

文章编号:1674-7054(2010)01-0062-05

巴西橡胶树 LIM 结构域基因克隆与生物信息学分析

吴 瑞^{1,2}, 朱家红², 张全琪², 张治礼^{2,3}

(1. 海南大学 农学院, 海南 海口 570228; 2. 中国热带农业科学院 热带作物生物技术研究所, 农业部热带生物技术重点开放实验室, 海南 海口 571101; 3. 海南省农业科学院, 海南 海口 571000)

摘 要: 根据乙烯利刺激巴西橡胶树 (*Hevea brasiliensis*) 胶乳差异表达 cDNA 文库的 EST 序列信息, 利用 RACE 技术从胶乳中成功克隆了 1 个含有 LIM 结构域的未知功能新基因 *HbLIM* 的全长 cDNA。生物信息学分析表明, 该基因全长 cDNA 由 1 016 bp 的碱基组成, 拥有 1 个 570 bp 的开放阅读框编码 189 个氨基酸残基; 推测氨基酸序列富含 Lys, Ser, Ala 和 Leu, 含有 2 个 LIM 结构域和 2 个 N-豆蔻酰化位置。序列相似性分析表明, *HbLIM* 编码的氨基酸序列与蓖麻、杨毛、葡萄相应的 LIM 基因的相似性高达 98%, 95% 和 92%。这些研究结果为深入探讨 LIM 结构域蛋白在巴西橡胶树乳管发育及乳管代谢中的功能奠定了基础。

关键词: 巴西橡胶树; 基因克隆; LIM 基因; 生物信息学

中图分类号: S 794.1; Q 785; Q 811.4

文献标志码: A

LIM 结构域蛋白是一个重要的发育调控因子, 广泛参与动物、植物以及微生物细胞的发育调控过程, 包括基因转录、细胞形态建成和信号传导等^[1]。在真核生物中, LIM 结构域蛋白通常含有 1 个或多个 LIM 结构域。植物 LIM 结构域蛋白存在于细胞质或细胞核中^[2], 富含半胱氨酸和组氨酸, 通常含有 1 个或多个 LIM 结构域及 1 个长的碳末端结构域且富含一段甘氨酸的重复^[3]。Baltz 证实, 向日葵花粉 LIM 结构域蛋白基因 *sp3* (后来命名为 *HaPLIM1*) 与花粉管生长有关^[4-5]。烟草中的 *NtLIM* 可能参与了激动蛋白细胞骨架的稳定^[6]。由于大多数细胞类型的细胞核中都能检测到向日葵 *HaWLIM1* 蛋白, 人们推测植物 LIM 蛋白可能与动物 LIM 蛋白一样也参与了基因转录, 其可能是通过聚集和稳定转录复合体起作用^[2], 随后这一推测在烟草中得到证实^[7-8]。乙烯利刺激橡胶树增产技术已被广泛应用于天然橡胶生产, 但关于乙烯利刺激橡胶树增产过程中的信号转导、增产机理等分子基础仍不清楚。为阐明乙烯利刺激橡胶树增产的分子基础, 本研究组成功构建了乙烯利诱导条件下的胶乳抑制性削减文库, 并获得了一批差异表达的 EST 序列, 其中 No. 123 EST 序列编码的蛋白和其他植物来源的 LIM 蛋白具有较高的同源性。这表明, 胶乳 LIM 基因的表达受到了乙烯利的调控, 并可能参与了乙烯利刺激橡胶树增产过程的分子调节。另外, 巴西橡胶树的乳管组织是合成和贮存胶乳的地方^[9], 由一类特化的细胞组成, LIM 结构域蛋白是否参与了乳管细胞内细胞骨架的稳定也有待证实。根据 No. 123 EST 的序列信息, 笔者利用 RACE-PCR 技术从胶乳中成功克隆了 1 个含有 LIM 结构域的未知功能新基因的全长 cDNA, 并对其序列结构和特征进行了分析。研究结果为深入探讨 LIM 结构域蛋白在巴西橡胶树乳管发育及乳管代谢中的功能奠定了基础。

1 材料与方法

1.1 材料 实验材料取于中国热带农业科学院实验场种植的巴西橡胶树 (*Hevea brasiliensis*) 热研 7-33-97。

收稿日期: 2009-10-26

基金项目: 国家自然科学基金 (30760917); 中央级公益性科研院所业务费专项资金 (ITBBZD0716)

