

西伯利亚蓼硫堇基因的克隆和序列分析

刘昌财^{1,2}, 刘关君^{1,*}, 刘桂丰¹, 李晓泽¹

¹东北林业大学林学院, 哈尔滨 150040; ²中国人民解放军 61699 部队, 湖北枝江 443200

摘要: 根据西伯利亚蓼抑制消减文库(SSH)中获得的硫堇(*THI*)基因的部分序列, 应用 RACE 技术克隆了具有 Poly A 的全长 cDNA 序列。基因全长 789 bp, 5' 非翻译区 90 bp, 3' 非翻译区 276 bp, 开放阅读框编码 140 个氨基酸。序列分析表明, 该编码蛋白与大多数植物 THI 蛋白前体高度相似, N 端具 24 个氨基酸的信号肽, 中间 46 个氨基酸为成熟 THI 部分, C 端的 70 个氨基酸为酸性多肽部分。西伯利亚蓼 THI 蛋白与丹参等双子叶植物 THI 蛋白有较高的同源性, 具保守的植物 THI 标签序列 C-C-X(5)-R-X(2)-[FY]-X(2)-C。此成熟 THI 蛋白带正电荷, 偏碱性, 推定可能具有抗病原微生物活性, 为一种新的植物 THI 蛋白, GenBank 登录号为 DQ981482。

关键词: 西伯利亚蓼; 硫堇; 基因克隆; RACE

Gene Cloning and Sequence Analysis of Thionins from *Polygonum sibiricum* Laxm.

LIU Chang-Cai^{1,2}, LIU Guan-Jun^{1,*}, LIU Gui-Feng¹, LI Xiao-Ze¹

¹College of Forestry, Northeast Forestry University, Harbin 150040, China; ²The No. 61699 Army of Chinese People's Liberation Army, ZhiJiang, Hubei 443200, China

Abstract: A partial sequence of a thionins (*THI*) gene was obtained from a random clone in *Polygonum sibiricum* SSH library and *THI* was cloned by using rapid amplification of cDNA end (RACE) technology. The acquired gene was 789 bp in full length, including 5' untranslated region of 90 bp, 3' untranslated region of 276 bp with Poly A, and an open reading frame (ORF) encoding 140 amino acid. Sequence analysis indicated that the protein, such as most of the plant THI, was a precursor which could be divided into three distinct domains, including an amino-terminal (N-terminal) signal peptide, a mature THI and a carboxy-terminal (C-terminal) acidic domain, which were composed of 24, 46 and 70 amino acid respectively. THI was found to be highly homologous with that from other dicotyledon. All of them possessed a conserved plant THI signature sequence of "C-C-X(5)-R-X(2)-[FY]-X(2)-C". The mature THI had positive charge and showed slightly high basic. As result, it was possibly involved in antimicrobial activity and was identified as a new plant THI. The gene accession nucleotide sequence number in GenBank was DQ981482.

Key words: *Polygonum sibiricum*; thionins (*THI*); gene cloning; RACE

硫堇(thionins, THI)是一类首先在小麦(*Triticum aestivum*)种子中发现的分子量为5 kDa左右的碱性蛋白,有45~50个氨基酸残基。其富含半胱氨酸,具3~4个位置保守的二硫键,是一种小分子多肽,在植物种子的胚乳、茎、根以及黄化或受病原菌胁迫的植物叶子中均有发现(Florack和Stiekema 1994)。植物THI多以大分子前体形式存在,包括信号肽、成熟THI和酸性多肽三部分,依据植物的种类、发生部位、总净电荷、氨基酸的数量和成熟蛋白质中二硫键的位置,THI至少可以分为4类(Garcia-Olmedo等1989)。前3种类型偏碱性,带正电荷,具有抗病原微生物的活性(Florack等1994),第4种类型如芸苔(*Brassica rapa*)的THI仅发生在开花部位,不带电

荷,无抗菌活性(Jung等2001)。不同类型的THI对细菌、真菌、酵母、动物和植物细胞的毒性作用已有较为详细的描述(Bohlmann和Apel 1991)。与普通抗生素作用机制所不同的是,它通过结合于真菌细胞特定细胞膜上的受体,增加膜透性,抑制糖的吸收,引起钾离子、磷酸根离子、蛋白和核酸的流出,从而起到抑菌作用,因而病原菌很难形成抗性生理小种(Florack和Stiekema 1994; Pelegrini和Franco 2005)。目前,已从小麦(Van Campenhout等1998)和拟南芥

