



研究报告

Research Report

光周期调控豆科植物生物钟的相关基因的鉴定

李宗飞^{1,2,3,4}✉, 张洁^{1,2}, 魏芳^{1,2}, 蔡梦蝶^{1,2}, 刘振鹏^{1,2}, 方宣钧^{1,2,3,4}✉

1 浙江农林大学暨阳学院生命研究所, 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

分子植物育种, 2017 年, 第 15 卷, 第 2 篇 doi: 10.5376/mpb.cn.2017.15.0002

本文首次以英文发表在 *Molecular Plant Breeding* 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, 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。

建议最佳引用格式:

Li Z.F, Zhang J., Wei F., Cai M.D., Liu Z.P., and Fang X.J., 2016, Identification of genes related to photoperiod regulating the circadian clock in legume plants, *Molecular Plant Breeding*, 15(1): 1- 20(doi: 10.5376/mpb. 2017.15.0001)

摘要 为了鉴定豆科植物中光周期调控生物钟的相关基因, 本研究采用生物信息学的方法对豆科中的光敏色素、隐花色素、LKP1/FKFI/LKP2、PIF3/PIF4/PIF5、SRR1、CO/COL1/COL2、TIC、XCT 和 FHY3 共 9 类光周期生物钟调控相关基因进行分析, 结果表明, 光周期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。而以拟南芥为参照, 百脉根和蒺藜苜蓿的光周期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基因组远小于大豆基因组造成的。

关键词 豆科植物, 光周期, 生物钟, 基因

Identification of Genes Related to Photoperiod that Regulating the Circadian Clock in Legume Plants

Li Zongfei^{1,2,3,4}✉, Zhang Jie^{1,2}, Wei Fang^{1,2}, Cai Mengdie^{1,2}, Liu Zhenpeng^{1,2}, Fang Xuanjun^{1,2,3,4}✉

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 this study, 9 kinds of genes related to photoperiod-mediated circadian clock were analyzed by bioinformatics methods in order to identify the genes related to photoperiod that regulate circadian clock in legume, which were Phytochrome, Cryptochrome, LKP1/FKFI/LKP2, PIF3/PIF4/PIF5, SRR1, CO/COL1/COL2, TIC, XCT and FHY3. The results showed that different degrees of differentiation occurred in the 3 species of *Arabidopsis thaliana* and the legumes. Using *Arabidopsis thaliana* as a reference, the differentiation degree of *Lotus japonicus* and *Medicago truncatula* genes related to photoperiod-mediated circadian regulation network was higher than that of *Glycine max*, which might be due to the sizes of *Lotus japonicus* and *Medicago truncatula* genome is far less than that of *Glycine max*.

Keywords Legume, Photoperiod, Circadian clock, Gene

生物钟几乎参与植物体所有的新陈代谢、生长发育的调控过程, 而且主要是通过控制基因的表达来实现的, 比如某些色素它们通过感知外界环境中的光信号来调控体内的生物钟的振荡节律, 同时, 它们自身的表达也受到内在的生物钟机制的精密调控(李宗飞等, 2015; Covington et al., 2008)。植物各器官的生长和发育都是由植物内在的生物钟控

制的, 例如, 生物钟在调控胚轴生长中的作用被认为是控制光信号输入的途径, 这种作用类似于控制CAB基因的表达(李宗飞等, 2015)。它可以协调各种生理活动, 使得体内的各种生理活动能够在适当的时机进行。因此, 对种植物生物钟调控网络进行研究有着非常重大的意义。

植物生物钟节律性这种内在的生理周期与地



球上24小时的光周期相似(Devlin and Kay, 2001)。而光则是外界环境中时间同步机制的一种极其重要的信号。所以研究者们就把鉴定光信号途径中的与生物钟相关的成员作为研究生物钟的一个重要的线索。在拟南芥中, 两类主要的光受体, 光敏色素(PHY)和隐花色素(CRY)都参与了由光周期控制的生物钟(Somers et al., 1998)。无论是光敏色素还是隐花色素, 都可以通过感知外界环境中的光信号来调控体内的生物钟的振荡节律, 同时其自身的表达也受到内在的生物钟机制的精密调控。表明在植物体内, 存在着一个调控光信号输入和设置生物钟节律的调控环。同时, 植物体内还存在着一系列的阀门机制来控制外界光信号输入生物钟系统。

在本研究中, 蒺藜苜蓿和百脉根是豆科中的模式植物, 而大豆是重要的经济作物, 这3个物种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。我们通过将得到的光周期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 光敏色素在豆科中直系同源基因的鉴定

在拟南芥中光敏色素编码的5个蛋白中, PHYB (Canto n, 1999)主要接受红光, 而PHYA(Franklin et al., 2007)和PHYE(Clack et al., 1994)则主要接受远红光。光敏色素以两种状态存在:一种为红光吸收型(Pr), 另一种为远红光吸收型(Pfr)。Pfr是生理激活型, 不稳定, Pr是生理失活型, 比较稳定。Pfr与Pr的状态可在光照下发生转型。PHYA是类型I(Johnson et al., 1994), 吸收红光转化成有活性的Pfr形式后会经历迅速的蛋白质水解过程;而其他4条基因是类型II, 具有高度的光稳定性, 光下稳定。

光引起的光敏色素在核一胞质内的运动与植物生理节律有关(Hanano et al., 2006)。有一系列的光敏色素参与了这个过程, 研究最为清楚的在细胞核内光敏色素相互作用因子是PIF3(phytochrome interacting factor 3) (McWatters et al., 2000)。PIF3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; 并进一步诱导下游的受光诱导的基因表达。

PHYA的结构域主要由4部分构成, 中间靠近N端的光敏色素结构域, 中间靠近C端是2个PAS结构域, PAS结构域是光周期调控的生物钟的相关因子的共同结构, 能够与含G-box顺式作用元件结合, 还有C端的HATPase结构域, 此外还有GAP和His结构域(图1)。

从构建的分子进化树上, 我们可以看到, 在拟南芥中, PHY可以分为2大类, PHYA, PHYC, PHYE为第一大类, PHYB, PHYD为第二大类。在第一大类中, PHYA和PHYC又聚为一小类(图1)。

PHYA在蒺藜苜蓿中有2个直系同源候选基因, 在百脉根中有一个直系同源候选基因, 而PHYB和PHYD在这两个物种各有一个直系同源候选基因, 但是MtPHYB与PHYB的相似性只有22.3%, 并且没有EST表达数据, 表明这个基因可能已经退化为假基因。MtPHYA1与MtPHYA2的相似性为100%, 而它们都在蒺藜苜蓿1号染色体上, MtPHYA1的起始位置为19 823 651 bp, 而MtPHYA2的起始位置为19 721 115 bp, 终止位置为19 726 501 bp, 2个基因在1号染色体上相差约100 000 bp(表1)。这两个基因可能近期的一个复制进化事件, 有可能是1号染色体包括MtPHYA在内部分基因的复制, 也可能是MtPHYA单独的基因发生复制转座子插入突变。

在大豆中, PHYA有4个直系同源候选基因, 而PHYB和PHYE各有2个直系同源候选基因(表1)。

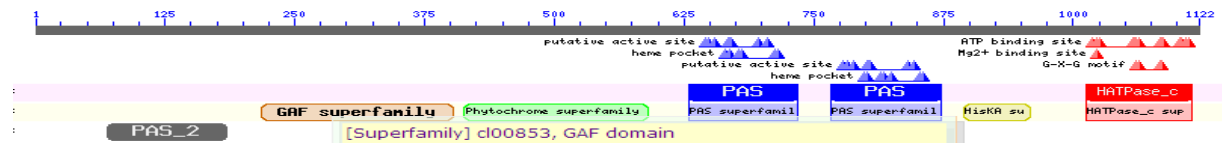
在这3个物种中, PHYA与PHYB的保守性在70%以上, 要高于PHYE的保守性, 后者为60%左右。