作者简介: 吴瑞 (1984-), 女 (彝), 云南个旧市人, 海南大学农学院 2007 级硕士研究生。

通信作者: 张治礼 (1970-), 男, 海南省农业科学院副院长, 副研究员, 博士。Email: zzl_catas@hotmail.com

1.2 试剂 RNA PCR (AMV) Ver. 3.0 试剂盒和 SMARTTM RACE 试剂盒 (Clontech) 购于 Clontech 公司, pGEM-T-Easy Vector, X-Gal, dNTPs 和 IPTG 购于 Promega 公司, DNA 回收试剂盒购于上海生工, 其他生化试剂和常规试剂均为超纯和分析纯。

1.3 RNA 提取及 cDNA 合成 参照张治礼等^[10] 的方法进行胶乳总 RNA 的提取, 用 Takara 公司的 RNA PCR (AMV) Ver. 3.0 试剂盒进行反转录合成 cDNA, 采用 SMARTTM RACE 试剂盒 (Clontech) 合成 5' cDNA。

1.4 基因克隆 乙烯利诱导条件下的胶乳抑制性削减文库中, 筛选获得的 No. 123 EST 序列与蓖麻、杨毛、葡萄等 LIM 家族成员具有较高的同源性。该序列具有终止密码子 TGA 和 ploy 结构。根据已知的 EST 序列设计特异性引物 GSP1 (5'-CCCTCAAGGGCTGCATAA-3') 与通用引物 UPM (5'AATACGACTCACTAT-AGGGCAAGCAGTGGTATCAACGCAGAGT-3') 搭配用于基因 5'端的克隆。PCR 扩增体系为 25 μ L 体系: 5' cDNA 模板 1.0 μ L, 10 $\times$ EX-Taq Buffer (含有 15 mmol $\cdot$ L⁻¹ MgCl₂) 2.5 μ L, dNTPs (2.5 mmol $\cdot$ L⁻¹ each) 0.5 μ L, Taq 酶 (5 U $\cdot$ μ L⁻¹) 0.5 μ L, 引物 1 和引物 UPM 各 1.0 μ L, ddH₂O 补足 25 μ L; PCR 反应条件为 94 $^{\circ}$ C 预变性 3 min, 94 $^{\circ}$ C 变性 30 s, 68 $^{\circ}$ C 退火 30 s, 72 $^{\circ}$ C 延伸 1 min, 35 个循环后 72 $^{\circ}$ C 再延伸 10 min。扩增出的 DNA 片段连接到 pGEM-T 载体上后送至天根生化科技有限公司进行测序。对所获得的序列和已有的 EST 序列进行拼接, 为了验证所拼接的序列同属一个基因, 分别在 5'端起始密码子的上游和 3'端终止密码子的下游设计特异引物进行 cDNA 的全长扩增, 然后将扩增的 DNA 片段连接到 pGEM-T 载体上进行测序。将测序结果与拼接结果进行比较。

1.5 生物信息学分析 采用 NCBI 站点提供的相关软件完成基因同源性分析和开放阅读框 (ORF) 识别; 氨基酸序列比对和进化树构建应用 DNAMAN 5.2 软件完成; 利用 Protparam 软件分析基因编码蛋白的氨基酸组成、相对分子质量、等电点等理化性质; ScanProsite 软件进行蛋白功能结构域预测; 利用 GOR 模型算法对 LIM 基因的氨基酸序列的二级结构进行在线预测; 利用网络服务器 TMHMM-2.0 对 LIM 基因的跨膜结构域分别进行在线预测和分析; 用 ExPASy 网络服务器的 ProtScale Sever 在线工具对 LIM 基因的氨基酸序列的疏水性/亲水性进行在线预测; 用 Psort 进行亚细胞定位; 用 Profun 分析蛋白质功能。所用相关分析软件的生物信息学网址如下:

软件 (software)	网址 (Web)
NCBI	http://www.ncbi.nlm.nih.gov/blas
Protparam	http://us.expasy.org/tools/protparam.html
ScanProsite	http://us.expasy.org/prosite
TMHMM-2.0	http://www.ch.embnet.org/software/TMPRED_form.html
Psort	http://www.Psort.org
Profun	http://www.cbs.dtu.dk/services/ProF