收稿 2006-08-24 修定 2006-12-26

资助 黑龙江省重点攻关项目(GB06B303)。

通讯作者(E-mail: liuguanjun2003@126.com; Tel: 0451-82190607)。

(*Arabidopsis thaliana*) (Epple等1995)等十余种植物中克隆了 *THI*。此种基因转入植物体内可提高植物的抗病力。Epple等(1997)首次报道, 内源 *THI* (*THI 2.1*) 受病原菌的入侵诱导, 转基因株系可增强拟南芥对病原真菌镰刀霉菌 (*Fusarium oxysporum*) 的抗性。燕麦 *THI* 转入水稻增强其对细菌的抗性 (Iwai等2002)。

西伯利亚蓼为蓼科植物, 多生于盐碱荒地。关于西伯利亚蓼的报道主要集中在形态学的研究, 陆静梅和李建东(1994)以扫描电镜观察西伯利亚蓼的解剖结构时, 发现其有抗盐碱的结构特征。其分子机制研究也有报道, 吕艳芳等(2006)报道, 经 NaHCO_3 溶液处理后西伯利亚蓼地下茎、茎和叶对外界盐碱胁迫的反应存在多样性。据此, 本文根据从 NaHCO_3 胁迫下西伯利亚蓼地下茎正向抑制消减文库 (SSH) 中获得的 *THI* 基因的部分序列, 采用 RACE 技术克隆了 *THI* 基因全长 cDNA 序列, 并对其及编码蛋白进行了序列分析。

材料与方法

以苗高 15~20 cm 的西伯利亚蓼 (*Polygonum sibiricum* Laxm.) 为试材, 用 3% NaHCO_3 胁迫处理 6 h, 构建西伯利亚蓼地下茎 SSH 文库。随机挑取克隆使用 MegaBACE-1000 DNA 序列分析仪进行测序 (本实验室自主完成), 经 GenBank 数据库进行 BLASTX 比对分析获得 1 个 225 bp 的 *THI* 基因片段 (图1), 推断此基因 5' 和 3' 端分别缺少约 99 和 105 个碱基, 为非全长 cDNA 序列基因。

按照 SDS/苯酚法提取西伯利亚蓼地下茎总 RNA。参照 SmartTM RACE cDNA 扩增试剂盒反转录合成 cDNA 第一链, 作为 RACE 模板。

根据已知序列设计 3' 端引物 THI-S, 序列为 5' ATGGTGTCGCTGAATACTGCAAGTTGGG 3', PCR 反应体系参照试剂盒 (SmartTM RACE cDNA Amplification Kit BD Biosciences) 说明书。反应条件为 94 °C 中变性 3 min, 然后按 94 °C 30 s、62 °C 1 min、72 °C 2 min 进行 35 轮的循环反应, 最后在 72 °C 下延伸 8 min。

根据上述方法将已获得的 3' RACE 片段在其 3' 非翻译区设计引物 THI-A 以获得基因全长, THI-A 引物序列为 5' ACAATCCAGGATCGCCGACAA-

CCTTC 3', PCR 反应体系同前。反应条件为 94 °C 变性 3 min, 然后按 94 °C 30 s、66 °C 1 min、72 °C 2 min 进行 40 轮的循环反应, 最后于 72 °C 下延伸 8 min。

PCR 扩增产物的特异性片段用宝生物工程 (大连) 有限公司胶回收试剂盒回收, 宝生物工程 (大连) 有限公司 pMD18-T 载体连接, 大肠杆菌 Top10 转化, 经 Amp⁺ 抗性及蓝白斑筛选, 随机挑取阳性克隆进行 PCR 鉴定, 由上海生工生物工程技术有限公司完成测序。

