

## 橡胶树半胱氨酸蛋白酶基因 *HbCP3* 的 克隆与表达特性分析

邹智, 杨礼富

(中国热带农业科学院橡胶研究所/农业部儋州热带作物科学观测实验站, 海南 儋州 571737)

**摘要:** 半胱氨酸蛋白酶是一类重要蛋白水解酶, 其广泛参与植物的生长发育及各种生物与非生物胁迫的应答。根据 EST 序列信息设计引物, 从橡胶树的根中分离到 1 条 1 056 bp 的 cDNA。该序列包含 1 个 1 023 bp 的开放阅读框(ORF), 13 bp 的 5' UTR 和 20 bp 的 3' UTR。ORF 预测编码 340 个氨基酸, 相对分子质量为 37 350, 等电点为 5.24, 属于液泡定位的分泌型蛋白。该蛋白含有 1 个 Inhibitor\_I29 和 1 个 Peptidase\_C1A 结构域, 可归为木瓜蛋白酶 C1 家族。该蛋白与蓖麻(*Ricinus communis*)和杨树(*Populus trichocarpa*)中同源蛋白的相似性都在 80%以上, 将基因命名为 *HbCP3*。表达分析显示, 在橡胶树的根、树皮、木质部、芽、叶片、叶柄和乳管等组织中, 基因仅在根中表达。

**关键词:** 橡胶树; 半胱氨酸蛋白酶; 根特异性表达

中图分类号: S794.101

文献标志码: A

文章编号: 1007-1032(2015)01-0053-05

### Molecular cloning and expression analysis on cysteine proteinase gene (*HbCP3*) from rubber tree (*Hevea brasiliensis* Muell. Arg.)

Zou Zhi, Yang Lifu

(Danzhou Investigation & Experiment Station of Tropical Crops, Ministry of Agriculture/Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Danzhou, Hainan 571737, China)

**Abstract:** Cysteine proteinases (CPs), a type of important hydrolysis proteases, have played an essential role in plant growth and in response to biotic or abiotic stresses. Based on the available EST sequence, a 1 056 bp cDNA named by *HbCP3* was segregated out by using specific primers from root tissues of rubber tree (*Hevea brasiliensis* Muell. Arg.). The cDNA contained a length of 1 023 bp ORF (open reading frame), 13 bp 5' UTR and 20 bp 3' UTR. Sequence analysis suggested that *HbCP3* had a deduced code of 340 amino acids with a theoretical molecular weight (Mw) of 37 350 and with isoelectric point (pI) of 5.24, and it belonged to a secretory protein located in vacuoles; The protein, containing one Inhibitor\_I29 domain and one Peptidase\_C1A domain, could be classified into the papain C1 family; Since the protein shared more than 80% similarity with its homologs *Ricinus communis* and *Populus trichocarpa*, it was hence named by *HbCP3*. Results of the expression analysis showed that among in all tissues, i.e., root, bark, xylem, bud, leaf, petiole and laticifer, *HbCP3* only exhibited in root.

**Keywords:** *Hevea brasiliensis*; cysteine proteinase; root-specific expression

半胱氨酸蛋白酶, 又叫巯基蛋白酶, 是一类催化活性位点、含有亲核半胱氨酸残基的蛋白水解酶。在植物中, 半胱氨酸蛋白酶广泛参与种子萌发、

幼苗发育、组织分化、衰老、细胞程序性死亡和胁迫应答等多种生物学过程<sup>[1-2]</sup>。根据蛋白的结构特征和进化关系, 植物半胱氨酸蛋白酶可分为木瓜蛋