2 结果与分析

2.1 巴西橡胶树 *HbLIM* 基因的克隆 在进行 5'RACE 扩增时, 获得了 1 条长约 500 bp 的片段。回收、克隆和测序后证实, 该扩增片段长度为 533 bp, 并含有起始密码子的 ATG, 将扩增获得的 5'端片段与已知的 EST 片段拼接后, 获得 1 个含有完整阅读框架的 cDNA 序列。根据拼接的 cDNA 序列在 5'端起始密码子的上游和 3'端终止密码子的下游设计 1 对特异性引物, 对基因的全长 cDNA 序列进行扩增验证, 测序结果与拼接结果一致, 表明本研究获得了基因的全长 cDNA 序列 (图 1)。该基因 cDNA 全长 1 016 bp, 其中 5'端非编码区 222 bp, 3'端非编码区 215 bp, 包含 1 个完整的开放阅读框序列 (ORF) 570 bp (GenBank 登录号为: GQ901953)。

2.2 生物信息学分析

2.2.1 *HbLIM* 编码蛋白的特征分析 *HbLIM* 基因编码 189 个氨基酸残基。对蛋白质结构域分析表明, 该蛋白中含有 2 个 LIM 结构域, 分别

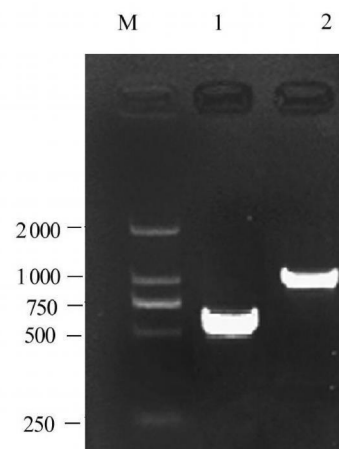


图 1 *HbLIM* 基因 cDNA 5'端及其全长的 PCR 扩增
(M. DL-2 000 marker; 1. 为 5' RACE 扩增产物; 2. cDNA 全长)

位于氨基酸 8~68 和 106~166 之间,在 54~59 和 152~157 位之间还分别含有 1 个 N-豆蔻酰化位置(图 2)。利用 ProtParam 软件预测,推测该蛋白的相对分子质量为 20.89 KD,理论等电点 9.04,属于稳定蛋白。该蛋白富含 Lys(23 个,占 12.2%)、Ser(22 个,占 11.6%)、Ala(15 个,占 8.5%)和 Leu(14 个,占 7.4%),没有 Trp, Pyl 和 Sec。

