

## 研究报告

## Research Report

# 豆科植物生物钟核心调控回路相关基因的鉴定

李宗飞<sup>1,2,3,4</sup>, 蔡梦蝶<sup>1,2</sup>, 刘振鹏<sup>1,2</sup>, 张洁<sup>1,2</sup>, 魏芳<sup>1,2</sup>, 方宣钧<sup>1,2,3,4</sup>

1 浙江农林大学暨阳学院生命科学研究部, 诸暨, 311800

2 翠溪生物技术研究院, 诸暨, 311800

3 海南省热带农业资源开发利用研究所, 三亚, 572025

4 广西大学生命科学与技术学院, 南宁, 530005

✉ 通讯作者: jim.xj.fang@hitar.org; ✉ 作者

豆科基因组学与遗传学, 2017 年, 第 8 卷, 第 1 篇 doi: 10.5376/lgg.cn.2017.08.0001

收稿日期: 2016 年 12 月 25 日

接受日期: 2017 年 01 月 17 日

发表日期: 2017 年 02 月 23 日

© 2017 BioPublisher 生命科学中文期刊出版平台

本文首次以英文发表在《Legume Genomics and Genetics》上。现依据版权所有人授权的许可协议, 采用 Creative Commons Attribution License 协议对其进行授权, 用中文再次发表与传播。只要对原作有恰当的引用, 版权所有人允许并同意第三方无条件的使用与传播。

引用格式:

Li et al., 2017, Identification of genes related to the key regulatory circuit of circadian clock in legume plant, Legume Genomics and Genetics, 08(1): 1-9 (doi: 10.5376/lgg.2017.08.0001)

**摘要** 为了揭示生物钟调控网络机制在豆科的 3 个物种(藜藜苜蓿, 百脉根和大豆)中的演变规律。本研究通过 Geneontology (GO)查找拟南芥生物钟的调控因子, 同时根据相关的文献进行补充和佐证。然后将得到的调控因子在全基因组水平上与藜藜苜蓿, 百脉根和大豆这三个物种进行比对, 进行相应的直系同源基因的鉴定、保守结构功能域的分析、多序列联配、系统进化树构建和进化分析以及基因的重新注释。本研究一共鉴定了 108 个大豆的生物钟相关候选基因, 51 个百脉根的生物钟相关候选基因, 34 个藜藜苜蓿的生物钟相关候选基因。我们从鉴定的直系同源基因的候选基因中选取 LHY/CCA1、PRR 和 CHE 基因构建了 3 个分子进化树, 分析表明, 生物钟调控途径在拟南芥和豆科的这 3 个物种中发生了不同程度的分化, 尤其是在一些关键性的成分比如生物钟调控的核心成分以及控制花期的关键性基因上。对以后研究豆科生物钟调控网络在实验阶段时的研究有一定的指导和参考价值。

**关键词** 豆科, 生物钟, 调控, 基因, 鉴定

## Identification of Genes Related to the Key Regulatory Circuit of Circadian Clock in Legume Plant

Li Zongfei<sup>1,2,3,4</sup>, Cai Mengdie<sup>1,2</sup>, Liu Zhenpeng<sup>1,2</sup>, Zhang Jie<sup>1,2</sup>, Wei Fang<sup>1,2</sup>, Fang Xuanjun<sup>1,2,3,4</sup>

1 Institute of Life Science, Jiyang College of Zhejiang A&F University, Zhuji, 311800

2 Cuixi Academy of Biotechnology, Zhuji, 311800

3 Hainan Institute of Tropical Agricultural Resources, Sanya, 572025

4 College of Life Sciences and Technology, Guangxi University, Nanning, 530005

✉ Corresponding author, jim.xj.fang@hitar.org; ✉ Authors

**Abstract** In order to reveal the evolution of the circadian clock regulation network in legume of three species (*Medicago truncatula*, *Lotus japonicus* and *Glycine max*). We find the regulation factor of the circadian clock of *Arabidopsis thaliana* through the Geneontology (GO), at the same time, according to the relevant literature, the regulation factor was supplemented and verified. Then regulation factor was obtained at the genome-wide level and *Medicago truncatula*, *Lotus japonicus* and *Glycine max* in the three species for comparison, identification of the corresponding direct homologous gene, analysis of the conserved functional domains, multiple sequence alignment, phylogenetic tree construction and evolution analysis and gene re-annotation. This study identified 108 *Glycine max* circadian clock related candidate genes, 51 *Lotus japonicus* circadian clock related candidate genes, 34 *Medicago truncatula* circadian clock related candidate genes. We selected LHY/CCA1, PRR and CHE genes from the candidate genes of the identified orthologous genes to construct 3 molecular phylogenetic trees, analysis showed that the circadian clock

regulation pathway in Arabidopsis and legume of the three species occurred different degree of differentiation, especially in some key components such as direct homologous gene of the circadian clock regulates and the key genes of the florescence control. It has certain guidance and reference value for the researching of the circadian clock control network of the legume in the experimental stage.

