从药用植物数据库(Medicinal Plant Genomics Resource, MPGR, <http://medicinalplantgenomics.msu.edu/>)和拟南芥信息资源数据库(The Arabidopsis Information Resource, TAIR, <http://www.arabidopsis.org/>)下载。下载的蛋白序列用Notepad++和Excel 2010软件进行编辑。若同一RsWRKY基因有不同的剪切方式,则选择序列最长的蛋白序列进行分析,并将蛋白质编号进行了简化处理,如rsa_locus_10660简化为rsa10660。

2 蛇根木WRKY转录因子的鉴定

从上述编辑好的蛋白序列中查找含有“WRKY”基序的序列,并分别统计基序WRKYGQK、WRKYGKK、WRKYGEK和WRKYGRN的数目,从而初步确定RsWRKY转录因子的数目。为确保所鉴定的WRKY转录因子的准确性,将上述推测的RsWRKY的蛋白序列利用NCBI保守域数据库(NCBI Conserved Domain Database, NCBI CDD)进行保守域分析,鉴定具有WRKY功能域的序列。最后,将含有WRKY功能域的序列在PlantTFcat(Plant Transcription Factor and Transcriptional Regulator Categorization and Analysis Tool)中进行比对分析,进一步验证其是否属于转录因子。

3 蛇根木RsWRKY转录因子的分类

参照Eulgem等(2000)WRKY结构域的定义和分组原则,利用MEGA 6.0软件(Tamura等2011)对蛇根木和拟南芥WRKY蛋白的65个氨基酸残基的WRKY域核心序列进行序列比对和聚类分析。进化树构建采用邻位相连法,Bootstrap设定为3 000个重复,进化距离的计算采用p距离法。

将蛇根木分属不同类的WRKY蛋白与相应类群的AtWRKY蛋白结构域的65个氨基酸残基进行比较分析,运用ClustalW 2.1软件(Larkin等2007)进行在线比对。含有2个WRKY域的蛋白分别标注N末端WRKY结构域和C末端WRKY结构域,各自进行比较分析,以鉴定N末端和C末端结构域。

4 蛇根木WRKY基序组成分析

RsWRKY蛋白的WRKY基序及锌指结构用MEME 4.8.0在线软件(<http://meme.nbcr.net/meme/intro.html>)进行分析。具体参数为:基序位点数量2~50;基序重复次数为任何(any);每个基序的长度为6~70个氨基酸残基;预测基序的数量最多为3。

5 蛇根木WRKY表达模式分析

蛇根木转录组测序的表达数据从MPGR数据库下载。RsWRKY和阿吗灵生物合成关键酶基因在根、茎、叶和花中表达数据系统聚类及其表达模式分析选用Cluster Project Version 1.0软件(<http://ibi.zju.edu.cn/software/clusterproject/>), 分类方法选择层次聚类法(hierarchical clustering), 凝聚(Agglomerative)中的Average linkage (UPGMA), 距离尺度(distance metric)选择欧几里得(Euclidian)。聚类结果由亮红到暗红的梯度表示基因在相应部位的表达量的高低, 黑色为不表达。

实验结果

1 蛇根木WRKY转录因子的鉴定

WRKY域是WRKY蛋白最显著的特征, 尤其是N端的保守七肽(Huang等2012)。蛇根木34 109个单拷贝蛋白序列中有64个含有WRKY基序, 其中48个蛋白含有典型七肽。进一步分析表明, 这48个蛋白中含有WRKYGQK、WRKYGKK和WRKYGRN基序的RsWRKY数目分别为43、4和1个。为了确保所鉴定RsWRKY蛋白的真实性, 将这48个可能属于WRKY家族的蛋白序列用NCBI CDD和PlantTFcat进一步分析, 确定了该48个蛋白都属于WRKY转录因子家族蛋白(蛋白序列下载网址ftp://ftp.plantbiology.msu.edu/pub/data/MPGR/Rauvolfia_serpentina/rsa.trans.estscan.pep.10072011.fasta.gz)。

将这48个RsWRKY转录因子与72个AtWRKY转录因子进行序列比对, 以检测RsWRKY序列是否完整。结果表明, 48个RsWRKY蛋白中, 序列最长的为rsa475, 有904个氨基酸残基; 序列最短的为rsa86988, 仅有96个氨基酸残基; 其中有37个蛋白含有1个完整的WRKY域, 7个蛋白序列含2个完整的WRKY域; 3个序列的WRKY域缺失C端的锌指结构, 它们分别为rsa86988、rsa17741和rsa475; 而rsa78817的WRKY域缺失部分N端序列(表1和图2)。

2 RsWRKY转录因子聚类分析

以WRKY域的65个氨基酸残基作为比对序列, 将蛇根木中48个RsWRKY蛋白与拟南芥已鉴定并分类的72个AtWRKY的WRKY域进行聚类分

析。若蛋白含有2个WRKY域, 则在基因名后边分别用N和C标注, 并作为独立的WRKY域分别进行分析。结果(图1)表明, 这48个RsWRKY可以分为三类。第1类群体(Group 1, G1)包括13个WRKY蛋白, 其中8个蛋白包含有2个WRKY域, 它们的N端WRKY域和C端WRKY域分别聚类于第1类群体的N端群(Group 1N, G1N)和C端群(Group 1C, G1C)中; 有5个WRKY蛋白序列只有1个WRKY域, 分别聚类于C端WRKY域(rsa16180、rsa33394、rsa35233和rsa78817)和N端WRKY域中(rsa8988)。第2类群体(Group 2, G2)包括29个蛋白, 分别和拟南芥的5个亚类聚在一起, 其中2个G2-a WRKY蛋白、4个G2-b、12个G2-c、5个G2-d和6个G2-e; 第3类群体(Group 3, G3)包括6个蛋白, 全部与拟南芥的G3-a类聚在一起, 未发现G3-b类成员。