用 BioEdit 软件的 Contig Assembly Program 进行电子拼接, 采用 NCBI 中的 BLASTX 程序进行同源性比较, 开放阅读框 (open reading frame, ORF) 程序查找序列的 ORF, 软件 Clustalw 1.8.1 进行多序列比较和分子进化树的绘制。SignalP 3.0 预测蛋白的信号肽及可能的切割位点。BioEdit 软件的 Kyte-Doolittle Scale Mean Hydrophobicity Profile 程序算法进行疏水性分析。软件 ProtParam (<http://au.expasy.org/tools/protparam.html>) 计算蛋白的分子量、净电荷和理论等电点。BLASTP 预测保守区。应用 Profile Search、Scan Prosite 和 Motif Scan 程序 (<http://au.expasy.org/prosite>) 进行轮廓、结构域和基序分析。

实验结果

1 西伯利亚蓼 *THI* 基因的 3' 端和 5' 端扩增

根据 SSH 文库中获得的 225 bp *THI* 片段序列, 设计引物 THI-S (图1), 采用 3' RACE 的方法扩增 *THI* 的 3' 端。经扩增, 得到 1 条约 400 bp 的 DNA 片段 (图2), 测序后比对证实为 *THI* 的 3' 端序列。

根据 3' RACE 扩增片段的测序结果, 在其非翻译区设计引物 THI-A (图1), 采用 5' RACE 的方法扩增完整的 *THI* 全长。经扩增得到 1 条约 800 bp 的 DNA 片段 (图3), 与所推测的片段大小一致, 测序后比对证实为 *THI* 的全长序列。

2 西伯利亚蓼 *THI* 基因的序列拼接和分析

经电子拼接获得带有 Poly A 的全长 cDNA 序列为 789 bp, 应用 NCBI 的 ORF 程序显示 5' 非翻译区为 90 bp, 3' 非翻译区为 276 bp; 开放读码框为 420 bp, 编码 140 个氨基酸 (图1)。

```

1   TCA ATA CAG ACA ATT AAT TAA GCT ACT ATA CCT ACC CTT CTC GAA CGC ACA AAG TAG GCT   60
61  AAG TAG TAA TTA ACC AAT CAA GAA AGA TCG ATG GAG GGC AAA ACT GTC ATT CAA AGC GTC   120
3                                     Met Glu Gly Lys Thr Val Ile Leu Ser Val   10
121 CTC CTG CTG AGC CTG GCG ATG GCG CAA TTT CAG GAA GAG GCA AAG ACC TCC TCC CAA ACC   180
131 Leu Leu Leu Ser Leu Val Met Ala Gln Phe Glu Val Glu Ala Lys Ser Cys Cys Glu Thr   240
181 ACC ACT GGC AGG AAC ATC TAT AAT AAC TGC GGT CTT GCG GGA GGC TCC AGG GAG CCG TGT   240
211 Thr Thr Ala Arg Asn Ile Tyr Asn Ser Cys Arg Leu Ala Gly Gly Ser Arg Glu Arg Cys   300
241 GCA TCA CTC TCT GGA TCC AAA CAT GTC ACT GGG AAC ACA TGC TCT CCG GGA TCG GAG AAA   300
51  Ala Ser Leu Ser Gly Cys Lys His Val Thr Gly Asn Thr Cys Ser Pro Gly Trp Glu Lys   360
381 TTT AAT GGC ATT CTT CAA GGT GAT GGT GAT GGT GTC GCT GAA TAC TCC AAG TCG GGC TCC   360
71  Phe Asn Ala Ile Leu Gln Gly Asp Gly Asp Gly Val Ala Glu Tyr Cys Gln Leu Gly Cys   390
141 GTA TCC TCT GTC TGT GGT GGC ATG AAC TTT CTC CAA AAC TCT GAC CCA GCT AAG ATT GTC   450
91  Val Ser Ser Val Cys Gly Ala Met Thr Phe Leu Gln Asn Ser Asp Pro Ala Lys Ile Val   480
421 AAT GGA GCG GTT GAA GAA TGT ACC AAC GCA TGT TCT GCT TTC TCC ATC AAG GGC TCT ATT   480
111 Asn Gly Ala Val Glu Glu Cys Thr Asn Ala Cys Ser Ala Phe Cys Ile Lys Gly Ser Ile   540
481 GTT GCA CTT GAA CTT GGC CAA TCC AAG GTC TAA GCG ATC CAA GCG GTC GGC GGT CAT CTT   540
131 Val Ala Pro Glu Leu Ala Glu Ser Thr Val End   580
541 GCA GAG TCC TGT TTT CCG CGT TTT CAA TAA ATG TCG TGT ATA TCC ATG ACA ACT TAT GCT   600
681 ACA CTC ATC TCG TCC TAC CTG CTA GCA TTA ATT CCG TTG TTT TCG GTT TGT GGC TCC ATG   660
661 TGT CAG TCG TCT CCA CCG AAG AGG TCG TCG GCG ATC CTG GAT TGT AAT GGC TCC TAC ATG   720
721 CTG CTA TCC ACA TAA AAG TCG AGT AAT AAT GTC TCT TTA TGA AAA AAA AAA AAA AAA   780
781 AAA AAA AAA   780
    
```