**Keywords** Legume, Circadian clock, Regulation, Gene, Identification

生物钟控制着植物体内大部分的生理活动, 相关的微阵列表达研究表明, 在拟南芥中, 有超过三分之一的基因的表达受到内在的生物钟机制的调控 (Barak et al., 2000; Covington et al., 2008), 而这一估计还是趋于保守的。植物叶片的运动, 气孔的闭合, 胚轴的延伸, 它可以协调各种生理活动, 使得体内的包括生长发育在内的各种生理活动能够在适当的时机进行 (Harmer et al., 2000; McClung et al., 2001)。以往对生物钟调控网络的研究主要集中在模式植物拟南芥上, 在其他植物中相关的研究还比较少见。在拟南芥已经鉴定过多个生物钟调控相关的基因, 并且揭示了一些生物钟调控生理途径的机制 (李宗飞等, 2015)。

蒺藜苜蓿、百脉根和大豆已先后完成了全基因组测序, 可以为我们的分析提供较为完整的基因组数据。在模式植物拟南芥中, 我们主要通过 GO 查找拟南芥生物钟的调控因子, 同时根据相关的文献进行补充和佐证。然后将得到的调控因子在全基因组水平上与蒺藜苜蓿, 百脉根和大豆这三个物种进行比对, 进行相应的生物信息学分析。

## 1 结果与分析

### 1.1 序列联配确定物种间的同源序列

通过检索模式植物拟南芥的生物钟, 生物钟调控, 生物钟输入等相关 GO 术语而得到 52 个基因, 其中有些具有多条上述 GO 注释, 去除多重注释的基因, 得到 42 个生物钟调控网络相关的基因, 再加上我们从其他文献中得到的 8 个基因, 我们在拟南芥中一共得到了 50 个生物钟相关的基因。这 50 个基因基本涵盖了拟南芥的生物钟调控网络的 3 个部分, 包括光信号输入的生物钟途径, 中央振荡器组成, 以及生物钟调控的一些生理活动。

我们将拟南芥全部基因的蛋白质序列和豆科的 3 个物种在全基因组水平上分别进行序列相似性比较 (表 1), 其中“Orthologs”项表示拟南芥中生物钟调控网络相关的基因在各个物种中直系同源候选基因的数目, “全长 ORF”项则表示有全长 ORF 候选基因的数目, “F-PUT”项则表示“全长 ORF”项中那些全长被 PUT 序列覆盖的基因数目。在大豆中一共鉴定了 108 个生物钟调控网络相关的候选基因, 其中

有全长 ORF 的有 95 个, 而有全长 PUT 序列配比的有 66 个。在蒺藜苜蓿中一共鉴定了 34 个生物钟调控网络相关的候选基因, 其中有全长 ORF 的有 33 个, 但是, 拥有全长 ORF 的候选基因中, 只有 7

个基因的全长有 PUT 序列配比。在百脉根中一共鉴定了 51 个生物钟调控网络相关的候选基因, 其中有全长 ORF 的有 25 个, 而其中全长 PUT 序列配比的也只有 7 个。

从总的情况来看, 在豆科这 3 个物种中, 拟南芥中生物钟调控网络相关的基因在大豆中对应的候选基因数目最多, 这可能与大豆的多倍体起源有关, 同时, 大豆中拥有全长 PUT 序列覆盖的候选基因的数目也要远远多于蒺藜苜蓿和百脉根中相应基因的数目。这可能是大豆中 EST 数据库中 EST 资源比蒺藜苜蓿和百脉根 EST 数据库中多造成的。

同时, *PRR9*, *PHYC*, *PHYD*, *LKP2*, *FSD1* (fe superoxide dismutase 1), *KAT2* (potassium channel in *Arabidopsis thaliana* 2), *LCL*, *FLC* (flowering locus c), *ARR4* (response regulator 4), *CHE*, *CO*, *COL1* (CONSTANS-LIKE 1) 这 12 个拟南芥中生物钟调控网络的相关基因在豆科的 3 个物种中均没有找到直系同源候选基因。对于这 12 个基因而言, 它们在豆科的直系同源基因已经在豆科的长期进化过程中丢失了, 因此, 这 12 个基因可能是造成豆科和十字花科分化的重要事件。

### 1.2 豆科中 *LHY/CCA1* 直系同源候选基因的鉴定

*LHY* 和 *CCA1* 是一类 R1 类 Myb 类型转录因子, 在白天达到表达的最高峰, 是生物钟中央振荡器的负调控因子, 其表达受到 *TOC1* 的正向调控。在夜晚, *LHY/CCA1* 受到 *PIF3* (phytochrome interacting factor 3) 的抑制, *PIF3* (McWatters et al., 2000) 与 *CCA1* 和 *LHY* 等基因启动子区的 G-box 结合, 进而抑制这些基因的表达; 而经红光照射后, *PHYB* 被活化并从细胞质进入细胞核中与 *PIF3* 结合, 解除 *PIF3* 对 *CCA1* 和 *LHY* 等基因的抑制; 并进一步诱导下游日间基因的表达并抑制夜间基因的表达。*LHY* 与 *TOC1* 等夜间基因的启动子的黑夜元件 (AAATATCT) 相结合而抑制其的转录表达。在白天, *CCA1/LHY* 表达水平的降低可以逐渐解除其对 *TOC1* 的抑制作用, 使得 *TOC1* 的 mRNA 表达量的增加。当 *TOC1* 的表达量达到一定量的时候, 又会正向调控 *CCA1* 和 *LHY* 的表达, 使得 *CCA1* 和 *LHY* 的表达量增加, 从而形成了一个振荡的生理循环。所以, 在拟南芥中由 *CCA1/LHY* 和 *TOC1* 构成的调控途径中, *TOC1* 首先正向调控 *LHY/CCA1* 的表达, 当 *LHY/CCA1* 达到一定的表达量, 又会负向抑制 *TOC1* 的表达。按照这个模型, MYB 转录因子 *CCA1*

表1 三个豆科植物中的50条候选基因的比对

Table 1 Comparison of 50 candidate genes from the three leguminous plants

物种 Species	Orthologs	全长 ORF	F-PUT
蒺藜苜蓿 <i>Medicago truncatula</i>	34	33	7
百脉根 <i>Lotus japonicus</i>	51	25	7
大豆 <i>Glycine max</i>	108	95	66