为确定蛇根木WRKY蛋白分类的准确性, 从拟南芥各类群体中随机选取一个WRKY蛋白和蛇根木相应群体WRKY蛋白的65个核心域氨基酸残基序列比对分析。如图2所示, 同一群体WRKY蛋白不仅在WRKY基序保守七肽和锌指结构上基本一致, 而且整个WRKY结构域中的65个氨基酸残基序列存在高度相似性并和相应群体的AtWRKY高度同源, 表明蛇根木WRKY蛋白的分类是正确的。

3 RsWRKY转录因子基序分析

为了确定RsWRKY结构域中的WRKY基序及锌指结构在不同群体中的保守性及多样性, 将WRKY结构域中包括WRKY基序的25个氨基酸残基及不同群体锌指结构的35个氨基酸残基分别运用MEME 4.8.0在线软件进行分析。结果表明, WRKYGQK七肽基序在不同群体的蛇根木WRKY结构域存在高度保守性, 只在G2-c和G2-d中发现一些变异序列(图2和3)。在G2-c中发现Q突变为K形成WRKYGKK; 在G2-d中Q突变为R, 而第2个K突变为N, 从而形成WRKYGRN。

在WRKY域中除N端保守的七肽基序外, C端还有一个典型的锌指结构。不同RsWRKY域锌指结构差异较大。第1、2类群锌指结构为C₂H₂型, 其中G1N锌指结构的类型为C-X₄-C-X₂₂-H-X₁-H(图4-A), G1C和G2-c锌指结构序列为C-X₄-C-X₂₃-H-X₁-H(图4-B), 第2类群除G2-c外, 其它均为C-X₅-C-X₂₃-H-X₁-H(图4-C); 第3类群锌指结构为C₂HC

表1 蛇根木WRKY转录因子信息
Table 1 WRKY transcription factors in *R. serpentine*

| 蛋白质简化编号 | 保守七肽 | 长度/aa | 锌指结构类型 | WRKY域数量 | 所属类群 |
|----------|---------|-------|-------------------------------|---------|------|
| rsa105 | WRKYGQK | 569 | C ₂ H ₂ | 2 | 1 |
| rsa10660 | WRKYGQK | 533 | C ₂ H ₂ | 2 | 1 |
| rsa16180 | WRKYGQK | 458 | C ₂ H ₂ | 1 | 1 |
| rsa2569 | WRKYGQK | 515 | C ₂ H ₂ | 2 | 1 |
| rsa2784 | WRKYGQK | 775 | C ₂ H ₂ | 2 | 1 |
| rsa33394 | WRKYGQK | 217 | C ₂ H ₂ | 1 | 1 |
| rsa35233 | WRKYGQK | 247 | C ₂ H ₂ | 1 | 1 |
| rsa3810 | WRKYGQK | 528 | C ₂ H ₂ | 2 | 1 |
| rsa4580 | WRKYGQK | 558 | C ₂ H ₂ | 2 | 1 |
| rsa475 | WRKYGQK | 904 | C ₂ H ₂ | 2 | 1 |
| rsa6153 | WRKYGQK | 596 | C ₂ H ₂ | 2 | 1 |
| rsa78817 | WRKYGQK | 105 | C ₂ H ₂ | 1 | 1 |
| rsa86988 | WRKYGQK | 96 | — | 1 | 1 |
| rsa11364 | WRKYGQK | 276 | C ₂ H ₂ | 1 | 2-a |
| rsa14559 | WRKYGQK | 388 | C ₂ H ₂ | 1 | 2-a |
| rsa1048 | WRKYGQK | 623 | C ₂ H ₂ | 1 | 2-b |
| rsa15618 | WRKYGQK | 652 | C ₂ H ₂ | 1 | 2-b |
| rsa39439 | WRKYGQK | 344 | C ₂ H ₂ | 1 | 2-b |
| rsa9668 | WRKYGQK | 480 | C ₂ H ₂ | 1 | 2-b |
| rsa10325 | WRKYGQK | 406 | C ₂ H ₂ | 1 | 2-c |
| rsa14679 | WRKYGQK | 223 | C ₂ H ₂ | 1 | 2-c |
| rsa19140 | WRKYGKK | 200 | C ₂ H ₂ | 1 | 2-c |
| rsa20939 | WRKYGKK | 193 | C ₂ H ₂ | 1 | 2-c |
| rsa24208 | WRKYGQK | 280 | C ₂ H ₂ | 1 | 2-c |
| rsa371 | WRKYGKK | 173 | C ₂ H ₂ | 1 | 2-c |
| rsa38206 | WRKYGQK | 188 | C ₂ H ₂ | 1 | 2-c |
| rsa6050 | WRKYGQK | 396 | C ₂ H ₂ | 1 | 2-c |
| rsa6058 | WRKYGQK | 301 | C ₂ H ₂ | 1 | 2-c |
| rsa66805 | WRKYGKK | 116 | C ₂ H ₂ | 1 | 2-c |
| rsa8054 | WRKYGQK | 221 | C ₂ H ₂ | 1 | 2-c |
| rsa9342 | WRKYGQK | 329 | C ₂ H ₂ | 1 | 2-c |
| rsa11023 | WRKYGRN | 433 | C ₂ H ₂ | 1 | 2-d |
| rsa12406 | WRKYGQK | 333 | C ₂ H ₂ | 1 | 2-d |
| rsa12469 | WRKYGQK | 392 | C ₂ H ₂ | 1 | 2-d |
| rsa3792 | WRKYGQK | 343 | C ₂ H ₂ | 1 | 2-d |
| rsa8531 | WRKYGQK | 91 | C ₂ H ₂ | 1 | 2-d |
| rsa143 | WRKYGQK | 268 | C ₂ H ₂ | 1 | 2-e |
| rsa15606 | WRKYGQK | 369 | C ₂ H ₂ | 1 | 2-e |
| rsa19677 | WRKYGQK | 295 | C ₂ H ₂ | 1 | 2-e |
| rsa23071 | WRKYGQK | 279 | C ₂ H ₂ | 1 | 2-e |
| rsa25734 | WRKYGQK | 203 | C ₂ H ₂ | 1 | 2-e |
| rsa27444 | WRKYGQK | 320 | C ₂ H ₂ | 1 | 2-e |
| rsa9353 | WRKYGQK | 237 | C ₂ HC | 1 | 3-a |
| rsa9788 | WRKYGQK | 318 | C ₂ HC | 1 | 3-a |
| rsa14878 | WRKYGQK | 312 | C ₂ HC | 1 | 3-a |
| rsa18329 | WRKYGQK | 141 | C ₂ HC | 1 | 3-a |
| rsa1882 | WRKYGQK | 431 | C ₂ HC | 1 | 3-a |
| rsa17741 | WRKYGQK | 196 | — | 1 | 3-a |

