

刺五加 Actin 基因 cDNA 全长序列的克隆与生物信息学分析*

柴丽花, 修乐山, 周 秘, 邢朝斌**

(河北联合大学生命科学学院 唐山 063000)

摘 要:目的:克隆刺五加 Actin 基因的全长 cDNA 序列,并对其进行生物信息学分析。方法:以刺五加的叶片为材料,提取总 RNA,逆转录为 cDNA,根据初步克隆到的刺五加 Actin 基因保守序列设计引物,利用套式 PCR 进行 3' 和 5'RACE 扩增,得到 Actin 基因的 3' 和 5' 端 cDNA 序列片段,拼接获得 cDNA 全长。将该序列进行 BLAST 比对、相似度分析,并对刺五加 Actin1 蛋白的二级结构和三级结构进行预测。结果:刺五加的 Actin 基因全长 1 507 bp,命名为 *ESActin1*,注册号 KC469585。该基因开放阅读框(ORF)长 1 134 bp,编码 377 个氨基酸,5' 端非编码区长 140 bp,3' 端非编码区长 233 bp。*ESActin1* 与 GenBank 中注册的其它植物肌动蛋白核苷酸序列的相似性在 75%以上,蛋白质序列的相似性在 94%以上。结论:首次报道了刺五加 Actin 基因的全长 cDNA 序列,为刺五加的分子生物学研究奠定基础。

关键词:刺五加 肌动蛋白 克隆 生物信息学分析

doi: 10.11842/wst.2013.09.007

中图分类号:R284.1

文献标识码:A

肌动蛋白(Actin)是一种普遍存在于真核生物中的重要蛋白质,主要构成细胞骨架和肌肉肌小节^[1],高等动植物中,肌动蛋白由多基因家族编码^[2]。这些高度相似的基因,由一个祖先基因通过多轮复制进化而来。肌动蛋白为单一多肽链的球状蛋白质(G-Actin),长 375~377 个氨基酸残基,分子量为 42 kD^[2]。不同物种的肌动蛋白氨基酸在序列上和长度上具有高度的保守性和同源性,表明肌动蛋白在植物体生命过程中的重要性。同时因其表达具有相对稳定的特点,因此常被用作基因表达分析时的内参基因^[3]。迄今为止,已在多种高等植物中克隆到 Actin 基因^[1]。然而,尚未见关于刺五加 Actin 基因克隆的相关报道。本研究在初步克隆到

的刺五加 Actin 基因保守序列的基础上,利用 RACE 技术扩增其全长序列,并进行生物信息学分析,以期对刺五加 Actin 基因的利用和其他基因的表达分析提供理论和技术支持。

1 材料与方法

1.1 材料

刺五加 *Eleutherococcus senticosus* (Rupr. et Maxim.)Harms 由河北联合大学邢朝斌副教授自黑龙江省穆棱市野外采集得到。实验前将刺五加叶片 3 枚剪下洗净杂质,晾干后用塑料膜封存,于 -70 ℃ 冻存作为提取总 RNA 的试材。

RevertAid™ First strand cDNA synthesis Kit(Fermentas 公司);3'-Full RACE Core Set with PrimeScript™ RTase、PrimeScript 逆转录酶、TDT、RNase

收稿日期:2013-03-16

修回日期:2013-04-11

* 国家自然科学基金委青年基金项目(30701086)刺五加关键酶基因单核苷酸多态性及其对刺五加苷含量的作用机制,负责人:邢朝斌;河北省自然科学基金项目(C2009001252)内生菌提高刺五加中药用成分含量的作用机制研究,负责人:邢朝斌;河北省自然科学基金-石药集团医药联合研究基金项目(H2012401006)刺五加鲨烯合酶基因家族及其对刺五加苷含量的作用机制,负责人:邢朝斌;河北联合大学培养基金(GP201306)关键酶基因家族对刺五加苷含量的作用机制,负责人:邢朝斌。