表2 LHY在豆科3个物种中的直系同源候选基因

Table 2 LHY in the direct candidate ortholog genes in three species of Legume

物种 Species	基因名 Gene name	基因号 ID No of gene	是否全长 If the gene is Full-Length	相似性 Similarity	EST 证据 evidence of EST
蒺藜苜蓿 <i>Medicago truncatula</i>	<i>MtLHY</i>	AC150443_50	否 No	35%	
百脉根 <i>Lotus japonicus</i>	<i>LjLHY</i>	chr3.CM0208.430.nd	否 No	24.7%	
大豆 <i>Glycine max</i>	<i>GmLHY1</i>	Glyma07g05410	是 Yes	42%	F
	<i>GmLHY2</i>	Glyma16g01980	是 Yes	42%	F
	<i>GmLHY3</i>	Glyma19g45030	是 Yes	46%	F
	<i>GmLHY4</i>	Glyma03g42260	是 Yes	44%	F

和 LHY 是作为转录调控环的负调控元件可以直接结合到 *TOC1*的启动子区域, 而 *TOC1*对 *CCA1*和 *LHY* 表达的正向调控不是直接的。

*LHY*在百脉根根和蒺藜苜蓿中都找到位于同一染色体上邻近的2个直系同源基因片段, 在大豆中找到了4个直系同源基因(表2)。在百脉根中, 有2个预测的基因片段可分别是 *chr3.CM0208.430.nd* 和 *chr3.CM0208.410.nd*, 在蒺藜苜蓿中, 这两个预测的基因片段分别是 *AC150443\_50*和 *AC150443\_49*。百脉根中的2个基因片段之间相差蒺藜苜蓿中2个基因片段, 但是这2个基因片段却无法拼接成一个完整基因, 通过与 *LHY* 基因的比对可看出, *chr3.CM0208.410.nd*和 *AC150443\_49*对应着 *LHY* 基因的 HTH 保守结构域, 而 *chr3.CM0208.430.nd* 和 *AC150443\_50*则对应着 *LHY* 基因的其余部分。出现这种情况, 可能有2个原因, 一个是基因组信息不全, 导致注释错误地将1个基因分成2个基因, 另外一种可能性就是在苜蓿属中, 在 *LHY* 基因的 HTH 保守结构域与其余部分插入了某个片段, 而导致了苜蓿属中 *LHY* 基因的功能丧失。我们将这4个基因片段分别去比对百脉根和蒺藜苜蓿中的 EST 库, 结果显示, *AC150443\_49*和 *AC150443\_50*的大部分均有 EST 对应, 而 *chr3.CM0208.410.nd* 则没有找到相应的 EST, 而 *chr3.CM0208.430.nd* 只有部分区域有 EST 与之对应。我们将 *chr3.CM0208.430.nd* 和 *AC150443\_50*分别作为百脉根和蒺藜苜蓿 *LHY* 基因的代表, 构建进化树, 但是, *chr3.CM0208.430.nd* 和 *AC150443\_50*在进化树上的位置只能作为一种参考。

*LHY* 在大豆中的4个直系同源候选基因均是全长, 与 *LHY* 的相似性为42%或者42%以上, 并且这4

个基因的全长中都有 PUT 序列覆盖, 不过 *GmLHY1* 和 *GmLHY2*对应的是同一个 PUT 序列, *GmLHY3*和 *GmLHY4*对应的是同一个 PUT 序列, 并且 *GmLHY2* 与 PUT 序列的相似性要大于 *GmLHY1*与 PUT 序列的相似性, *GmLHY3*要大于 *GmLHY4*。所以在大豆中至少有2个 *GmLHY* 基因的表达。

*LHY*基因与 *CCA1*同属于 R1类 MYB 转录因子, 并且互为旁系同源基因。LHY 蛋白的 HTH 保守结构域非常明显。值得注意的是, 在 NCBI 的 CDD 结构域分析将 LHY 蛋白该保守区域命名为 SNAT 结构域。而 Pfam 数据库的比对结果表明该结构域为 HTH 保守结构域, 所以, 综合以上信息, 本研究将该结构域命名为 HTH。HTH 结构域和 SNAT 结构域同属于 MYB 转录因子家族的保守结构域, 不同的是, HTH 结构域结合的是 DNA 的特定区域, 而 SNAT 结构域是蛋白质相互作用的识别区域。根据 Pfam 中的比对和图1结果, 在 LHY 蛋白中, 除了 HTH 这个高度保守性的区域外, 从拟南芥 LHY 蛋白对应的93到473的位点也具有相对保守性, 对应着 Pfam B 类数据库中的120568结构。从氨基酸序列比对图可以看出, HTH 的结构域对于维持 LHY 的功能是非常重要的。即使是在单子叶植物和双子叶植物分离之前, 已经和 *LHY* 基因分离的拟南芥的 *CCA1*基因在这一区域也是非常保守的。蒺藜苜蓿的情况比较特殊, 它的注释信息中 *LHY* 基因的 HTH 结构域的缺失的原因目前还不清楚, 有可能是注释信息错误, 也有可能是本身 *LHY* 发生了某种形式的突变而导致了 *LHY* 基因的失活。

拟南芥中的 *CCA1*基因与 *LHY* 基因发生分离的重要事件就是在330位点到430位点之间, *CCA1*基因发生了4小段氨基酸的缺失。而 *LHY* 基因在十字花

