

doi:10.3969/j.issn.1672-5565.2013.03.11

# 油菜蔗糖转化酶基因的电子克隆和生物信息学分析

苏宁, 杨万年\*

(华中师范大学生命科学学院, 湖北 武汉 430079)

**摘要:**运用电子克隆技术获得油菜中一个蔗糖转化酶基因 cDNA 序列,同时根据此段序列设计引物以油菜 cDNA 为模板进行扩增,经测序得到证实。采用生物信息学方法,对该基因编码蛋白从氨基酸组成、基本理化性质、跨膜结构域、信号肽导肽、疏水性/亲水性、二级结构、亚细胞定位等方面进行了预测和分析。结果表明:该基因 cDNA 序列长度为 2 150bp,包含一个 1 779bp 开放阅读框,编码 592 个氨基酸;该编码蛋白含有蔗糖转化酶的多个典型的保守结构域。同源比对分析显示,该基因编码的氨基酸序列与拟南芥等植物的蔗糖转化酶基因具有高度的相似性,进一步确定该蛋白为蔗糖蛋白酶。研究结果为该基因进一步的实验克隆,表达分析,功能鉴定奠定基础。

**关键词:**油菜;蔗糖转化酶基因;电子克隆;生物信息学

**中图分类号:**Q785 **文献标识码:**A **文章编号:**1672-5565(2013)-03-224-09

## In silico cloning and bioinformatics analysis of invertase gene in *Brassica napus*

SU Ning, YANG Wan-nian\*

(College of Life Sciences Central China Normal University, Wuhan 430079, China)

**Abstract:** A Invertase gene was obtained from *Brassica napus* by using the in silico cloning technique. At the same time we amplified this sequence with rape cDNA as a template using primers designed according to the sequence of cDNA. Sequencing confirmed the reliability of the results. The protein of this gene was also predicted and studied through the bioinformatics analysis method, including the composition of amino acid sequence, basic physical and chemical properties, transmembrane region, signal peptide, guide peptide, hydrophobicity and hydrophilicity, secondary structure of protein, subcellular localization and so on. The results showed that the cDNA length of the gene was 2150bp, which included a 1 779bp open reading frame(ORF) and encoded 592 amino acids residue. The encoding protein of the INV gene possessed several typical conserved domains. Homology comparison and phylogenetic analysis showed that the amino acid encoded by INV gene in *Brassica napus* was highly homologous with those encoded by INV gene in *Arabidopsis thaliana* and so on. The above results determine the protein is invertase and these will provide the basis for molecular cloning, expression analysis and function identification of INV genes in *Brassica napus* further.

**Keywords:** *Brassica Napus*; Invertase Gene; In Silico Cloning; Bioinformatics

蔗糖转化酶 Invertase ( $\beta$ -fructofuranosidase, EC 3.2.1.26) 是一种最常见的酶,存在于酵母、细菌和植物中。在植物糖代谢中发挥着重要的作用,能够不可逆地催化蔗糖的水解生成葡萄糖和果糖。在植物中,根据其亚细胞定位,可以分为细胞壁蔗糖转化

酶、液泡蔗糖转化酶以及细胞质蔗糖转化酶。根据其最适 pH 可分为两大类:酸性蔗糖转化酶和碱性蔗糖转化酶(或中性蔗糖转化酶)<sup>[1]</sup>。碱性蔗糖转化酶存在于细胞质中。酸性蔗糖转化酶分为可溶性酸性蔗糖转化酶和细胞壁束缚蔗糖转化酶:可溶性

收稿日期:2013-02-08;修回日期:2013-03-30.

基金资助:国家自然科学基金(31270316)项目资助。

作者简介:苏宁,女,在读硕士,研究方向:植物发育与分子生物学;E-mail: ladysn2010@163.com.

\* 通讯作者:杨万年,男,教授,从事植物发育与逆境适应的分子机理、植物发育的功能基因组学研究;E-mail: yangwn@mail.ccnu.edu.cn.

酸性蔗糖转化酶是一种液泡酶,存在于植物细胞的液泡中,催化蔗糖水解成己糖,起着调节植物组织中糖的积累和液泡中蔗糖利用的作用。细胞壁束缚的酸性 Inv 主要参与韧皮部质外体卸载时蔗糖的分解,以保持库-源之间蔗糖的浓度梯度。在调控植物衰老及果实发育中起重要作用<sup>[2]</sup>。

关于植物蔗糖转化酶方面的研究很多,但是在油菜方面展开的工作却较少。油菜是我国主要的蜜源植物,而花蜜是一种糖类复合体,主要成分包括蔗糖、葡萄糖和果糖。所以,蔗糖的积累和代谢在调节油菜花蜜的产量方面也有着重要的作用。曾有研究表明拟南芥花中蔗糖转化酶基因突变后会造花蜜合成受阻,可溶性糖含量下降<sup>[3]</sup>;另有研究者通过 RNA 干扰的方式抑制了番茄中蔗糖转化酶抑制蛋白,使番茄的蔗糖转化酶的活性增强,果实中的糖分含量显著增加<sup>[4]</sup>。这一探索为我们提供了启示,利用转基因方法构建超表达植株增加油菜产蜜量也是具有可行性的。因此,本文以模式植物拟南芥中的一个蔗糖转化酶基因为研究对象,利用电子克隆的方法,获得油菜中的一个蔗糖转化酶基因的 cDNA 序列,并采用生物信息学软件对该基因结构、功能等进行综合预测分析,为进一步克隆油菜蔗糖转化酶基因,研究其在油菜发育过程中的调节作用,改善油菜的产蜜量提供理论参考。