** 通讯作者:邢朝斌,副教授,主要研究方向:分子生药学研究。

(TaKaRa 公司);RNAprep Pure Plant Kit、普通琼脂糖凝胶 DNA 回收试剂盒、PGM-T 质粒(TIANGEN 公司);大肠杆菌 TOP10 感受态细胞(北京天根生化科技有限公司);PCR 产物纯化试剂盒(Biomiga 公司)。

1.2 方法

1.2.1 提取刺五加总 RNA

按照 RNAprep Pure Plant Kit 说明书要求,称取 0.1 g 刺五加冻存叶片提取总 RNA,取总 RNA 1 μ L 用 1%琼脂糖凝胶电泳验证,结果表明总 RNA 提取成功,将最后得到的总 RNA 于 -70°C 保存备用。

1.2.2 5' RACE 扩增获得刺五加 Actin 基因 cDNA 的 5' 末端

根据已克隆的刺五加 Actin 基因保守区的 cDNA 序列设计 5' RACE 特异性引物 ACT52:5'-AGGGCATCAGTGAGATCACGAC-3' 和 ACT51:5'-GCCCATCCCAACCATAACACCA-3', 锚定引物 AP 长:5'-AAGCAGTGGTATCAACGCAGAGTACGCGGGGGGGG-3', 通用上游接头引物 AP 短:5'-AAGCAGTGGTATCCAACGCAGAGT-3'^[4]。取提取的总 RNA 3 μ L, 根据 RevertAid™ First strand cDNA synthesis Kit 说明书进行逆转录反应。以 AUP1:5'-GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTTT-3' 为引物,合成第一条 cDNA 链。逆转录完成后,根据说明书要求将 25 μ L 逆转录产物用 RNase H 处理 1 h,再用 PCR 产物纯化试剂盒纯化产物。根据 TDT 说明书要求将纯化产物进行 TDT 末端加尾,加尾 2 h。再根据 PCR 纯化试剂盒说明纯化加尾产物,得到加尾的 cDNA。利用巢式 PCR 和降落 PCR 技术进行 5' RACE 扩增,第一轮 PCR 以 AP 长和 ACT52 为引物,反应体系:加尾产物 1 μ L, dNTP 4 μ L, AP 长 1 μ L, ACT 52 引物 1 μ L, LABuf II 2.5 μ L, LA Taq 0.3 μ L, 水 15.2 μ L。反应条件为:95 $^{\circ}\text{C}$ 预变性 3 min, 95 $^{\circ}\text{C}$ 变性 30 s, 62 $^{\circ}\text{C}$ 退火 30 s, 72 $^{\circ}\text{C}$ 延伸 2 min 30 s, 2 个循环, 95 $^{\circ}\text{C}$ 变性 30 s, 60 $^{\circ}\text{C}$ 退火 30 s, 72 $^{\circ}\text{C}$ 延伸 2 min 30 s, 2 个循环, 95 $^{\circ}\text{C}$ 变性 30 s, 58 $^{\circ}\text{C}$ 退火 30 s, 72 $^{\circ}\text{C}$ 延伸 2 min 30 s, 2 个循环, 95 $^{\circ}\text{C}$ 变性 30 s, 56 $^{\circ}\text{C}$ 退火 30 s, 72 $^{\circ}\text{C}$ 延伸 2 min 30 s, 2 个循环, 95 $^{\circ}\text{C}$ 变性 30 s, 54 $^{\circ}\text{C}$ 退火 30 s, 72 $^{\circ}\text{C}$ 延伸 2 min 30 s, 27 个循环, 72 $^{\circ}\text{C}$ 最后延伸 10 min。将第一轮 PCR 产物稀释 100 倍后,取 1.0 μ L 做模板,以 AP 短和 ACT51 为引物进行第二轮 PCR,反应体系和条件与第一轮 PCR 相同,反应完成后,用 1%琼脂糖

凝胶电泳验证后,再用 1.2%琼脂糖凝胶电泳,并根据琼脂糖凝胶 DNA 回收试剂盒说明书切胶回收纯化反应产物,克隆至 PGM-T 质粒载体后将其转化到大肠杆菌 TOP10 感受态细胞中,提取重组质粒后送至上海 Life 公司测序。