白酶 C1、天冬氨酸内肽酶 C13、胱冬肽酶 C14、钙依赖半胱氨酸蛋白酶 C2、泛素 C 端水解酶 C12 和泛素特异蛋白酶 C19 六大家族。目前,针对木瓜蛋白酶开展的研究较多<sup>[1]</sup>。橡胶树(*Hevea brasiliensis* Muell. Arg.)隶属于大戟科,是一种原产于亚马逊河流域的热带雨林乔木,由其生产的橡胶是天然橡胶的唯一商业来源。目前,橡胶树主要移植于东南亚、东亚和非洲等的部分国家和地区<sup>[3-4]</sup>。这些植胶区多为不发达地区,关于橡胶树分子生物学的研究相对滞后。橡胶树半胱氨酸蛋白酶基因 *HbCP1* 在乳管、树皮和叶片中均有表达,且在胶乳中基因受割胶和乙烯利诱导<sup>[5]</sup>。本文报道 1 个根特异性表达的半胱氨酸蛋白酶基因 *HbCP3*,通过对其序列特征和表达特性进行分析,以期为进一步的功能研究与利用提供参考。

## 1 材料与方法

### 1.1 材料

#### 1.1.1 基因克隆与表达分析所用材料

橡胶树无性系热研 7-33-97 的组培苗由中国热带农业科学院橡胶研究所育种课题组提供。组培苗装袋栽于大棚。选择长势一致、都具有 4 个叶蓬且最上层叶片完全绿化的植株,待下层叶片出现明显黄化后(即用 SPAD-502 便携式叶绿素测定仪测得最底层叶片的叶绿素含量约降为最上层叶片的 1/2),分别采集根、树皮、木质部、芽、4 个叶蓬的叶片、最上层叶片的叶柄和来自叶柄的胶乳(乳管细胞的细胞质成分)作为试验分析材料。

#### 1.1.2 菌株、质粒、酶、试剂盒及生化试剂

菌株、质粒、酶、试剂盒及生化试剂详见文献[6]。引物合成和基因测定委托北京三博远志生物技术有限责任公司完成。

### 1.2 方法

#### 1.2.1 总 RNA 的提取和 cDNA 第一链的合成

胶乳和其他组织总 RNA 的提取参照文献[6-7];cDNA 第一链的合成参照 PrimeScript® RT reagent Kit With gDNA Eraser 试剂盒说明书。

#### 1.2.2 基因克隆

以从根转录组文库中筛选到的 EST(未发表)作

为查询序列,应用 BLASTn 程序<sup>[8]</sup>同源搜索橡胶树的基因组序列(热研 7-33-97,未发表),然后根据序列信息,采用 Primer 5 在基因编码区的两侧设计引物对 *HbCP3F*(5'-AGTACCACTTCCCATGGCATAAC-3')和 *HbCP3R*(5'-TCTTGTGTGATTCTGAATAATTATGCGGTAGG-3')。PCR 扩增、产物克隆与测序参照文献[6]。

#### 1.2.3 序列的生物信息学分析

多序列比对和进化树的构建采用 ClustalW2<sup>[9]</sup>和 MEGA 4.0<sup>[10]</sup>(用 Neighbor-Joining 法,bootstrap 值设为 1 000);蛋白的一级结构分析采用 ProtParam 和 ProtScale(<http://web.expasy.org/>);保守结构域分析、翻译后修饰、亚细胞定位分析和跨膜结构域的预测参照文献[6]。

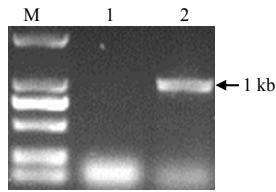
#### 1.2.4 基因的表达特性分析

根据测序结果,设计引物对 *HbCP3Fq*(5'-ATGA GAAGGAGATGCGTTACCA-3')和 *HbCP3Rq*(5'-TAGCTTGTATGGTTTGCCTGC-3');以 18SrRNA(18S F: 5'-GCTCGAAGACGATCAGATACC-3'; 18S R: 5'-TTCAGCCTTGC GACCATAC-3')作为参照基因,进行半定量分析,72 °C 延伸时间设为 20 s,目的基因和 18SrRNA 的扩增循环数分别设为 25 个和 18 个。