而PHYB, PHYD在豆科中的3个物种均没有找到直系同源基因, 表明在拟南芥中, 这2个基因实在十字花科和豆科分开后, 可能通过基因组复制产生的。

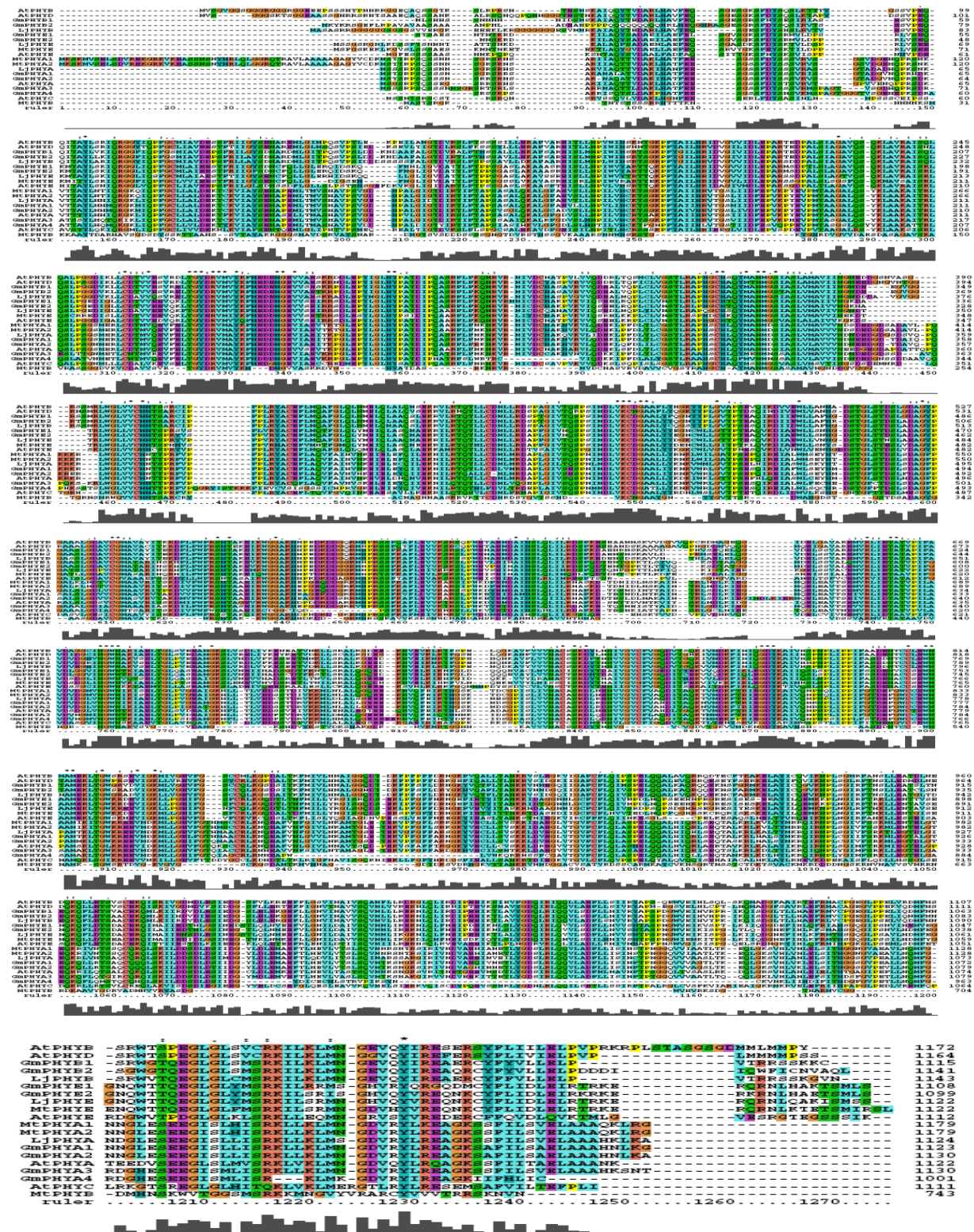
PHY的5个同源基因在十字花科和豆科中的进化历史如下: 在十字花科和豆科分离之前, PHY的祖先基因先复制, 然后分化为PHYB和PHYE, 然后PHYE再通过复制, 分为PHYA和PHYE。在十字花科和豆科分开后, 在拟南芥中, PHYB通过复制, 分化为PHYB和PHYD, 而PHYA通过复制, 分化为PHYA和PHYC。而在蒺藜苜蓿和百脉根中, PHYA, PHYB和PHYE可能在全基因组复制后, 又丢失了复制的一份, 而保持了这3个基因的单拷贝。而在大豆中, PHYA, PHYB和PHYE都经历了2轮复制, 但是PHYB和PHYE在一次复制事件后, 又丢失了一份拷贝。GmPHYA1和GmPHYA2是第二次复制事件产



A

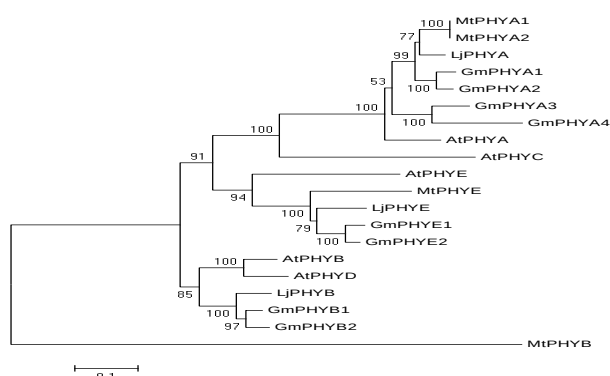


B





C



D

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. AlPHYB																			
2. AlPHYD	0.123																		
3. GmPHYB1	0.212	0.246																	
4. GmPHYB2	0.238	0.267	0.063																
5. LjPHYB	0.223	0.248	0.093	0.109															
6. GmPHYE1	0.458	0.470	0.419	0.437	0.437														
7. GmPHYE2	0.449	0.460	0.408	0.428	0.436	0.053													
8. LjPHYE	0.460	0.474	0.414	0.428	0.430	0.157	0.143												
9. MlPHYE	0.523	0.531	0.503	0.523	0.521	0.248	0.240	0.244											
10. AlPHYE	0.491	0.526	0.460	0.472	0.479	0.417	0.401	0.412	0.482										
11. MlPHYA1	0.577	0.593	0.556	0.567	0.575	0.604	0.604	0.607	0.680	0.686									
12. MlPHYA2	0.577	0.593	0.556	0.567	0.575	0.604	0.604	0.607	0.680	0.686	0.000								
13. LjPHYA	0.561	0.575	0.551	0.566	0.564	0.607	0.601	0.615	0.695	0.686	0.088	0.088							
14. GmPHYA1	0.577	0.596	0.567	0.572	0.585	0.629	0.618	0.634	0.707	0.698	0.116	0.116	0.100						
15. GmPHYA2	0.572	0.598	0.564	0.567	0.580	0.612	0.604	0.620	0.692	0.695	0.123	0.123	0.109	0.056					
16. AlPHYA	0.564	0.564	0.546	0.564	0.564	0.582	0.580	0.648	0.680	0.680	0.169	0.169	0.169	0.185	0.194				
17. GmPHYA3	0.609	0.634	0.596	0.601	0.623	0.643	0.640	0.654	0.716	0.707	0.205	0.205	0.203	0.205	0.210	0.244			
18. GmPHYA4	0.680	0.704	0.668	0.663	0.677	0.707	0.701	0.734	0.772	0.788	0.314	0.314	0.302	0.306	0.296	0.356	0.201		
19. AlPHYC	0.607	0.623	0.609	0.626	0.601	0.680	0.683	0.685	0.710	0.713	0.572	0.572	0.561	0.601	0.588	0.564	0.601	0.701	
20. MlPHYB	1.268	1.273	1.192	1.172	1.202	1.361	1.350	1.385	1.471	1.350	1.524	1.524	1.484	1.511	1.511	1.545	1.497	1.617	1.484

图1 PHY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: PHY蛋白保守结构域分析; B: PHY氨基酸序列比对; C: PHY蛋白的分子进化树; D: 4个物种之间PHY蛋白的进化距离

Figure 1 Conservation domain analysis, protein sequence alignment, molecular evolution tree and evolutionary distance of PHY proteins

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

表1 光敏色素基因在豆科3个物种中的直系同源候选基因

Table 1 The ortholog Candidate genes of PHY in three species of Legume

物种	基因名	基因号	是否全长	相似性	EST 证据
Species	Gene name	ID No. of gene	If the gene is Full-Length	Similarity	Evidence of EST
蒺藜苜蓿 <i>Medicago truncatula</i>	<i>MtPHYA1</i>	AC166743_2	是	75.1%	E
			Yes		
	<i>MtPHYA2</i>	AC148406_27	是	75.1%	E
			Yes		
	<i>MtPHYB</i>	AC158464_25	是	22.3%	N
			Yes		
	<i>MtPHYE</i>	AC152185_37	是	56.6%	E
			Yes		
百脉根 <i>Lotus japonicus</i>	<i>LjPHYA</i>	chr5_CM0048_230_nc	是	78.6%	E
			Yes		
	<i>LjPHYB</i>	chr6_CM1829_170_nc	是	73.0%	E
			Yes		
	<i>LjPHYE</i>	chr6_CM0037_1520_nc	是	61.0%	E
			Yes		
大豆 <i>Glycine max</i>	<i>GmPHYA1</i>	Glyma20g22160	是	78.4%	F
			Yes		
	<i>GmPHYA2</i>	Glyma10g28170	是	76.9%	F
			Yes		
	<i>GmPHYA3</i>	<i>Glyma19g41210</i>	是	73.7%	E
			Yes		
	<i>GmPHYA4</i>	<i>Glyma03g38620</i>	否	56.2%	N
			No		



续表 1

Continuing Table 1

物种	基因名	基因号	是否全长	相似性	EST 证据
Species	Gene name	ID No. of gene	If the gene is Full-Length	Similarity	Evidence of EST
大豆 <i>Glycine max</i>	GmPHYB1	<i>Glyma09g03990</i>	是 Yes	73.1%	F
	GmPHYB2	<i>Glyma15g14980</i>	是 Yes	71.2%	E
	GmPHYE1	<i>Glyma15g23400</i>	否 No	60.1%	E
	GmPHYE2	<i>Glyma09g11600</i>	是 Yes	61.4%	E

生的, 2 者的进化距离为 0.056, 而 *GmPHYB1* 和 *GmPHYB2* 的进化距离为 0.063, *GmPHYE1* 和 *GmPHYE2* 的进化距离为 0.053, 表明这 2 个基因的拷贝是第二次复制事件产生的。

1.2 隐花色素在豆科中直系同源基因的鉴定

CRY 作为生物钟的光受体, 能够接受蓝光, *CRY1* 和 *CRY2* 的功能是冗余的, 作为蓝光介导生物钟的受体。*CRY1* 和 *CRY2* 接受不同强度的蓝光信号, *CRY1* 在蓝光下稳定, 主要接受高强度的蓝光, 而 *CRY2* 在蓝光条件下迅速降解, 主要接受低强度的蓝光, 蓝光诱导下 *CRY2* 的降解可能与 *ZTL* 有关。

在拟南芥中隐花色素有 *CRY1*、*CRY2* 和 *CRY3*, 其中 *CRY3* 与 *CRY1*、*CRY2* 的分歧较远。*CRY1* 和 *CRY2* 的功能研究已知, *CRY3* 与前两者之间的相似性较低, 在 30% 以下, 并且缺失了 DNS 关键结构域, 表明它和 *CRY2* 和 *CRY1* 的分化时代较远, 可能已经退化为假基因或者演化成了其它的功能(图 2)。*CRY1*、*CRY2* 在蒺藜苜蓿中各鉴定了一个直系同源基因, *CRY1* 在百脉根中鉴定有 3 个直系同源基因, 其中只有 1 个是全长基因, *CRY2* 在百脉根中鉴定了 2 个直系同源基因(表 2)。

CRY1 在大豆中鉴定了 4 个直系同源基因, *CRY2* 在大豆中鉴定了 2 个直系同源基因。*CRY1*、*CRY2* 基因在拟南芥和豆科中这 3 个物种中的保守性较高, 表明 *CRY* 基因对植物来说是非常重要的。*CRY1* 在拟南芥和豆科 3 个物种中的保守性要大于 *CRY2* 的保守性, 前者在 75% 以上, 而后者在 60% 左右。而 *CRY3* 在豆科的 3 个物种均找不到直系同源基因。

