

生物信息学预测植物 miRNAs 的方法

张廷婷¹, 胡述浩², 闫彩霞¹, 赵小波¹, 单世华¹, 姜林平³

(¹山东省花生研究所, 青岛 266100; ²山东大学威海分校, 威海 264209; ³海阳市农业技术推广站, 山东海阳 265100)

摘要: MicroRNA 又称 miRNA, 是一类广泛存在于动植物体内、大小为 21 ~ 25nt 左右的内源性非编码单链小分子 RNA。在植物体中, miRNA 主要负责调控转录后基因的表达, 在机体发育、生长和应答胁迫方面发挥着重要的调节作用。近年来, miRNA 已成为分子生物学领域的研究热点。本文介绍了几种发现和研究 miRNA 的方法, 重点阐述了生物信息学预测 miRNA 方法的流程和策略, 并对各过程中采用的方法及信息学软件的优劣性进行比较和分析, 为预测 miRNA 提供新的思路。

关键词: 生物信息学; 植物; miRNA 预测

Bioinformatics Prediction of MicroRNAs in Plant

ZHANG Ting-ting¹, HU Shu-hao², YAN Cai-xia¹, ZHAO Xiao-bo¹, SHAN Shi-hua¹, JIANG Lin-ping³

(¹Shandong Peanut Research Institute, Qingdao 266100; ²Shandong University at Weihai, Weihai 264209;

³Haiyang Agricultural Technology Extension Station, Haiyang Shandong 265100)

Abstract: MicroRNAs (miRNAs) are endogenous non-coding single-strand molecules with 21-25 nucleotides in length, and they are widely existed in plants and animals. In plant, miRNAs have been confirmed that they negatively regulate gene expression at the post-transcriptional level and play an important regulatory role in plant growth and development, and response to various biotic and abiotic stresses. In recent years, miRNA has become the hot-spot in modern molecule biology. In this paper, several methods for identification and characterization of plant miRNAs were introduced, and further the processes and strategies for miRNA bioinformatics prediction were reviewed with emphasis, as well as investigation and analysis of some bioinformatics softwares. This paper will provide new ideas for miRNA identification in plant.

Key words: bioinformatics; plant; miRNA identification

MicroRNA 又称 miRNA, 是一类广泛存在于动物、植物、病毒等多种有机体内、大小为 21 ~ 25nt 左右的内源性非编码单链小分子 RNA^[1]。1993 年, R. C. Lee 等^[2]首次在秀丽小杆线虫 (*Caenorhabditis elegans*) 中发现了一种 22nt 的非编码 RNA-lin-4, 它参与调控胚胎后期的发育。2000 年, 果蝇中 miRNA-let7 的发现, 掀起了寻找 miRNA 的热潮。2002 年, 一些研究组相继在拟南芥中利用直接克隆技术克隆得到 100 多条 miRNAs, 这些 miRNAs 在多个植

物物种中具有高度保守性^[3-4]。

miRNA 分子的合成和作用机制十分复杂。首先, miRNA 分子的初级转录物 pri-miRNA 在细胞核内由含有核糖核酸酶 III (RNA 酶)-Drosha 的多重复合物介导, 加工产生长度为 60 ~ 120 个核苷酸的发夹结构, 即 miRNA 的前体结构 pre-miRNA^[5]。然后, 在核转运蛋白的作用下, 将 pre-miRNA 运送到细胞质, 在胞质内由 RNA 酶 III-Dicer 加工形成 18 ~ 25 个核苷酸的成熟 miRNA 分子, 与互补链结合成双螺

收稿日期: 2014-03-07 修回日期: 2014-05-20 网络出版日期: 2014-12-11

URL: <http://www.cnki.net/kcms/detail/11.4996.S.20141211.2205.011.html>

基金项目: 现代农业产业技术体系建设专项 (CARS-14); 山东省现代农业产业技术体系项目 (SDAIT-05-011-03); 山东省农业生物资源创新利用研究课题

第一作者研究方向为花生品种资源创新利用。E-mail: tingtingzh@163.com

通信作者: 单世华, 研究方向为花生品种资源创新利用。E-mail: shhshan@sina.com

旋结构。之后,miRNA 序列加入到介导 miRNA 活性的 RNA 诱导的沉默复合物 RISC (RNA-induced silencing complex) 中,与靶基因结合行使功能^[6-7]。

现在的模型假定成熟的 miRNA 组装进入沉默 RISC 后,以碱基互补配对的方式与靶 mRNA 结合,若完全互补或近似完全互补将引起靶基因互补区的特异性断裂,从而导致基因沉默,若互补性低则会抑制靶基因的翻译^[1]。miRNA 介导的基因调控的翻译抑制现象在动物中很常见,同时也存在于植物中,但其机制尚未清楚。而通过 RISC 指导的靶 mRNA 的切割则在植物中普遍存在^[8]。

自从在拟南芥中发现了 miRNAs 之后,对 miRNAs 的鉴定和功能分析成为了研究热点,并在过去的 10 年研究中取得了显著的研究成果,近年来,miRNA 的研究也已经成为分子生物学的热点领域,其研究也由最初在模式生物中只发现 1~2 个小分子 RNA 发展到在不同物种中大量发现 miRNA,而且内容不断深入以及相关理论和实验技术也逐渐得到完善。

在众多发现和研究 miRNA 的方法中,生物信息学方法以其方便、快速和经济的优势受到许多研究组的青睐。实验证明,生物信息学方法是预测和发现新的 miRNA 的一个有效方法,它是以基因组序列和计算机程序鉴定为基础的方法^[9]。目前,通过各种计算机软件以及其他计算工具已经成功地预测和鉴定了动、植物中大多数 miRNAs^[10]。因此,生物信息学方法预测 miRNA 的技术创新,可以大大提高发现和研究 miRNA 的效率。

生物信息学方法主要是利用某种算法,通过设计筛选标准,建立评分系统,对候选序列进行选择,miRNAs 在植物中的高度保守性及其前体的二级结构特征为在其他物种中同源预测 miRNAs 提供了依据。伴随着物种测序工作的逐步开展和序列信息的日益丰富,利用生物信息学发现和预测 miRNAs 序列的方法已经成为简单、高效的策略之一^[11-13]。该方法预测 miRNAs 具有方便、快速和经济的优点,但是大部分策略仅能预测得到保守的 miRNAs 序列,同时也会出现假阳性序列,需要通过进一步的实验验证,来完善预测和研究。

1 miRNA 的预测策略

miRNA 预测的理论依据往往是利用 miRNA 物种间的保守性。在 miRNA 研究的发展历程中出现了众多优秀的预测方案,各自从不同的角度着手,利

用 miRNA 独特的性质完成预测工作。

E. C. Lai 等^[14]通过对 miRNA 保守性的研究发现,在形成发夹结构时,miRNA 臂(近乎)完美保守;而 miRNA* 发夹结构中另一条互补的臂只需保留与 miRNA 形成发夹结构的序列即可。E. Berezikov 等^[15]使用 phylogenetic shadowing 方法对 miRNA-miRNA* 区域(pre-miRNA 发夹结构)进行扫描,得到近千条候选 miRNA,超过 69 个家族,最终通过实验方法确定了 16 条成熟的 miRNA。