## 2 结果与分析

### 2.1 基因克隆结果

在前期构建的橡胶树根转录组文库(未发表)中,获得 1 条 1 305 bp 的 EST,该序列包含 1 个 1 023 bp 的 ORF,其编码蛋白与其他植物的半胱氨酸蛋白酶同源。通过将 EST 及原始读段(read)比对到橡胶树的基因组上,发现基因全长 1 393 bp,含有 1 个 90 bp 的内含子,而 EST 的 3' 端存在 2 bp 的多 A 加尾序列。以 *HbCP3F* 和 *HbCP3R* 为引物,以根反转录的 cDNA 作为模板进行 PCR 扩增,成功获得 1 条约 1 000 bp 的清晰条带(图 1)。测序结果表明,除不含内含子外,PCR 扩增到的序列与基因组的对应区域完全一致(图 2)。因前期研究组已经克隆了另一个半胱氨酸蛋白酶基因 *HbCP2*(邹智等,待发表),故将新基因命名为 *HbCP3*。



M Takara DL 2 000 ; 1 无模板的阴性对照 ; 2 以根反转录 cDNA 作为模板的 PCR 扩增。

图 1 *HbCP2* 基因的 PCR 扩增

Fig.1 PCR amplification of *HbCP3* gene

```

1 agtaccacttcccATGGCATAACGGCCCAACATCGAGGTATCTCTTTGGCATTGATATT
1 M A Y T A Q H R G I S L A L I F
61 CATTTTGGGAGCTTTGGCTTCTCAAGCCATAGCTCGTACTCTCCAAGATGCATCAATGTA
17 I L G A L A S Q A I A R T L Q D A S M Y
121 CGACAGGCATGAGGAATGGATGACTCGTTATGGACGTGATATACAAGGATGCTAATGAGAA
37 D R H E E W M T R Y G R V Y K D A N E K
181 GGAGATGCGTTACAAATATTCAAGGAAAATGATAGCGCATAGAATCTTTAAACAAAGA
57 E M R Y Q I F K E N V E R I E S F N K D
241 TGCAGCAAACCATACAAGCTAGGCATCAATAAGTTTGCAGATCTTACTAACGAAGAATT
77 A G K P Y K L G I N K F A D L T N E E F
301 CAAAACCAAGAATAAGATTCAAGGGCATATGTGTTCTGCACAAGTAGGTCCTTTCAG
97 K T T R N R F K G H M C S A Q V G P F R
361 ATATGAAAATGTTTCTGCAGTACCATCTTCCATGGACTGGAGAAAGAGGGAGCAGTTAC
117 Y E N V S A V P S S M D W R K K G A V T
421 AGCCATTAAGGACCAAGGCCAATGTGtaagtgtccaactttccactactagagaacaac
137 A I K D Q G Q C G
481 caggtcatcagtaattatatttttagttttatgaataataaaaatgtatacagGATC
146 S
541 CTGCTGGGCATTTTCTGCTGTGGCAGCTGTGGAAGGAATTACTCAGTACGAAAGTGGAAA
147 C W A F S A V A A V E G I T Q L A S G K
601 ATTGATCTCACTTTCTGAGCAAGAAGTGGTTGATGTGACACTAAGGGAGAGGATCAAGG
167 L I S L S E Q E L V D C D T K G E D Q G
661 CTGCAAGGTGGCCTAATGGATGATGCTTTAAATTCAATGAACAAAATCAAGGCCATAAC
187 C Q G G L M D D A F K F I E Q N Q G L T
721 GACTGAGGCTAATCTTATGAGGATCTGATGGAACATGCAACCAAGAAGGAAGC
207 T E A N Y P Y E G S D G T C N T K K E A
781 CAACATGCGACCAAGATAACAGGTTACGAAAGATGTTCCAGCCAACAATGAGGCAGCATT
227 N H A A K I T G Y E D V P A N N E A A L
841 AATGAGAGCTGCTCAAAGCAACAGTGTGCTTGTCCATCGATGCTGGTGGATTTCGAGTT
247 M R A A A K Q P V A V A I D A G G F E F
901 TCAGTCTATTCAAGTGGCATTTTTACAGGATCTTGTGGCACTGAAGTATGATCGGTGT
267 Q F Y S S G I F T G S C G T E L D H G V
961 TGCTGCTTGGGTATGGAGAGCAATGGAATGAATTATGGCTTGTGAAAACACTCATG
287 A A V G Y G E S N G M N Y W L V K N S W
1021 GGGCACACAATGGGGCAGGAAGGGTACATAAGGATGAGAAAGATGTTGATGCAAAGGA
307 G T Q W G E E G Y I R M Q K D V D A K E
1081 AGGGCTTGGCGCATTGCTATGCAAGCATCTTATCTACCGCATAAttattcagaatcac
327 G L C G I A M Q A S Y P T A *
1141 acaagagccactcgaagccatcctaagtaattagttcatatataatagtttctgtgta
1201 tgtgtgcttggtttagtaactaccatggttggtaacctcagctagtaggtgtacagtc
1261 ctcgtgcaagatacaggtttgcatatcagcaattattatgtaataatgtaataatctc
1321 tatattactttcttttctgctcaagagtttgatgcaataattgttctcacaactcctc
1381 ttatattatggaa