1.2.3 3' RACE 扩增获得刺五加 Actin 基因 cDNA 的 3' 末端

根据 ACT 基因片段的保守区 cDNA 序列,应用 Primer premier 5.0 软件设计 3' RACE 特异性外侧引物 ACT32:5'-CCACGAGACTACTTACAACCTCC-3' 和内侧引物 ACT31:5'-CGGATCGTATGAGCAAGGAA-3',根据 3'-Full RACE Core Set with Prime-Script™ RTase 说明书进行 3' RACE。利用套式 PCR 原理,以 3' RACE outer primer:5'-TACCGTCGTTC-CAGTAGTGATTT-3' 和 ACT32 引物进行 outer PCR 反应,反应体系:逆转录液 2.5 μ L, 1 $\times$ cDNA Dilution Buffer II 7.5 μ L, ACT32 引物 2 μ L, 3' RACE Outer Primer (10 $\mu\text{mol}\cdot\text{L}^{-1}$) 2 μ L, 10 $\times$ LA PCR Buffer II 5 μ L, TaKaRa LA Taq (5 U $\cdot\mu\text{L}^{-1}$) 0.3 μ L, dd H₂O 30.7 μ L, 反应条件:94 $^{\circ}\text{C}$ 预变性 3 min, 94 $^{\circ}\text{C}$ 变性 1 min, 55 $^{\circ}\text{C}$ 退火 30 s, 72 $^{\circ}\text{C}$ 延伸 3 min, 20 个循环, 72 $^{\circ}\text{C}$ 最后延伸 10 min。取 outer PCR 产物 1 μ L, 以 3' RACE inner primer 和 ACT31 进行套式 PCR 的 inner PCR 反应,反应体系:outer PCR 产物 1 μ L, dNTP Mix (2.5 mmol $\cdot\text{L}^{-1}$ each) 8 μ L, 10 $\times$ LA PCR Buffer II 5 μ L, Takara LA Taq (5 U $\cdot\mu\text{L}^{-1}$) 0.5 μ L, 3' RACE Inner Primer (10 $\mu\text{mol}\cdot\text{L}^{-1}$) 2 μ L, ACT31 引物 2 μ L, dd H₂O 31.5 μ L, 反应条件:94 $^{\circ}\text{C}$ 预变性 3 min, 94 $^{\circ}\text{C}$ 变性 1 min, 55 $^{\circ}\text{C}$ 退火 30 s, 72 $^{\circ}\text{C}$ 延伸 2 min 30 s, 20 个循环, 72 $^{\circ}\text{C}$ 最后延伸 10 min。套式 PCR 反应完成后,以 1.2%琼脂糖凝胶电泳,根据琼脂糖凝胶 DNA 回收试剂盒说明书切胶回收纯化反应产物,克隆至 PGM-T 质粒载体后将其转化到大肠杆菌 TOP10 感受态细胞,挑取阳性克隆质粒进行测序。

1.2.4 Actin 基因全长的拼接与 PCR 验证

利用 DNAMAN 6.0 软件将 Actin 基因的保守序列和 5' 末端序列及 3' 末端序列进行拼接,得到刺五加 Actin 基因的 cDNA 全长序列,根据该序列设计扩增包含 Actin 基因起始密码子的上游引物 AC-TQS:5'-CGGAAGTTGGCTCGGTTTG-3' 和包含终止密码子的下游引物 ACTQX:5'-GGACCCTGGCAACGAAGAC-3'。以“1.2.2”中逆转录得到的 cDNA 为