除了 miRNA 自身的保守性,其对应靶位点的保守性同样可以作为筛选候选 miRNA 的依据,植物 miRNA 会与靶基因完美或近似完美互补,从而对靶位点进行切割,而动物 miRNA 与靶位点互补性很低,只会造成靶基因的沉默,因此该方法只能局限于植物 miRNA 方面的研究。A. Adai 等^[16]将转录产物和包含有 miRNA 的非编码区域结合,并从中筛选出成熟 miRNA 与靶基因的结合。该方法的优点是,可以回避物种保守性带来的限制。同时它也存在一定缺点,即会产生数量巨大的候选序列,还需要进一步设置条件进行筛选。

之后,出现采用序列匹配和研究结构信息的方法,M. Legendre 等^[17]使用 ERPIN 程序捕获前体序列和二级结构信息,进而完成预测。X. Wang 等^[18]采用了同类型的 RNAforester 软件。这类软件可以探测到传统 BLAST 方法检测不到的信息。

I. Bentwich 等^[19]开发了一种新的预测和鉴定 miRNA 的方法。对于保守和非保守序列都可以使用,该方法是扫描人体基因组中全部的发夹结构,然后去除含有编码蛋白和重复覆盖的发夹结构,然后设置筛选条件,得到候选序列。

另一类途径是利用已知 mRNA 和 RNA 的结构作为突破口。X. Xie 等^[20]系统分析了人类、鼠、狗等物种的基因组,得到一系列位于启动子和 3'-UTR 区域内的修饰序列;并注意到这些修饰强烈倾向于 8 个碱基和以‘U’开始,并且 miRNA 往往以‘A’开始,而动物 miRNA 中心区域的长度也与修饰相似。因此,研究者猜想这些修饰作为 miRNA 的沉默位点,通过在全基因组范围进行 8-mer 修饰的配对检索,并对二级结构进行预测和分析,最终得到了 129 个候选 miRNA 序列。而 K. C. Miranda 等^[21]开发了一种模式匹配的方法预测 pre-miRNA 和 target,在成熟 miRNA 序列中发现了多种长度的修饰——patterns,然后设计算法,构建软件完成预测,这种方法有 75% 以上的成功率。

随着研究的深入,通过对热力学稳定性的分析,E. Bonnet 等^[22]发现 pre-miRNA 普遍含有较高的 MFEs,而 S. Washietl 等^[23]研究出另一种从整体上研究 RNA 的方法,后来,该研究组对非编码 RNA 开始感兴趣,在移除编码蛋白区域后,通过多物种保守序列评分和热力学结构稳定性计算(Z-score 和 MFE)得到候选序列。随着物种测序工作的逐步开展,miRBase 数据库的建立,以及生物信息学手段的不断丰富,该方法也日臻完善,最终形成一套完善的体系。

上述这些预测 miRNA 的策略都是从 miRNA 的自身性质出发,当更好地了解 Droscha 或 Dicer 这 2 种酶的作用机制之后,miRNA 的预测工作将会有进一步的提升,所以关于 Droscha 和 Dicer 的研究也一直是 miRNA 研究工作中的一个热点领域。

2 miRNA 的预测流程

表达序列标签(EST, expressed sequence tag)技术是生物信息学预测 miRNA 方法中一种比较常用而且相对成熟的手段。该方法是利用 miRNA 序列在物种间的高度保守性进行同源比较,然后对 miRNA 前体的二级结构的各项数据评分,设置筛选标准,最终获得候选 miRNA 序列,然后通过软件可以完成对靶基因的预测工作。预测流程如下。

首先,从 miRBase 数据库(<http://www.mirbase.org/>)中下载动物、植物或同源性较高物种的成熟 miRNA 序列作为预测的模板,由于 miRNA 的高度保守性,这些物种中的成熟 miRNA 序列包含有大量重复序列需要清除,可以使用 DNASTAR 5.0 中的 Assemble 功能去除成熟 miRNA 序列的重复。

其次,从 NCBI 或其他数据库中下载 EST 序列和 GSS 序列,作为候选序列。使用模式匹配软件-PatScan 将重复后的成熟 miRNAs 序列和 EST、GSS 序列进行模式匹配,要求碱基错配数在 3 以内,无插入和缺失。同时也可以使用 NCBI 下 BLAST 功能完成该过程。众所周知,miRNA 是一类非编码小 RNA 序列,因此提取筛选结果中的上下游各 600nt 序列,进行 BLASTX 分析,设置 E 值:1e-6,去除能够编码蛋白的序列,保留非编码序列。

最后,使用 RNA 结构预测和分析软件,对保留下来的序列进行筛选,并最终得到候选 miRNA 序列及其 miRNA 前体序列。筛选标准^[11-12,24]如下:(1)miRNA 前体能折叠成发夹结构;(2)成熟 miRNA 与相同家族已知 miRNA 位于同一臂上;(3)miRNA

前体二级结构要有绝对值较高的最小折叠自由能(MFE, minimal folding free energy)和较高的最小折叠自由能系数(MFEI, minimal folding free energy index);(4)miRNA 与 miRNA* 的碱基错配数小于等于 4,无缺口或较大的环。完成该步骤比较常用的软件有 RNAhybrid、RNAfold 和在线软件 Mfold(<http://www.bioinform.rpi.edu/applications/mfold/old/rna>)。RNAhybrid 的核心算法是在一个大的潜在靶基因 RNA 分子中需找能量上有力的杂交位点,可以完成在大的靶 RNA 分子中预测 miRNA 的多个潜在结合位点^[25]。RNAfold 是 Vienna RNA Package(<http://www.tbi.univie.ac.at/~ivo/RNA/index.html>)里的一种通过计算 MFE 来预测 RNA 二级结构的软件。它的算法和理论支持主要来源于 M. Zuker 等^[26],Mfold 也是采用 Zuker 算法作为核心,使用该程序可以完成对植物 miRNA 前体 MFE, miRNA 前体的长度,碱基 A、C、G、U、(A + U)的含量百分比以及最小折叠自由能系数(MFEI),同时,可以将 miRNA 前体结构图形化,界面友好,操作便利,受到广大研究者的青睐。

3 靶基因的确认

研究 miRNA 功能的最关键一步就是 miRNA 靶基因的确认工作。植物 miRNA 序列与靶基因之间完美(近乎完美)互补的方式,使得靶基因的预测工作变得相对简单。目前为止,已有众多成熟的靶基因预测软件可供研究人员选择,如:PatScan, miRU, miRNAassist, psRNAtarget, TargetAlign, TargetFinder 等。

PatScan 是最早用于靶基因预测软件之一^[27],它主要是以 miRNA 与靶基因的互补性为基础,研究者可以自行定义模式匹配的要求,包括错配数范围,是否允许 GU 配对,是否允许插入与缺失的存在,从而获得预测结果,但是该方法缺乏对于大量系统的靶位点分析能力。

F. L. Xie 等^[10]开发了以 BLAST 为基础的靶位点预测程序,该程序操作简单,学习起来相对容易,但是使用 BLAST 研究 miRNA 的靶位点却存在一定的争议,原因在于它不能够区分 GU 摆动,容易导致预测结果的缺失,影响结果的完整性。

Y. Zhang^[28]为研究 miRNA 靶基因,开发了一种在线工具——miRU,它是以 Smith-Waterman 算法为基础设计出的软件。miRU 可以完成对潜在靶基因的预测工作,但存在一定的局限性,可供选择的数据库太少,仅有少数模式植物的基因组序列供使用,无