—: 未检测。

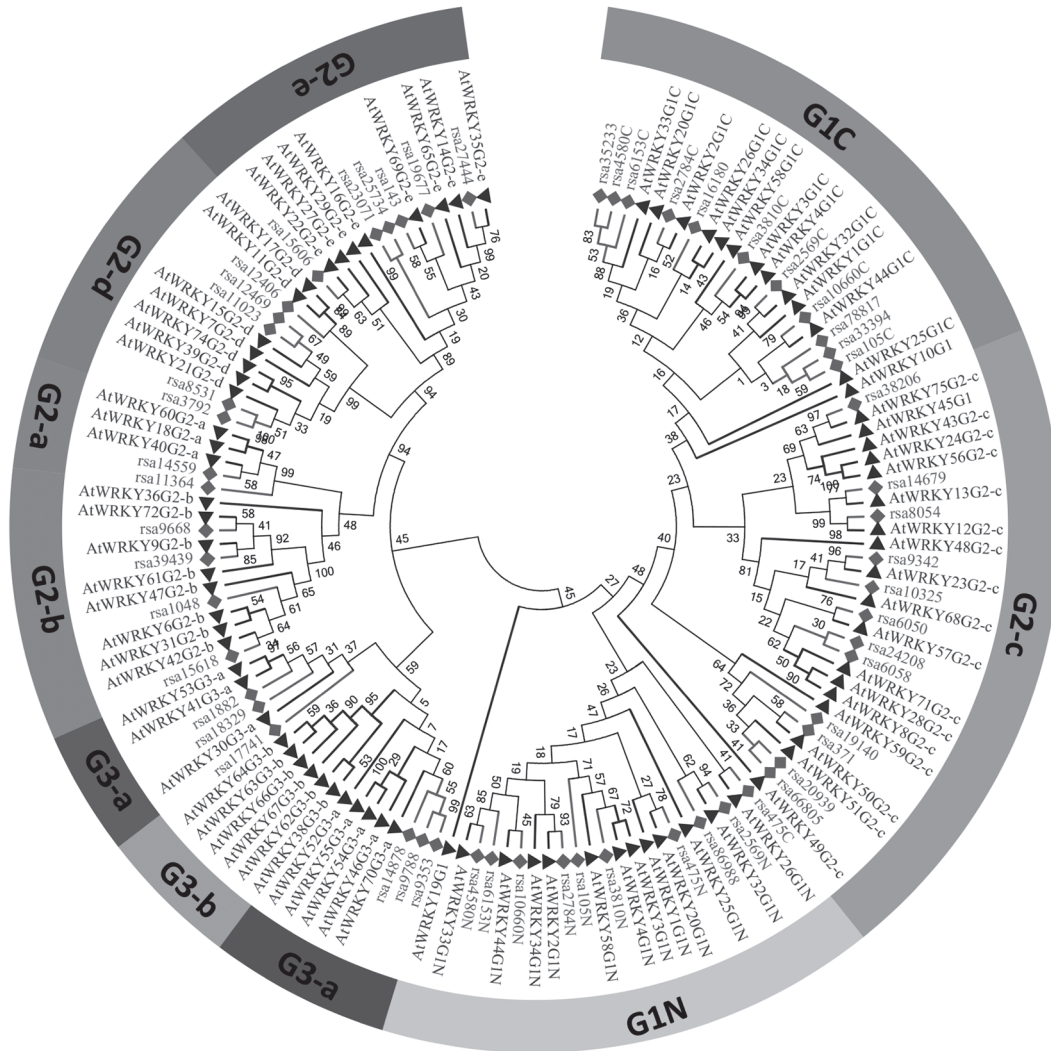


图1 蛇根木与拟南芥WRKY域的系统进化树

Fig.1 Phylogenetic tree of WRKY domains between *R. serpentina* and *Arabidopsis*

型, G3-a和G3-b的序列均为C-X₇-C-X₂₃-H-X₁-C (图4-D), 在蛇根木中仅有G3-a类。48个RsWRKY蛋白序列中rsa86988和rsa17741两种蛋白的WRKY域锌指结构不完整, rsa475 C端WRKY域的锌指结构也不完整, 这些都需进一步测序证实(图2); rsa6153的C端WRKY域不具有典型的锌指结构特征(图2)。