模板,PCR 扩增 Actin 基因全长。PCR 反应体系:LA Taq 酶 0.5 μL ,10 $\times$ LA Taq Buffer II 5 μL ,2.5 mmol $\cdot\text{L}^{-1}$ dNTPs 8 μL ,模板 cDNA 1 μL ,引物 AC-TQS 和 ACTQX 各 2 μL ,dd H₂O 31.5 μL 。反应条件 94 $^{\circ}\text{C}$ 预变性 5 min,94 $^{\circ}\text{C}$ 变性 1 min,56 $^{\circ}\text{C}$ 退火 30 s,72 $^{\circ}\text{C}$ 延伸 1 min 20 s,35 个循环,72 $^{\circ}\text{C}$ 最后延伸 10 min。用 1.2%琼脂糖凝胶电泳 PCR 产物并回收,将该产物克隆到 PGM-T 载体后转化到 TOP10 感受态细胞并测序。

1.2.5 生物信息学与系统进化分析

参照罗红梅等^[5]和辛天怡等^[6]方法,利用在线软件 ProtParam、SOPMA、ProtScale、MotifScan、TMHMM 2.0 和 SWISS-MODEL 进行生物信息学分析。使用 MEGA 5.05 软件中的邻接法(Neighbor-Joining)法构建系统发育树。

2 结果与分析

2.1 刺五加 ESActin1 全长 cDNA 的克隆

以刺五加 cDNA 的加尾产物为模板,利用巢式 PCR 和降落 PCR 技术进行 5' RACE 扩增,电泳得到长度为 290 bp 的 Actin 基因片段。利用套式 PCR 技术进行 3' RACE 扩增,电泳得到长度为 433 bp 的 Actin 蛋白基因片段。将测序所得序列与已克隆的刺五加 Actin 基因的保守序列进行拼接,根据所得 cDNA 全长设计扩增包含 Actin 基因起始密码子和终止密码子的上、下游引物 ACTQS 和 ACTQX,PCR 扩增得到长度为 1 259 bp 的 1 条带,与预期结果相符。该片段测序结果与 Actin 基因 5'端和 3'端 cDNA 序列与保守序列拼接结果完全相同。结果如图 1 所示。

2.2 刺五加 Actin 的氨基酸序列和蛋白结构分析

刺五加 Actin 基因全长 1 507 bp(GenBank 登录号 KC469585,命名为 ESActin1),该基因 5'端非编码区(5'UTR)长 140 bp,3'端非编码区(3'UTR)长 233 bp,终止密码子为 TAA,ORF 长 1 134 bp,推测编码 377 个氨基酸,3'端有 Poly A 尾。利用 ProtParam 软件预测 ESActin1 基因编码的蛋白质分子量为 41.67 kD,理论等电点(PI)为 5.39^[7]。将刺五加的 Actin 基因序列与 NCBI 数据库中 18 种植物的 Actin 基因序列进行在线 Blast 对比,结果表明,不同植物的 Actin 基因具有高度的保守性,仅在少数位点有差异。

利用在线软件 SOPMA 预测蛋白的二级结构发现,刺五加 Actin 蛋白含有 35.81%的 α -螺旋,22.02%的延伸链,7.96%的 β -转角和 34.22%的随意卷曲。刺五加 Actin 蛋白的三级结构见图 2,该蛋白由 4 个亚基组成。通过 SignalP 4.0 软件系统分析信号肽序列的结果表明,刺五加 Actin 蛋白没有发现信号肽及其剪切位点,该蛋白是非分泌型蛋白^[8]。利用 ProtScale 软件对刺五加 Actin 蛋白的亲水性/疏水性分析结果表明,该蛋白亲水性氨基酸含量较高,具有较强的亲水性。通过 TMHMM 2.0 软件分析发现刺五加 Actin 蛋白不存在跨膜区域,该蛋白被定位于细胞质中,且发挥重要作用。

利用 MotifScan 分析结果显示,刺五加 Actin 蛋白序列具有多个蛋白修饰位点:1 个 N-糖基化位点(14~17 位:NGTG),1 个 cAMP-环磷酸腺苷和