科和豆科分离的重要事件就是在190位点到290位点之间也发生了4小段氨基酸的缺失。在大豆中,从图1中可以得知, *GmLHY1*和 *GmLHY2*聚为一类, *GmLHY3*和 *GmLHY4*聚为一类。与 *GmTOC1*的相关的数据相比, *GmLHY* 基因显得更加保守。这也可以从2个同源序列的联配图也可以看出来, *LHY* 基因在同源基因的变化要比 *TOC1*小得多。从氨基酸

需类比对图中可以看出, *GmLHY1*, *GmLHY2*的分离发生的主要事件在70位点到90位点之间通过某种进化机制引入了20个氨基酸, 这个区域位于 HTH 保守结构域内, 应该是一个比较重大的进化事件, 其结果目前尚不得而知。有趣的是, 这与 *GmTOC1* 与 *GmTOC2*的分离情形类似。这也许说明了在相同生理途径中不同基因的趋同进化。

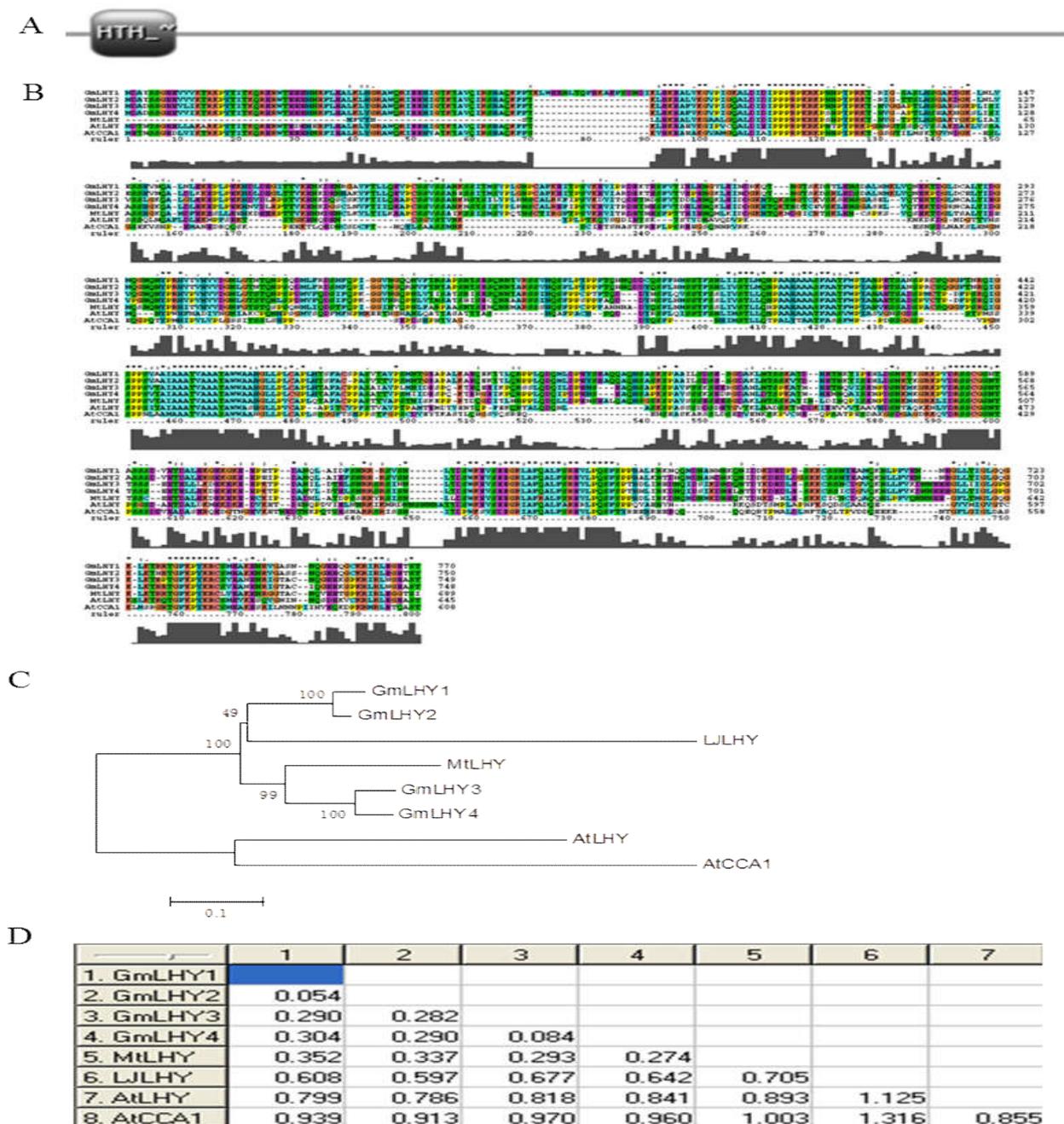


图1 LHY的保守结构域, 多序列联配图, 分子进化树和进化距离

注: A: LHY保守结构域分析; B: LHY多序列联配; C: LHY蛋白的分子进化树; D: 4个物种之间LHY蛋白的进化距离

Figure 1 The LHY conservation domain, sequences alignment figures, molecular phylogenetic tree and evolutionary distance

Note: A: LHY conservation domain analysis; B: LHY sequences alignment; C: Molecular evolution tree of LHY proteins; D: The distance of LHY protein between four species

表 3 PRR 基因家族在豆科 3 个物种中的直系同源候选基因

Table 3 The direct candidate ortholog genes of PRR gene family in three species of Legume