图1 西伯利亚蓼 *THI* 基因全长 cDNA 序列和推导的氨基酸序列

Fig. 1 Nucleotide and deduced amino acid sequences of *THI* cDNA from *P. sibiricum*

下划线部分为 24 个氨基酸的信号肽；底纹部分为 46 个氨基酸的成熟 *THI* 蛋白；其它部分氨基酸为酸性多肽部分。方框表示起始密码子；End 表示终止子；箭头之间核苷酸部分为从 SSH 库中获得长度为 225 bp 的 *THI* 基因部分序列；加粗部分 329~356 bp 和 680~705 bp 分别为引物 *THI*-S 和 *THI*-A。

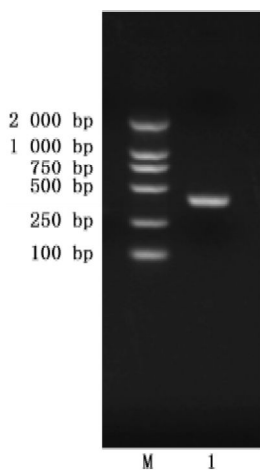


图2 *THI* 的 3' RACE 扩增产物

Fig. 2 3' RACE product of *THI* cDNA

M: DL2000 分子量; 1: 3' RACE 扩增产物。

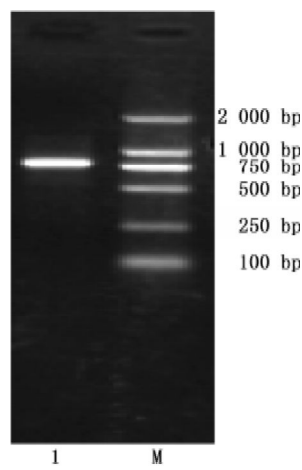


图3 *THI* 的 5' RACE 扩增产物

Fig. 3 5' RACE product of *THI* cDNA

M: DL2000 分子量; 1: 5' RACE 扩增产物。

经疏水性、SignalP 3.0、ProtParam和Scan Profile程序分析发现,编码蛋白的等电点(IEP)

为 6.5, 分子量(MW)为 14.6 kDa。N端第 1~24 位氨基酸区段疏水性很强,第 24 位 Ala 和第 25 位 Lys

为信号肽切割位点, IEP 为 4.5, MW 为 2.6 kDa, 这一段具典型信号肽特征, 为信号肽部分。第 25~70 位氨基酸区段为成熟 THI 部分, 亲水性较强, 具 6 个半胱氨酸, 占成熟 THI 氨基酸数量的 13%, 带正电荷, 呈碱性, IEP 为 9.15, MW 为 4.9 kDa。C 端第 71~140 位氨基酸区段为酸性多肽部分, 含 2 段范围较广疏水性较强的氨基酸部分, IEP 为 4.2, MW 为 7.1 kDa。这些特征与大多数植物 THI 前体一致, 推断 THI 在西伯利亚蓼地下茎中以此三部分构成的前体形式存在(图 1)。

Motif Scan 程序分析显示, 成熟 THI 部分的第 3、4、16、26、32 和 40 位的 6 个半胱氨酸(C)残基是植物 THI 极为保守的 3 个二硫键的形成位点, 第 10 和 13 位氨基酸分别为植物 THI 保守的精氨酸(R)和芳香族氨基酸[这里为酪氨酸(Y)], 第 8~10 和 15~17 位氨基酸分别为 2 个蛋白激酶 C