法采用用户提供的候选序列作为研究对象。为弥补 Smith-Waterman 算法的不足, X. Dai 等^[29]在此基础上开发了一种高通量 miRNA 靶基因预测软件, 命名为 psRNA target。该软件具有 2 大功能特点: (1) 通过以证明的评分方式将 miRNA 和 mRNA 进行互补配对; (2) 通过计算非配对能 (UPE, unpaired energy) 评价靶位点亲和性。同时, 该程序可以由用户自己提供 miRNA 序列或 mRNA 序列, 或者使用在线提供的数据库, 进行配对分析, 寻找是否有合适的靶基因位点。

为改善对 miRNA 靶基因的预测工作, F. L. Xie 等^[30]开发出一类 Smith-Waterman 算法的预测软件——Target align。该软件使用 Dynamic 程序作为评分机制, 研究者可以根据自身要求, 灵活地设置参数 (图 1)。与其他预测软件相比, 该程序具有很强的敏感性和精确度。

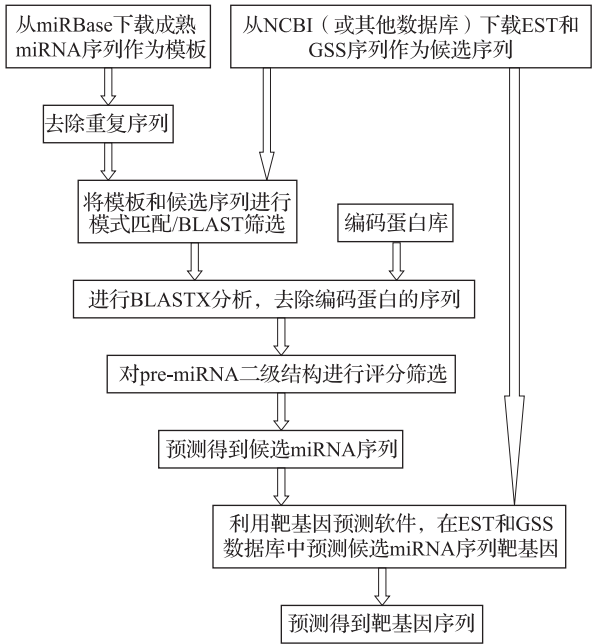


图 1 生物信息学预测 miRNA 的流程

Fig. 1 Flow of miRNA identification by bioinformatics

4 结论与展望

miRNA 已成为分子生物学领域的研究热点。本文介绍了几种发现和研究 miRNA 的方法, 重点阐述了生物信息学预测 miRNA 方法的流程和策略, 并对各过程中采用的方法及信息学软件的优劣性进行比较和分析, 为预测 miRNA 提供了新的思路。近年来, 生物信息方法因其快速、高效、经济的优势, 已经成为预测获得 miRNA 及其靶基因的强有力工具。随着对 miRNA 形成机制的不断研究, 以及与靶位点

作用方式的深入了解, miRNA 的预测效率和方法将不断改善, 成熟的 miRNA 预测流程也将极大提高研究工作的效率。尽管 miRNA 生物信息学预测工具不断地更新和完善, 但先进且综合的植物 miRNA 分析工具还比较缺乏, 而且研究中, 算法、工具以及软件运用方式的差异也会引起植物 miRNA 预测和分析的不确定^[31], 因此, 合理选择和运用先进的预测分析方法, 伴随生物信息资源的日益丰富, 生物信息学方法预测和分析植物 miRNA 将会取得更大发展。

参考文献

[1] Bartel D. MicroRNAs: genomics, biogenesis, mechanism, and function [J]. Cell, 2004, 116: 281-297

[2] Lee R C, Feinbaum R L, Ambros V. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14 [J]. Cell, 1993, 75 (5): 843-854

[3] Reinhart B J, Weinstein E G, Rhoades M W, et al. MicroRNAs in plants [J]. Genes Dev, 2002, 16: 1616-1626

[4] Jones-Rhoades M W, David P, Bartel D P. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA [J]. Molecul Cell, 2004, 787-799

[5] Lee Y, Ahn C, Han J, et al. The nuclear rnase III drosha initiates microRNA processing [J]. Nature, 2003, 425: 415-419

[6] Denli A M, Tops B B, Plasterk R H, et al. Processing of primary microRNAs by the microprocessor complex [J]. Nature, 2004, 432: 231-235

[7] Gregory R I, Yan K P, Amuthan G, et al. The microprocessor complex mediates the genesis of microRNAs [J]. Nature, 2004, 432: 235-240

[8] Yekta S, Shih I H, Bartel D P. MicroRNA-directed cleavage of HOXB8 mRNA [J]. Science, 2004, 304: 594-596

[9] Li W, Ruan K. MicroRNA detection by microarray [J]. Anal Bioanal Chem, 2009, 394 (4): 1117-1124

[10] Xie F L, Huang S Q, Guo K, et al. Computational identification of novel microRNAs and targets in *Brassica napus* [J]. Febs Letters, 2007, 581 (7): 1464-1474

[11] Zhang B H, Pan X P, Wang Q L, et al. Identification and characterization of new plant microRNAs using EST analysis [J]. Cell Res, 2005, 15 (5): 336-360

[12] Zhang B H, Pan X P, Stellwag E J. Identification of soybean microRNAs and their targets [J]. Planta, 2008, 229 (1): 161-182

[13] Taylor P F, Zhang B H. Identification of plant microRNAs using expressed sequence tag analysis [J]. Meth Molecul Biol, 2011, 678: 13-25

[14] Lai E C, Tomancak P, Williams R W, et al. Computational identification of Drosophila microRNA genes [J]. Genome Biol, 2003, 4 (7): R42

[15] Berezikov E, Guryev V, Belt J V D, et al. Phylogenetic shadowing and computational identification of human microRNA genes [J]. Cell, 2005, 120: 21-24

[16] Adai A, Johnson C, Mlotshwa S, et al. Computational prediction of miRNAs in *Arabidopsis thaliana* [J]. Genome Res, 2005, 15 (1): 78-91

[17] Legendre M, Gautheret D. Sequence determinants in human polyadenylation site selection [J]. BMC Genomics, 2003, 4 (1): 7

[18] Wang X, Zhang J, Li F, et al. MicroRNA identification based on sequence and structure alignment [J]. Bioinformatics, 2005, 21 (18): 3610-3614

大豆 *PEBP* 基因家族的初步分析

张礼凤¹, 徐 冉¹, 张彦威¹, 戴海英¹, 王彩洁¹, 李 伟¹, 张 军^{1,2}

(¹ 山东省农业科学院作物研究所, 济南 250100; ² 滨州职业技术学院, 山东滨州 256603)

摘要: 磷脂酰乙醇胺结合蛋白 (PEBP, phosphatidyl ethanolamine-binding protein) 基因家族在动物、植物和微生物中广泛存在, 在控制植物开花和种子休眠中起重要作用。本研究对大豆 *PEBP* 基因家族进行了分析, 发现了 27 个大豆 *PEBP* 基因的候选序列, 其中 16 个具有完整 PEBP 结构域的全长序列被认为是大豆 *GmPEBP* 家族基因。*GmPEBP* 基因分布在 9 条染色体上, 基因结构高度保守。通过系统发生分析, 可将大豆 *GmPEBP* 基因家族成员分为 FT-like、TFL1-like 和 MFT-like 3 个亚族, 并且发现 *GmPEBP* 家族成员数目按照大豆物种特异性的方式进行了扩张。对重复基因的 *K_s* 分析表明, 绝大多数重复基因主要由 5900 万年前和 1300 万年前的大豆基因组复制所致。