物种 Species	基因名 Gene name	基因号 ID No of gene	是否全长 If the gene is Full-Length	相似性 Similarity	EST 证据 evidence of EST
蒺藜苜蓿 <i>Medicago truncatula</i>	<i>MtPRR2</i>	AC144516_23	是 Yes	46.8%	E
	<i>MtPRR5a</i>	AC150443_48	是 Yes	26.0%	E
	<i>MtPRR5b</i>	CR940305_20	是 Yes	30.7%	E
	<i>MtPRR7</i>	AC149306_58	是 Yes	36.6%	E
百脉根 <i>Lotus japonicus</i>	<i>LjPRR1</i>	chr4_CM0087.600.nc	否 No	46.0%	E
	<i>LjPRR2a</i>	chr2_CM0021_770_nc	是 Yes	44.5%	N
	<i>LjPRR2b</i>	chr4_CM0831_100_nd	是 Yes	44.3%	N
	<i>LjPRR5a</i>	chr1_CM0105_560_nc	是 Yes	25.7%	N
	<i>LjPRR5b</i>	LjSGA_092681_1	否 No		N
	<i>LjPRR7a</i>	chr3_LjT05P05_60_nd	否 No	37.9%	N
	<i>LjPRR7b</i>	chr5_LjT08O17_130_nd	否 No	16.3%	N
大豆 <i>Glycine max</i>	<i>GmPRR1a</i>	Glyma04g33110	是 Yes	47%	F
	<i>GmPRR1b</i>	Glyma06g21120	是 Yes	47%	F
	<i>GmPRR1c</i>	Glyma05g00880	是 Yes	47%	E
	<i>GmPRR1d</i>	Glyma17g11040	是 Yes	56%	E
	<i>GmPRR2a</i>	Glyma01g40900	是 Yes	44.9%	E
	<i>GmPRR2b</i>	Glyma17g16360	是 Yes	46.0%	E
	<i>GmPRR2c</i>	Glyma05g06070	是 Yes	47.3%	E
	<i>GmPRR2d</i>	Glyma11g04440	是 Yes	32.9%	F
	<i>GmPRR3</i>	Glyma11g15580	是 Yes	25.1%	F
	<i>GmPRR5a</i>	Glyma04g40640	是 Yes	32.2%	F
	<i>GmPRR5b</i>	Glyma06g14150	是 Yes	34.4%	E
	<i>GmPRR5c</i>	Glyma16g02050	是 Yes	24.2%	F
	<i>GmPRR5d</i>	Glyma07g05530	是 Yes	24.1%	
	<i>GmPRR5e</i>	Glyma19g44970	是 Yes	24.2%	
	<i>GmPRR7a</i>	Glyma10g05520	否 No	40.5%	F
	<i>GmPRR7b</i>	Glyma12g07860	是 Yes	29.7%	
	<i>GmPRR7c</i>	Glyma13g19870	是 Yes	18.8%	

### 1.3 PRR家族候选基因的鉴定

拟南芥的 PRR (pseudo-response regulator, 伪反应调控因子) 家族与生物钟有着密切的联系, PRR1(又名 TOC1), 3, 5, 7, 9 被称为生物钟五重奏, 在控制生物钟的节奏方面有着重要的作用, 此外, PRR2 在拟南芥的生物钟也起着一定的作用。PRR9 在豆科中没有找到直系同源基因, 而 PRR5 在豆科中的同源基较多, 并且根据构建的分子进化树, 蒺藜苜蓿和大豆的部分 PRR5 候选基因对应着拟南芥没有的基因(表 3)。对于 PRR 基因数目的变异, 是否由拟南芥中的五重奏在其他植物中变为三重奏、七重奏, 还需要更多的研究。在拟南芥的 PRR 家族中, PRR1 和 PRR2、PRR3、PRR5、PRR7、PRR9 的亲缘关系较远, 结果表明, PRR 基因对生物钟的周期, 某个阶段以及幅度, 开花时间以及对红光控制的下胚轴的延伸的敏感性都有影响。PRR 转录物有节律性的积累, 顺序为 PRR9-PRR7-PRR5-PRR3-TOC1。

其表达最高峰出现在黄昏后 2 个小时到 10 个小时从 PRR9 到 TOC1。研究表明 PRR 参与了生物钟生理途径中的光信号输入途径但是却并不是产生生物钟的节律性振荡所必须的。拟南芥中的 PRR7 和 PRR9 基因也在温度介导的生物钟系统中发挥作用, 并且 2 者的功能有重叠。

从保守结构域分析和同源序列联配示意图我们可以看出, 植物 PRR 类基因有 2 个比较保守的区域, 分别是 REC 区域和 CCT 区域(图 2)。REC 区域由 2 个部分组成, 一个是组氨酸蛋白激酶部分, 包含磷酸化位点, 另一个是蛋白应答调节部分, 包括分子间相互作用的识别位点和可能在同源基因或者自身形成二聚化的位点。CCT 区域存在于 PRR 的 C-端, 负责传递光信号, 与 B-BOX 型锌指结构, GATA 型锌指结构, 以及 REC 结构发生作用以调控植物的光信号传导途径。所以, CCT 保守区域也存在于光信号途径的其他蛋白中, 比如 CO, COL 蛋白。