```

编码区用大写的碱基字母标识, 碱基下面为推测编码氨基酸; 由 EST 界定的转录非翻译区用小写的碱基字母标识, 其中包括 5' UTR、内含子和 3' UTR; PCR 扩增所使用的引物 HbCP3F 和 HbCP3R 用下划线标识。

图 2 *HbCP3* 基因的核酸及其推测编码的氨基酸序列

Fig.2 Nucleotide sequence and deduced amino acids of *HbCP3* gene

## 2.2 基因序列的生物信息学分析

### 2.2.1 *HbCP3* 基因及其编码蛋白的一级结构

序列分析结果表明, *HbCP3* 编码区的 GC 含量为 44.18%, 预测编码 340 个 aa, 相对分子质量为 37 350, 等电点为 5.24, pH 7.0 时的带电荷数为 -8.46。就氨基酸组成而言, 在构成该蛋白的 20 种氨基酸中, 丙氨酸(A)含量最高(11.8%), 甘氨酸(G)含量次之(10.0%), 组氨酸(H)含量最低(1.5%); 就构成氨基酸的理化特性而言, 该蛋白包含 35 个强碱性氨基酸(K、R), 44 个强酸性氨基酸(D、E), 109 个疏水氨基酸(A、I、L、F、W、V)和 95 个极性氨基酸(N、C、Q、S、T、Y), 脂肪族指数为 63.26, 不稳定系数(II)为 20.69; 蛋白疏水性最强的残基是 14 位的亮氨酸(L, 分值为 3.189), 亲水性最高的残基是 55 位的谷氨酸(E, 分值为 -2.511), 总平均疏水指数为 -0.458。除 N-端外, 蛋白的其他区域主要表现为亲水性。