4 RsWRKY基因表达分析

基因表达模式是研究基因功能的重要手段。为深入了解蛇根木RsWRKY基因的功能, 将48个RsWRKY分别在幼嫩和成熟的根、茎、叶及花中表达的RNA-Seq数据进行系统聚类及分析。图5显示, 48个RsWRKY在不同组织中的表达存在很大差异, 大多数RsWRKY在根和成熟的叶片中表达量

相对较高, 而在茎和花中的表达量较低。根据聚类分析结果, 48个RsWRKY的表达模式可分为5大类群: 第I类群RsWRKY基因在根、茎、叶和花中都表达, 但在根中表达量最高; 第II类群在根、茎、叶和花中都表达, 但表达量除根中外其他组织均较低; 第III类群在根、茎、叶和花等器官中都不表达或表达量极低; 第IV类群主要在成熟叶片中表达; 第V类群在根、茎、叶和花等器官表达量都较高(图5)。

5 RsWRKY与阿吗灵生物合成关键酶基因的共表达分析

为了初步了解蛇根木RsWRKY转录因子与主要药用有效成分生物合成之间的关系, 进行了Rs-

| | | |
|--------------|---|----|
| Group 1N | | |
| AtWRKY3G1N | ADKPADDGYNWRKYGQKQVKGSDFPSYKCTHPACPVKKKVERS LDGQVTEI IYKQGNHHEL | 65 |
| rsa3810N | VDKPADDGYNWRKYGQKQVKGSEYPRSYKCTHQNCPVKKKVERSHDQVTEI IYKQGNHQP | 65 |
| rsa105N | LDKPADDGYNWRKYGQKLVKAKEHPRSYKCTHLNCPVKKKVERAPDGHVTEI IYKQGNHHEMP | 65 |
| rsa4580N | EQRKSDDGYNWRKYGQKQVKGSENPRSYKCTFPSCPTKKKVERNLEGI TEI IYKGNHNHPK | 65 |
| rsa6153N | ERRRSEDGYNWRKYGQKQVKGSENPRSYKCTYPNCPVKKKVERS LDGQI TEI IYKGNHNHPK | 65 |
| rsa10660N | GDRPSYDGYNWRKYGQKQVKGSEYPRSYKCTHPNCPVKKKVERS LDGQIAE IYKGEHNHPK | 65 |
| rsa2784N | GGAPAEADGYNWRKYGQKQVKGSEYPRSYKCTHPNCPVKKKVERS QEGHI TEI IYKGAHNHPK | 65 |
| rsa475N | QEKALDDGYSWRKYGQKLVKGSEYPRSYKCTYPNCEVKKIFERSPDGQI TEI IYKGS HDHPK | 65 |
| rsa2569N | PRTYPYDGYNWRKYGQKQVKGSRYSYRCTYADCSAKKIECS DSHSNHVI EIVYRSRHNHP | 65 |
| rsa86988 | PTPRTPDGYNWRKYGQX----- | 65 |
| Group 1C | | |
| AtWRKY2G1C | DVDILDDGYRWRKYGQKQVKGPNPRSYKCTAPGCTVRKHVERASHDLKSVITTYEGKHNHDVP | 65 |
| rsa16180 | EIDVLLDDGYRWRKYGQKQVKGPNPRSYKCTAGCPVRKHVERASEDIKSVITTYEGKHNHEVP | 65 |
| rsa33394 | EVXLLDDGKWRKYGQKQVKGPNPRSYKCTSVGCNVRKHVERSTDYKAVVTTYEGKHNHDIP | 65 |
| rsa35233 | DIDILDDGYRWRKYGQKQVKGPNPRSYKCTYNGCPVRKHVERASHDLRAVITTYEGKHNHP | 65 |
| rsa78817 | -----RWRKYGQKQVKGPNPRSYRCTYSGCNVRKQVERSADPKAVITSYEGKHSHE I- | 65 |
| rsa105C | EVDLDDGKWRKYGQKQVKGPNPRSYRCTYAGCNVRKHVERASTDPKAVITTYEGKHNHGIP | 65 |
| rsa10660C | EVDLDDGKWRKYGQKQVKGPNPRSYRCTSLKCNVRKQVVERTSDDPTAFITTYEGKHNHP | 65 |
| rsa2569C | DVGSIGDGYRWRKYGQKQVKGPNPRSYRCTYAGCPVRKHVERASHDLKSNVAFI IYKGRHDHDM | 65 |
| rsa2784C | EVDLDDGKWRKYGQKQVKGPNPRSYKCTAGCTVRKHVERASHDLKSVITTYEGKHNHDVP | 65 |
| rsa3810C | EVDLDDGKWRKYGQKQVKGPNPRSYRCTNPGCNVRKHVERAASDPKAVITTYEGKHS HDVP | 65 |
| rsa4580C | DIDILDDGKWRKYGQKQVKGPNPRSYKCTYAGCPVRKHVERASHDLRAVITTYEGKHNHDVP | 65 |
| rsa475C | SERSSDDGYNWRKYGQKLVKXX | 65 |
| rsa6153C | DIDILDDGKWRKYGQKQVKGPNPRSYKCTNPGCPVRKHVERASHDLRAVITTYEASTNHDVP | 65 |
| Group 2-a | | |
| AtWRKY40G2-a | TTLVVKDGYQWRKYGQKQVTRDNPSPRAYFKCACAPSCSVKKKQVRSVEDQSVLVATYEGEHNHP | 65 |
| rsa11364 | KSLVVKDGYHWRKYGQKQVTRDNPSPRAYFKCSFAPCTQVKKKQVRSIGDPSILVATYEGEHNHQ | 65 |
| rsa14559 | TSLIVKDGQYQWRKYGQKQVTRDNPSPRAYFKCSFAPCTCPVKKKQVRSVEDQSVLVATYEGEHNHP | 65 |
| Group 2-b | | |
| AtWRKY61G2-b | ETPTMNDGQWRKYGQKQVKGPNPRSYRCTIAASCVRKQVQRCEDEMSILISTYEGTHNHPL | 65 |
| rsa1048 | EAAMISDGCQWRKYGQKQVKGPNPRSYRCTMVGCGPVRKQVQRSAEDRSILITTYEGHHTHPL | 65 |
| rsa15618 | EAPMIDGQCQWRKYGQKQVKGPNPRSYRCTMAAGCPVRKQVQRCAEDRTILITTYEGNHNHP | 65 |
| rsa39439 | EAATMNDGQWRKYGQKQVKGPNPRSYRCTVAPGCPVRKQVQRCEDEMSILITTYEGTHNHPL | 65 |