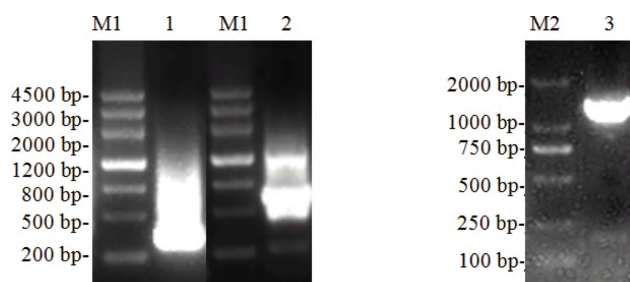


图 1 刺五加 Actin 基因的克隆

注:1.刺五加 Actin 基因的 5'克隆 2.Actin 基因的 3'克隆 3.Actin 基因全长的 PCR 扩增;M1. DL 4500 DNA marker;M2. DL 2000 DNA marker。



图 2 刺五加 Actin 蛋白三级结构同源建模

cGMP-环磷酸鸟苷蛋白激酶磷酸化位点(337~340 : RKYS) 4 个酪蛋白激酶 II 磷酸化位点 (79~82 : SNWD、204~207 :TTAE、236~239 :SSVE、360~363 : SKGE) 8 个 N-豆蔻酰化 (N-myristoylation) 位点 (17~22 :GMVKAG、44~49 :GVMVGM、50~55 :GQK-DAY、76~81 :GIVSNW、160~165 :GVSHTV、247~252 : GQVITI、270~275 :GMEAAG、345~350 :GSILAS) ,第 55~65 位氨基酸(YVGDEAQSQRG)、第 358~366 位氨基酸(WISKGEYDE)和第 106~118 位氨基酸(LL-TEAPLNPKANR)为肌动蛋白家族的标志性序列。表明刺五加 Actin 蛋白可能在蛋白水平上存在复杂的调控和修饰作用。

2.3 刺五加 Actin 蛋白的分子系统进化分析

将刺五加 Actin 蛋白的氨基酸序列与 NCBI 数据库中已登载的 20 个物种的 Actin 蛋白的氨基酸

序列进行多序列对比,并采用邻接法将对对比结果绘制成 Actin 基因系统进化树,见图 3。对比结果与系统进化树分析均显示,刺五加与桑树、油桐、蓖麻、橡胶树、葡萄和番木瓜亲缘关系较近,聚为一支,其他双子叶植物聚为一支,裸子植物、单子叶植物分别聚为一支,哺乳动物人和小家鼠聚为一支,该进化树符合传统分类结果。

3 讨论

本实验根据已克隆的刺五加 Actin 基因的保守区序列,利用 RACE 技术获得了该基因的 cDNA 全长序列。由于一般方法较难得到基因 5'末端序列,本实验采用改良的 5' RACE 技术^[9],并将 TDT 加尾时间设为 2 h,成功获得该基因的 5'末端未知序列。