### 2.2.2 蛋白的保守结构域、翻译后修饰、亚细胞定位及跨膜结构域分析

结构域搜索显示, HbCP3 含有 1 个 Inhibitor\_I29 (39-96 位, PF08246, E 值为  $2.31e^{-24}$ ) 和 1 个 Peptidase\_C1 (123-339 位, PF00112, E 值为  $1.96e^{-125}$ , 其中包含保守的半胱氨酸、组氨酸和天冬氨酰活性位点) 结构域(图 3)。MotifScan 分析显示, 蛋白的 28-36 位为酪氨酸磷酸化位点; 9-14、19-24、133-138、142-147、186-191、215-220、272-277 和 307-312 位为 N-豆蔻酰化位点; 29-32、34-37、92-95、125-128、169-172、180-183、222-225、233-236 和 280-283 位为酪氨酸蛋白激酶 II 磷酸化位点; 98-100、164-166 和 222-224 位为蛋白激酶 C 磷酸化位点; 119-122 位为天冬氨酰糖基化位点; 141-152 位为巯基蛋白酶半胱氨酸活性位点; 282-292 位为巯基蛋白酶组氨酸活性位点; 299-318 位为巯基蛋白酶天冬氨酰活性位点; 38-48 位为类似 FARP 的结构域; 176-198 位为类似 LCCL 的结构域。TMHMM 显示, HbCP3 的 N 端含有 1 个跨膜螺旋, 其中 9-26 位为跨膜区。TargetP 和 SignalP 分析表明, HbCP3 是一个分泌型蛋白, 其中 1-27 位为信号肽, 其酶切位点在 27 和 28 位之间(图 3)。

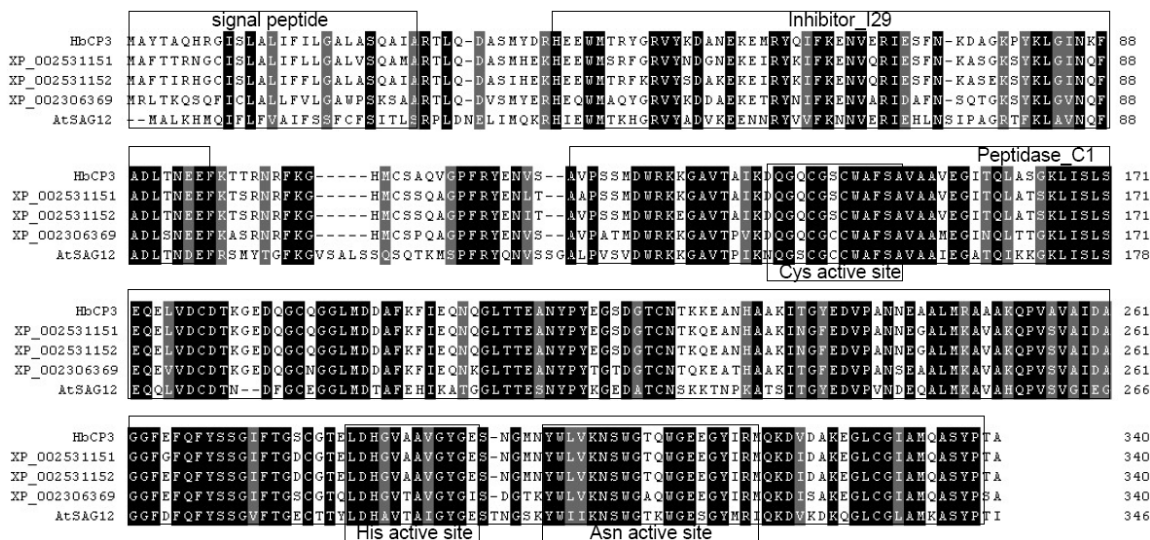


图 3 HbCP3 与其他物种中同源蛋白的多重比对结果

Fig.3 Multiple sequence alignment of HbCP3 with its homologs from other species

2.2.3 同源比对与进化分析

同源分析表明, HbCP3 与蓖麻(*Ricinus communis*, XP\_002531151, XP\_002531152)、杨树(*Populus trichocarpa*, XP\_002306369)中同源蛋白的相似性都在 80%以上, 而与拟南芥 SAGH12(AT5G45890)、橡胶树 CP1(BG33750)及本研究组克隆到 HbCP2(邹智等, 未发表)的相似性分别为 61%、50%和 89.4%, 显示出较高的保守性(图 4)。