在大豆的 4 个 *CRY1* 基因中, *GmCRY1a* 和

GmCRY1b 聚为一类, *GmCRY1c* 和 *GmCRY1d* 聚为一类, 对它们的氨基酸序列的进化距离的分析, 发现 *GmCRY1a* 和 *GmCRY1b* 的进化距离为 0.024, *GmCRY1c* 和 *GmCRY1d* 的进化距离为 0.016, 表明 *GmCRY1c* 和 *GmCRY1d* 面临的选择的压力要大于 *GmCRY1a* 和 *GmCRY1b*, 因而更加保守, 这个也可以从这 4 个基因与拟南芥中的 *CRY1* 基因的进化距离看出来。*GmCRY1a*, *GmCRY1b*, *GmCRY1c*, *GmCRY1d* 与拟南芥的进化距离分别为 0.160, 0.172, 0.148, 0.145。从这个意义上来说, *GmCRY1c* 和 *GmCRY1d* 更有可能是在大豆种行使调节抑制下胚轴伸长和蓝光引导昼夜节律时钟的功能。而 *MtCRY1* 和 *LjCRY1a* 与 *AtCRY1* 的进化距离分别为 0.133, 0.139, 这种高度保守性表明它们可能在蒺藜苜蓿和百脉根中行使抑制下胚轴伸长和蓝光引导昼夜节律时钟的功能。

1.3 LKP1/FKFI/LKP2 家族在豆科中直系同源候选基因的鉴定

相关的研究表明, *ZTL* (ZEITLUPE) 又名 *LKP1*, *FKF1* (FLAVIN-BINDING, KELCH REPEAT, F-BOX 1), *LKP2* (LOV, KELCHPROTEIN 2) 可能参与了光信号途径控制的生物钟生理途径。*F-box* 结构域出现在作为受体的蛋白质中, 并且能够将特定的底物连接到蛋白质泛素连接酶以降解 (Craig, 2001)。*ZTL* 蛋白质家族的这些结构域的独特结合特性表明, 他们可能介导某种依赖于光的降解生物钟成分的生理过程, *TOC1* 就是 *ZTL* 的底物之一 (图 3)。

这些基因在光输入途径的作用可能是由它们与光受体的直接相互作用决定的。*ZTL* 在体内被发

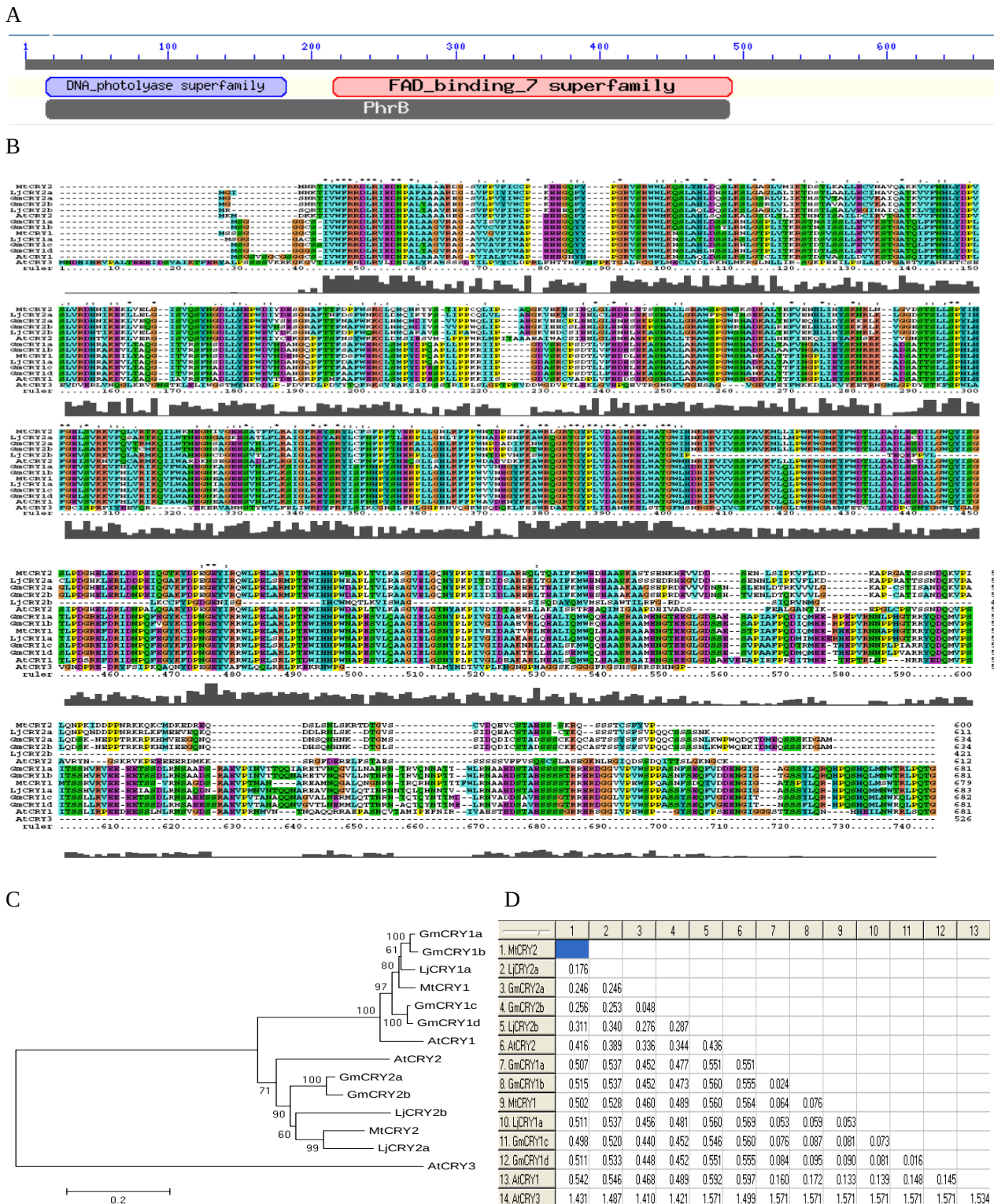


图 2 CRY 蛋白的保守功能结构域分析, 氨基酸序列比对, 分子进化树和进化距离

注: A: CRY 蛋白保守结构域分析; B: CRY 氨基酸序列比对; C: CRY 蛋白的分子进化树; D: 4 个物种之间 CRY 蛋白的进化距离

Figure 2 Conservation domain analysis, protein sequence alignment, molecular evolution tree and evolutionary distance of CRY proteins

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



表2 隐花色素在豆科3个物种中的直系同源候选基因

Table 2 The forecast Orthologs genes of CRY 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>MtCRY1</i>	AC174468_20	是 Yes	77.3%	E
	<i>MtCRY2</i>	AC122171_26	是 Yes	58.9%	F
百脉根 <i>Lotus japonicus</i>	<i>LjCRY1a</i>	LjT09A12_100_nd	是 Yes	77.4%	E
	<i>LjCRY1b</i>	LjSGA_056301.1	否 No		N
	<i>LjCRY1c</i>	LjSGA_011915.1	否 No		N
	<i>LjCRY2a</i>	chr6.CM1613.50.nc	是 Yes	59.5%	E
	<i>LjCRY2b</i>	chr5.CM0071.480.nd	是 Yes	45.2%	E
			是 Yes		
大豆 <i>Glycine max</i>	<i>GmCRY1a</i>	Glyma13g01810	是 Yes	77.4%	F
	<i>GmCRY1b</i>	Glyma14g35020	是 Yes	75.9%	F
	<i>GmCRY1c</i>	Glyma06g10830	是 Yes	76.1%	F
	<i>GmCRY1d</i>	Glyma04g11010	是 Yes	76.1%	F
	<i>GmCRY2a</i>	Glyma20g35220	是 Yes	63.0%	F
	<i>GmCRY2b</i>	Glyma10g32390	是 Yes	63.0%	F

现与PHYB和CRY1基因有相互作用。ZTL还可以与TOC1和PRR5结合物结合，TOC1和PRR5与ZTL的结合与它们的高度磷酸化状态有关。体内的实验表明，PRR5与ZTL结合体的降解受到蓝光的抑制，TOC1与PRR3在体内相互作用，而它们的磷酸化可以促进它们的结合。另外，ZTL与PRR3都可以与TOC1的N端结合，这就表明，PRR3可能通过磷酸化与TOC1结合从而保护TOC1免受ZTL降解。

LKP2在豆科中的3个物种中没有直系同源基因，所以LKP2可能是在十字花科和豆科分开以后由某种机制导致LKP基因在拟南芥中的复制，功能分化

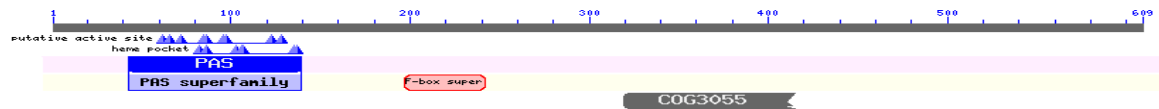
而产生的，拟南芥中，LKP1和LKP2氨基酸序列相似性为72.1%(表3)。

LKP1在蒺藜苜蓿和百脉根中各有一个直系同源基因，相似性分别为82.8%和75.5%，而在大豆中有4个直系同源候选基因。除了GmLKP1c的相似性只有68.0%之外，其他三个与拟南芥中LKP1的相似性都在85%左右。PUT序列的数据表明，MtLKP1和GmLKP1a，GmLKP1b，GmLKP1c全长在各自体内都有表达。而LjLKP1却暂时没有找到EST表达的数据，其余LKP1的相似性也偏低，只有75.5%。