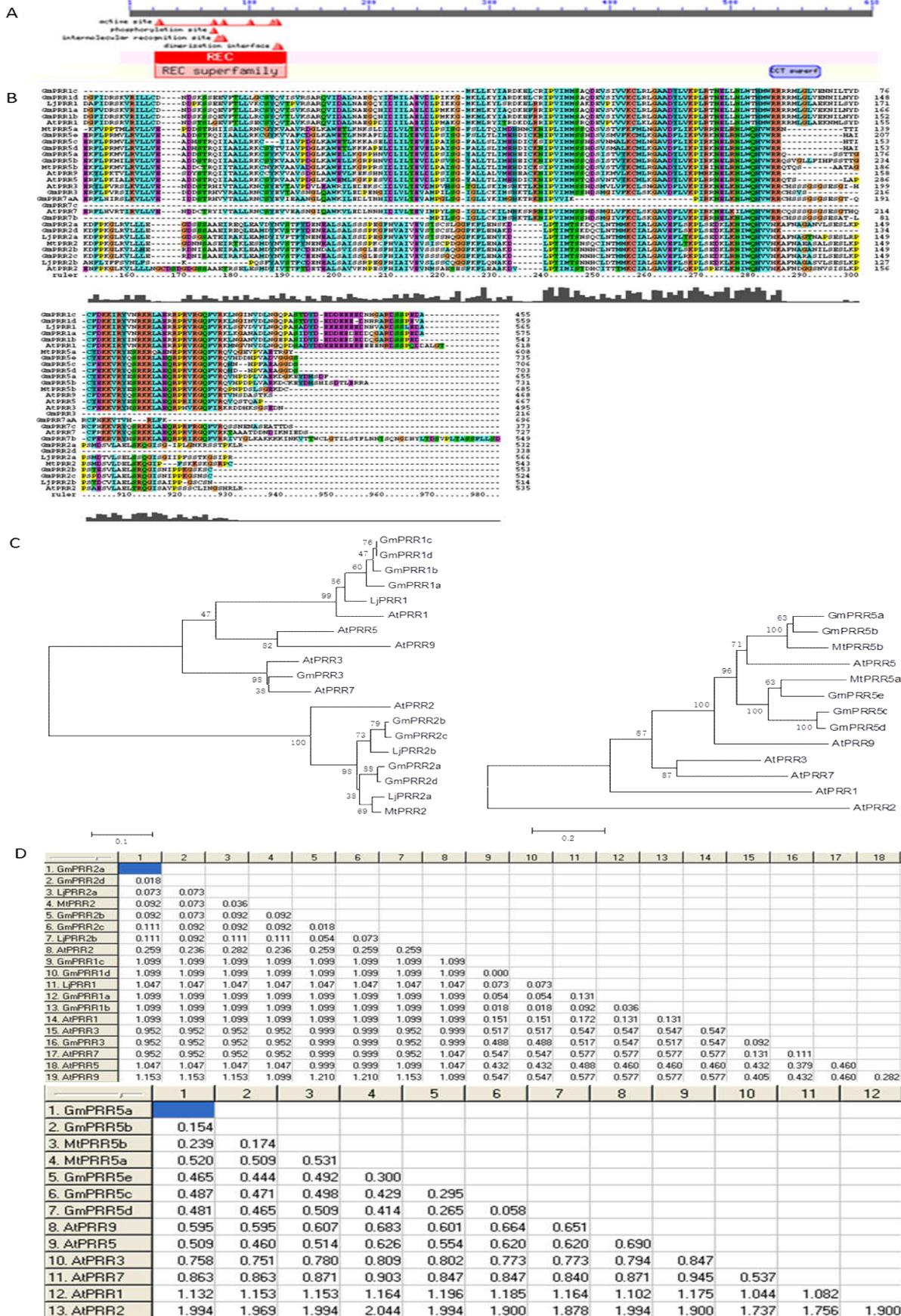


图2 PRR蛋白的保守功能结构域分析, 氨基酸序列比对, 分子进化树和进化距离

注: A: PRR蛋白保守结构域分析; B: PRR氨基酸序列比对; C: PRR蛋白的分子进化树; D: 4个物种之间PRR蛋白的进化距离  
Figure 2 Conserved functional domain analysis, amino acid alignment, molecular evolution tree and evolutionary distance of PRR proteins

Note: A: Conserved functional domain analysis of PRR protein; B: PRR amino acid sequences alignment; C: Molecular evolution tree of PRR proteins; D: The distance of PRR protein between four species

PRR1和PRR5在蒺藜苜蓿中没有找到直系同源候选基因, PRR2, PRR7在蒺藜苜蓿中各有一个直系同源基因, PRR2在蒺藜苜蓿中有2个直系同源候选基因。这4个直系同源候选基因在蒺藜苜蓿中均有部分表达。

PRR1在百脉根有1个直系同源候选基因, 而PRR2, PRR5, PRR7在百脉根中均有2个直系同源候选基因, 其中只有PRR1在百脉根中有部分表达。

PRR1, PRR2在大豆中均有4个直系同源候选基因, PRR5在大豆中有5个直系同源候选基因, PRR7有3个直系同源候选基因, PRR3在大豆中的直系同源候选基因则亲缘关系较远。

在大豆的4个PRR1基因中, *GmPRR1a*和*GmPRR1b*聚为一类, *GmPRR1c*和*GmPRR1d*聚为一类。*GmPRR1c*的一个显著的特点就是在REC保守结构域中丢失了部分活化位点和磷酸化位点, 这种进化事件发生在*GmPRR1c*和*GmPRR1d*分离时。这种结构上的突变有可能使*GmPRR1c*失去生物活性功能而突变为假基因, 也有可能是另外一种进化的结果, *GmPRR1c*和*GmPRR1d*的祖先基因在复制后, 发生亚功能化(Subfunctional), 使得*GmPRR1c*丧失部分活性, 但是还保留了与其他分子的作用的识别位点和二聚化位点, 可能与其他的*GmPRR1*基因形成二聚体, 起辅助作用。从表3中可看出*GmPRR1c*和*GmPRR1d*的进化距离(0.096)远大于*GmPRR1a*和*GmPRR1b*的进化距离(0.052), 表明*GmPRR1c*和*GmPRR1d*分离的时间远早于*GmPRR1a*和*GmPRR1b*分离的时间, 这就意味着*GmPRR1a*和*GmPRR1b*所面临的选择压力大于*GmPRR1c*和*GmPRR1d*所面临的选择压力。这也可以从同源序列的联配(图2)比对上看出来, *GmPRR1c*和*GmPRR1d*的结构域发生几处缺失, 说明他们面临的选择压力少。