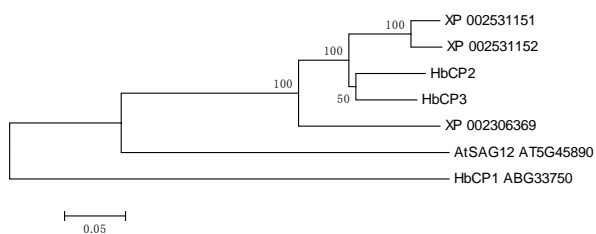


图 4 HbCP3 蛋白的进化分析

Fig.4 Phylogenetic analysis of HbCP3 and its homologs from other species

2.3 基因的表达特性分析

由于 NCBI 公共数据库中已释放了一些橡胶树芽、乳管、树皮、叶片的转录组数据<sup>[11-14]</sup>, 因此, 笔者首先对这些数据进行了同源搜索, 结果都未找到同源读段, 这表明 *HbCP3* 不在这些组织中表达, 或其表达丰度太低而测序深度不够。在此基础上, 本研究中以相同浓度的反转录 cDNA(包括根、树皮、木质部、芽、4 个叶蓬的叶片、叶柄、胶乳和乙烯利诱导的衰老叶片)作为模板, 用内参 18SrRNA 及目标基因的定量引物分别进行 PCR 扩增, 结果显示, 18SrRNA 的表达比较稳定, 而 *HbCP3* 仅在根组织中被检测到(图 5)。



P1、P2、P3、P4 分别表示从下往上的 4 个叶蓬的叶片; P4ETH 表示被乙烯利诱导的 P4 衰老叶片。

图 5 HbCP3 基因的表达特性分析

Fig.5 Expression profiles of HbCP3 gene

3 结论与讨论

半胱氨酸蛋白酶因其在植物生长发育及胁迫应答等生物学过程中广泛参与蛋白的降解及实现氮素循环利用而备受关注。HbCP3 与本研究组克隆到的另一个半胱氨酸蛋白酶基因 *HbCP2* 的相似性为 89.4%, 而与另外一个研究组报道的 *HbCP1*<sup>[5]</sup>相

比, 虽然两者在核酸水平上无明显同源, 但在蛋白水平上, 二者的相似性高达 50%。与 HbCP1 蛋白一样, HbCP3 包含 1 个 Inhibitor\_I29 和 1 个 Peptidase\_C1a 结构域, 隶属于木瓜蛋白酶 C1 家族。不同的是, HbCP3 只含有 340 个氨基酸, 比 HbCP1 的 457 个氨基酸少, 同时, HbCP3 缺乏 HbCP1 N

端的 RTX 和 C 端的 GRAN 结构域。表达分析显示, *HbCP3* 的表达特性与 *HbCP1* 明显不同, *HbCP1* 在乳管、树皮和叶片中都有表达, 且在胶乳中的表达丰度明显高于树皮和叶片<sup>[5]</sup>。在本研究所分析的根、树皮、木质部、芽、叶片、叶柄和乳管等组织中, *HbCP3* 仅在根中被检测到, 表现出明显的组织特异性, 而这得到了转录组测序<sup>[11-14]</sup>与半定量 RT-PCR 的一致证实。由于 *HbCP3* 与拟南芥半胱氨酸蛋白酶中 *AtSAGH12* 的相似性最高, 而 *AtSAGH12* 是一个衰老叶片特异表达的基因<sup>[15-16]</sup>, 因而笔者特意分析了 *HbCP3* 在不同衰老程度叶片以及用乙烯利诱导的衰老叶片中的表达特性, 结果显示基因在这些材料中也都没有得到扩增, 这进一步证实 *HbCP3* 是一个根特异性表达的基因(至少在本研究所分析的各种组织中)。生物信息学分析显示, *HbCP3* 的不稳定系数小于 40, 总平均疏水指数小于零, 且 N 端含有一段与 *AtSAGH12* 类似的分泌型信号肽, 而 *AtSAGH12* 定位于液泡<sup>[17]</sup>, 这表明 *HbCP3* 为液泡定位的、稳定的亲水性蛋白。