肌动蛋白是真核生物中最古老的保守蛋白之

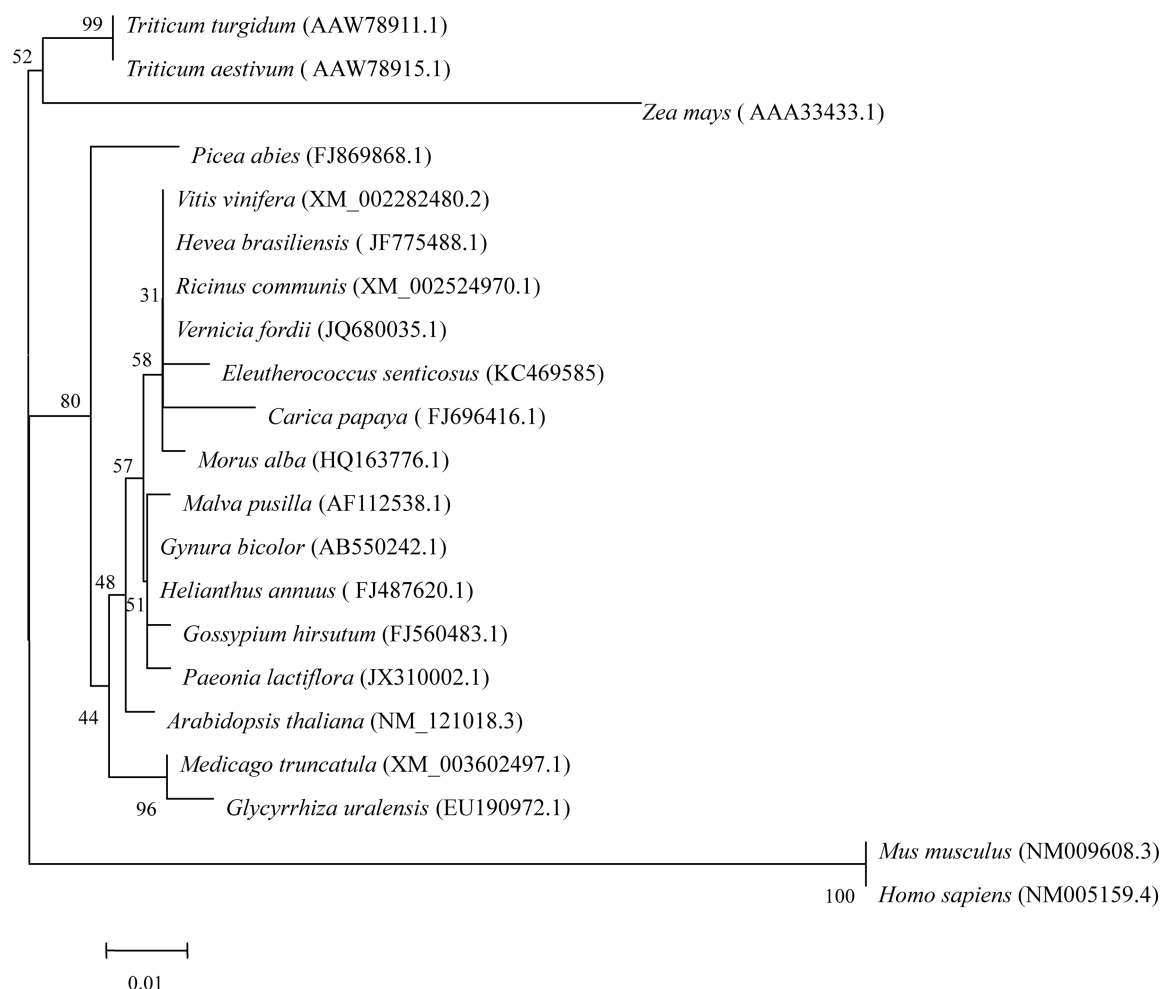


图3 Actin 蛋白的系统进化树

参考文献

- 1 吴永娜,胡静,王引权,等. 当归肌动蛋白基因片段的克隆及序列分析. 中草药, 2012, 43(12): 2485~2489.
- 2 王芳. 甘草肌动蛋白基因 GuActin2 的克隆和表达分析. 植物生理学通讯, 2009, 45(10): 995~1000.
- 3 孙美莲,王云生,杨冬青,等. 茶树实时荧光定量 PCR 分析中内参基因的选择. 植物学报, 2010, 45(5): 579~587.
- 4 邢朝斌,龙月红,何闪,等. 刺五加法尼基焦磷酸合酶基因的克隆、生物信息学及表达分析. 中国中药杂志, 2012, 37(12): 1725~1730.
- 5 辛天怡,姚辉,罗焜,等. 羌活药材 ITS/ITS2 条形码鉴定及其稳定性与准确性研究. 药学学报, 2012, 47(8): 1098~1105.
- 6 罗红梅,宋经元,李雪莹,等. 人参皂苷合成生物学关键元件 HGMR 基因克隆与表达分析. 药学学报, 2013, 48(2): 219~227.
- 7 Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the ExPASy Server. In The Proteomics Protocols Handbook. Totowa: Humana Press, 2005.
- 8 Bendtsen J D, Nielsen H, von Heijne G, et al. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol, 2004, 340(4): 783~795.
- 9 罗聪,何新华,陈虎,等. 一种高效获取基因 5' 末端的 RACE 方法. 植物生理学报, 2011, 47(4): 409~414.
- 10 陈颖,王刚,赵俊霞,等. 高等植物体内的肌动蛋白. 生物学报, 2003, 38(1): 13~15.
- 11 孔卫青,杨金宏. 桑树肌动蛋白 actin 基因全长序列的克隆与分析. 广西植物, 2012, 32(3): 362~366.
- 12 杨亚军,王新超,马春雷. 茶树肌动蛋白基因 (*CsActin1*) 全长 cDNA 克隆与生物信息学分析. 植物研究, 2009, 29(3): 339~345.
- 13 Nicot N, Hausman J F, Hoffmann L, et al. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. J Exp Bot, 2005, 56(421): 2907~2914.