氨基酸序列比对发现,不同蛋白区段序列的保守性变化很大(图 2)。N-端和 2 个 LIM 结构域保守性较高,LIM 区间序列保守性较低,C-端区域的保守性最低。

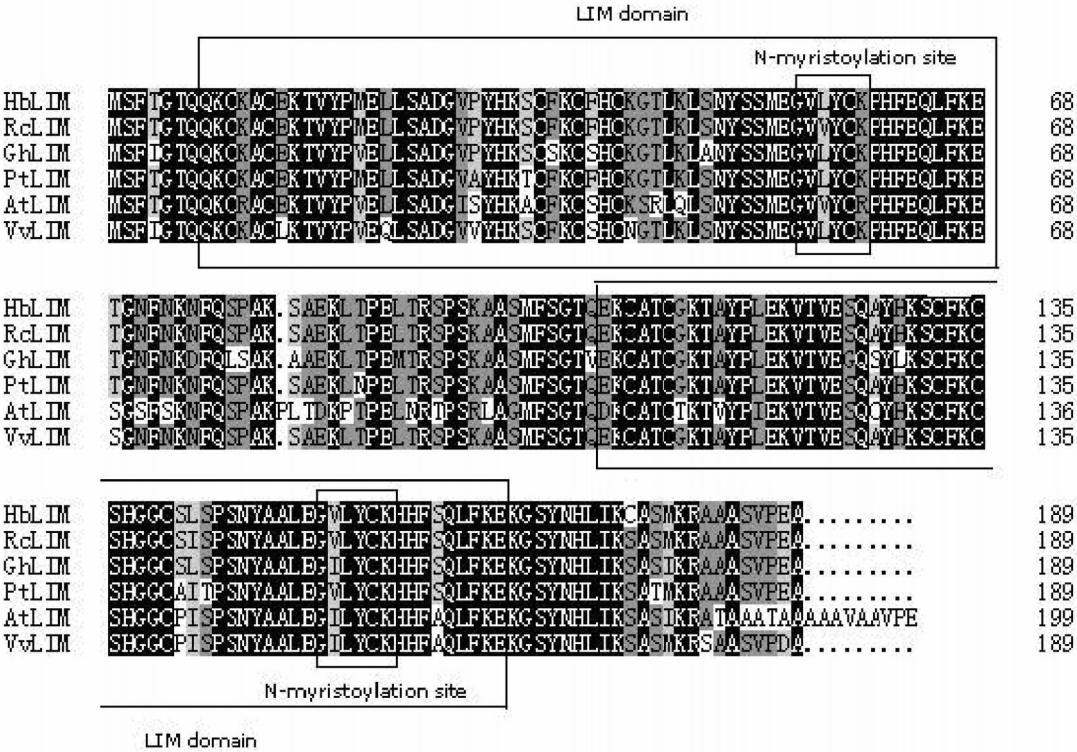


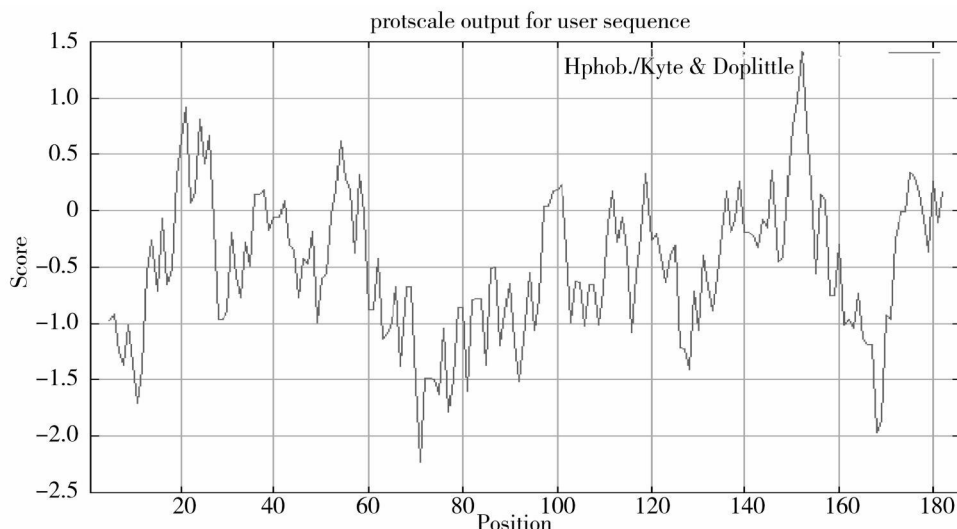
图 2 几种不同来源的 LIM 蛋白的氨基酸序列比对

(HbLIM from *Hevea brasiliensis* ACX47456; RcLIM from *Ricinus communis* XP002530411; GhLIM from *Gossypium hirsutum* AAL38006; PtLIM from *Populus trichocarpa* XP002312139; AtLIM from *Arabidopsis thaliana*, NP191136; VvLIM from *Vitis vinifera*, XP002269537; LIM 结构域和豆蔻酰化位点分别用实线框标记)

2.2.2 HbLIM 基因的二级结构预测和跨膜区预测 利用 GOR 算法对 LIM 基因氨基酸序列的二级结构进行在线预测,结果显示 HbLIM 由 14.52%的 α 螺旋、16.13% β 折叠和 69.35% 不规则的卷曲组成, α 螺旋主要存在于序列的中间和末端,而 β 折叠则贯穿于整个序列中。利用网络服务器 TMHMM-2.0 的在线工具对 LIM 的跨膜结构域进行在线预测和分析,结果表明,该基因编码蛋白无明显的跨膜结构。

2.2.3 HbLIM 的氨基酸序列的疏水性/亲水性预测和分析 依据氨基酸分值越低亲水性越强,分值越高疏水性越强的规律,用 ExPASy 网络服务器的 ProtScale Sever 在线工具对 HbLIM 基因的氨基酸序列的疏水性/亲水性进行在线预测,结果如图 2 所示,在 20~30,50~60 和 145~155 位氨基酸之间有 3 个明显的疏水区,绝大部分为亲水区,推测其可能属于亲水性蛋白(见图 3)。