综上所述, *HbCP3* 很可能是一个行使 *AtSAGH12* 类似功能的半胱氨酸蛋白酶, 至于其组织特异性分化是否赋予其新的功能或生物学意义还有待研究。

#### 参考文献:

- [1] Grudkowska M, Zagdańska B. Multifunctional role of plant cysteine proteinases[J]. Acta Biochim Pol, 2004, 51(3): 609-624.
- [2] 闫龙凤, 杨青川, 韩建国, 等. 植物半胱氨酸蛋白酶研究进展[J]. 草业学报, 2005, 14(5): 11-19.
- [3] 吴春太, 苏运, 魏孝民, 等. 橡胶树无性系叶绿素含量与光合速率和白粉病抗性的关系[J]. 湖南农业大学学报: 自然科学版, 2011, 37(4): 381-384.
- [4] 吴春太, 马征宇, 刘汉文, 等. 橡胶 RRIM600 的产量与产量构成因素的通径分析[J]. 湖南农业大学学报: 自然科学版, 2014, 40(5): 478-482.
- [5] Peng S Q, Zhu J H, Li H L, et al. Cloning and characterization of a novel cysteine protease gene (*HbCP1*) from *Hevea brasiliensis*[J]. J Biosci, 2008, 33(5): 681-690.
- [6] 邹智, 杨礼富, 安峰, 等. 橡胶树 *AtCAB1* 同源基因的克隆及其在稳定与衰老期叶片中的差异分析[J]. 热带农业科学, 2013, 33(4): 30-35.
- [7] 邹智, 杨礼富. 巴西橡胶树线粒体 50S 核糖体蛋白 L21 cDNA 的克隆与分析[J]. 生物技术通报, 2011, 22(3): 100-104.
- [8] Altschul S F, Madden T L, Schäffer A A, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs [J]. Nucleic Acids Res, 1997, 25(17): 3389-3402.
- [9] Thompson J D, Higgins D G, Gibson T J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice [J]. Nucleic Acids Res, 1994, 22(22): 4673-4680.
- [10] Tamura K, Dudley J, Nei M, et al. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 [J]. Mol Biol Evol, 2007, 24(8): 1596-1599.
- [11] Xia Z, Xu H, Zhai J, et al. RNA-Seq analysis and *de novo* transcriptome assembly of *Hevea brasiliensis* [J]. Plant Mol Biol, 2011, 77(3): 299-308.
- [12] Triwitayakorn K, Chatkulkawin P, Kanjanawattanawong S, et al. Transcriptome sequencing of *Hevea brasiliensis* for development of microsatellite markers and construction of a genetic linkage map [J]. DNA Res, 2011, 18(6): 471-482.
- [13] Chow K S, Mat-Isa M N, Bahari A, et al. Metabolic routes affecting rubber biosynthesis in *Hevea brasiliensis* latex [J]. J Exp Bot, 2011, 63(5): 1863-1871.
- [14] Li D, Deng Z, Qin B, et al. *De novo* assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.) [J]. BMC Genomics, 2012, 13: 192.
- [15] Lohman K N, Gan S S, John M C, et al. Molecular analysis of natural leaf senescence in *Arabidopsis thaliana* [J]. Physiol Plant, 1994, 92(2): 322-328.
- [16] Noh Y S, Amasino R M. Identification of a promoter region responsible for the senescence-specific expression of SAG12 [J]. Plant Mol Biol, 1999, 41(2): 181-194.
- [17] Otegui M S, Noh Y S, Martínez D E, et al. Senescence-associated vacuoles with intense proteolytic activity develop in leaves of *Arabidopsis* and soybean [J]. Plant J, 2005, 41(6): 831-844.

责任编辑: 苏爱华

英文编辑: 王 库