关键词: 大豆; *PEBP* 基因家族; 进化分析

Preliminary Analysis of the *PEBP* Gene Family in Soybean (*Glycine max*)

ZHANG Li-feng¹, XU Ran¹, ZHANG Yan-wei¹, DAI Hai-ying¹, WANG Cai-jie¹, LI Wei¹, ZHANG Jun^{1,2}

(¹ Crop Institute, Shandong Academy of Agricultural Sciences, Ji'nan 250100;

² Binzhou Vocational College, Binzhou Shandong 256603)

Abstract: The *PEBP* gene family encoding proteins with similarity to phosphatidyl ethanolamine binding proteins was found in the animals, plants, and microbes. The *PEBP* genes played important roles on regulation of flowering time and seed dormancy. We presented here a systematic analysis of the *PEBP* gene family in soybean (*Glycine max*). At the level of genome-wide screening, we identified 27 *PEBP* candidate genes from soybean and considered 16 genes with full length of coding sequence as *GmPEBP* genes. Sequence analysis showed the *GmPEBP* genes had conserved genomic structures and were distributed on 9 soybean chromosomes. Phylogenetic analysis showed that the *GmPEBP* family could be grouped into three subfamilies, named FT-like, TFL1-like, and MFT-like, and suggested that species-specific expansion contributed to the evolution of this family. Analysis of synonymous substitution rates indicated most duplicated pairs were from the whole genome duplication occurred at 59 and 13 million years ago.

Key words: soybean; *PEBP* gene family; phylogenetic analysis

磷脂酰乙醇胺结合蛋白 (PEBP, phosphatidyl ethanolamine-binding protein) 基因家族广泛存在于动物、植物和微生物中, 均具有保守的 PBP (PF01161) 结构域。在模式植物拟南芥中共发现了 6 个 PEBP 成员, 分别为 *FT* (*Flowering Locus T*)、*TFL1* (*Terminal Flower1*)、*TSF* (*Twin Sister of FT*)、*MFT* (*Mother of FT* and *TFL1*)、*BFT* (*Brother of FT and TFL1*) 和 *ATC* (*Arabidopsis thaliana CENTRORADIALIS homologue*)。*FT* 作为成花素在叶片中表达, 经韧皮部输送至茎分生组织, 与 *FD* (*Flowering Lotus D*) 互作, 激活花分生组织基因 *API* 表达, 从而促进开花^[1-3]。*TSF* 的功能与 *FT* 基因相似, 过量表达 *TSF* 的拟南芥开花提前^[4]。

TFL1 与 *FT* 功能相反,抑制茎顶端分生组织形成花原基,延迟开花,维持无限生长习性^[5-6]。*MFT* 主要在种子中表达,通过调控 ABA 和 GA 信号途径,在种子发育和萌发中起重要作用^[7-8]。在短日照植物水稻中,*Hd3a* (*Heading date 3a*) 为 *FT* 的同源基因。在水稻和拟南芥中过量表达 *Hd3a* 基因,转基因植株的开花期均比野生型明显提前^[9-10]。

大豆对光周期反应敏感,是典型的短日照植物。栽培大豆主要有 2 种生长类型,一类为有限生长习性,另一类为无限生长习性^[11]。遗传分析表明有限和无限生长习性主要受 1 对基因 (*dt1/dt1* 或 *Dt1/Dt1*) 控制。*Dt1* (Glyma19g37890.1) 编码蛋白含有 PEBP 结构域,转入拟南芥 *tf1* 突变体 (开花提前),能够恢复开花延迟表型,因而被认为是拟南芥 *TFL1* 的同源基因^[12]。本研究旨在通过对大豆 *PEBP* 家族基因的生物信息学和分子进化研究,为进一步明确其功能奠定基础。

1 材料与方法

1.1 *GmPEBP* 基因的鉴定

根据拟南芥中已知的 6 个 *PEBP* 基因编码的蛋白序列,利用 BlastP 软件检索从 Phytozome v 9.1 下载的大豆基因组数据库 (<http://www.phytozome.net/soybean.php>),去除冗余后得到的序列被认为是蛋白候选序列,经 Pfam (<http://pfam.janelia.org/>) 软件分析后,将含有 PEBP 结构域的蛋白序列作为 *GmPEBP* 基因家族成员。

1.2 *GmPEBP* 基因的进化分析

利用 Clustal X 软件对大豆 *GmPEBP* 基因进行多序列比对,利用 MEGA 4.0 软件中邻近算法 (NJ, neighbor-joining) 构建系统发生树,bootstrap 值设为 1000。*GmPEBP* 同源基因的 K_s 值采用 PAML 4 软件计算^[13]。利用 K_s 估算重复基因的片段重复事件发生的时间,计算方法为: $T = K_s / 2\lambda$, λ 为分子替换速率。大豆的 λ 采用 5.17×10^{-9} 个替换/同义替换位点/年^[14]。

1.3 *GmPEBP* 基因的电子表达分析

大豆 SoySeq 数据库 (<http://soybase.org/soyseq/>) 中贮存了 14 种大豆组织的 RNA-Seq 序列。利用 *GmPEBP* 基因的编码序列检索 SoySeq 数据库,得到 *GmPEBP* 基因在 14 种组织中的表达谱^[15]。

2 结果与分析

2.1 大豆 *GmPEBP* 家族基因的挖掘

利用拟南芥中已知的 *PEBP* 基因对大豆基因

组数据库进行检索,共发现 27 条候选序列,经 Pfam 分析,其中 11 条序列的 PEBP 结构域不完整,分别为 Glyma02g07650.1、Glyma05g34030.2、Glyma08g05651.1、Glyma08g47816.1、Glyma08g47823.1、Glyma12g30940.2、Glyma12g30940.3、Glyma13g39360.2、Glyma18g53680.2、Glyma18g53680.3 和 Glyma20g02845.1。其余 16 条含有完整 PEBP 结构域的全长序列被认为是大豆 *GmPEBP* 基因,编码 172 ~ 190 个氨基酸,但大部分 *GmPEBP* 蛋白含有的氨基酸在 175 个左右 (表 1)。

大豆 *GmPEBP* 基因的结构高度保守,16 个 *GmPEBP* 基因均含有 3 个内含子、4 个外显子,其中第 2 和第 3 外显子的长度亦非常保守,均为 62 bp 和 41 bp,第 1 和第 4 外显子的保守性较弱,长度范围分别为 189 ~ 249 bp 和 218 ~ 230 bp。内含子的长度差异很大,Glyma16g26660.1 的第 3 内含子长度最长,为 3154 bp, Glyma16g32080.4 的第 2 内含子长度最短,仅有 88 bp。内含子的长度差异直接导致 *GmPEBP* 基因的长度差异十分明显,最短的 Glyma16g32080.4 长度为 1002 bp,最长的 Glyma18g53680.1 长度为 4999 bp。

大豆 *GmPEBP* 基因分布在 9 条染色体上,其中 16 号染色体上含有 5 个 *GmPEBP* 基因,19 号染色体有 3 个 *GmPEBP* 基因,18 号染色体上有 2 个 *GmPEBP* 基因,3、5、8、9、10 和 13 号染色体上各有 1 个 *GmPEBP* 基因。