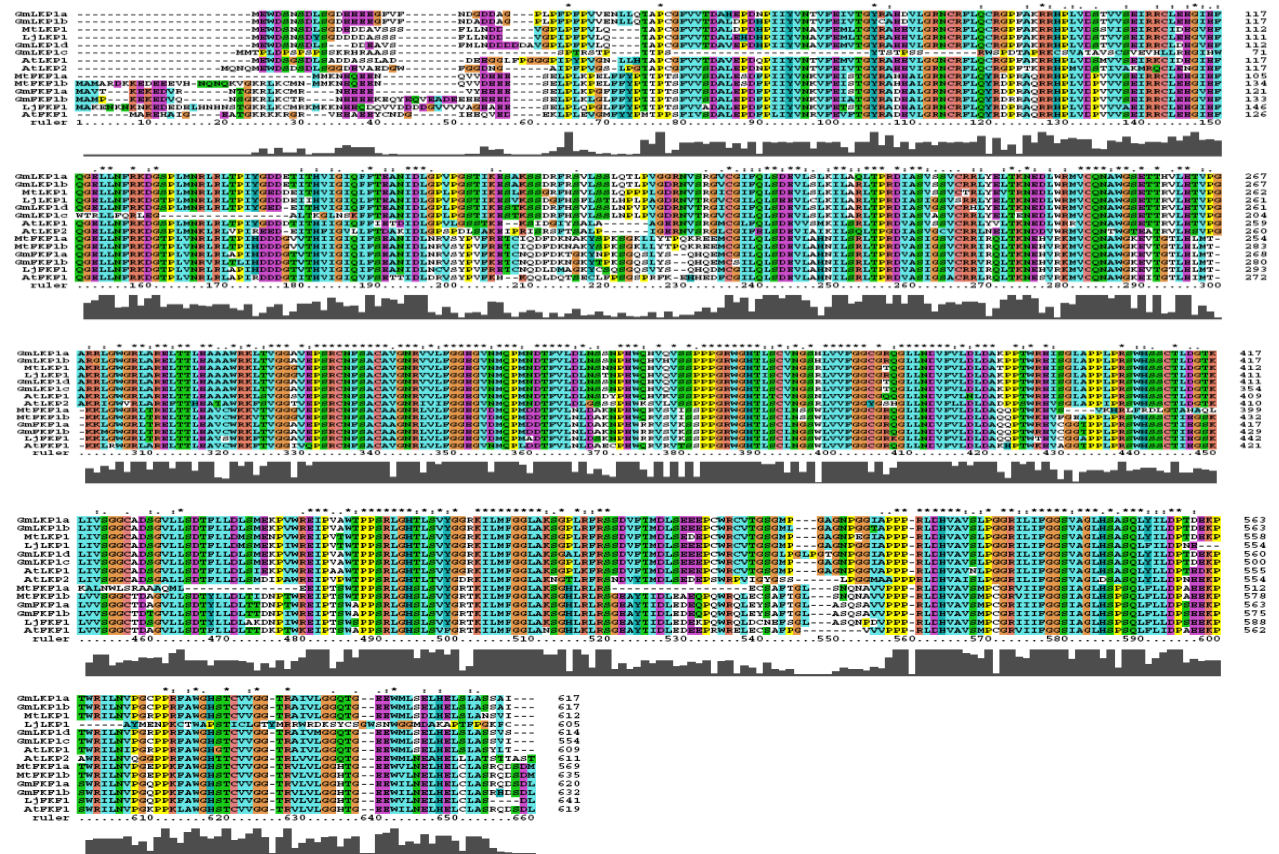
FKF1在百脉根中只有一个直系同源基因，在蒺



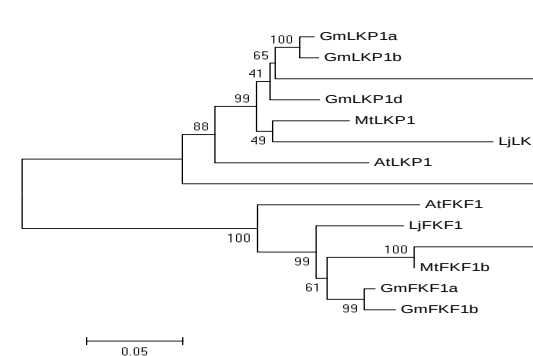
A



B



C



D

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. GmLKP1a	1												
2. GmLKP1b	0.017	1											
3. MfLKP1	0.080	0.084	1										
4. LjLKP1	0.167	0.172	0.154	1									
5. GmLKP1d	0.053	0.057	0.053	0.149	1								
6. GmLKP1c	0.157	0.159	0.177	0.284	0.154	1							
7. AtLKP1	0.127	0.130	0.139	0.237	0.137	0.251	1						
8. AtLKP2	0.251	0.256	0.264	0.336	0.256	0.373	0.278	1					
9. MfFKF1a	0.417	0.417	0.443	0.505	0.427	0.537	0.450	0.541	1				
10. MfFKF1b	0.352	0.352	0.376	0.434	0.361	0.464	0.382	0.474	0.062	1			
11. GmFKF1a	0.333	0.333	0.361	0.421	0.342	0.454	0.370	0.450	0.132	0.068	1		
12. GmFKF1b	0.342	0.342	0.370	0.430	0.352	0.464	0.379	0.460	0.144	0.080	0.022	1	
13. LjFKF1	0.345	0.345	0.382	0.440	0.358	0.460	0.382	0.443	0.167	0.101	0.068	0.080	1
14. AtFKF1	0.355	0.358	0.392	0.443	0.367	0.484	0.367	0.474	0.221	0.154	0.144	0.159	0.174

图 3 LKP1/LKP2/FKF1 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: LKP1/LKP2/FKF1 蛋白保守结构域分析; B: LKP1/LKP2/FKF1 氨基酸序列比对; C: LKP1/LKP2/FKF1 蛋白的分子进化树; D: 4 个物种之间 LKP1/LKP2/FKF1 蛋白的进化距离

Figure 3 Conservation domain analysis, protein sequence alignment, molecular evolution tree and evolutionary distance of LKP1/LKP2/FKF1 proteins

Note: A: Conservation domain analysis of LKP1/LKP2/FKF1 protein; B: LKP1/LKP2/FKF1 protein sequences alignment; C: Molecular evolution tree of LKP1/LKP2/FKF1 proteins; D: The distance of LKP1/LKP2/FKF1 protein between four species



表3 LKP1/LKP2/FKF1在豆科3个物种中的直系同源候选基因

Table 3 The forecast Orthologs genes of LKP1/LKP2/FKF1in three species of Legume

物种	基因名	基因号	是否全长	相似性	EST 证据
Species	Gene name	ID No. of gene	If the gene is Full-Length	Similarity	Evidence of EST
蒺藜苜蓿 <i>Medicago truncatula</i>	<i>MtLKP1</i>	AC148970_35	是 Yes	82.8%	F
	<i>MtFKF1a</i>	AC202355_8	是 Yes	69.7%	E
	<i>MtFKF1b</i>	AC140104_1	是 Yes	77.9%	E
百脉根 <i>Lotus japonicus</i>	<i>LjLKP1</i>	chr6.LjT36L14.70.nd	否 No	75.5%	N
	<i>LjFKF1</i>	chr4.CM0042.1420.nc	是 Yes	76.1%	E
大豆 <i>Glycine max</i>	<i>GmLKP1a</i>	Glyma13g00860	是 Yes	84.8%	F
	<i>GmLKP1b</i>	Glyma17g06950	是 Yes	84.5%	F
	<i>GmLKP1c</i>	Glyma15g17480	是 Yes	68.0%	F
	<i>GmLKP1d</i>	Glyma09g06220	是 Yes	83.4%	E
	<i>GmFKF1a</i>	Glyma05g34530	是 Yes	81.2%	F
	<i>GmFKF1b</i>	Glyma08g05130	是 Yes	79.0%	F

蒺藜苜蓿和大豆中各有2个直系同源基因。*FKF1*在蒺藜苜蓿和百脉根中的直系同源基因只有部分表达数据，而大豆中的两个直系同源基因则在体内全长表达。

从氨基酸需类比对图中可以看出，*GmLKP1c*的蛋白质序列在N端保守区域发生2处大的缺失，而*MtFKF1a*也在C端保守区域发生3处明显的缺失或者突变。

1.4 PIF3/PIF4/PIF5 基因在豆科中直系同源候选基因的鉴定

PIF(PHYTOCHROME-INTERACTING FACTOR)家族是一类与光敏色素有相互作用的因子。其中，*PIF3/PIF4/PIF5*是一类在细胞核内光敏色素的相互作用因子。*PIF3*，*PIF4*和*PIF5*在拟南芥中与光敏色素一起作用来调节植物应对远红光的反

应。*PIF3*，*PIF4*和*PIF5*蛋白质的主要结构域是由碱性氨基酸组成的螺旋一环一螺旋(bHLH)结构中的转录因子。

其中 *PIF3* 蛋白质还含有 1 个 PAS(Per-Arn-Sim-like domain)结构域，能与G-box 顺式作用元件结合，主要调节下游含有G-Box的 MYB类转录物因子，比如*LHY/CCA1*。相关研究表明，*PIF3*与*LHY/CCA1*等基因启动子区的G-box结合以抑制这些基因的表达，而这种抑制作用可以被 *PHYB*所解除。

*PIF5*编码一个新的bHLH类转录因子，属于类 *PIF3*转录因子家族。*PIF5*与*TOC1*存在着相互作用，负向调控*PHYB*的表达，而其蛋白质表达水平又受到*PHYB*的调控。*PIF4*的作用与之类似，*PIF5*还参与了植物的阴影回避生理过程。