## 1 材料和方法

### 1.1 电子克隆获得油菜蔗糖转化酶基因序列

表1 分析基因结构功能所用到的网站及软件

Table 1 The websites and softwares occupied to analyse the gene structure and its function

检索内容 Items	分析工具 Softwares
开放阅读框	<a href="http://www.ncbi.nlm.nih.gov/gorf/orfig.cgi">http://www.ncbi.nlm.nih.gov/gorf/orfig.cgi</a>
理化性质	<a href="http://web.expasy.org/cgi-bin/protparam/protparam">http://web.expasy.org/cgi-bin/protparam/protparam</a>
跨膜结构预测	<a href="http://www.cbs.dtu.dk/services/TMHMM/">http://www.cbs.dtu.dk/services/TMHMM/</a>
信号肽预测分析	<a href="http://www.cbs.dtu.dk/services/SignalP/">http://www.cbs.dtu.dk/services/SignalP/</a>
导肽预测分析	<a href="http://www.cbs.dtu.dk/services/TargetP/">http://www.cbs.dtu.dk/services/TargetP/</a>
疏水性/亲水性预测和分析	<a href="http://web.expasy.org/protscale/">http://web.expasy.org/protscale/</a>
二级结构预测	<a href="http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html">http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html</a>
亚细胞定位	<a href="http://psort.hgc.jp/form.html">http://psort.hgc.jp/form.html</a>
保守结构域分析	<a href="http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi">http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi</a>

## 2 结果与分析

### 2.1 油菜蔗糖转化酶基因的开放阅读框及其编码蛋白

以拟南芥蔗糖转化酶基因 AtCWINV4 为探针,在油菜 EST 序列数据库中同源检索后得到一系列探

以拟南芥中蔗糖转化酶 AtCWINV4 的蛋白序列为探针,利用 NCBI 的 blastp 工具搜索,得到油菜中的同源蛋白序列 AC3466101,对应的核苷酸序列为 GQ146458.1,进一步在油菜的 EST 数据库中同源检索,得到一系列高度相似的 EST 序列。再用 DNAs-tar 软件对这些序列进行拼接,获得重叠群 contig。将 contig 重叠群在 GenBank 的 EST 数据库中进行检索验证,直至没有新的 EST 序列出现,从而得到最后的新生序列。

### 1.2 以油菜 cDNA 为模板扩增并测序检测获得的新生序列

根据新生序列设计特异性引物(正向引物: TACACCAGTTCTACACACACACAATATC; 反向引物: TTATCTCGCAATTACTTTACCATCACT),提取油菜花的 RNA 逆转录为 cDNA,以此为模板扩增新生序列的开放阅读框(AtCWINV4 是在拟南芥中花中特异性表达的基因,我们以 AtCWINV4 为探针序列获得新生序列,推测其在油菜花中的表达会相对高些,故选用油菜花为材料)。将扩增获得的序列进行胶回收连接 pGEM-T 进行测序检测,运用 vector NTI 对测序结果与电子克隆获得的新生序列进行比对。

### 1.3 油菜蔗糖转化酶基因的生物信息学分析

利用在线生物信息学分析软件对基因的编码蛋白进行氨基酸组成、基本理化性质、跨膜结构域、信号肽导肽、疏水性/亲水性、二级结构、亚细胞定位、保守结构域等方面的预测与分析(见表1)。

针序列高度相似的 EST 序列,去除冗余序列后,剩下 48 条。利用 DNAsStar 软件对所获 48 条序列进行拼接分析,最终电子克隆得到一条长度为 2 150bp 的重叠群。ORFfinder 软件分析显示,该 cDNA 序列包含一个 1 779bp 的完整的开放阅读框,编码 592 个氨基酸,起始密码子 ATG 位于第 91 位,终止密码子 TAA