| rsa9668 | DTPTMNDGQWRKYGQKQVKGPNPRSYRCTVAPNCPVRKQVQRCAEDMSILITTYEGTHNHPL | 65 |
| Group 2-c | | |
| AtWRKY75G2-c | QVDILDDGYRWRKYGQKQVKGPNPRSYRCTYGGCNVKKQVQRLTVDQEVVTTYEGVHSHPIE | 65 |
| rsa10325 | EVDHLLDDGYRWRKYGQKQVKGPNPRSYRCTSAACGVKKRVERSDDPSIVVTTYEGTHTHPCP | 65 |
| rsa14679 | DVDVLLDDGKWRKYGQKQVKGPNPRSYRCTQDNCRVKKRVERLAEDPRMIVITTYEGRHIHSPS | 65 |
| rsa19140 | DLETLLDDGKWRKYGQKQVKGPNPRSYRCTSVGCKVKKRTERDREDSYLLITTYEGIHNNKSP | 65 |
| rsa20939 | KLDIMDDGFKWRKYGQKQVKGPNPRSYRCTYKSNLGNVKKRVERDREDSRFVITTYEGMHNHES | 65 |
| rsa24208 | EVDHLEDGYRWRKYGQKQVKGPNPRSYRCTTQKCPVKKRVERSFQDPSIVITTYEGTHNHHP | 65 |
| rsa371 | EVEVLLDDGKWRKYGQKQVKGPNPRSYRCSVEGCPVKKRVERDKEDPRYVITTYEGIHNNQIP | 65 |
| rsa3206 | QVDILDDGYRWRKYGQKQVKGPNPRSYRCTYQGCNVKKQVQRLSKDEGVVTTYEGMHSHP | 65 |
| rsa6050 | EVDHLEDGYRWRKYGQKQVKGPNPRSYRCTNTKCTVKKRVERSEDPTIVITTYEGQHCHPTI | 65 |
| rsa6058 | EVDHLEDGYRWRKYGQKQVKGPNPRSYRCTYQKCTVKKRVERSFQDPSVITTYEGQHNNHCP | 65 |
| rsa66805 | QLEVMDDGKWRKYGQKQVKGPNPRSYRNYFKCSGGCNVKKRVERDLDPNYVITTYEGMHNHAL- | 65 |
| rsa8054 | DVDVLLDDGKWRKYGQKQVKGPNPRSYRCTHNNCRVKKRVERLEDSDCRMVITTYEGHNNHSP | 65 |
| rsa9342 | EVDHLEDGYRWRKYGQKQVKGPNPRSYRCTNATCNVKKRVERSFNDPSVVVTTYEGQHTHSP | 65 |
| Group 2-d | | |
| AtWRKY17G2-d | IADIPPDEYSWRKYGQKQVKGPNPRSYRCTYKGSHPHPRGYKCSYRGCPCARKHVERALDDSTML | 65 |
| rsa11023 | MADIPPDDYSWRKYGRNPVKGSPHPRGYKCSYRGCPCARKHVERALDDPSMLIVTYEGEHNHSL | 65 |
| rsa12406 | IADIPPDEYSWRKYGQKQVKGSPYPRGYKCSYRGCPCARKHVERATDDPKMLIVTYEGEHRHTQ | 65 |
| rsa12469 | MADIPPDDYSWRKYGQKQVKGSPHPRGYKCSYRGCPCARKHVERALDDPAMLIVTYEGDHNHSH | 65 |
| rsa3792 | LADIPPDEYSWRKYGQKQVKGSPHPRGYKCSYRGCPCARKHVERCLEDPMSMLMVTYEGEHNHPR | 65 |
| rsa8531 | LADIPPDEYSWRKYGQKQVKGSPHPRGYKCSYRGCPCARKHVERCLEDPMSMLMVTYEGEHNHPR | 65 |
| Group 2-e | | |
| AtWRKY14G2-e | GEVVPDLWAWRKYGQKQVKGSPHPRGYRCSYKGSARKQVERSRTDPNMLVITYTSEHNHPW | 65 |
| rsa143 | EVYPPDPSWSWRKYGQKQVKGSPYPRGYRCSYKGCPCARKQVERS RHDPTKLLITYTSEHNHPL | 65 |
| rsa15606 | AEALSSDMWSWRKYGQKQVKGSPYPRGYRCSYKGLARKQVERNRSDPGMFIVTYTAEHNHPM | 65 |
| rsa19677 | MGAPPSDWSAWRKYGQKQVKGSPYPRGYRCSYKGCPCARKQVERS RVDPSMLVITYTSEHNHPW | 65 |
| rsa23071 | NEGPPSDCWSWRKYGQKQVKGSPYPRGYRCSYKGCPCARKQVERC RDTDPVLIITYTSSHNHAG | 65 |
| rsa25734 | NEGPPSDCWSWRKYGQKQVKGSPYPRGYRCSYKGCPCARKQVERC RDTASMLIITYTCSHNHPG | 65 |
| rsa27444 | GEVVPDLWAWRKYGQKQVKGSPYPRGYRCSYKGCPCARKQVERS RDTDPNMLVITYTSEHNHPW | 65 |
| Group 3-a | | |
| AtWRKY30G3-a | VDRTLDDGFSWRKYGQKQVKGAFPRGYRCTYRKSQGCATKQVQRSDENQMLLEISYRGIHSC | 65 |
| rsa9353 | TPDSFGDGHAWRKYGQKQVKGAFPRGYRCTHFKFDGQCQASKQVQKIQDDPPLYRTTYNGQHTC | 65 |
| rsa9788 | TSTLIDDGHAWRKYGQKQVILHADHPRNYFRCTHFKFDGQCQATKQVQMIQDNPPLYRTTYHGRHTC | 65 |
| rsa14878 | LSSTMQDEYAWRKYGQKEILHSKFPRCYFRCTHRNEQGCKATKQVQMIKQDPLVYQTTYFGGHTC | 65 |
| rsa18329 | LEGPTDDGYSWRKYGQKQVILGARYPRSYRCTYRLVQNCWATKQVQRSDNDPTVETIYKGAHTC | 65 |
| rsa1882 | LEGPLDDGHSWRKYGQKQVILGAKFPRAYRCTHRNTQGCLATKQVQRSDSDSIFEVTYKGRHSC | 65 |
| rsa17741 | LEGPLDDGYSWRKYGQKQVILGAKYPR----- | 65 |