*PIF4*和*PIF5*还在生物钟调控光周期控制的胚

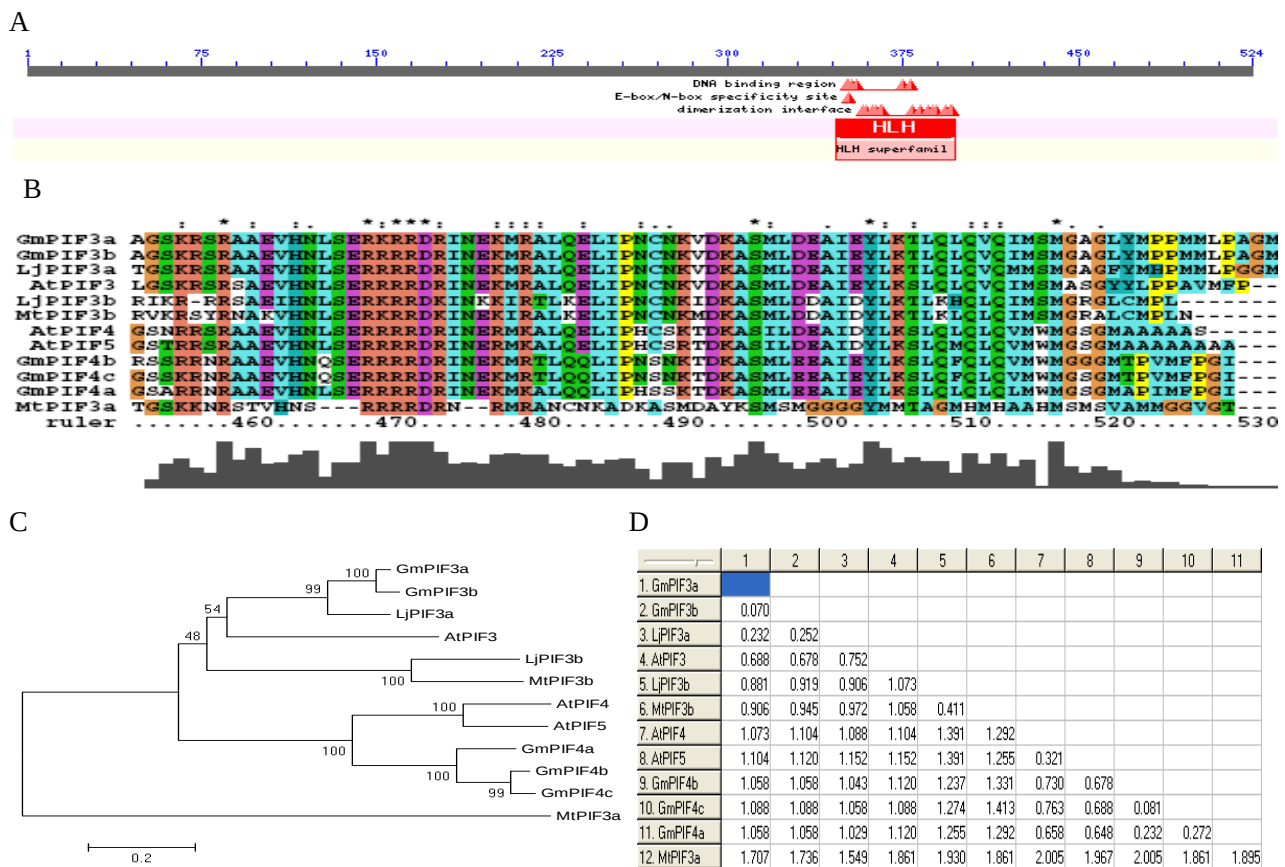


图4 PIF3/PIF4/PIF5 蛋白的保守功能结构域分析, 氨基酸序列比对, 分子进化树和进化距离

注: A: PIF3/PIF4/PIF5蛋白保守结构域分析; B: PIF3/PIF4/PIF5氨基酸序列比对; C: PIF3/PIF4/PIF5蛋白的分子进化树; D: 4个物种之间PIF3/PIF4/PIF5蛋白的进化距离

Figure 4 Conservation domain analysis, protein sequence alignment, molecular evolution tree and evolutionary distance of PIF3/PIF4/PIF5 proteins

Note: A: Conservation domain analysis of PIF3/PIF4/PIF5 protein; B: PIF3/PIF4/PIF5 protein sequences alignment; C: Molecular evolution tree of PIF3/PIF4/PIF5 proteins; D: The distance of PIF3/PIF4/PIF5 protein between four species

表4 PIF3/PIF4/PIF5 在豆科3个物种中的直系同源候选基因

Table 4 The forecast Orthologs genes of PIF3/PIF4/PIF5 in three species of Legume

物种	基因名	基因号	是否全长	相似性	EST 证据
Species	Gene name	ID No. of gene	If the gene is Full-Length	Similarity	Evidence of EST
蒺藜苜蓿 <i>Medicago truncatula</i>	<i>MtPIF3a</i>	AC157894_8	是	16.1%	N
			Yes		
	<i>MtPIF3b</i>	AC137831_27	是	20.7%	E
百脉根 <i>Lotus japonicus</i>	<i>LjPIF3a</i>	chr5_CM0048_80_nd	否		N
			No		
	<i>LjPIF3b</i>	chr1_CM0122_50_nc	是	23.0%	N
			Yes		
	<i>LjPIF4</i>	LjSGA_030731	否		N
			No		



续表 4

Continuing Table 4

物种	基因名	基因号	是否全长	相似性	EST 证据
Species	Gene name	ID No. of gene	If the gene is Full-Length	Similarity	Evidence of EST
大豆 <i>Glycine max</i>	<i>GmPIF3a</i>	Glyma10g28290	是 Yes	25.0%	F
	<i>GmPIF3b</i>	Glyma20g22280	是 Yes	24.1%	F
	<i>GmPIF3c</i>	Glyma03g38670	否 No		N
	<i>GmPIF3d</i>	Glyma19g41260	否 No		N
	<i>GmPIF4a</i>	Glyma08g41620	是 Yes	33.4%	F
	<i>GmPIF4b</i>	Glyma02g45150	是 Yes	30.2%	F
	<i>GmPIF4c</i>	Glyma14g03600	是 Yes	26.0%	E

轴伸长的生理途径中起着重要的作用, *PIF4*和*PIF5*也可以与光敏色素相互作用。在夜晚, *PIF4*和*PIF5*需要短日照条件来刺激其表达量的增加, 而长日照条件则不足以促进其表达。所以, 光周期控制的胚轴延伸与短日照条件下*PIF4*与*PIF5*的积累有关, 这是由内在的生物钟机制和外在的光周期机制一致性决定的。

*PIF3*在蒺藜苜蓿和百脉根中各有2个直系同源候选基因, 在大豆中有4个直系同源候选基因。但是, 豆科中*PIF3*的相似性与拟南芥中的*PIF3*相似性较低(表4)。但是经过分子进化树的分析, 我们发现*MtPIF3a*, *MtPIF3b*和*LjPIF3b*均是*PIF3*的旁系同源基因, 这将会在后面进行分析(图4)。

*PIF3*在大豆中有4个直系同源候选基因, 其中*GmPIF3a*和*GmPIF3b*均是全长, 并且基因编码序列全长都有PUT序列覆盖。而*GmPIF3c*和*GmPIF3d*均不是全长, 而且没有EST表达数据, 表明这2个基因有可能在长期的进化的过程已经沦为没有功能的假基因。

*PIF4*在蒺藜苜蓿中没有找到直系同源候选基因, 在百脉根中有一个直系同源候选基因片段, 但是没有EST表达数据。

*PIF4*在大豆中有3个直系同源候选基因。并且*GmPIF4a*和*GmPIF4b*的基因全长都有PUT序列覆盖,

*GmPIF4c*也有EST部分表达数据。

*PIF5*在豆科的3个物种中均没有找到直系同源候选基因, 可能表明*PIF5*的直系同源基因在豆科的长期进化过程中丢失了。

*PIF3*在豆科中的进化历史比较复杂, *MtPIF3a*与其他*PIF3/PIF4/PIF5*的亲缘关系较远, 与*PIF3*的相似性只有16.1%, 而*MtPIF3b*和*LjPIF3b*也位于*PIF3*的外层, *MtPIF3a*可能是蒺藜苜蓿中*PIF*家族的另外一个基因。在长期的进化过程中, *MtPIF3a*在拟南芥中的直系同源基因已经丢失。当然也有可能是另一种可能性, 这将会在下面分析到。*PIF3*在以拟南芥为代表的十字花科和豆科中的进化历史可能如下: 在十字花科和豆科分离之前, *PIF3*基因就发生一次复制事件, 复制为*PIF3A*和*PIF3B*。十字花科和豆科分离之后, *PIF3B*在拟南芥中丢失, 而在豆科3个物种蒺藜苜蓿, 百脉根和大豆中得以保存, 分别是*MtPIF3b*, *LjPIF3b*, *GmPIF3c*和*GmPIF3d*。*PIF3A*在蒺藜苜蓿中为*MtPIF3a*, 可能由于某种机制, 发生了功能较大的分化, 其相应的功能由*MtPIF3b*替代。而*PIF3A*在大豆中发生1次复制, 复制为*GmPIF3a*和*GmPIF3b*。

*PIF4*的进化历史比较简单, 其在蒺藜苜蓿中已经丢失, 而在百脉根中也可能已经退化。而在大豆中则经历了2次复制事件, 并且丢失了1个基因拷贝。



表5 SSR1在豆科2个物种中的直系同源候选基因

Table 5 The forecast Orthologs genes of SSR1 in two species of Legume

物种	基因名	基因号	是否全长	相似性	EST 证据
Species	Gene name	ID No. of gene	If the gene is Full-Length	Similarity	Evidence of EST
百脉根 <i>Lotus japonicus</i>	<i>LjSRR1a</i>	LjSGA_013417	是	51.6%	F
			Yes		
	<i>LjSRR1b</i>	LjSGA_032322	否		E
大豆 <i>Glycine max</i>	<i>GmSRR1a</i>	Glyma11g21500.1	是	53.3%	F
			Yes		
	<i>GmSRR1b</i>	Glyma06g08770.2	是		F
			Yes		