位于第 1 876 位。

该基因的开放阅读框为：

**ATG**GGTATCTCAAATATGTTTCTGTGTTATTATTGCTTGTACTACTCAGTTTAAAGCAGTAATAACATCAAAGGGGT  
TAAAGCTTTTCATCATGTTTACGAAAATTTGCAATCTCAATCAGTCGAGTCAGTGGATCATCTTCACAGAACCGCT  
TTTACTTTCAACCTCCAAAGCATTGGATTAACGATCCAAATGGTCCAGTATACTACAAAGGTTTCTACCATCTCT  
TCTACCAGTATAACACCAAGGGTGC GGTTTGGGGCAACATTGTGTGGGCACATTTCGGTCTCCAAAGACTTAGTA  
AATTGGGAGGCTCTTGAACATGCCCTCTATCCCTCCAAATGGTTTGACATTAAGGTACATGGTCCGGTTCAATA  
ACAATCGTACCGGGAAAAGGACCGATTATCCTCTATACAGGTGTTAACCAAACGAAACTCAAATCCAAAACCTA  
CGCAATCCCAAAGGACCCCTCAGACCCATACCTAAGAAAATGGATTAAACCAGATGATAACCCGCTCGTAATGCC  
AGACTATAACATGAACGGTTTCAGCTTTCCGTGACCCGACCACGGCATGGTTCTCCAAAGACGGGCATTGGAGAA  
CCGTGGTTGGTTCAAAAAGAAAGCACAGAGGGATTGCTTATATCTACAGAACCCGAGATTTTAAAGCATTGGGT  
AAAGGTAAGCACCCGGTTCCTCCAAAGAATCAACCGGTATGTGGGAGTGTCTGATTTCTTCCCGGTTTCCAC  
AACCGATTTCCAAAACGTTTGGACTTGGATTACACCGGTTCGAACACAAAACACGTGTTGAAAGTTAGCTTGG  
ACATAACCCGTTTTGAGTATTACACGGTGGGGAAGTACGATCCTAAGAAAGGAGAAGTACGTACCAAACGGTGAT  
ACACCCGATGGTTGGGACGGTTTGGAGATTGATTATGGTAACTTCTATGCTTCCAAGACATTCTTTGACTACAAA  
AAGAATAGAAGAATCTTGTGGGGTTGGGCCAATGAATCGGACACCGTTGAAGATGATATCTCCAAGGGTTGGGC  
TGGTCTTCAGGTGATTCCGAGAACGGTGCTTCTTGATGCAAACAAGAAGCAACTAGTGTTTTGGCCTATTGAAG  
AAATAGAGTCATTAAGAAGTAACTACGTACGAATGAACAACAAAACATCAAGACGGGTCAACGCTTAGAAGTC  
AAAGGAATCACTCTGCTCAGCCGATGTGGAAAGTACATTCAACGTCGGACAATGTCTAGATAAAGCTGAGGA  
ATTCGACCCGAGCTACACATTCAAACCGTTAGATTTGTGTAAAATAAAAAGGCTCGAATGTGACAGGTGGTGTGG  
ACTTTTCGGTTTGTATCACATTGGCCACTCCTGATTTGGAAGAGTACACTCCTGTCTTCTTTAGAGTCTTCAAAGA  
CACTTCTACAGATAAGCCTAAGGTTCTCATGTGCTCCGATGCTAGACCTTCATCGTTGAAGCAAGACAGGGGTCC  
ACTCAAACAAGATAGGATGTATAAACCATCGTTTGTGGCTTTGTGGATGTTGATCTATCTGATGGAAGGATCTCT  
CTAAGGAGTTTGATTGATCACTCGGTAGTAGAGAGCTTTGGAGCTTTAGGGAAAACAGTATAACGTCAAGAGT  
GTATCCAGTGAAAGCAGTGAAAGGGAAATGCACATTTGTATGTGTTTAAACAATGGAACACAGACTGTAAACATAG  
AGAGTCTTGATGCATGGAGCATGGAGAAGCCTTTCAGATGATGAATAATGGAGCTCTCT**AAA**

翻译的氨基酸序列为：

MGISNIVSVLLLLVLLSLSNNIKGVKAFHHVYENLQSQSVESVDHLHRTAFHFQPPKHWINDPNGPVYKGFYHLF  
YQYN TKGAVWGNIVWAHSVSKDLVNWEALEHALYPSKWF DIKGTWSGSITIVPGKGPXILYTG VNNQNETQIQNYAIP  
XDPSDPYLRKWKIPDDNPLVMPDYTMNGSAFRDPTTAWFSKDGHWRTVVGSKRKRHRGIAYIYRSRDFKHVVKGK  
HPVHSKESTGMWECPDFFPVSTTDFQNLDDLDTGSNTKHV LKVS LDIRFEYYTVGKYDPKKEKYVPNGDTPDG  
WDGLRFDYGNFYASKTFDYKKNRRILWGWANESDTVEDDISKGWAGLQVIPRTVLLDANKKQLVFWPIEIESLRS  
NYVRMNNKNIKTGQRLEVKGITPAQADVETFNVGQCLDKAEEDPSYTFKPLDLCKIKGSNVTGGVGPFLITLAT  
PDLEEYTPVFFRVFKDSTDKPKVLMCSDARPSSLKQDRGPLKQDRMYKPSFAGFVDV DLS DGRISLRS LIDHSVVE  
SFGALGKTVITSRVYPVKAVKGNHLYVFNNGTQTVNIESLDAWSMEKPLQMMNNGAL\*

## 2.2 以油菜 cDNA 扩增新生序列开放阅读框

根据获得的新生序列设计引物,以油菜花 CDNA 为模板进行扩增新生序列的开放阅读框,将扩增的片段进行胶回收后连接 PGEM-T,挑了两个单克隆 T2、T4,进行测序检测。将测序反馈结果与电子克隆获得的新生序列进行比较,结果见图 1。

如图 1 所示,1 910bp 大小的片段为目标序列,包含了 1 779bp 的开放阅读框,clipseq#3、clipseq#4 分别为 T2、T4。比对结果显示 T2、T4 与目标序列完全一致,表明了电子克隆获得的油菜蔗糖转化酶基因确实是存在的。