图2 蛇根木WRKY蛋白与同类似南芥WRKY蛋白核心域的氨基酸序列比对分析
 Fig.2 Alignment analysis of core WRKY domain amino acid sequences between RsWRKYs
 in *R. serpentina* and AtWRKYs in *Arabidopsis*



图3 蛇根木48个WRKY蛋白中WRKY基序的保守性分析

Fig.3 Analysis of conserved WRKY motif in 48 WRKY proteins of *R. serpentina*

图4 蛇根木48个WRKY蛋白锌指结构基序的保守性分析

Fig.4 Analysis of conserved zinc finger motif in 48 WRKY proteins in *R. serpentina*

*WRKY*与阿吗灵合成关键基因的共表达分析。检索发现,蛇根木基因组中有33个基因参与阿吗灵的生物合成,其中包括2个色氨酸脱羧酶(*rsa2786*和*rsa682*)、8个异胡豆苷合酶(*rsa1050*、*rsa12498*、*rsa16766*、*rsa17326*、*rsa2950*、*rsa30240*、*rsa45750*和*rsa55277*)、2个异胡豆苷糖苷酶(*rsa36278*和*rsa518*)、4个聚精液素醛酯酶(*rsa3656*、*rsa3920*、*rsa5984*和*rsa7002*)、4个维诺任碱合酶(*rsa13551*、*rsa2969*、*rsa33082*和*rsa9123*)、3个霹雳萝芙木碱还原酶(*rsa4803*、*rsa800*和*rsa9784*)、5个萝卡辛糖苷酶(*rsa40125*、*rsa5373*、*rsa77*、*rsa90555*和*rsa9952*)和5个乙酰阿吗碱酯酶(*rsa23962*、*rsa25052*、*rsa27432*、*rsa3992*和*rsa517*)。初步分

析表明,上述33个基因中,有13个基因在根、茎、叶和花中的表达量均较低或不表达(数据未发表)。其余20个关键酶基因根据表达模式的不同可以分为四个类群:第I类群包括2个关键酶基因(*rsa4803*和*rsa9123*),在所有组织中表达量都较高。其中,*rsa4803*在花中的表达量相对较低,和*RsWRKY*基因*rsa9788*的表达模式相似;而*rsa9123*在所有组织表达都很高,和*RsWRKY*第V类群的*rsa475*表达模式一致。第II和第III类群在根、茎、叶和花中都表达,尽管整体表达量较第I类低,但它们在蛇根木主要药用部位根中具有相对较高的表达量,和*RsWRKY*基因中的第II类群表达模式相似。其中*rsa77*在老根中特异性表达,和*RsWRKY*基

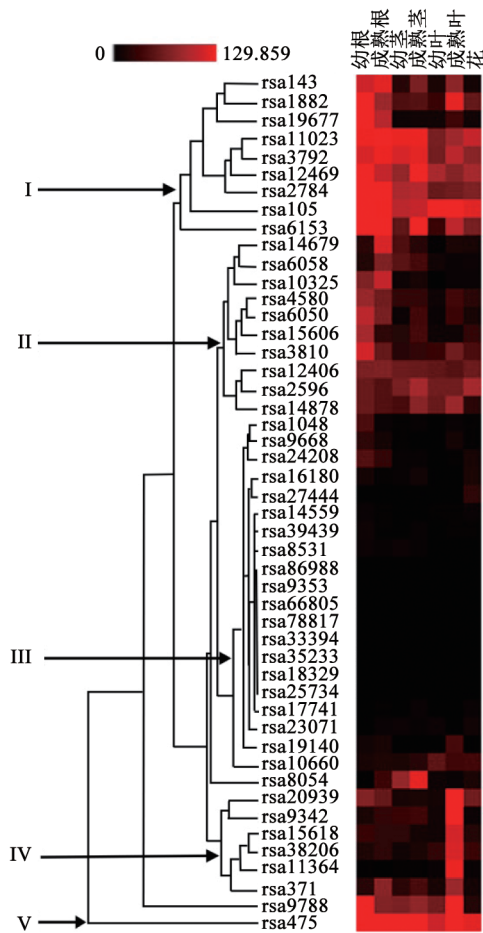


图5 48个*RsWRKY*基因的系统聚类及表达谱
Fig.5 Hierarchical clustering and expression profiles of 48 *RsWRKY* genes