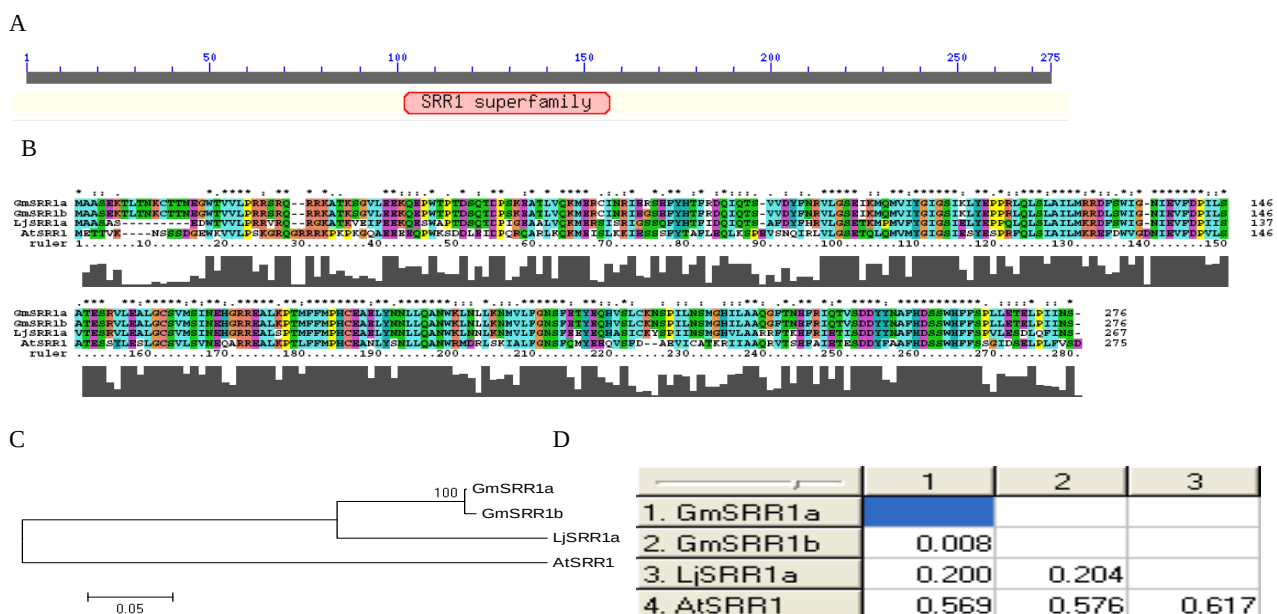


图 5 SSR1 蛋白的保守功能结构域分析, 氨基酸序列比对, 分子进化树和进化距离

注: A: SSR1蛋白保守结构域分析; B: SSR1氨基酸序列比对; C: SSR1蛋白的分子进化树; D: 4个物种之间SSR1蛋白的进化距离

Figure 5 Conservation domain analysis, protein sequence alignment, molecular evolution tree and evolutionary distance of SSR1 proteins

Note: A: Conservation domain analysis of SSR1 protein; B: SSR1 protein sequences alignment; C: Molecular evolution tree of SSR1 proteins; D: The distance of SSR1 protein between four species

总之, 在豆科的长期进化过程中, 拟南芥中的 $PIF5$ 在豆科的3个物种中均找不到直系同源基因, 而 $PIF3$ 和 $PIF4$ 中的直系同源基因已经在蒺藜苜蓿中丢失或者演化为其它功能的基因。而在大豆中均有相应功能的基因存在; $PIF3$ 和 $PIF4$ 在百脉根中的情形则介于蒺藜苜蓿和大豆之间。

1.5 $SRR1$ 在豆科中直系同源候选基因的鉴定

$SRR1$ (sensitivity to red light reduced 1)的生物钟的功能和 $PIF3$ 类似, 都与 $PHYB$ 介导的光信号传导有关, 是生物钟中央振荡器产生正常节律所必需的调控因子。光信号途径和生物钟的之间的联系最初



也在SRR1(sensitivity to red light reduced 1)基因中找到了证据, *srr1*突变株系中在PHYB介导的信号途径和普通的生物钟信号输出途径中存在着某种缺陷, 表明SRR1基因是正常的生物钟震荡功能所必须的。分子遗传学上的研究*srr1*突变株系在光照条件一致的情况下, 会改变很多生物钟输出的成分。

SRR1基因在蒺藜苜蓿中没有找到直系同源候选基因, 在百脉根有2个直系同源基因, 其中*LjSRR1a*为全长, 并且全长有PUT序列覆盖, 而但是*LjSRR1b*不是全长(表5)。

SRR1在大豆也有2个直系同源基因, 相似性在53%以上, 并且全长都有PUT序列覆盖。GmSRR1a和GmSRR1b的进化距离只有0.008, 表明这2个基因是在大豆属与苜蓿属种分离后, 最近一次复制事件的(图5)。

1.6 CO/COL1/COL2 在豆科中直系同源候选基因的鉴定

CO基因编码一个核蛋白, 这个蛋白包含了一个在TOC1蛋白和CO-like蛋白中都具有的CCT结构域, 还包含了2个B-box型的锌指区域, 这2个区域可能接到了蛋白质之间的相互作用。在长日照条件下, CO基因的表达在白天的最后时期以及夜晚时达到最大。但是在短日照条件下, 它的表达模式就发生了改变, 大部分的表达都发生在夜晚。CO的这种与光周期相一致的表达被认为在长日照条件下诱导开花的生理过程中发挥了重要的作用。CO的这种表达机制似乎与它在控制FT(FLOWERING LOCUS T)的活性有关。FT是一个及其重要的催化开花的基因, 并且它的表达与CO的活性有着紧密的联系。在*toc1-1*突变株系中, CO的表达模式发生了改变, 无论是在长日照还是在短日照条件下, 这种改变导致了对光周期不敏感的早花现象。

在拟南芥中, CO与COL2的亲缘关系比CO与COL1的亲缘关系更近一些。拟南芥是长日照植物, 而百脉根, 大豆, 蒺藜苜蓿都是短日照植物, 而CO则是拟南芥由生物钟调控的光周期控制开花的生理途径中一个非常重要的基因, CO和COL1在蒺藜苜蓿, 百脉根, 大豆, 中都找不到直系同源基因, 也说明了CO基因是控制拟南芥在短日照条件下开花的一个关键性的基因。

在蒺藜苜蓿中没有找到COL1, COL2的直系同源基因, 相似性最高的基因为AC127169_10, 在拟

南芥中对应的是CONSTANS-like B-box zinc finger protein(表6)。

在百脉根中, COL2只有1个直系同源候选基因片段, 相似性为50.1%, 有EST证据。

从氨基酸需类比对图中可以看出, CO, COL1和COL2基因在N端和C端分别有1段比较保守的区域, 而从PUT证据来看, GmCOL2a和GmCOL2b有可能在大豆中有表达的。而GmCOL2c和GmCOL2d在大都中还没有找到表达的数据(图6)。

在大豆中, COL2有4个直系同源候选基因, 相似性在50%以上。其中GmCOL2a和GmCOL2b有全长或接近全长的PUT序列的支持。

1.7 TIC 在豆科中直系同源候选基因的鉴定

TIC(time for coffee), 主要是抑制光敏色素介导的光信号输入进植物的生物钟系统, 与ELF3的功能类似, 充当生物钟的阀门的效应, 使中央振荡器的节律性不被外界的高强度光所干扰。TIC的在下游的调控目标主要是LHY。

进一步的研究表明, 在*tic*突变株系中, 一些受到生物钟控制的性状表现为无节律性, 表明其在生物钟生理途径中具有一定的功能。与ELF3在夜晚的功能相似, TIC在白天具有控制生物钟某个开光的功能。对*tic/elf3*双突变株系的完全无节律性性状表明, 这两个基因在生物钟生理途径中具有重要的功能。光信号途径和生物钟之间的联系最初也在SRR1(sensitivity to red light reduced 1)基因中找到了证据, *srr1*突变株系中在PHYB介导的信号途径和普通的生物钟信号输出途径中存在着某种缺陷 (Staiger et al., 2003), 表明SRR1基因是正常的生物钟震荡功能所必须的。

TIC基因在蒺藜苜蓿中有一个直系同源候选基因, 其相似性为40.4%, 有EST证据(表7)。

TIC基因在百脉根中有4个直系同源候选基因, 但是都不是全长。其中*LjTIC1*和*LjTIC2*有EST数据支持。而*LjTIC1*和*LjTIC4*在CM0357的位置比较邻近, 因此我们推测*LjTIC1*和*LjTIC4*就是百脉根中一个TIC基因由于基因组信息不全而被注释成了2个基因片段(表7)。TIC基因在大豆中有4个直系同源候选基因, 其中GmTIC1, GmTIC2, GmTIC3, 是全长, 并且有EST支持。而GmTIC4不是全长, 并且没有EST支持(表7)。

*LjTIC1*虽然不是全长, 但是其翻译的蛋白质序

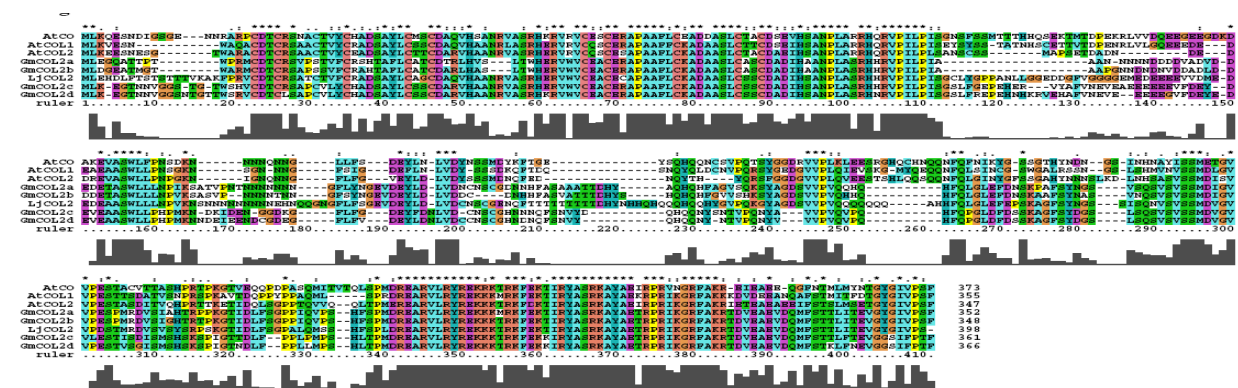


表6 CO/COL1/COL2在豆科2个物种中的直系同源候选基因

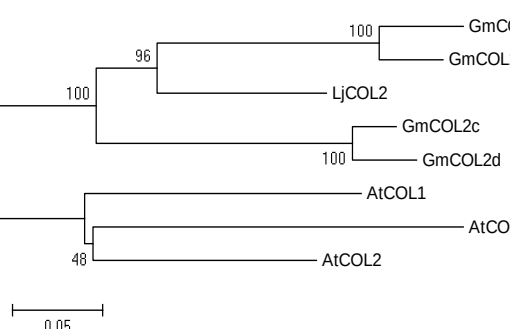
Table 6 The forecast Orthologs genes of CO/COL1/COL2 in two species of Legume