磷酸化位点(protein kinase C phosphorylation site) TAR 和 SCR, 以上为显著的植物 THI 标签(plant thionins signature) C-C-X(5)-R-X(2)-[FY]-X(2)-C 序列特征(图 4)。C 端酸性多肽部分第 19~24 位氨基酸区段为许多植物 THI 保守的酰基化序列 GCVSSV 位点(图 5)。

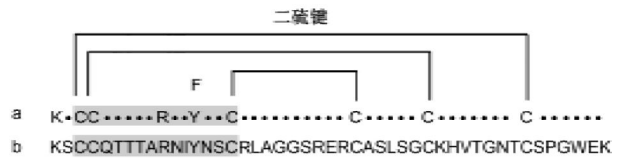


图4 西伯利亚蓼和其他植物成熟 THI 保守的氨基酸序列比较

Fig. 4 Comparison on the conserved amino acid sequences of the mature THI between *P. sibiricum* and other plants
底纹部分为植物 THI 标签; a: 植物成熟 THI 保守的氨基酸序列; b: 西伯利亚蓼成熟 THI 氨基酸序列。

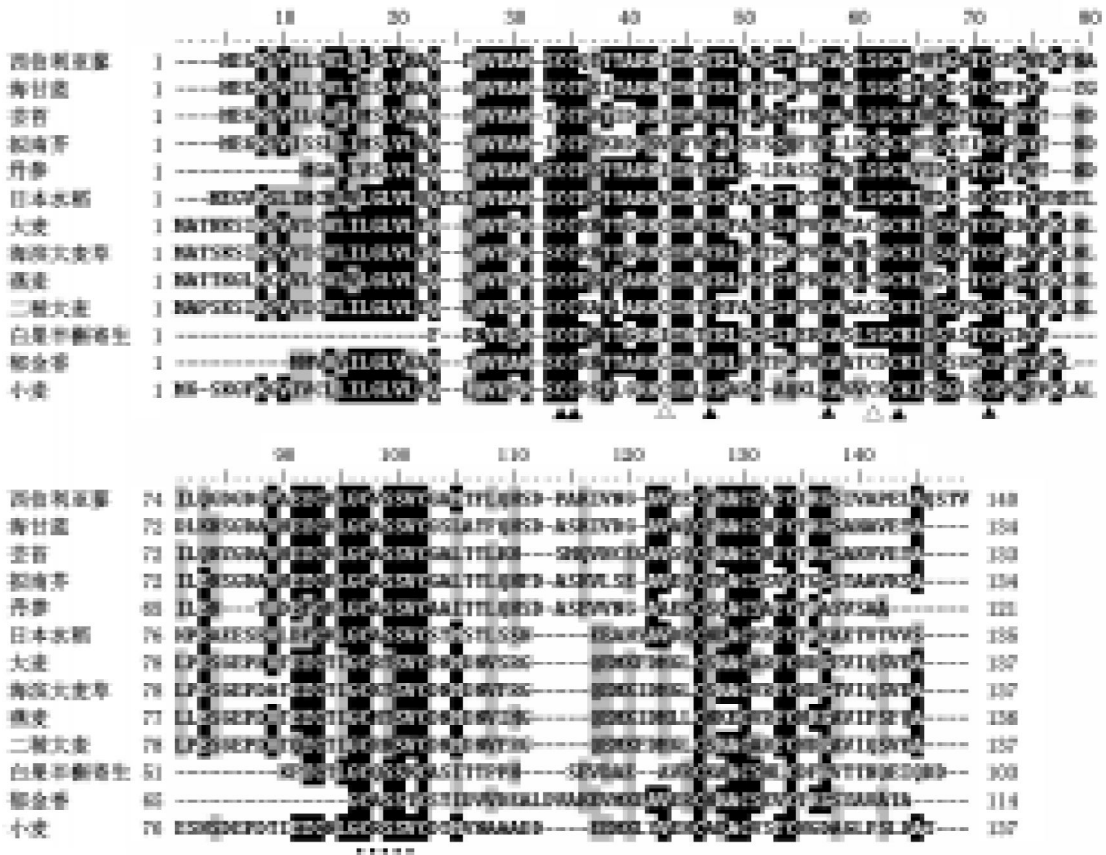


图5 西伯利亚蓼和其它12种植物的THI蛋白前体序列比较

Fig. 5 Comparison on the amino acid sequences of THI precursor between *P. sibiricum* and other 12 plants