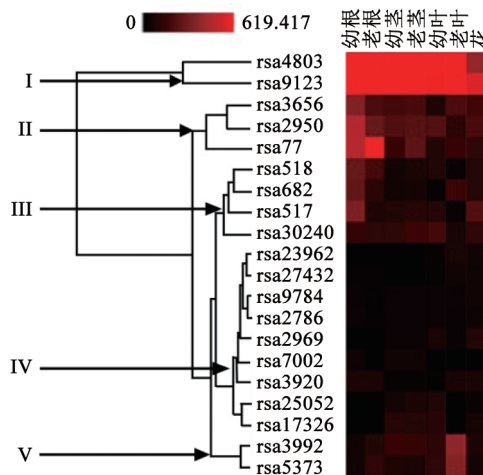


图6 蛇根木中阿吗灵生物合成关键基因的表达谱
Fig.6 Expression profiles of the key genes involved in ajmaline biosynthesis in *R. serpentina*

因第II类群中的rsa10325表达模式相似;第IV类群在根、茎、叶和花等器官中都表达量极低,而且没有明显特异性表达,和*RsWRKY*基因第III类群表达模式相似;第V类群只包括2个基因,尽管在所有组织都有表达,但在成熟叶中表达量较高,这和*RsWRKY*基因第IV类群表达模式相似。

讨论

本研究从34 109个蛇根木单拷贝蛋白序列中鉴定出48个*RsWRKY*,占总蛋白的0.14%。这与拟南芥(Eulgem等2000)、水稻(Wu等2005;孙利军等2014)、玉米(Wei等2012)和西红柿(Huang等2012)等相比,无论是WRKY蛋白总数还是所占的比例都相对较少。其主要原因可能是蛇根木基因组测序尚未完成,一些WRKY还未被发现。其次,蛇根木WRKY基因单拷贝的居多,而上述其它植物中多数WRKY基因发生了复制,导致其大量扩张(杨致荣等2013)。再次,蛇根木中除了已鉴定的48个*RsWRKY*外,还有一些序列包含一段与WRKY域高度相似的序列,但这些序列都不具有WRKY域典型的保守七肽和锌指结构特征,它们是否属于WRKY蛋白还有待于进一步的实验验证(数据未发表)。

依据WRKY保守域的不同,蛇根木中48个*RsWRKY*可以分为三大类(图1、2和表1)。第1类群体包括13个成员,其中具有第1类群体所具有典型的结构特征即含有2个WRKY域的蛋白有8个,其余5个只含有1个WRKY域。除了rsa8698聚类于G1N中,其余4个(rsa35233、rsa16180、rsa28817和rsa33394)聚类于G1C中。但rsa8698序列不完整,需测序重新确定其WRKY域的个数。这些只具有1个WRKY域的第1类蛋白可能源于2个WRKY域家族,而在进化的过程中丢失掉1个WRKY域(Ling等2011)。关于第1类家族中只有1个WRKY域的现象已有不少报道。如在拟南芥中,第1类群体中成员AtWRKY10和AtWRKY19只包括1个WRKY域(Ling等2011),在本文中它们分别聚类于第1类群的C端域群(G1C)和N端域群(G1N)中。单子叶植物中的第1类群体中只有1个WRKY域的现象更普遍,如玉米中第1类中的16个成员只含有1个WRKY域(Wei等2012)。

根据WRKY域的分类,蛇根木*RsWRKY*系统

进化树可分为6个分枝(图1), 这与宋辉和南志标(2014)的研究结果一致。其中G2-a和G2-b在一个分支上, G2-d和G2-e在另一个分支上, 这与Zhang和Wang (2005)对5种植物WRKY基因系统发育关系的研究结果一致。G1C和G2-c分别处于不同的分枝, 但是它们聚类在一个大分枝上, 部分G2-c类并没有聚在G2-c分支上, 而是聚在G1C和G2-c大分枝上, 表明部分G2-c类基因和G1C有共同的起源(宋辉和南志标2014)。G3类蛋白单独聚在一个分支上, 蛇根木G3类WRKY蛋白只有5个, 占总数10%, 而且全部属于G3-a类。这与水稻和拟南芥明显不同(拟南芥有11个, 占总数的20%; 水稻有36个, 占30%), 表明该类蛋白基因的保守性, 而且无扩增现象。这与黄瓜中G3类成员鉴定结果一致(Ling等2011)。而在松树中, 83个PmWRKY蛋白未检测到G3类成员, 说明了G3类WRKY在植物进化过程中具有较大的变异, 不同种属植物中该类群成员的功能有较高的歧化性(杨致荣等2013)。

蛇根木RsWRKY转录因子的分离和鉴定为研究RsWRKY介导的蛇根木次生代谢调控网路奠定了基础。蛇根木药用有效成分主要集中在根、茎和叶中。而蛇根木第I类群和第IV类群RsWRKY基因分别在根和成熟叶片中高表达, 第V类群在根、茎、叶和花等器官表达量都较高(图5), 因此这3个类群是未来研究的重点。RsWRKY转录因子通过调控下游基因的表达参与植物次生产物合成的调控。已鉴定的48个RsWRKY基因中的部分基因表达谱与阿吗灵合成关键基因的表达谱类似, 可能调控了阿吗灵的生物合成, 但这些RsWRKY基因的具体功能尤其是对蛇根木药用有效成分合成途径的调控机制需进一步研究。