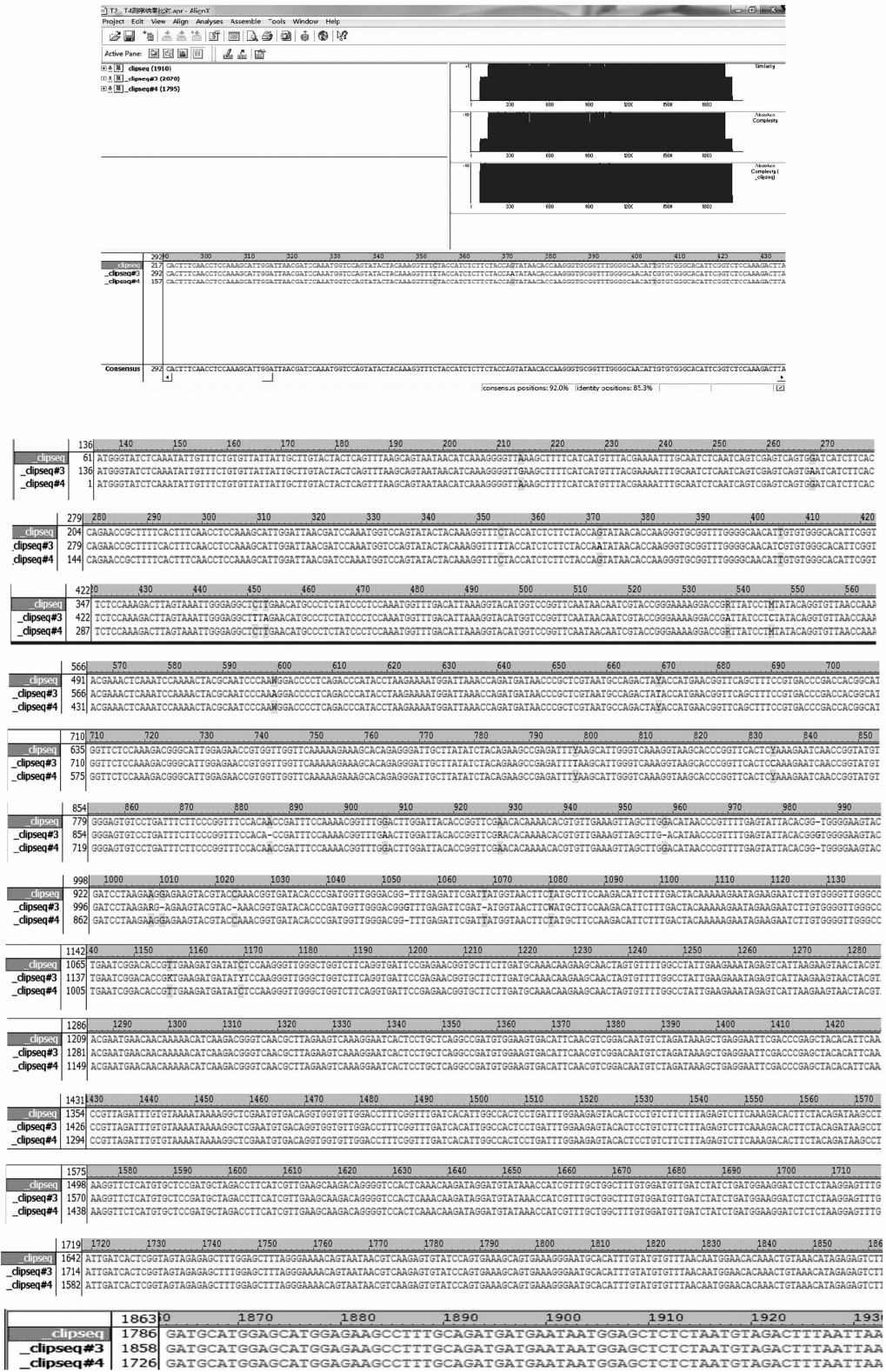


图 1 扩增的开放阅读框序列与目标开放阅读框序列的比较

Fig. 1 Comparison of amplified sequence of the open reading frame and the target open reading frame sequence

2.3 蛋白的一级结构及理化性质预测

利用 ProtParam 预测该蛋白的基本理化性质,表明该蛋白相对分子量为 67 333. 7Da,理论等电点 (PI) 为 8.61,带正点残基 (Arg + Lys) 总数为 70,带

负电残基 (Asp + Glu) 总数为 65 (见表 2)。理论推导半衰期为 30h,不稳定参数是 28.45,属于稳定蛋白<sup>[5]</sup>。该蛋白脂溶指数为 75.83,总的亲水性平均系数为-0.457,预测该蛋白属于亲水蛋白。

表 2 油菜蔗糖转化酶基因编码蛋白的氨基酸组成

Table 2 The amino acid composition of encoded protein by INV in Brassica napus

氨基酸	比例 (%)	氨基酸	比例 (%)	氨基酸	比例 (%)	氨基酸	比例 (%)	氨基酸	比例 (%)
Ala (A)	4.2	Gln (Q)	2.9	Leu (L)	7.6	Ser (S)	7.3	Ply (O)	0
Arg (R)	3.9	Glu (E)	3.9	Gly (G)	7.3	Thr (T)	5.9	Ser (U)	0
Asn (N)	5.6	Gly (G)	7.3	Met (M)	1.7	Trp (W)	2.9	Asx (B)	0
Asp (D)	7.1	His (H)	2.9	Phe (F)	4.9	Tyr (Y)	4.6	Glx (Z)	0
Cys (C)	0.7	Ile (I)	4.7	Pro (P)	5.7	Val (V)	8.1	Xaa (X)	0.3