2.2.4 序列比较和系统进化分析 将 GenBank 中登录的玉米(*Zea mays*, NP001147673),拟南芥(*Arabidopsis thaliana*, NP191136),棉花(*Gossypium hirsutum*, AAL38006),烟草(*Nicotiana tabacum*, AAD56951),蓖麻(*Ricinus communis*, XP002530411),番茄(*Solanum lycopersicum*, AAX95760),葡萄(*Vitis vinifera*, XP002269537),海岸松(*Pinus pinaster*, ACA33841),向日葵(*Helianthus annuus*, AAD56959),油菜(*Brassica napus*, ABB51614),麝香百合(*Lilium longiflorum*, ABY21316),蓝桉(*Eucalyptus globulus*, BAD91878),水稻(*Oryza sativa* NP001053405),大豆(*Glycine max* ACU23614)等来源的 LIM 结构域基因与来源于橡胶树(*Hevea brasiliensis*)的 LIM 结构域基因进行比较,构建了系统进化树(图 4)。结果表明,不同植物来源的 LIM 可以分成明显的 2 组,其中巴西橡胶树 HbLIM 与蓖麻的 RcLIM 同源性最近。

图 3 *HbLIM* 的氨基酸序列的疏水性 / 亲水性分析

2.2.5 *HbLIM* 编码蛋白的亚细胞定位和蛋白质功能分类 用 Psort 分析了 *HbLIM* 的亚细胞定位,结果显示,该蛋白定位于细胞质的可能性为 65%,定位于线粒体基质的可能性为 10%。ProtFun 预测显示,该基因编码蛋白具有转录、信号转导和转录调控功能的可能性分别为 0.08,0.073 和 0.328。

3 讨 论

LIM 是 3 种转录因子 Lin - 11, Isl - 1 和 Mec - 3 名称首字母的缩写^[11]。目前发现的 LIM 蛋白已有 60 多种,它们分别参与了多种基因的转录调控和细胞的发育与分化^[12]。LIM 蛋白对胚胎发育、细胞分化和细胞骨架形成的调节主要通过蛋白质与蛋白质之间的相互作用实现,还可能参与介导了转录因子、信号通路蛋白和细胞骨架蛋白相关蛋白质间的相互作用。

巴西橡胶树乳管是合成和贮存胶乳的组织,由维管形成层分化而来^[9]。已有的研究表明,肌动蛋白细胞骨架可能在巴西橡胶树乳管排胶和乳管堵塞中起重要作用^[13]。动物细胞中 LIM 结构域蛋白 CRP (cysteine-rich protein) 和 MLP (Muscle LIM Protein) 可作为蛋白质识别的界面而介导蛋白质的相互作用^[14],能与斑联蛋白和 α - 辅肌动蛋白相结合,是肌动蛋白胞质骨架形成的必要成分^[15]。巴西橡胶树 *HbLIM* 与其他植物 LIM 结构域蛋白一样,与动物的 CRP/MIP 蛋白具有相当数量的相同序列,暗示植物 LIM 基因可能具有与 CRP 类似的功能。笔者推测认为,*HbLIM* 可能通过蛋白与蛋白之间的相互作用参与了巴西橡胶树乳管发育、排胶和乳管堵塞。LIM 蛋白具有转录、信号转导和转录调控等功能。施用乙烯利可以延缓巴西橡胶树乳管堵塞和延长排胶时间,提高胶乳产量。*HbLIM* 的转录受到乙烯利的诱导,但 *HbLIM* 是否参与了胶乳代谢以及乙烯利刺激胶乳生产的分子过程尚不清楚。

参考文献:

- [1] DAWID I B, TOYAMA R, TAIRA M. LIM domain protein[J]. C R Acad Sci [III], 1995, 318:295 - 306.
- [2] MUNDEL C, BALTZ R, ELIASSEN A. A LIM-domain protein from sunflower is localized to the cytoplasm and/or nucleus in a wide variety of tissues and is associated with the phragmoplast in dividing cells[J]. Plant Mol. Biol. ,2000,42:291 - 302.
- [3] ELIASSEN A, GASS N, MUNDEL C. A Molecular and expression analysis of a LIM protein gene family from owerling plants [J]. Mol. Gen. Genet. ,2000,264:257 - 267.