2.2 大豆 *GmPEBP* 家族基因的进化分析

为了确定大豆基因组中 *GmPEBP* 基因家族的分子进化规律和系统发生关联,本研究利用邻接法构建了大豆 *GmPEBP* 蛋白和拟南芥中 6 个已知 PEBP 蛋白的系统发生树,发现 *GmPEBP* 基因分为 3 个亚家族,即 FT-like、TFL1-like 和 MFT-like。其中 FT-like 亚家族包含 9 个 *GmPEBP* 基因,TFL1-like 亚家族包含 6 个 *GmPEBP* 基因,而 MFT-like 只含有 1 个 *GmPEBP* 基因 (图 1)。每个亚家族中都含有拟南芥和大豆的 *PEBP* 基因,表明大豆和拟南芥拥有共同的 *PEBP* 基因祖先。在 FT-like 亚家族中,拟南芥含有 2 个基因,大豆含有 9 个 *GmPEBP* 基因,其中 4 对 *GmPEBP* 基因为旁系同源基因;在 TFL1-like 亚家族中,拟南芥含有 3 个基因,大豆含有 6 个 *GmPEBP* 基因,为 3 对旁系同源基因,表明 FT-like 和 TFL1-like 亚家族在大豆和拟南芥分离之后按照大豆物种特异性的方式进行了扩张。

表 1 大豆 *GmPEBP* 基因的鉴定

Table 1 Identification of *GmPEBP* genes in soybean

基因 Gene	外显子 1 (bp) Exon1	内含子 1 (bp) Intron1	外显子 2 (bp) Exon2	内含子 2 (bp) Intron2	外显子 3 (bp) Exon3	内含子 3 (bp) Intron3	外显子 4 (bp) Exon4	蛋白长度 (aa) Protein length	染色体 Chromosome
Glyma16g04840.1	201	181	62	110	41	1481	224	175	16
Glyma19g28390.1	201	96	62	105	41	1507	224	175	19
Glyma16g26660.1	201	792	62	127	41	3154	227	176	16
Glyma16g26690.1	201	187	62	131	41	2165	227	176	16
Glyma08g47810.1	201	122	62	680	41	169	224	175	8
Glyma18g53680.1	201	166	62	2987	41	1315	227	176	18
Glyma18g53690.1	189	145	62	1441	41	179	230	173	18
Glyma16g04830.1	195	133	62	391	41	231	221	172	16
Glyma19g28400.1	195	142	62	375	41	1601	221	172	19
Glyma09g26550.1	201	117	62	93	41	327	218	173	9
Glyma16g32080.4	198	121	62	88	41	274	218	172	16
Glyma03g35250.1	201	210	62	472	41	99	218	173	3
Glyma19g37890.1	201	215	62	483	41	95	218	173	19
Glyma10g08340.1	201	186	62	500	41	117	221	174	10
Glyma13g22030.1	198	147	62	486	41	1892	221	173	13
Glyma05g34030.1	249	94	62	92	41	993	221	190	5

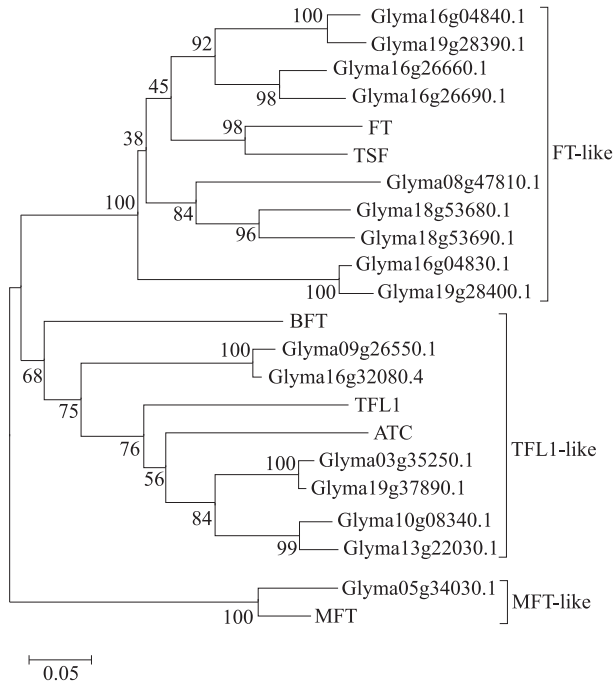


图 1 大豆 *GmPEBP* 家族基因的进化分析

Fig.1 Phylogenetic analysis of soybean *GmPEBP* gene family

2.3 大豆 *GmPEBP* 基因的多序列比对

在 TFL1-like 亚家族中,6 个 *GmPEBP* 基因的序列相似性在 70.52%~97.11% 之间。旁系同源基因 Glyma03g35250.1-Glyma19g37890.1、Glyma09g26550.1-Glyma16g32080.4 和 Glyma10g08340.1-Glyma13g22030.1 的序列相似性很高,分别为 97.11%、97.11% 和 93.10%,3 对大豆旁系同源基因与其在拟南芥中的直系同源基因 *TFL1* 的序列相似性分别为 75.87%、69.94% 和 64.11%。在 *TFL1* 中,决定其功能的关键氨基酸为第 88 位的组氨酸(His88),其在 6 个大豆成员中亦高度保守(图 2)。

在 FT-like 亚家族中,9 个 *GmPEBP* 基因的序列相似性的变异范围较大,为 57.99%~96.51%,与拟南芥中的 *FT* 基因的序列相似性在 63.53%~75.00% 之间。4 对旁系同源基因 Glyma16g04830.1-Glyma19g28400.1、Glyma16g04840.1-Glyma19g28390.1、Glyma16g26660.1-Glyma16g26690.1 和 Glyma18g53680.1-Glyma18g53690.1 的序列相似性较高,分别为 96.51%、94.29%、91.43% 和 76.02%。在决定 *FT* 基因功能的第 85 位关键氨基酸位点上,9 个大豆 *GmPEBP* 基因都呈现为酪氨酸(Tyr85)(图 2)。