由表 2 可知,蛋白含量较多的为 Val(8.1%), Lys(7.9%) 和 Leu(7.6%), 含量较低的是 Cys(0.7%), Xaa(0.3%); 不含 Ply、Ser、Asx、Glx 氨基酸。

## 2.4 跨膜分析

跨膜结构域是膜中蛋白与膜脂结合的主要部

位,一般由 20 个左右的疏水氨基酸组成,形成螺旋与膜脂相结合。预测和分析跨膜结构域对认识蛋白的结构、功能、分类及在细胞中的作用均有一定的意义。利用 Exspasy 软件中的 TMHMM Server2.0 工具对该蛋白的氨基酸序列进行跨膜区预测,表明该蛋白不是膜蛋白,基本无跨膜螺旋区(见图 2)。

```
# Sequence Length: 592
# Sequence Number of predicted TMHs: 0
# Sequence Exp number of AAs in TMHs: 3.56528
# Sequence Exp number, first 60 AAs: 3.54939
# Sequence Total prob of N-in: 0.18602
Sequence TMHMM2.0 outside 1 592
```

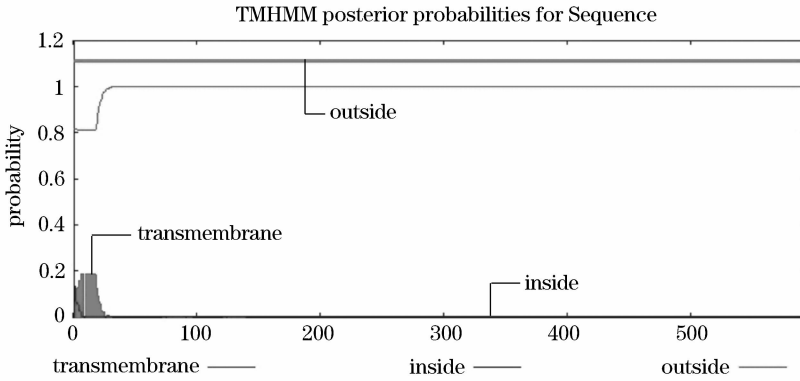


图 2 油菜蔗糖转化酶基因编码蛋白的跨膜结构域分析

Fig.2 Transmembrane analysis of the encoded protein by INV in Brassica napus

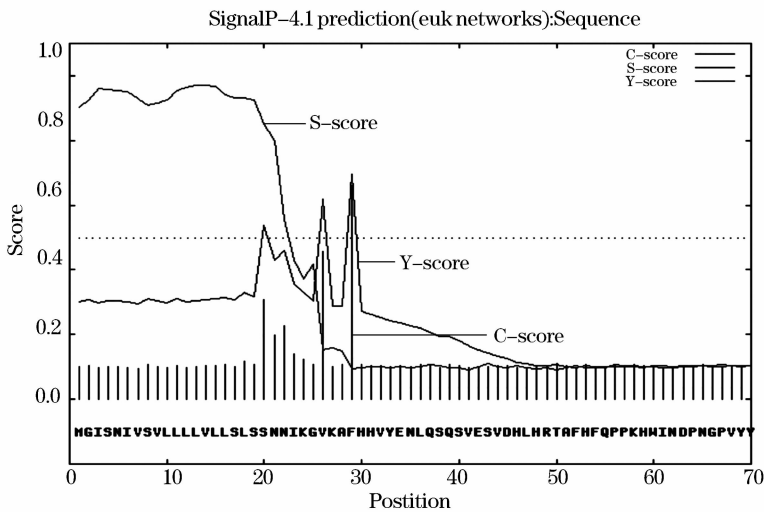


图 3 蔗糖转化酶基因编码蛋白的信号肽预测

Fig.3 Signal peptide prediction of encoded protein by INV in Brassica napus

注:C score:原始剪切位点的分值;S score:信号肽的分值;Y score:综合剪切位点的分值

Notes:C score:Scores of Putative cleavage site;S score:Scores of Signal peptide;Y score:Scores of Synthesis cleavage site

### 2.5 信号肽及导肽预测分析

信号肽常指新合成多肽链中用于指导蛋白质的跨膜转移(定位)的N-末端的氨基酸序列。曾有研究表明,酸性蔗糖转化酶的N-末端上游有一个长达100个氨基酸的序列,很可能组成一个信号肽和一个前肽,可能起蛋白质折叠、蛋白定位和酶活调节作用<sup>[1]</sup>。利用SignalP 4.1 Server 预测该蛋白的信号肽,结果如图3。第29位Phe具有最高原始剪切位

点的分值0.660和最高综合剪切位点的分值0.695;第13位Leu具有最高信号肽分值为0.972。该蛋白的平均信号肽分值为0.778,可推测该蛋白存在信号肽。

同时利用TargetP1.1Server进行导肽分析也证实了其作为叶绿体转运蛋白(分值为0.014)和线粒体转运蛋白(分值为0.035)的可能性比较小,该蛋白存在信号肽(分值为0.992)(见图4)。

```
### targetp v1.1 prediction results #####
Number of query sequences: 1
Cleavage site predictions not included.
Using PLANT networks.
```