▲ 为双子叶植物硫磺形成 3 个二硫键位置保守的六个半胱氨酸残基; 为大多数植物硫磺前体保守的酰基化序列位点; △ 为大多数单子叶植物硫磺具有的另外 2 个位置保守的半胱氨酸残基。

3 西伯利亚蓼与其它植物的THI蛋白的同源性比较

经BLASTX分析,发现此种基因与GenBank中其它植物海甘蓝(*Crambe abyssinica*)、芸苔(*Brassica rapa* ssp. *pekinensis*)、拟南芥(*Arabidopsis thaliana*)、丹参(*Salvia miltiorrhiza*)、白果槲寄生(*Viscum album*)、日本水稻(*Oryza sativa*)、大麦(*Hordeum vulgare*)、海滨大麦草(*Hordeum marinum*)、燕麦(*Avena sativa*)、栽培二稜大麦(*Hordeum vulgare*

ssp. *vulgare*)、郁金香(*Tulipa gesneriana*)和小麦(*Triticum aestivum*)的THI基因具较高的同源性,同源性在54%~72%之间。

软件Clustalw 1.8.1对此种基因相近的12个物种的THI蛋白序列进行多序列比对(图5),绘制进化树(图6)。分析表明,THI蛋白的部分序列较为保守,西伯利亚蓼与丹参、海甘蓝突变体等双子叶植物THI蛋白在进化关系上较近,而与其他

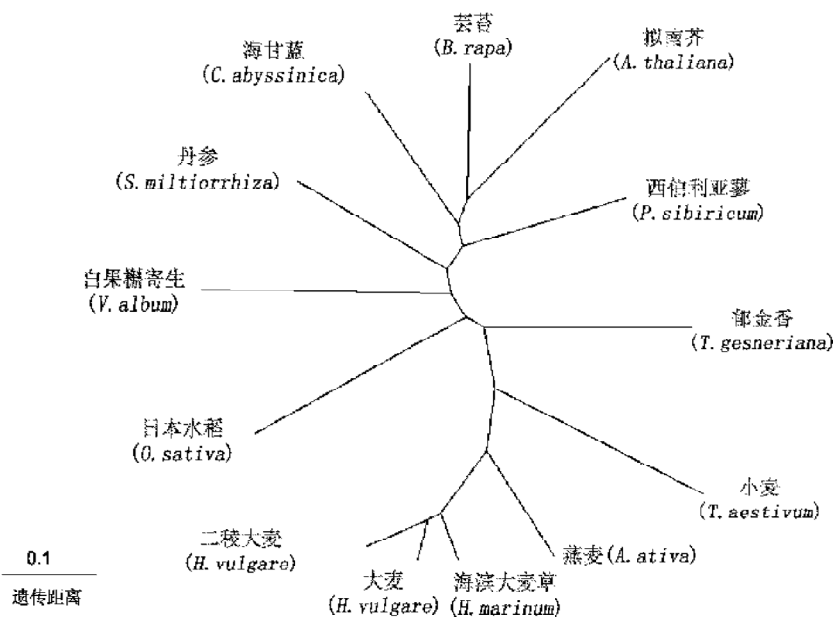


图6 植物THI蛋白的系统进化树

Fig.6 Evolutionary tree of plant THI

如日本水稻、小麦等单子叶植物THI蛋白进化关系较远。

讨 论

RACE技术是一种快速扩增cDNA 5'和3'末端的简单而有效的方法(Schaefer 1995)。结合SSH文库进行RACE,一般的步骤是根据一段已知序列的非末端cDNA片段同时设计3' RACE和5' RACE特异性引物分别扩增cDNA的3'和5'末端,将3'和5'末端扩增产物进行电子拼接全长,再在非翻译区设计特异性引物扩增到真正的全长cDNA,才能获得插入目的片段的克隆载体。本文只通过3' RACE扩增后,根据获得的3' cDNA序列,直接在3'非翻译区设计特异性引物经5' RACE扩增出完