#### 1.4 CHE家族候选基因的鉴定

CHE是一种转录因子TCP家族的成员。以往的研究中虽然发现了当CCA1和LHY积累到一定程度时, 就会通过直接结合到TOC1的启动子区域来抑制TOC1的表达, 但是与之相反, 在TOC1蛋白中, 却没有发现相关的功能结构域来调控LHY和CCA1的表达(图3)。为了解决这个问题, Prunedo-Paz等(2009)用酵母单杂交技术来筛选与LHY和CCA1基因调控区域结合的转录因子, 结果

发现TCP转录因子I类家族专一的结合到CCA1基因起始密码子前面的一段长约171bp的区域, 命名为CHE, 但是CHE蛋白却不结合到LHY基因的相关区域。CHE蛋白结合CCA1基因区域的序列为GGNCCCAC, 命名为TBS, 进一步的研究表明, 在白天, CCA1和LHY的高表达促进了CHE的表达, 而在夜晚, CHE的表达量不断增加, 从而将CCA1和LHY的表达量抑制到最低。在白天结束的时候, TOC1会通过结合CHE, 来重新开始下一个循环。

CHE基因在豆科的3个物种中均没有找到CHE的直系同源候选基因。蒺藜苜蓿、百脉根和大豆与CHE亲缘关系最近的旁系同源基因分别为AC151824\_21、chr5.CM1439.250.Nd和Glyma10g43190, 它们在拟南芥中的直系同源均为AT5G23280。这表明有可能在豆科中AT5G23280可能代替了CHE来调控CCA1的表达, 或者是在豆科中, AT5G23280调控LHY的表达。



图3 CHE蛋白保守结构域分析

Figure 3 Conserved domain analysis of CHE protein

## 2 讨论

本研究采用生物信息学的方法对豆科中的生物钟调控网络进行了分析。利用拟南芥中Geneontology(GO)注释和最新的文献获得了50个构成生物钟调控网络关键基因, 涉及光照和温度等生物钟输入部分, 中央振荡器以及生物钟输出等方面。我们一共鉴定了108个大豆的生物钟相关候选基因, 51个百脉根的生物钟相关候选基因, 34个蒺藜苜蓿的生物钟相关候选基因。

我们从鉴定的直系同源基因的候选基因中选取LHY/CCA1、PRR和CHE基因构建了3个分子进化树, 并结合功能结构域和多序列联配的结果, 对这些基因的功能和进化趋势进行了分析。我们的分析表明, 生物钟调控途径在拟南芥和豆科的这3个物种中发生了不同程度的分化, 尤其是在一些关键性的成分比如生物钟调控的核心成分以及控制花期的关键性基因上。而以拟南芥为参照, 百脉根和蒺藜苜蓿的生物钟调控网络相关基因的分化程度要大于大豆, 这可能是由于百脉根和蒺藜苜蓿的基因组远小于大豆。

表 4 4 个物种进行生物信息学分析相应的数据库和网址

Table 4 Bioinformatics analyse the four species' corresponding database and website name and website

物种 Species	网站名称 Site Name	网址 Website
拟南芥 <i>Arabidopsis</i>	TAIR	<a href="http://www.arabidopsis.org/">http://www.arabidopsis.org/</a>
蒺藜苜蓿 <i>Medicago truncatula</i>	MtGDB UrMeLDB	<a href="http://www.plantgdb.org/MtGDB/">http://www.plantgdb.org/MtGDB/</a> <a href="http://mips.gsf.de/proj/plant/jsf/medi/index.jsp">http://mips.gsf.de/proj/plant/jsf/medi/index.jsp</a>
百脉根 <i>Lotus Japonicus</i>	Medicago miyakogusa.jp LjGDB	<a href="http://www.kazusa.or.jp/lotus/">http://www.kazusa.or.jp/lotus/</a> <a href="http://www.plantgdb.org/LjGDB/">http://www.plantgdb.org/LjGDB/</a>
大豆 <i>Glycine max</i>	GmGDB Phytozome- Glycine max	<a href="http://www.plantgdb.org/GmGDB/">http://www.plantgdb.org/GmGDB/</a> <a href="http://www.phytozome.net/Glycine_max.php">http://www.phytozome.net/Glycine_max.php</a>

表 5 植物生物钟调控网络相关的 GO 术语和 GO 号

Table 5 GO item and GO number of plant circadian clock regulated network

生物钟调控网络相关 GO 术语 GO item of plant circadian clock regulated network	GO 号 GO Number
regulation of circadian rhythm	GO0042752
positive regulation of circadian rhythm	GO0042753
entrainment of circadian clock by photoperiod	GO0043153
entrainment of circadian clock	GO0009649
positive regulation of circadian rhythm	GO0042753
circadian rhythm	GO0007623
circadian regulation of gene expression	GO0032922
circadian regulation of calcium ion oscillation	GO0010617

### 3 材料与方法

#### 3.1 基因组数据来源与获得

蒺藜苜蓿, 百脉根和大豆的原始数据的获得和分析以在线为主。表 4 中列出了 4 个物种进行在线生物信息学分析的基因组数据库及相应网站。

#### 3.2 利用 GO (Geneontology) 获取拟南芥生物钟调控网络相关基因信息

我们在 Gene Ontology 上搜寻 biological clock, circadian clock, circadian rhythm 等关键词, 得到表 5 中的 Go 术语, 再进入 TAIR 数据库, 获得生物钟相关基因的序列及其他相关信息。