Name	Len	cTP	mTP	SP	other	Loc	RC
Sequence	592	0.014	0.035	0.992	0.019	S	1
cutoff		0.000	0.000	0.000	0.000		

图4 油菜蔗糖转化酶基因编码蛋白的导肽预测

Fig.4 Guide peptide prediction of encoded protein by INV in Brassica napus

### 2.6 蛋白的疏水性/亲水性预测和分析

氨基酸疏水性和亲水性在形成和维持蛋白质高级结构中起着重要的作用,具有疏水性的氨基酸在蛋白质的三维结构中趋向于朝向内部,而具有亲水性的氨基酸更常以朝蛋白质表面的状态被发现。对疏水性/亲水性的预测和分析,为蛋白质高级结构的

预测和功能分析提供理论依据。

采用ProtScale分析此蛋白的疏水性/亲水性结果表明(见图5),亲水性最强分值为-2.978,位于多肽链的第290位的Lys;疏水性最强分值为3.500,位于多肽链第10位的Leu。总体来看,整条多肽链表现出较强的亲水性,推测其为可溶性蛋白。

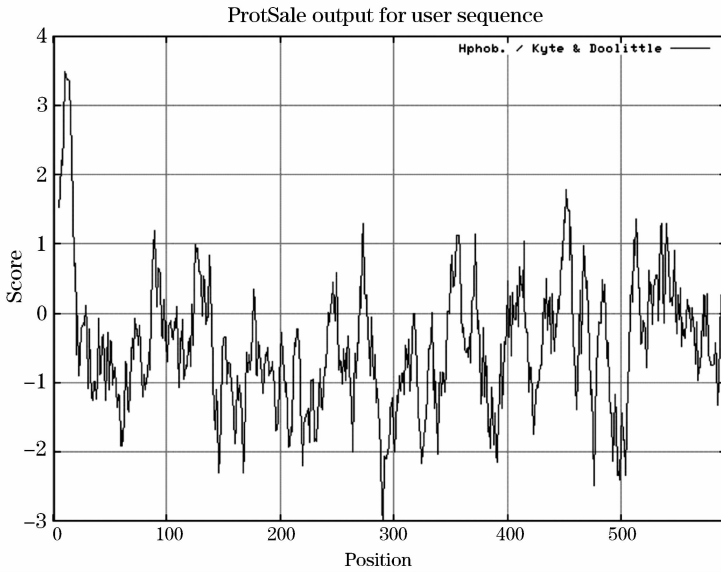


图5 油菜蔗糖转化酶基因编码蛋白氨基酸疏水性/亲水性预测

Fig.5 Prediction the hydrophobic/hydrophilic of encoded protein by INV in Brassica napus

### 2.7 二级结构预测分析

运用SOPMA法对该蛋白的二级结构进行在线分析,结果如图6。分析发现,该蛋白的二级结构主要有三种形式组成, $\alpha$ -螺旋(Alpha helix),延伸链

(Extended strand)和无规则卷曲(Random coil),具体组成见表3。其中无规则卷曲是该蛋白二级结构中最多的结构元件, $\alpha$ -螺旋和延伸链则分散在蛋白质中。

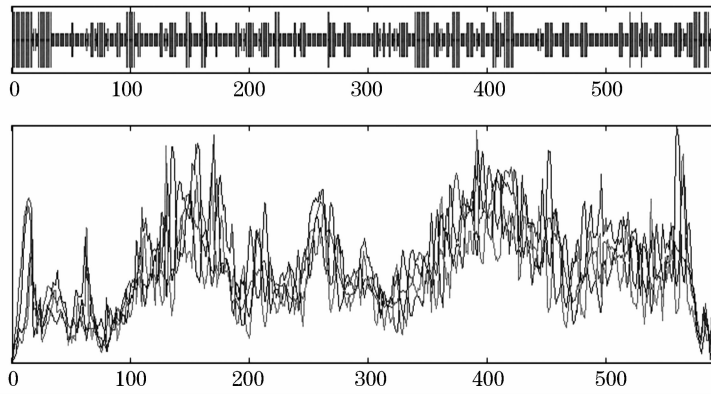


图 6 油菜蔗糖转化酶基因编码蛋白的二级结构预测

Fig. 6 Prediction of the secondary structure of encoded protein by INV in Brassica napus

表 3 油菜蔗糖转化酶基因编码蛋白  
二级结构各种构象组成

Table 3 The composition and percentage of the secondary structure of encoded protein by INV in Brassica napus

二级结构类型	氨基酸数目(个)	百分比(%)
α-螺旋	89	15.03
延伸链	151	25.51
无规则卷曲	316	53.38
β转角	36	6.08

## 2.8 亚细胞定位

通过对该蛋白的亚细胞定位情况进行分析,结果如表 4,发现该蛋白在质膜外行使功能的可能性最大,概率为 0.820,我们推测该蛋白可能位于细胞壁上。

## 2.9 保守结构域分析

利用 NCBI 中的保守结构域数据库(Conserved Domain Database, CDD)分析发现该蛋白含有 active

site, substrate binding site, GH32\_B\_Fructosidase 结合位点和保守域,属于 GH43\_62\_32\_68 superfamily 家族(见图 7),与探针序列(拟南芥 AtCWINV4)编码蛋白的保守结构域一致,属于蔗糖转化酶。