整的THI cDNA序列,不但可减少克隆全长基因的时间和提高效率,而且可降低实验费用,这从SSH文库中克隆基因来说有一定的参考价值。

大多数单子叶植物成熟THI部分有8个位置一致的半胱氨酸残基,形成4个二硫键(Schrader-Fischert和Apel 1993)。生物信息方法分析表明,西伯利亚蓼THI在相同位置上有6个半胱氨酸,形成3个二硫键,这与双子叶植物THI的特征一致,显示其在亲缘关系上与其它双子叶植物的THI较为接近。

根据Garcia-Olmedo等(1989)对植物THI的分类,西伯利亚蓼THI应与白果槲寄生等THI同属于第3种类型,推定此种THI具抗病原微生物活性。另外,西伯利亚蓼THI片段是从碱胁迫下西

伯利亚蓼地下茎的 SSH 文库中获得的, 在碱胁迫下西伯利亚蓼地下茎的 THI 是否同时参与抗碱和抵抗真菌的作用还有待进一步研究。

参考文献

- 陆静梅, 李建东 (1994). 西伯利亚蓼解剖结构的扫描电镜观察. 东北师大学报(自然科学版), 3: 83~87
- 吕艳芳, 王大海, 刘关君, 杨传平, 周颖 (2006). 盐碱胁迫下西伯利亚蓼体内无机离子的变化. 广西植物, 26 (3): 304~307
- Bohlmann H, Apel K (1991). Thionins. Annu Rev Plant Physiol Plant Mol Biol, 42: 227~240
- Epple P, Apel K, Bohlmann H (1995). An *Arabidopsis thaliana* thionin gene is inducible via a signal transduction pathway different from that for pathogenesis-related proteins. Plant Physiol, 109: 813~820
- Epple P, Apel K, Bohlmann H (1997). Overexpression of an endogenous thionin enhances resistance of *Arabidopsis* against *Fusarium oxysporum*. Plant Cell, 9: 509~520
- Florack DEA, Dirkse WG, Visser B, Heidekamp F, Stiekema WJ (1994). Expression of biologically active hordothionins in tobacco, effects of pre- and pro-sequences at the amino and carboxyl termini of the hordothionin precursor on mature protein expression and sorting. Plant Mol Biol, 24: 83~96
- Florack DEA, Stiekema WJ (1994). Thionins: properties, possible biological roles and mechanisms of action. Plant Mol Biol, 26: 25~37
- Garcia-Olmedo F, Rodriguez-Palenzuela P, Hernandez-Lucas C, Ponz F, Marafia C, Carmona MJ, Lopez-Fando J, Fernandez JA, Carbonero P (1989). The thionins: a protein family that includes purothionins, viscotoxins and crambins. Oxford Surv Plant Mol Cell Biol, 6: 31~60
- Jung BG, Choi YO, Lee KO, Chi YH, Kang SS, Lee SY (2001). Molecular cloning and characterization of a flower-specific thionin in chinese cabbage. J Biochem Mol Biol, 34 (3): 201~205
- Iwai T, Kaku H, Honkura R, Nakamura S, Ochiai H, Sasaki T, Ohashi Y (2002). Enhanced resistance to seed-transmitted bacterial diseases in transgenic rice plants overproducing an oat cell-wall-bound thionin. Mol Plant Microbe Interact, 15: 515~521
- Pelegrini PB, Franco OL (2005). Plant γ -thionins: novel insights on the mechanism of action of a multi-functional class of defense proteins. Int J Biochem Cell Biol, 37: 2239~2253
- Schaefer BC (1995). Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends. Anal Biochem, 227: 255~273
- Schrader-Fischer G, Apel K (1993). cDNA-derived identification of novel thionin precursors in *Viscum album* that contain highly divergent thionin domains but conserved signal and acidic polypeptide domains. Plant Mol Biol, 23: 1233~1242
- Van Campenhout, Sági L, Stappen VJ, Volckaert G (1998). Characterisation of type-I thionin loci from the A, B, D and R genomes of wheat and rye. Theor Appl Genet, 96: 80~86