物种	基因名	基因号	是否全长	相似性	EST 证据
Species	Gene name	ID No. of gene	If the gene is Full-Length	Similarity	Evidence of EST
百脉根	<i>LjCOL2</i>	chr1_CM0215_30_nd	否	50.1%	E
<i>Lotus japonicus</i>			No		
大豆	<i>GmCOL2a</i>	Glyma18g51320	是	51.7%	F
<i>Glycine max</i>			Yes		
	<i>GmCOL2b</i>	Glyma08g28370	是	56.3%	F
			Yes		
	<i>GmCOL2c</i>	Glyma13g07030	是	53.2%	N
			Yes		
	<i>GmCOL2d</i>	Glyma19g05170	是	51.9%	N
			Yes		

A



B



C

	1	2	3	4	5	6	7
1. AtCO							
2. AtCOL1	0.354						
3. AtCOL2	0.330	0.292					
4. GmCOL2a	0.589	0.523	0.478				
5. GmCOL2b	0.571	0.518	0.446	0.082			
6. LjCOL2	0.473	0.435	0.440	0.252	0.265		
7. GmCOL2c	0.553	0.467	0.462	0.364	0.354	0.292	
8. GmCOL2d	0.535	0.484	0.467	0.388	0.373	0.311	0.060

图 6 CO/COL1/COL2 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: CO/COL1/COL2蛋白保守结构域分析; B: CO/COL1/COL2氨基酸序列比对; C: CO/COL1/COL2蛋白的分子进化树; D: 4个物种之间CO/COL1/COL2蛋白的进化距离

Figure 6 Conservation domain analysis, protein sequence alignment, molecular evolution tree and evolutionary distance of CO/COL1/COL2 proteins

Note: A: Conservation domain analysis of CO/COL1/COL2 protein; B: CO/COL1/COL2 protein sequences alignment; C: Molecular evolution tree of CO/COL1/COL2 proteins; D: The distance of CO/COL1/COL2 protein between four species

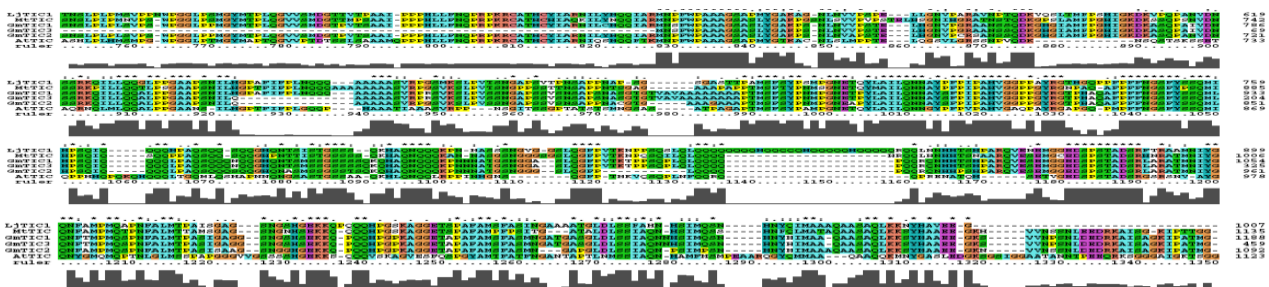


表7 TIC在豆科3个物种中的直系同源候选基因

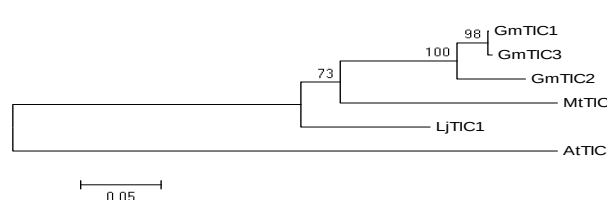
Table 7 The forecast Orthologs genes of TIC in three species of Legume

物种	基因名	基因号	是否全长	相似性	EST 证据
Species	Gene name	ID No. of gene	If the gene is Full-Length	Similarity	Evidence of EST
蒺藜苜蓿 <i>Medicago truncatula</i>	<i>MtTIC</i>	AC137838_22	是	40.4%	E
			Yes		
	<i>LjTIC1</i>	chr5.CM0357.1060.nd	否	26.7%	E
			No		
<i>Lotus japonicus</i>	<i>LjTIC2</i>	chr3.CM0711.580.nd	否		E
			No		
	<i>LjTIC3</i>	LjSGA_039130.1	否		N
			No		
大豆 <i>Glycine max</i>	<i>LjTIC4</i>	chr5.CM0357.1040.nd	否		E
			No		
	<i>GmTIC1</i>	Glyma20g28260.1	是	37.0%	E
			Yes		
	<i>GmTIC2</i>	Glyma10g39490	是	40.1%	E
			Yes		
	<i>GmTIC3</i>	Glyma20g28260.2	是	22.8%	E
			Yes		
	<i>GmTIC4</i>	Glyma06g10030	否		N
			No		

A



B



C

	1	2	3	4	5
1. LjTIC1					
2. MtTIC	0.220				
3. GmTIC1	0.204	0.224			
4. GmTIC3	0.207	0.227	0.003		
5. GmTIC2	0.224	0.255	0.064	0.061	
6. AtTIC	0.598	0.696	0.628	0.633	0.648

图 7 TIC 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: TIC蛋白保守结构域分析; B: TIC氨基酸序列比对; C: TIC蛋白的分子进化树; D: 4个物种之间TIC蛋白的进化距离

Figure 7 Conservation domain analysis, protein sequence alignment, molecular evolution tree and evolutionary distance of TIC proteins

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



列最长的也含有1 007个氨基酸。所以我们认为其大致可以代表其可能的全长基因序列来构建进化树和计算进化距离(图7)。

从氨基酸需类比对图中可以明显看出, 物种间序列的变异很多有相当一部分发生在谷氨酰胺重复的区域, 尤其是LjTIC蛋白序列在保守区域比其他物种的TIC蛋白质序列多了4个重复的“QQQQH”序列, 而在其周围也有谷氨酰胺的多个重复序列。

1.8 XCT 在豆科中直系同源候选基因的鉴定

Martin(2008)在拟南芥中鉴定了XCT(XAP5 CIRCADIAN TIMEKEEPER)基因, 该基因在应对光照调节生物钟和光形态发生生理过程中有着重要的作用。XCT对植物固有的生物钟功能有着重要的作用。xct突变株系在各种条件下, 都表现为振荡周

期缩短。更有趣的是, 当光的特性和波长不同的, XCT应对光的作用是不同的甚至是相反的。xct突变株系对红光特别敏感, 但是对蓝光的反应却是正常的。与之相反, xct突变株系对胚轴延伸的抑制作用对蓝光敏感, 对红光却不敏感。XCT可能在总体上协调植物应对光照来协调植物生长发育的各种生理活动, XCT包含一个XAP保守结构域, 这个结构域在高等真核生物钟很保守, 表明它在高等真核生物的功能是非常重要的。

XCT基因在拟南芥和豆科的蒺藜苜蓿, 百脉根, 大豆中的相似性非常高, 表明该基因对植物维持基本功能有着非常重要的作用, 可能不仅仅限于生物钟调控途径(表8)。如图8所示, XCT有一个非常保守的XAP5结构域。而同时, 我们也从图8看到, GmXCT1在保守区缺失了很大一部分序列。

表8 XCT在豆科3个物种中的直系同源候选基因

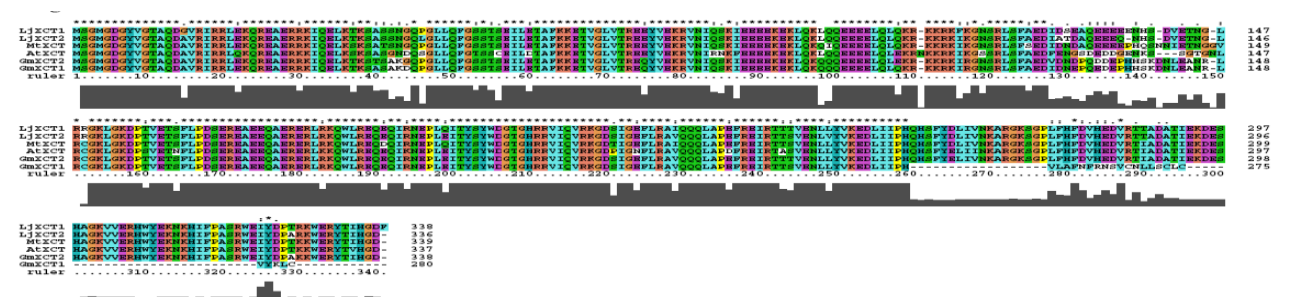
Table 8 The forecast Orthologs genes of XCT in three species of Legume

物种	基因名	基因号	是否全长	相似性	EST 证据
Species	Gene name	ID No. of gene	If the gene is Full-Length	Similarity	Evidence of EST
蒺藜苜蓿	<i>MtXCT</i>	AC140035_48	是	84.7%	E
<i>Medicago truncatula</i>			Yes		
百脉根	<i>LjXCT1</i>	chr2.CM0099.260.nd	是	84.7%	F
<i>Lotus japonicus</i>			Yes		
	<i>LjXCT2</i>	chr4.CM0042.1230.nc	是	85.5%	F
大豆	<i>GmXCT1</i>	Glyma08g04680	是	63.5%	F
<i>Glycine max</i>			Yes		
	<i>GmXCT2</i>	Glyma05g35040	是	84.7%	F
			Yes		