参考文献

- 宋辉, 南志标(2014). 蒺藜苜蓿全基因组中蒺藜苜蓿全基因组中WRKY转录因子的鉴定与分析. 遗传, 36 (2): 152~168
- 孙利军, 黄磊, 李大勇, 张慧娟, 宋凤鸣(2014). 水稻OsWRKY转录因子对非生物胁迫响应的重叠表达特性分析. 植物生理学报, (50) 11: 1651~1658
- 杨致荣, 王兴春, 薛金爱, 孟令芝, 李润植(2013). 药用植物长春花WRKY转录因子的鉴定及表达谱分析. 生物工程学报, 29 (6): 785~802
- Bayer A, Ma X, Stockigt J (2004). Acetyltransfer in natural product biosynthesis functional cloning and molecular analysis of vinorine synthase. Bioorg Med Chem, 12: 2787~2795
- Bracher D, Kutchan TM (1992). Strictosidine synthase from *Rauvolfia serpentina*: analysis of a gene involved in indole alkaloid biosynthesis. Arch Biochem Biophys, 294 (2): 717~723
- Brand LH, Fischer NM, Harter K, Kohlbacher O, Wanke D (2013). Elucidating the evolutionary conserved DNA-binding specificities of WRKY transcription factors by molecular dynamics and *in vitro* binding assays. Nucleic Acids Res, 41 (21): 9764~9778
- Dogru E, Warzecha H, Seibel F, Haebel S, Lottspeich F, Stockigt J (2000). The gene encoding polyneuridine aldehyde esterase of monoterpenoid indole alkaloid biosynthesis in plants is an ortholog of the alpha/beta-hydroxylase super family. Eur J Biochem, 267: 1397~1406
- Dou L, Zhang X, Pang C, Song M, Wei H, Fan S, Yu S (2014). Genome-wide analysis of the WRKY gene family in cotton. Mol Genet Genomics, 289: 1103~1121
- Eulgem T, Rushton PJ, Robatzek S, Somssich IE (2000). The WRKY superfamily of plant transcription factors. Trends Plant Sci, 5 (5): 199~206
- Huang S, Gao Y, Liu J, Peng X, Niu X, Fei Z, Cao S, Liu Y (2012). Genome-wide analysis of WRKY transcription factors in *Solanum lycopersicum*. Mol Genet Genomics, 287: 495~513
- Kutchan TM, Hampp N, Lottspeich F, Beyreuther K, Zenk MH (1988). The cDNA clone for strictosidine synthase from *Rauvolfia serpentina*. DNA sequence determination and expression in *Escherichia coli*. FEBS Lett, 237 (1-2): 40~44
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R et al (2007). Clustal W and Clustal X version 2.0. Bioinformatics, 23 (21): 2947~2948
- Li HL, Zhang LB, Guo D, Li CZ, Peng SQ (2012). Identification and expression profiles of the WRKY transcription factor family in *Ricinus communis*. Gene, 503: 248~253
- Ling J, Jiang W, Zhang Y, Yu H, Mao Z, Gu X, Huang S, Xie B (2011). Genome-wide analysis of WRKY gene family in *Cucumis sativus*. BMC Genomics, 12: 471
- Srivastava A, Tripathi AK, Pandey R, Verma RK, Gupta MM (2006). Quantitative determination of reserpine, ajmaline, and ajmalicine in *Rauvolfia serpentina* by reversed-phase high-performance liquid chromatography. J Chromatogr Sci, 44: 557~560
- Sun L, Ruppert M, Sheludko Y, Warzecha H, Zhao Y, Stockigt J (2008). Purification, cloning, functional expression and characterization of perakine reductase: the first example from the AKR enzyme family, extending the alkaloidal network of the plant *Rauvolfia*. Plant Mol Biol, 67: 455~467
- Suttipanta N, Pattanaik S, Kulshrestha M, Patra B, Singh S, Yuan L (2011). The transcription factor CrWRKY1 positively regulates the terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. Plant Physiol, 157: 2081~2093
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol, 28 (10): 2731~2739
- Wang L, Zhu W, Fang L, Sun X, Su L, Liang Z, Wang N, Londo JP, Li S, Xin H (2014). Genome-wide identification of WRKY family

- genes and their response to cold stress in *Vitis vinifera*. *BMC Plant Biol*, 14: 103
- Warzecha H, Gerasimenko I, Kutchan TM, Stockigt J (2000). Molecular cloning and functional bacterial expression of a plant glucosidase specifically involved in alkaloid biosynthesis. *Phytochemistry*, 54: 657~666
- Wei KF, Chen J, Chen YF, Wu LJ, Xie DX (2012). Molecular phylogenetic and expression analysis of the complete WRKY transcription factor family in maize. *DNA Res*, 19 (2): 153~164
- Wen F, Zhu H, Li P, Jiang M, Mao W, Ong C, Chu Z (2014). Genome-wide evolutionary characterization and expression analyses of WRKY family genes in *Brachypodium distachyon*. *DNA Res*, 21 (3): 327~339
- Wu KL, Guo ZJ, Wang HH, Li J (2005). The WRKY family of transcription factors in rice and Arabidopsis and their origins. *DNA Res*, 12 (1): 9~26
- Yang Z, Patra B, Li R, Pattanaik S, Yuan L (2013). Promoter analysis reveals cis-regulatory motifs associated with the expression of the WRKY transcription factor CrWRKY1 in *Catharanthus roseus*. *Planta*, 238: 1039~1049
- Yu J, Chen Q (2014). Antitumor activities of *Rauwolfia vomitoria* extract and potentiation of gemcitabine effects against pancreatic cancer. *Integr Cancer Ther*, 13 (3): 217~225
- Zhang Y, Wang L (2005). The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. *BMC Evol Biol*, 5: 1