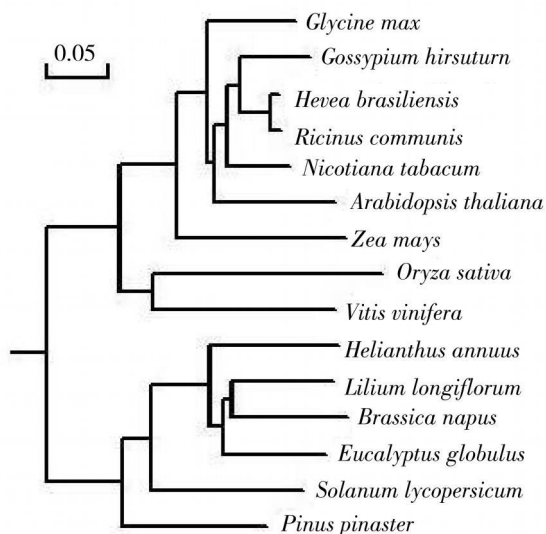


图 4 LIM 结构域蛋白的系统进化分析

- [4] BALTZ R, DOMON C, PILLAY D. A Characterization of a pollen-specific cDNA from sunflower encoding a zinc finger protein [J]. *Plant J.*, 1992, 2: 713 – 721.
- [5] BALTZ R, EVRARD J L, DOMON C. A LIM motif is present in a pollen-soecific protein [J]. *Plant Cell.*, 1992, 4: 1465 – 1466.
- [6] THOMAS C, HOFFMANN C, DIETERLE M. Tobacco WLIM1 is a novel F-actin binding protein involved in actin cytoskeleton remodeling [J]. *Plant Cell.*, 2006, 18: 2194 – 2206.
- [7] KAWAOKA A, KAOFTIEN P, YOSHIDA K. Functional analysis of tobacco LIM protein Ntlim1 involved in lignin biosynthesis [J]. *Plant J.*, 2000, 22: 289 – 301.
- [8] KAWAOKA A, EBINUMA H. Transcriptional control of lignin biosynthesis by tobacco LIM protein [J]. *Phytochemistry*, 2001, 857: 1 149 – 1 157.
- [9] 赵中奇. 橡胶树乳管发育与胶量关系研究 [J]. *云南农业大学学报*, 1998, 13 (4): 397 – 400.
- [10] ZHANG Z L, YANG Y, LIU K C. A rapid and efficient protocol for total RNA isolation from latex of *Hevea brasiliensis* [J]. *Chinese Bulletin of Botany*, 2007, 24: 516 – 520.
- [11] OSTENDORFF H P, PEIRANO R I, PETERS M A. Ubiquitination-dependent cofactor exchange on LIM homeodomain transcription factors [J]. *Nature*, 2002, 416: 99 – 103.
- [12] 高政权, 孟春晓, 吴继林, 等. 巴西橡胶树乳管肌动蛋白细胞骨架与采胶的关系 [J]. *热带作物学报*, 2003, 23 (3): 22 – 26.
- [13] SCHEICHEL K L, BECKERLE M C. The domain is a modular protein-binding interface [J]. *Cell*, 1994, 79: 211 – 219.
- [14] POMIES P, LOUIS H A, BECKERLE M C. CRP1 a LIM domain protein implicated in muscle differentiation interacts with a lpha-actinin [J]. *J. Cell Biol.*, 1997, 139: 157 – 168.
- [15] ALAN D AGULNICK, MASANORI TAIRA, JOSEPH J BREEN. Interactions of the LIM-domain-binding factor Ldb1 with homeodomain proteins [J]. *Nature*, 1996, 385: 270 – 272.

Cloning and Bioinformatics Analysis of a LIM Domain Protein Gene from *Hevea brasiliensis*

WU Rui^{1,2}, ZHU Jia-hong², ZHANG Quan-qi², ZHANG Zhi-li^{2,3}

(1. College of Agriculture, Hainan University, Haikou 570228, China;

2. Key Laboratory of Ministry of Agriculture for Tropical Crop Biotechnology, Institute of Tropical Biosciences and Biotechnology,

Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China;

3. Hainan Academy of Agricultural Sciences, Haikou 571100; China)

Abstract: Based on the information of EST sequences from latex SSH cDNA library stimulated by ethephon, RACE was used to isolate the full length cDNA of an unknown functional novel gene with two LIM domains from *H. brasiliensis*, *HbLIM*. Bioinformatics analysis results showed which comprises of 1 016bp with a 570bp ORF and encodes a 189 amino acid residues. The deduced amino acid sequence with two LIM domains and two N-myristoylation sites is rich in Lys, Ser, Ala and Leu and has high identity of 98%, 95% and 92% to the homologues from *Pollen specific*, *Populus trichocarpa* and *Vitis vinifera*. These results were very important for elucidation of the functions of LIM-domain protein in latexifer development and metabolism in *H. brasiliensis*.

Key words: *Hevea brasiliensis*; Gene cloning; LIM gene; bioinformatical analysis