表 4 油菜蔗糖转化酶亚细胞定位结果

Table 4 Subcellular location of encoded protein by INV in Brassica napus

存在部位	概率
胞外 outside	0.820
微体(过氧化物酶体) microbody (peroxisome)	0.315
内质网膜 endoplasmic reticulum (membrane)	0.100
内质网腔 endoplasmic reticulum (lumen)	0.100

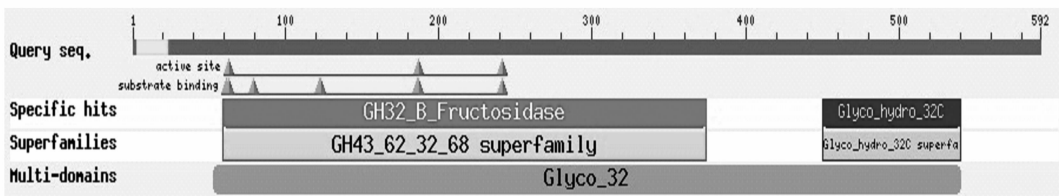


图 7 油菜蔗糖转化酶保守结构域分析

Fig. 7 Conserved domain of encoded protein by INV in Brassica napus

## 2.10 油菜蔗糖转化酶蛋白的氨基酸同源性分析

对油菜蔗糖转化酶蛋白的氨基酸在 GegeBank 数据库中的 Blastp 中进行同源性分析,选用 E 值为 0 的几个物种进行比较。发现该编码蛋白的氨基酸序列与拟南芥、蓖麻、木薯、胡萝卜、番木瓜、烟草、马铃薯、番茄中的蔗糖转化酶的氨基酸序列相似性分别为 99%, 98%, 97%, 97%, 98%, 96%, 97%, 96%。

这些物种均包含了酸性蔗糖转化酶多个保守结构域: Y/FHFQP, WINDPNGP, YHLFYQ, WAHS, IL-YTG, FRDPTTAW, WECPDFFPV<sup>[6]</sup>。从 genedoc 软件对油菜蔗糖转化酶蛋白的氨基酸序列和其他物种的蔗糖转化酶蛋白的氨基酸序列进行多重比对结果(见图 8)及几个物种的蔗糖转化酶的系统进化树(见图 9)进一步证明本研究电子克隆获得的目的基

因确实为蔗糖转化酶基因。

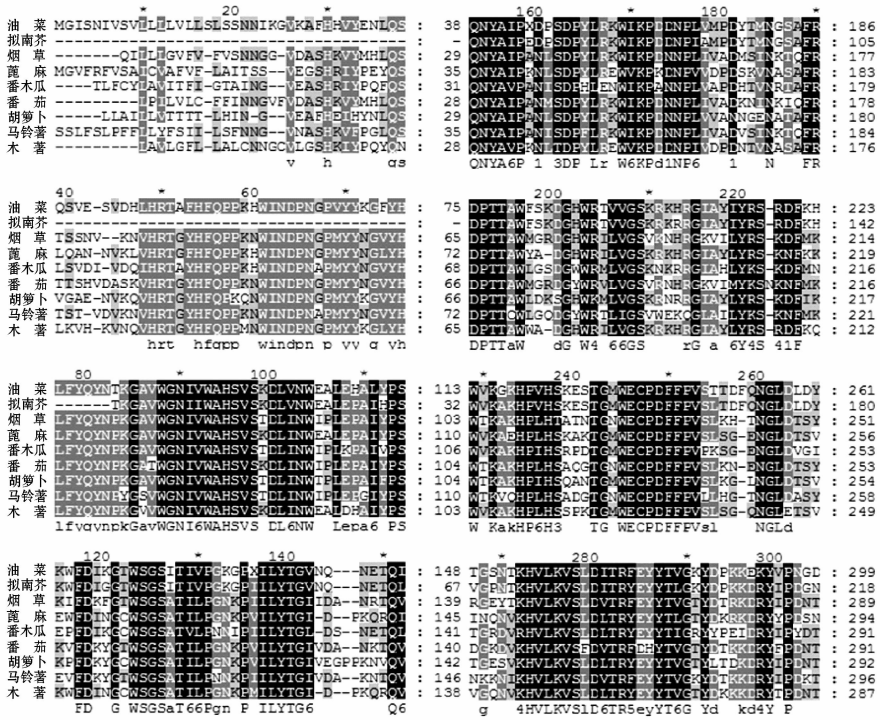


图 8 油菜蔗糖转化酶与其他物种蔗糖转化酶的氨基酸序列比对

Fig. 8 Homology analysis of INV endocodes amino acid sequences and those from other plant species