#### 3.3 直系同源基因的鉴定

用拟南芥中的相关基因的氨基酸序列在豆科这三个物种各自的基因组数据库上进行 Blastp 或者 tBlastn, 选取得分在 100 以上, E 值在 -30 以下的序列为候选序列, 将比对结果中相似性最高的序列在 NCBI 中反向 Blastp 拟南芥的蛋白质库以确定是否是直系同源基因, 如果不是, 则表明该基因在该物种没有直系同源基因, 如果是, 则以该序列为标准, 进一步确定其它直系同源基因。

#### 3.4 保守结构功能域的分析

综合利用 NCBI 上的 CDD (Marchler-Bauer, 2007) (Conserved Domain Search) 对目标序列进行保守结构域的分析。

#### 3.5 多序列联配, 系统进化树构建和进化分析

利用 ClustalX 程序 (Devlin and Kay, 2001), 对 4 个物种中的直系同源候选基因进行氨基酸序列比对, 采用默认参数。用 MEGA4 软件 (Johnson, 2001; Paloma, 2005) 的 NJ 方法构建系统进化树, 自展开 1000 次, 计算进化距离。同时, 根据构建的进化树, 我们可以大致推算某个基因在豆科和十字花科之间的进化历史。

#### 3.6 表达数据的获得和分析

PlantGDB 的 PUT 序列是经过聚类去除冗余并部分拼接为全长 cDNA 的高质量 EST 序列。将直系同源候选基因的核苷酸序列 BLASTN 比对 PUT 序列, 如果匹配上的 PUT 序列相似性大于 95%, 碱基长度长于 200bp, E 值小于  $1E-30$ , 我们则认为该 PUT 能够证明其匹配的直系同源候选基因部分是真实表达的。我们将 EST 证据分为 3 类: F, E, N, F 代表整个候选基因有 PUT 序列覆盖, 或者两端有 PUT 序列覆盖并且整个基因 80% 的序列被 PUT 序列覆盖。N 表示该基因找不到符合要求的配比 PUT 序列, E 则表示该基因的某个部分或者几个部分有符合要求的

PUT序列覆盖, 但是还没有达到F的标准。

### 3.7 基因的重新注释

在鉴定豆科中3个物种同源候选基因时, 由于基因组信息不全或者注释错误, 会出现没有全长ORF的基因。我们将该部分ORF的基因所对应的基因组区域序列在GenScan上重新注释, 得到的全长ORF的基因, 并将得到的ORF序列重新Blast比对物种相应的PlantGDB的基因组序列, 以确认其是否对应着原始的基因组区域; 同时将ORF序列翻译的氨基酸序列反向Blastp拟南芥的蛋白质数据库以确认其是否是相关的直系同源基因。

### 作者贡献

李宗飞是本实验的执行人, 负责实验设计、实施、数据分析及论文初稿写作; 蔡梦蝶、刘振鹏参与数据分析、初稿的形成、修改; 魏芳和张洁参与稿件修改、翻译及文献的校对; 方宣钧博士是研究项目的构思者, 指导论文的写作及修改。全体作者阅读并同意最终的文本。

### 致谢

本研究由诸暨市翠溪生物技术研究院(生命科学与生物技术创新基金, No: 201601201)资助。宣佳为本论文稿件进行全文的英文评审。

### 参考文献

- Barak S., Tobin E.M., Andronis C., Sugano S., and Green R.M., 2000, All in good time: The Arabidopsis circadian clock, Trends Plant Sci, 5: 517-522
- Covington M.F., Maloof J.N., Straume M., Kay S.A., and Harmer S.L., 2008, Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and

- development, Genome Biol, 9(8): 1
- Devlin P.F., and Kay S.A., 2001, Circadian photoperception, Annu. Rev. Physiol, 63: 677-694
- Harmer S.L., Hogenesch J.B., Straume M., Chang H., Han B., Zhu T., Wang X., Kreps J.A., and Kay S.A., 2000, Orchestrated transcription of key pathways in Arabidopsis by the circadian clock, Science 290: 2110-2113
- Johnson C.H., 2001, Endogenous time keepers in photosynthetic organisms, Annu Rev Physiol, 63: 695-728
- Li Z.F., Zhang J., Liu Z.P., and Fang X.J., 2015, Gene Regulation Network of Biological Clock in Plant, Fenzi Zhiwu Yuzhong (online) (Molecular Plant Breeding), 13(1): 1001-1008
- Li Z.F., Zhuo W., Liu Z.P., and Fang X.J., 2015, Effects of Gene Regulation of Circadian Clock on Plant Growth and Development, Douke Jiyinzuxue Yu Yichuanxue (online) (Legume Genomics and Genetics), 6(1): 1-4
- Marchler-Bauer A., 2007, CDD: a conserved domain database for interactive domain family analysis, Nucleic Acids Res, 35(1): 237-240
- McClung C.R., 2001, Circadian rhythms in plants, Annual review of plant biology, 52(1): 139-162
- McWatters H.G., Bastow R.M., Hall A., and Millar A.J., 2000, The ELF3 zeitnehmer regulates light signalling to the circadian clock, Nature 408: 716-720
- Paloma M., 2005, Circadian clock signaling in Arabidopsis thaliana: from gene expression to physiology and development, Int. J. Dev. Biol., 49: 491-500
- Pruneda-Paz J.L., Breton G., Para A., and Kay S.A., 2009, A functional genomics approach reveals CHE as a component of the Arabidopsis circadian clock, Science, 323(5920): 1481-1485