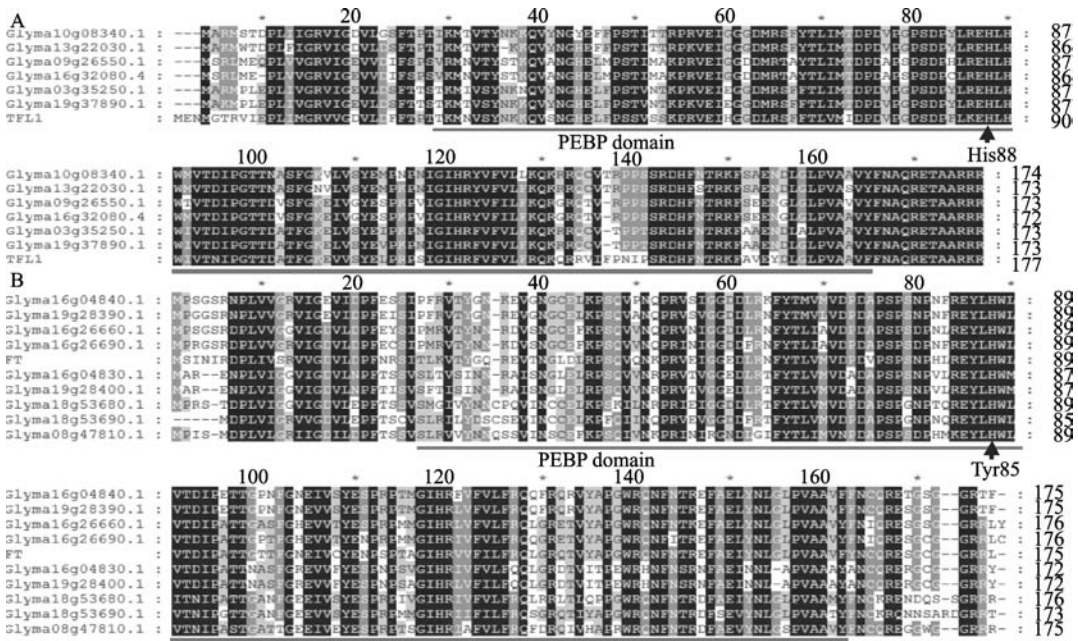


图 2 大豆 *GmPEBP* 基因 TFL1-like (A) 和 FT-like (B) 亚家族的蛋白质序列对比

Fig. 2 Alignment of *GmPEBP* protein sequences in TFL1-like (A) and FT-like (B) subfamilies

2.4 大豆 *GmPEBP* 片段重复事件发生时间分析

本研究中,利用同义替换率(K_s)来估算大豆 *Gm-PEBP* 旁系同源基因的片段重复事件发生的时间(表

2)。在 TFL1-like 亚家族中, Glyma09g26550. 1-Glyma16g32080. 4、Glyma03g35250. 1-Glyma 19g37890. 1 和 Glyma10g08340. 1-Glyma13g22030. 1 重复基因对

表 2 大豆 *GmPEBP* 基因旁系同源基因的片段重复事件发生的时间

Table 2 Estimation of the date of duplication events in *GmPEBP* paralogs

亚家族	重复对	同义位点数	非同义位点数	同义替换率	时间 (MYA)
Subfamily	Duplicated pair	S	N	K_s	Date
FT-like	Glyma16g04840. 1-Glyma19g28390. 1	126. 8	398. 2	0. 12	11. 61
	Glyma16g26660. 1-Glyma16g26690. 1	137. 9	390. 1	0. 32	30. 95
	Glyma18g53680. 1-Glyma18g53690. 1	131. 4	384. 6	0. 50	48. 36
	Glyma16g04830. 1-Glyma19g28400. 1	135. 4	380. 6	0. 07	6. 77
	Glyma16g04840. 1-Glyma16g26660. 1	122. 4	402. 6	0. 81	78. 34
	Glyma16g04840. 1-Glyma16g26690. 1	121. 7	403. 3	0. 76	73. 50
	Glyma19g28390. 1-Glyma16g26660. 1	126. 1	398. 9	0. 72	69. 63
TFL1-like	Glyma19g28390. 1-Glyma16g26690. 1	128. 0	397. 0	0. 71	68. 67
	Glyma09g26550. 1-Glyma16g32080. 4	141. 9	374. 1	0. 15	14. 51
	Glyma03g35250. 1-Glyma19g37890. 1	156. 1	362. 9	0. 03	2. 90
	Glyma10g08340. 1-Glyma13g22030. 1	137. 1	381. 9	0. 07	6. 77
	Glyma03g35250. 1-Glyma10g08340. 1	135. 2	383. 8	0. 41	39. 65
	Glyma03g35250. 1-Glyma13g22030. 1	133. 3	382. 7	0. 47	45. 45
	Glyma19g37890. 1-Glyma10g08340. 1	136. 3	382. 7	0. 42	40. 62
	Glyma19g37890. 1-Glyma13g22030. 1	132. 7	383. 3	0. 48	46. 42

内的 K_s 值较低,分别为 0.15、0.03 和 0.07,预测片段重复时间发生在距今 1451 万年、290 万年和 677 万年之前;而重复基因对之间 K_s 值较高的 Glyma03g35250.1-Glyma10g08340.1、Glyma03g35250.1-Glyma13g22030.1、Glyma19g37890.1-Glyma10g08340.1 和 Glyma19g37890.1-Glyma13g22030.1 分别为 0.41、0.47、0.42 和 0.48,预测片段重复时间发生在距今 3965 万年、4545 万年、4062 万年和 4642 万年之前。在 FT-like 亚家族中, Glyma16g04840.1-Glyma19g28390.1、Glyma16g26660.1-Glyma16g26690.1、Glyma18g53680.1-Glyma18g53690.1 和 Glyma16g04830.1-Glyma19g28400.1 重复基因对内的 K_s 值同样较低,分别为 0.12、0.32、0.50 和 0.07,而重复基因对之间的 K_s 值较高的 Glyma16g04840.1-Glyma16g26660.1、Glyma16g04840.1-Glyma16g26690.1、Glyma19g28390.1-Glyma16g26660.1 和 Glyma19g28390.1-Glyma16g26690.1 分别为 0.81、0.76、0.72 和 0.71。

表 3 大豆 *GmPEBP* 基因的电子表达谱

Table 3 Expression profiles of soybean *GmPEBP* genes based on *in silico* analysis

基因 Gene	嫩叶 Young leaf	花 Flower	荚 Pod	荚皮 Pod shell 10DAF	荚皮 Pod shell 14DAF	种子 Seed 10DAF	种子 Seed 14DAF	种子 Seed 21DAF	种子 Seed 25DAF	种子 Seed 28DAF	种子 Seed 35DAF	种子 Seed 42DAF	根 Root	根瘤 Nodule
Glyma19g28400.1				2		2								
Glyma19g37890.1	1						1							1
Glyma16g26660.1	3	20	4	11	7	1								
Glyma16g04830.1		3	2	4	3	6	14							
Glyma16g32080.4		48	2	4			1	7	10	21	29	11		
Glyma18g53680.1						2	1							
Glyma09g26550.1		4	4		1	2			1		2		6	
Glyma03g35250.1				3	1	1				1			1	
Glyma05g34030.1	2	1	1			32	211	190	795	1064	1233	1097	3	4

DAF:Days after flowering

3 讨论

PEBP 基因在植物中属于一个较小的基因家族,在植物的成花转变、种子发育和萌发中发挥重要作用。目前,已在拟南芥、水稻、玉米和大豆等多种植物中分离了众多 *PEBP* 基因。拟南芥中含有 6 个 *PEBP* 成员,其中 *FT*、*TFL1* 和 *MFT* 的功能比较清楚。*FT* 基因促进开花,*TFL1* 基因抑制开花,*MFT* 则在种子萌发中发挥作用。水稻中含有 19 个 *PEBP*

2.5 大豆 *GmPEBP* 的电子表达谱

目前 SoySeq 数据库 (<http://soybase.org/soyseq/>) 中贮存了 14 个不同大豆组织的 RNA-Seq 序列,为在 mRNA 水平上研究大豆 *GmPEBP* 基因的表达提供了数据基础。利用 *GmPEBP* 基因检索 SoySeq 数据库,发现 7 个 *GmPEBP* 基因 (Glyma08g47810.1、Glyma19g28390.1、Glyma16g04840.1、Glyma16g26690.1、Glyma18g53690.1、Glyma13g22030.1 和 Glyma10g08340.1) 在 14 个大豆组织中没有检测到匹配的 RNA-Seq 序列,其余 9 个 *GmPEBP* 基因在多种组织中都有表达,但表达丰度差异很大 (表 3)。由表 3 可以看出, Glyma05g34030.1 (MFT-like 亚家族) 除了荚皮中没有表达以外,在其余 12 个组织中都有表达,尤其在种子中的表达量很高,且随着种子的发育进程表达量逐渐增高。Glyma16g26660.1 主要在花和荚皮中表达,而 Glyma16g32080.4 主要在花和种子中表达。