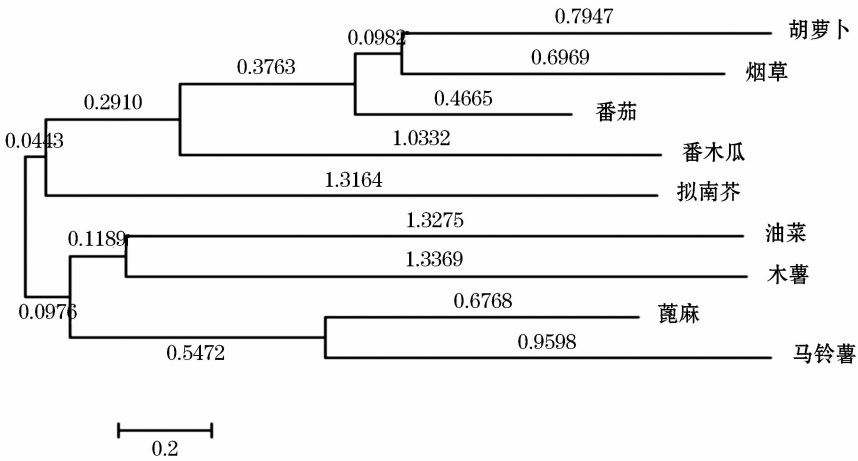


图 9 九个物种蔗糖转化酶蛋白序列进化树

Fig. 9 Phylogenetic tree of invertase protein of nine plant species

### 3 讨论

电子克隆技术是近年来基于基因组计划和 EST 计划 (Expressed Sequence Tag, 表达序列标签) 的实施而发展起来的新型基因克隆技术。其原理是利用不同物种的同类基因之间存在序列保守性<sup>[7-8]</sup>, 技术核心则是运用生物信息学技术组装延伸 ESTs 序列, 获得基因的部分乃至全长 cDNA 序列进一步利用 RT-PCR 的方法进行克隆分析、验证。相较于传统的基因克隆方法, 电子克隆具有成本低、效率高、针对性强和技术要求低的优势<sup>[9-10]</sup>。利用电子克隆

技术获得新基因虽然受到 EST 的数量和质量的限制, 但是近年来油菜基因组测序的逐渐完善, 为本研究提供了良好的材料基础。

本研究采用生物信息学手段获得了油菜中的一个蔗糖转化酶基因, 具有 1 779bp 的完整阅读框, 编码 592 个氨基酸, 是一种稳定蛋白。不具跨膜结构, 存在信号肽和导肽, 可能为可溶性分泌蛋白。同源分析显示该基因具有多个典型的蔗糖转化酶结构域, 与其他植物的蔗糖转化酶基因具有高度相似性, 保守性较强。同时, 根据目标序列设计特异性引物,



以油菜花 cDNA 为模板扩增新生序列的开放阅读框,测序比对结果显示与电子克隆获得的蔗糖转化酶基因的开放阅读框完全一致,强有力地证明了运用电子克隆技术去发现新基因是可行的,对于本研究而言预测的油菜蔗糖转化酶基因也确实存在的。本文选取的是油菜花中的 cDNA,因为探针序列拟南芥的 AtCWINV4 是在花中特异性表达的,我们推测电子克隆获得的油菜蔗糖转化酶基因在油菜花中的表达也可能相对高些。而该基因在油菜其他组织中的表达值得我们进一步探索,今后将以此为方向,对油菜蔗糖转化酶基因的表达、功能鉴定、实际应用展开更深入的研究。

### 参考文献(References)

- [1] 李肖粟,王建设,张根发. 植物蔗糖转化酶及其基因表达调控研究进展[J]. 园艺学报,2008,35(9):1384-1392.
- [2] 刘慧英,朱祝军. 转化酶在高等植物蔗糖代谢中的作用研究进展[J]. 植物学通报,2002,19(6):666-674.
- [3] Jeffrey M. Ruhlmann, Brian W. Kram and Clay J. Carter. CELL WALL INVERTASE 4 is required for nectar production in Arabidopsis[J]. Journal of Experimental Botany, 2010,61(2): 395-404.
- [4] Jin Y, Ni DA, Ruan YL. Posttranslational elevation of cell wall invertase activity by silencing its inhibitor in tomato delays leaf senescence and increases seed weight and fruit hexose level[J]. Plant Cell,2009,21(7):2072-2089.
- [5] 陈珊珊,郭晋隆,李国印,阙友雄,许莉萍甘蔗过氧化氢酶基因的电子克隆及生物信息学分析[J]. 生物信息学,2012,10(1):65-70.
- [6] Xuemei Ji, Wim Van den Ende, Andre Van Laere, et al. Structure, Evolution, and Expression of the Two Invertase Gene Families of Rice[J]. Journal of Molecular Evolution, 2005,(60):615 - 634.
- [7] 高静,程汉,蔡海滨,安泽伟,黄华孙巴西橡胶树 HbRD22 基因的电子克隆与生物信息学分析[J]. 热带作物学报,2010,31(9):1439-1444.
- [8] 冯浩咏,苗向阳,罗绪刚. 生物信息学在发现新基因方面的应用[J]. 生物技术通报,2007,(5):1-5.
- [9] 王冬冬,朱延明,李勇,李杰,柏锡电子克隆技术及其在植物基因工程中的应用[J]. 东北农业大学学报. 2006,37(3):403-408.
- [10] 胡骥,萧浪涛. 生物信息学在新基因全长 cDNA 电子克隆中的应用[J]. 生物技术通报,2007,(4):93-96.