A



B



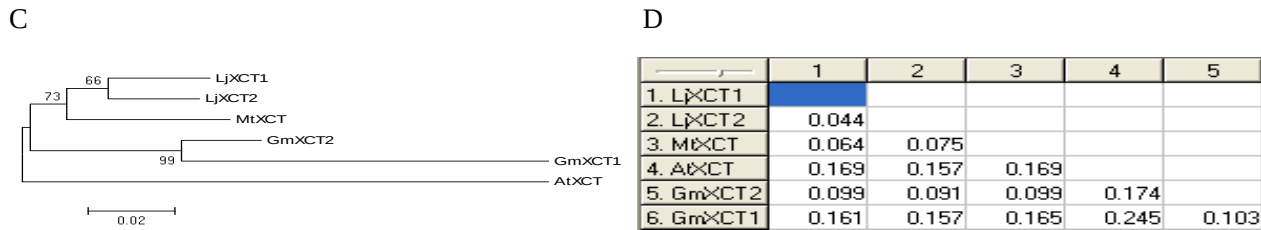


图 8 XCT 蛋白的保守功能结构域分析, 氨基酸序列比对, 分子进化树和进化距离

注: A: XCT蛋白保守结构域分析; B: XCT氨基酸序列比对; C: XCT蛋白的分子进化树; D: 4个物种之间XCT蛋白的进化距离

Figure 8 Conservation domain analysis, protein sequence alignment, molecular evolution tree and evolutionary distance of XCT proteins

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

表9 LHY在豆科3个物种中的直系同源候选基因

Table 9 The forecast Orthologs genes of LHY in three species of Legume

物种	基因名	基因号	是否全长	相似性	EST 证据
Species	Gene name	ID No. of gene	If the gene is Full-Length	Similarity	Evidence of EST
大豆	<i>GmFHY3a</i>	Glyma06g47210	是	63.9%	F
<i>Glycine max</i>			Yes		
	<i>GmFHY3b</i>	Glyma04g14850	是	63.9%	E
			Yes		

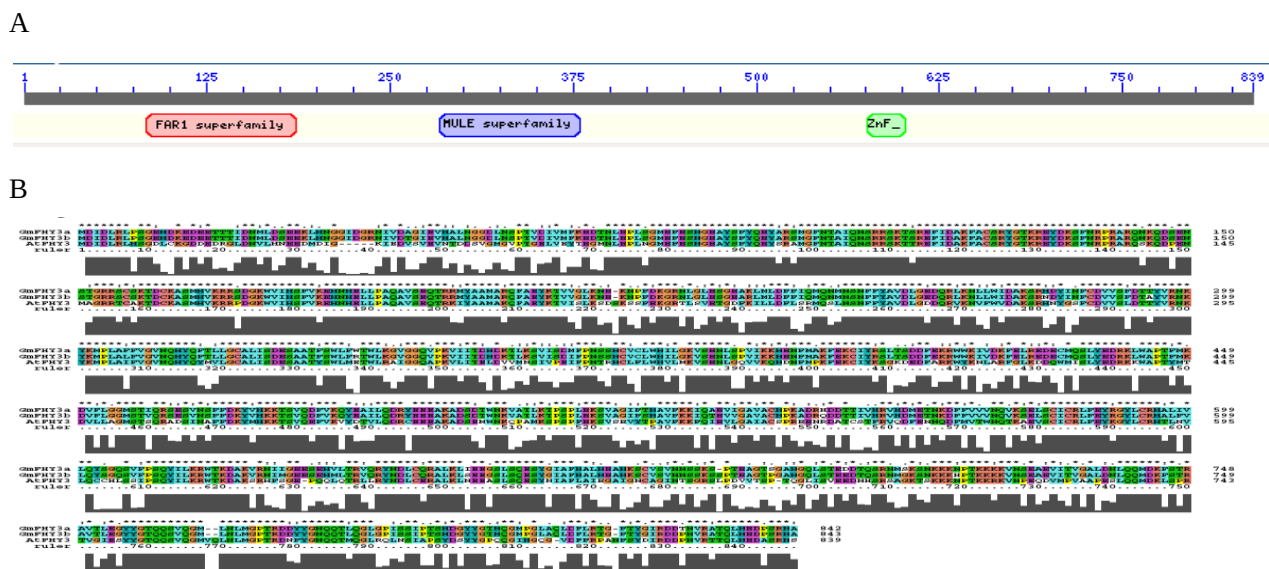


图 9 FHY3 蛋白的保守功能结构域分析和氨基酸序列比对

注: A: FHY3 蛋白保守结构域分析; B: FHY3 氨基酸序列比对

Figure 9 Conservation domain analysis, protein sequence alignment of FHY3 proteins

Note: A: Conservation domain analysis of FHY3 protein; B: FHY3 protein sequences alignment

1.9 *FHY3* 在豆科中直系同源候选基因的鉴定

FHY3(far-red elongated hypocoty l3)是PHYA光

信号传导途径中的一员。*FHY3*和*FAR1*是一对同源蛋白, 它们都有一个明显的与突变了了的转录子同源



的序列。它们作为一种新发现的转录因子, 对激活 *FHY1* 和 *FHL* 的表达起着关键性的作用。而 *FHY1* 和 *FHL* 的表达对光诱导下 *PHYA* 在核中的积累以及随后的光信号传导是必需的。

FHY3 有 3 个明显的结构域, 两端都是保守的锌指结构域, 其中 N 端的是 C2H2 型锌指结构域, 属于 FAR1 亚家族结构域, 对结合 DNA 以及介导光信号的传递有着重要的作用, 而中间的核心转录区域属于 MULE 亚家族, 它和 C 端的 SWIM 型锌指结构域一起在转录水平上调控 *FHY3* 的活性, 并且可以与自身形成同源二聚体或者与 FAR1 形成异源二聚体。并且, 这种形成同源和异源二聚体的能力与 *FHY3* 的活性相关 (Rongcheng et al., 2008) (图 9)。

在红光条件下, 对 2 个不同的 *fhy3* 等位基因株系对胚轴延伸的抑制作用明显加强, 并且胚轴延伸的生物钟节律性被打断。*FHY3* 对接收红光来进行生物钟的重新设定有着非常的作用, 尤其是在在一天中的早期阶段维持生物钟节律性 (Allen et al., 2006)。

FHY3 在蒺藜苜蓿和百脉根中均没有找到直系同源基因, 在大豆中有 2 个直系同源候选基因, 这 2 个基因之间的相似性为 97.3%, 与 *FHY3* 的相似性都为 63.9%。表明 *GmFHY3a* 和 *GmFHY3b* 是最近发生的一次基因复制事件而产生的 2 个同源基因 (表 9)。其中 *GmFHY3a* 在体内全长表达, 而 *GmFHY3b* 只有部分表达的 PUT 数据

2 讨论

本部分采用生物信息学的方法对豆科中的光

周期生物钟调控网络进行了分析。我们一共鉴定了 41 个大豆的生物钟相关候选基因, 22 个百脉根的生物钟相关候选基因, 13 个蒺藜苜蓿的生物钟相关候选基因。我们根据鉴定的直系同源基因的候选基因构建了 9 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, 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。而以拟南芥为参照, 百脉根和蒺藜苜蓿的光周期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由于其基因组远小于大豆。

从进化的角度来看, 不同物种间直系同源基因的功能都是从相同祖先基因分化而来, 一般来说它们功能最为接近, 我们通过同源基因所鉴定的候选基因是在目前没有实验数据支持的情况下能推知的与拟南芥相关基因功能最为类似的基因。与传统的分子生物的方法相比, 生物信息学更为便捷, 更具有目的性, 而且可以在全基因的水平上进行分析, 所得出的结论也更加全面, 我们的研究可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参考价值。

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 10 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。

表 10 4 个物种进行生物信息学分析相应的数据库和网址

Table 10 The Name and Website of the Databank of four species

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



3.2 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3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。

3.3 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。

3.4 多序列联配, 系统进化树构建和进化分析

利用ClustalX程序, 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软件的NJ方法构建系统进化树的构建, 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。

3.5 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。

3.6 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的错误, 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; 张洁、刘振鹏参与数据分析、初稿的形成、修改; 蔡梦蝶和魏芳参与稿件修改、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创新基金《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, 在此表示感谢。

参考文献

- Allen T., Koustenis A., Theodorou G., Somers D.E., Kay S.A., Whitelam G.C., and Devlin P.F., 2006, Arabidopsis *PHY3* specifically gates phytochrome signaling to the circadian clock, *Plant Cell*, 18(10): 2506-16
- Canto ´ n F.R., 1999, Quail PH Both *phyA* and *phyB* mediate light-imposed repression of *PHYA* gene expression in Arabidopsis. *Plant Physiol*, 121: 1207-1215
- Clack T., Mathews S., and Sharrock R.A., 1994, The phytochrome apoprotein family in Arabidopsis is encoded by five genes: the sequences and expression of *PHYD* and *PHYE*, *Plant Mol. Biol*, 25: 413-427
- 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
- Craig K.L., and Tyers M., 2001, The F-box: A new motif for ubiquitin dependent proteolysis in cell cycle regulation and signal transduction. *Prog. Biophys. Mol. Biol*, 72(3): 299-328
- Franklin K.A., Allen T., and Whitelam G.C., 2007, Phytochrome A is an irradiance-dependent red light sensor, *Plant J*, 50(1): 108-17
- Hanano S., Domagalska M.A., Nagy F., and Davis S.J., 2006, Multiple phytohormones influence distinct parameters of the plant circadian clock, *Genes Cells*, 11(12): 1381-1392
- Johnson E., Bradley J.M., Harberd N.P., and Whitelam G.C., 1994, Photoresponses of light-grown *phyA* mutants of Arabidopsis (phytochrome A is required for the perception of daylength extensions). *Plant Physiol*, 105: 141-149



- Li Z.F., Zhang J., Liu Z.P., and Fang X.J., 2015a, 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., 2015b, 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
- Lin R.C., Teng Y.B., Park H.J., Ding L., Black C., Fang P., and Wang H.Y., 2008, Discrete and Essential Roles of the Multiple Domains of Arabidopsis FHY3 in Mediating Phytochrome A Signal Transduction, *Plant Physiology*, 148: 981-992
- Marchler-Bauer A., 2007, CDD: a conserved domain database for interactive domain family analysis, *Nucleic Acids Res*, 35(1): 237-240
- McWatters H.G., Bastow R.M., and Hall A., 2000, Millar A.J. The ELF3 zeitnehmer regulates light signalling to the circadian clock, *Nature*, 408: 716-720
- Staiger D., Allenbach L., Salathia N., Fiechter V., Davis S.J., Millar A.J., Chory J., and Fankhauser C., 2003, The Arabidopsis SRR1 gene mediates PHYB signaling and is required for normal circadian clock function, *Genes Dev*, 17: 256-268