基因,玉米中至少含有 25 个 *PEBP* 成员^[16]。本研究利用拟南芥 *PEBP* 蛋白序列为探针,对大豆的基因组数据库 (<http://www.phytozome.net/soybean.php>) 进行检索,共获得了 16 个具有完整 *PEBP* 结构域的全长序列,分布在大豆 9 条染色体上。

进化分析表明,大部分植物 *PEBP* 基因分为 3 个亚家族,即 FT-like、TFL1-like 和 MFT-like。在本研究中,16 个 *GmPEBP* 基因亦归属于 3 个亚家族,其中 FT-like 亚家族中成员最多,含有 9 个 *GmPEBP* 基因;TFL1-like 成员次之,含有 6 个 *GmPEBP* 基因;

而 MFT-like 亚家族成员最少,只有 1 个 *GmPEBP* 基因。O. N. Danilevskaya 等^[17]发现玉米中含有 25 个 *PEBP* 成员,其中属于 FT-like 亚家族的基因有 15 个,TFL1-like 亚家族的基因有 6 个,MFT-like 亚家族的基因有 3 个,另外 1 个 *PEBP* 成员因只含有第 4 外显子而被认为是假基因,其中 *ZCN8* 为 *FT* 基因的同源基因。

Z. Tian 等^[12]鉴定了 4 个大豆 *TFL1* 同源基因,分别为 Glyma03g35250. 1、Glyma19g37890. 1、Glyma10g08340. 1 和 Glyma13g22030. 1,其中 Glyma19g37890. 1 (*GmTfl1*) 位于控制无限结莢习性的 *Dt1* 位点内,主要在幼根、幼叶和花中表达,而 Glyma13g22030. 1 主要在幼根、幼茎和芽中表达,在花中没有表达。在拟南芥 *tfl1-1* 突变体(开花提前)中表达 Glyma19g37890. 1,转基因拟南芥与野生型相似,表现出开花推迟的特征。F. Kong 等^[18]发现了 10 个 *FT* 同源基因,其中 *GmFT2a* (Glyma16g26 660. 1) 和 *GmFT5a* (Glyma16g04830. 1) 在叶、茎、茎尖、子叶、根、花芽和未成熟种子中都有较强表达,尤其在叶片中表达量更高。过量表达 *GmFT2a* 或 *GmFT5a* 的拟南芥开花提前,预示其具有 *FT* 基因的功能。C. Fan 等^[19]在大豆基因组中鉴定了 6 个 *FT* 同源基因(*GmFTL1-6*)和 4 个 *TSF* 同源基因(*GmTSF1-4*),发现多数 *GmFTL* 在营养生长阶段的叶、根和茎尖等组织中的表达较低,而在生殖生长阶段的叶、根、花和茎尖等组织中的表达增强。过量表达 *GmFTL1-6* 的拟南芥花期提前,表明 6 个 *GmFTL* 基因都具有 *FT* 基因的功能,而过量表达 *GmTSF3*、*GmTSF4* 和 *GmPEBP21* 的拟南芥的花期并没有改变。本研究中,Glyma16g26660. 1 主要在花和荚皮中表达,表明 Glyma16g26660. 1 不但具有 *FT* 基因促进开花的功能,同时可能在荚皮发育中亦起作用;Glyma16g32080. 4 主要在花和种子中表达,目前功能尚不明确,预测其在花期调控和种子发育中起作用;Glyma05g34030. 1 主要在种子中表达,且随发育进程表达提高,预测其在种子发育和萌发中起作用。

PEBP 家族基因的结构非常保守,大都含有 3 个内含子和 4 个外显子,尤其第 2 和第 3 外显子的长度高度保守,分别为 62 bp 和 41 bp。本研究中的 16 个 *GmPEBP* 基因,都含有 4 个外显子和 3 个内含子,外显子 2 和外显子 3 也均为 62 bp 和 41 bp,而在 25 个玉米 *PEBP* 成员中,有 2 个基因(*ZCN2* 和 *ZCN20*)的结构发生变异,外显子 1 和 2 发生融合,只含有 2 个内含子和 3 个外显子^[17]。

FT 和 *TFL1* 的氨基酸序列相似性较高,但功能相反。*TFL1* 蛋白第 88 位的组氨酸残基(His88)和 *FT* 蛋白第 85 位的酪氨酸残基(Tyr85)对其相反功能起决定作用^[20]。蛋白质结构分析表明,*FT* 和 *TFL1* 蛋白的晶体结构高度相似,形成一个口袋状的结构,在口袋入口处,*TFL1* 的 His88/Asp144 能形成氢键,而 *FT* 的 Tyr85/Gln140 不能形成氢键,由此造成 *TFL1* 和 *FT* 的功能差异^[21]。在 *TFL1*-like 亚家族中,6 个 *GmPEBP* 成员在 His88 上高度保守,均为组氨酸;在 *FT*-like 亚家族中,9 个 *GmPEBP* 成员在 Tyr85 上亦高度保守,均为酪氨酸(图 2)。

基因重复对基因家族的进化具有重要作用,可以产生新基因、新功能或者亚功能。基因的复制主要通过片段重复、串联重复和逆转录转座及其他转座事件来实现。在植物基因组中,最重要的基因扩张模式为串联重复和长片段重复^[22]。本研究中,4 对 *GmPEBP* 基因 (Glyma16g04830. 1-Glyma16g04840. 1、Glyma16g26660. 1-Glyma16g26690. 1、Glyma18g 53680. 1-Glyma18g53690. 1 和 Glyma19g28390. 1-Glyma19g28400. 1) 出现在染色体的同一区段或相邻区段,为串联重复事件所致。3 对 *GmPEBP* 基因 (Glyma09g26550. 1-Glyma16g32080. 4、Glyma03g35250. 1-Glyma19g37890. 1 和 Glyma10g08340. 1-Glyma13g22030. 1) 出现在 2 条染色体的同源区,可能为片段重复事件所致。研究表明大豆基因组在大约 59 百万年前和 13 百万年前经历了 2 次基因组复制,大豆第 1 次基因组复制发生的时间与被子植物中的蝶形花亚科(papilionoideae)的起源时间相邻(约 59 百万年),第 2 次基因组复制发生的时间恰为大豆特有的异源四倍体化的时间(allotetraploidy event)^[14]。同源基因 K_s 值在 0.06 ~ 0.39(均值为 0.13)之间,表明基因复制来自于 13 百万年前的基因组复制,而 K_s 值在 0.40 ~ 0.80(均值为 0.59)之间,则表明基因复制来自于 59 百万年前的基因组复制^[14]。本研究中来自 *TFL1*-like 亚家族中的 3 对重复基因对内的 K_s 值均远低于 0.39,而重复基因对之间的 K_s 值在 0.40 ~ 0.80 之间(表 2),表明 *TFL1*-like 亚家族中的重复基因对内的基因复制来自于 13 百万年前的大豆基因组异源四倍体化过程,而重复基因对之间的基因复制来自于 59 百万年前的基因组复制。来自 *FT*-like 亚家族的 4 对重复基因中, Glyma18g53680. 1-Glyma18g53690. 1 的 K_s 值处于 0.40 ~ 0.80 之间,表明片段重复发生于 59 百万年前的基因组复制,而其余 3 对重复基因对内的 K_s 值均处于 0.06 ~ 0.39 之间,