一,由一个大的、高度保守的基因编码,一般编码 375~377 个氨基酸^[1,10],本研究克隆的刺五加肌动蛋白由 377 个氨基酸组成,其核苷酸序列具有肌动蛋白基因家族的特征序列,与桑树(登录号 HQ163776.1)、陆地棉(登录号 FJ560483.1) Actin 基因序列同源性为 87% 和 89%,氨基酸序列同源性均为 99%,但隶属于五加科的刺五加与隶属于桑科的桑树和锦葵科的陆地棉属于不同科属,因此进一步反映出肌动蛋白的高度保守性,这同其参与构成细胞骨架的生理功能密切相关。通过比较发现,虽然刺五加与橡胶树(登录号 JF775488.1)、蓖麻(登录号 XM002524970.1)和桑树^[11]氨基酸序列相似度为 99%,但 Actin 蛋白序列中决定 α -螺旋、无规则卷曲、 β -转角和延伸链的氨基酸含量在不同植物间有明显区别,形成不同的肌动蛋白,在相应结合蛋白作用下,参与不同植物的生理活动^[12],表明植物 Actin 蛋白在进化过程中既具有保守性,又能在不同植物中起到相似作用。

由于 Actin 基因在各个器官中表达较稳定,常被用做实时定量 PCR 研究中的内参基因^[3,13],本实验首次克隆出刺五加的肌动蛋白基因并对该基因进行了序列分析,为进一步研究刺五加肌动蛋白家族其它基因及利用 RT-PCR 技术研究肌动蛋白基因在刺五加不同器官中的表达奠定基础。

Clone of Full-length cDNA Sequence of Actin Gene of *Eleutherococcus Senticosus* and Bioinformatics Analysis

Chai Lihua, Xiu Leshan, Zhou Mi, Xing Zhaobin

(College of Life Science, Hebei United University, Tangshan 063000, China)

Abstract: This study was aimed to clone the full-length cDNA sequence of actin gene of *Eleutherococcus senticosus*. And bioinformatics analysis was used. The total RNA was isolated from leaves of *E. senticosus*, and cDNA was synthesized by reverse transcription of total RNA. Primers were designed according to the conserved sequence that had been cloned of Actin of *E. senticosus*. Then, the 3' and 5' cDNA fragments were cloned by nested PCR. The full-length gene was obtained by gene splicing method. Sequencing results were compared and treated with similarity analysis by blast analysis in the GenBank. Protein secondary structure and tertiary structure of Actin of *E. senticosus* was predicted by online software. The results showed that the full-length cDNA of Actin of *E. senticosus* is 1507 bp, which named *EsActin1*, GenBank accession No. KC469585. The conserved sequence, which contained a 1134 bp open reading frame that encoding a 377 amino acid residues, a 5'-UTR

of 140 bp and a 3'-UTR of 233 bp. Homologous alignment showed that it shared over 75% nucleotide sequence similarity and over 94% amino acid sequences similarity with Actins in other plants. It was concluded that this study first isolated and reported the full-length cDNA sequence of actin gene of *E. senticosus*, and laid a foundation for the molecular biology research of *E. senticosus*.

Keywords: *Eleutherococcus senticosus*, Actin, clone, bioinformatics analysis

(责任编辑 叶丽萍 张志华, 责任译审 王 晶)