表明片段重复发生于 13 百万年前的大豆基因组异源四倍体化过程。由此,推测 *GmPEBP* 基因在大豆中的扩张可能与 2 次基因组复制有关。

参考文献

[1] Abe M, Kobayashi Y, Yamamoto S, et al. FD, a bZIP protein mediating signals from the floral pathway integrator *FT* at the shoot apex [J]. *Science*, 2005, 309: 1052-1056

[2] Wigge P A, Kim M C, Jaeger K E, et al. Integration of spatial and temporal information during floral induction in *Arabidopsis* [J]. *Science*, 2005, 309: 1056-1059

[3] Cobesier L, Vincent C, Jang S, et al. FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis* [J]. *Science*, 2007, 316: 1030-1033

[4] Yamaguchi A, Kobayashi Y, Goto K, et al. *TWIN SISTER OF FT (TSF)* acts as a floral pathway integrator redundantly with *FT* [J]. *Plant Cell*, 2005, 46: 1175-1189

[5] Bradley D, Ratcliffe O, Vincent C, et al. Inflorescence commitment and architecture in *Arabidopsis* [J]. *Science*, 1997, 275: 80-83

[6] Ratcliffe O J, Amaya I, Vincent C A, et al. A common mechanism controls the life cycle and architecture of plants [J]. *Development*, 1998, 125: 1609-1615

[7] Xi W, Liu C, Hou X, et al. *MOTHER OF FT AND TFLI* regulates seed germination through a negative feedback loop modulating ABA signaling in *Arabidopsis* [J]. *Plant Cell*, 2010, 22: 1733-1748

[8] Nakamura S, Abe F, Kawahigashi H, et al. A wheat homolog of *MOTHER OF FT AND TFLI* acts in the regulation of germination [J]. *Plant Cell*, 2011, 23: 3215-3229

[9] Kojima S, Takahashi Y, Kobayashi Y, et al. *Hd3a*, a rice ortholog of the *Arabidopsis FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions [J]. *Plant Cell Physiol*, 2002, 43: 1096-1105

[10] 孙洪波, 贾贞, 韩天富. *PEBP* 家族基因在植物发育调控中的作用 [J]. *植物生理学通讯*, 2009, 45(8): 739-747

[11] 胡国玉, 李杰坤, 黄志平, 等. 不同结荚习性夏大豆种质的农艺表现及其与产量的相关分析 [J]. *植物遗传资源学报*, 2014, 15(2): 417-422

[12] Tian Z, Wang X, Lee R, et al. Artificial selection for determinate growth habit in soybean [J]. *Proc Natl Acad Sci USA*, 2010, 107: 8563-8568

[13] Yang Z. PAML 4: phylogenetic analysis by maximum likelihood [J]. *Mol Biol Evol*, 2007, 24: 1586-1591

[14] Schmutz J, Cannon S B, Schlueter J, et al. Genome sequence of the palaeopolyploid soybean [J]. *Nature*, 2010, 463: 178-183

[15] Severin A J, Woody J L, Bolon Y, et al. RNA-Seq atlas of *Glycine max*: a guide to the soybean transcriptome [J]. *BMC Plant Biol*, 2010, 10: 160

[16] 付建新, 王琳琳, 戴思兰. *FT/TFLI* 基因家族调控高等植物生长发育的分子机理 [J]. *分子植物育种*, 2011(9): 1662-1672

[17] Danilevskaia O N, Meng X, Hou Z, et al. A genomic and expression compendium of the expanded *PEBP* gene family from maize [J]. *Plant Physiol*, 2008, 146: 250-264

[18] Kong F, Liu B, Xia Z, et al. Twocoordinately regulated homologs of *FLOWERING LOCUS T* are involved in the control of photoperiodic flowering in soybean [J]. *Plant Physiol*, 2010, 154: 1220-1231

[19] Fan C, Hu R, Zhang X, et al. Conserved CO-FT regulons contribute to the photoperiod flowering control in soybean [J]. *BMC Plant Biol*, 2014, 14: 9

[20] Hanzawa Y, Money T, Bradley D. A single amino acid converts a repressor to an activator of flowering [J]. *Proc Natl Acad Sci USA*, 2005, 102: 7748-7753

[21] Ahn J H, Miller D, Winter V J, et al. A divergent external loop confers antagonistic activity on floral regulators *FT* and *TFLI* [J]. *EMBO J*, 2006, 25: 605-614

[22] Kong H, Landherr L L, Frohlich M W, et al. Patterns of gene duplication in the plant *SKPI* gene family in angiosperms: evidence for multiple mechanisms of rapid gene birth [J]. *Plant J*, 2007, 50: 873-885

(上接第 150 页)

[19] Benwich I, Aviel A, Karov Y, et al. Identification of hundreds of conserved and nonconserved human microRNAs [J]. *Nat Genet*, 2005, 37(7): 766-770

[20] Xie X, Lu J, Kulbokas E J, et al. Systematic discovery of regulatory motifs in human promoters and 3'-UTRs by comparison of several mammals [J]. *Nature*, 2005, 434(7031): 338-345

[21] Miranda K C, Huynh T, Tay Y, et al. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes [J]. *Cell*, 2006, 126(6): 1203-1217

[22] Bonnet E, Wuyts J, Rouze P, et al. Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences [J]. *Bioinformatics*, 2004, 20(17): 2911-2917

[23] Washietl S, Hofacker I L, Lukasser M, et al. Mapping of conserved RNA secondary structures predicts thousands of functional non-coding RNAs in the human genome [J]. *Nat Biotechnol*, 2005, 23(11): 1383-1390

[24] Zhang B H, Pan X P, Cannon C H, et al. Conservation and divergence of plant microRNA genes [J]. *Plant J*, 2006, 46(2): 243-259

[25] Zuker M, Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information [J]. *Nucleic Acids Res*, 1982, 9: 133-148

[26] Zuker M. Mfold web server for nucleic acid folding and hybridization prediction [J]. *Nucleic Acids Res*, 2003, 31(13): 3406-3415

[27] Dsouza M, Larsenal N, Overbeek R. Searching for patterns in genomic data [J]. *Trends Genet*, 1997, 13: 49-78

[28] Zhang Y. miRU: an automated plant miRNA target prediction server [J]. *Nucleic Acids Res*, 2005, 33: 701-704

[29] Dai X, Zhao P X. psRNATarget: a plant small RNA target analysis server [J]. *Nucleic Acids Res*, 2011, 39: 155-159

[30] Xie F L, Zhang B H. Target-align: a tool for plant microRNA target identification [J]. *Bioinformatics*, 2010, 26(23): 3002-3003

[31] 蔡蕊, 未晓巍, 武慧, 等. 植物 microRNA 的生物信息学预测与分析 [J]. *植物遗传资源学报*, 2013, 14(3): 565-570