

玉米纹枯病菌 (*Rhizoctonia solani* AG-1-IA) β -1,4-内切纤维素酶基因的生物信息学分析

14 组

李海云 乔芬
李乾楠 郭继元



报告内容

一

研究背景

二

β -1, 4-内切纤维素酶基因的生物信息学分析

三

结论



一. 研究背景

症状



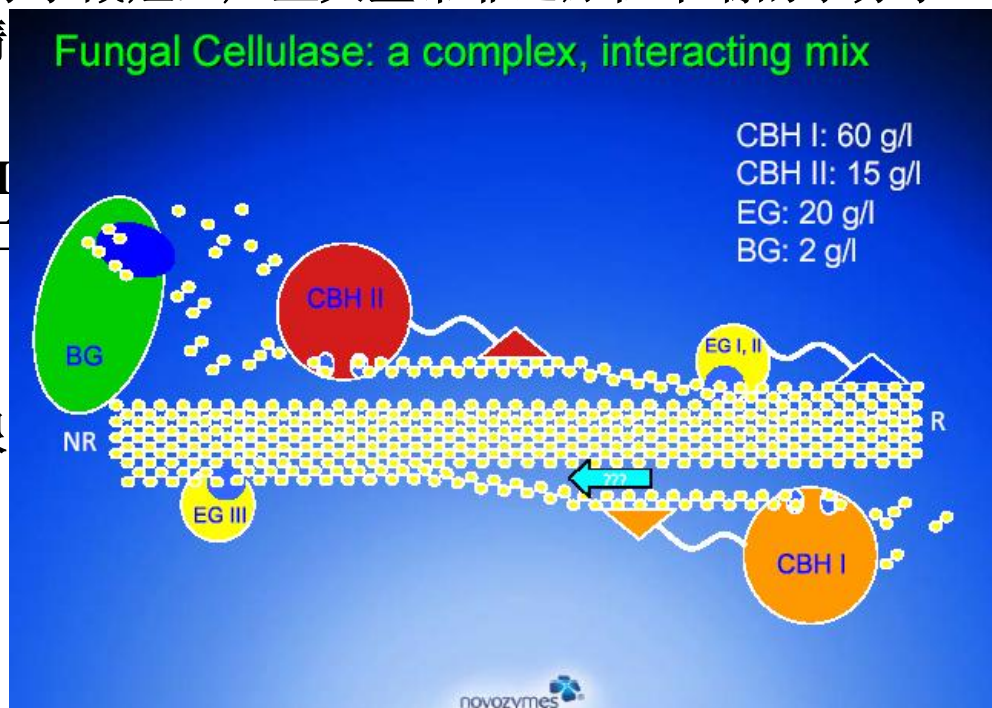
- ◆ 多数研究者认为，引起玉米纹枯病的病原菌有立枯丝核菌 (*Rhizoctonia solani*)、禾谷丝核菌 (*R. cerealis*) 和玉蜀黍丝核菌 (*R. zaeae*)，其中玉蜀黍丝核菌常危害果穗导致穗腐，禾谷丝核菌主要侵害小麦，而玉米纹枯病的主要病原菌是立枯丝核菌，立枯丝核菌的AG-1-IA菌丝融合群为引致玉米纹枯病的优势病原菌。
- ◆ 近几十年的研究成果表明，植物病原真菌的致病因子主要有酶、毒素和激素三大类。其中，纤维素酶是纤维素生物降解过程很重要的一组酶，主要是降解植物细胞壁的纤维素组分，导致组织坏死，利于病原菌的侵入，在植物病原真菌的致病性中起重要作用。

◆ 纤维素酶是一类能够水解纤维素的 β -D-糖苷键生成葡萄糖的多组分酶的总称。纤维素酶属于糖苷水解酶类，主要来自于真菌和细菌。根据其催化反应功能的不同可分为：

(1) 内切葡萄糖苷酶(简称EG)：作用于纤维素分子内部的非结晶区，随机水解 β -1,4-糖苷键，将长链纤维素分子截短，产生大量带非还原性末端的小分子纤维素。主要产物是纤维糊精

(2) 外切葡萄糖苷酶（简称CBH）
糖苷键，每次切下一个纤维二糖

(3) β -葡萄糖苷酶（简称BG）：
纤维二糖和纤维三糖的水解很
，这种酶的专一性比较差。



植物病原真菌中已克隆的纤维素酶基因

Source	Designation	Type	Family	Accession number
<i>Gaeumannomyces graminis</i>		Glucanase	3	GGU35463
<i>Leptosphaeria maculans</i>	Lema_P062490	Glucanase	3	CBX90257
<i>Rhizoctonia solani</i> AG3		Glucanase	3	DQ926702
<i>Humicola grisea</i>	egl2	Endoglucanase	5	D84470
<i>Humicola insolens</i>	CMC3	Endoglucanase	5	X76046
<i>Penicillium brasilianum</i>	cel5c	Endoglucanase	5	EU379561
<i>Penicillium pinophilum</i>		Endoglucanase	5	D84470
<i>Postia placenta</i> Mad-698-R		Endoglucanase	5	XM_002476990
<i>Macrophomina phaseolina</i>		Endoglucanase	5	U14948
<i>Malbranchea cinnamomea</i>	cel6a	Eellobiohydrolase	6	CQ838174
<i>Humicola insolens</i>	avi2	Cellulase	6	BAB39154
<i>Leptosphaeria maculans</i>	Lema_P047310	Cellobiohydrolase	6	CBX92025
<i>Gibberella zeae</i>		Cellulase	6	AAQ72468
<i>Fusarium oxysporum</i>	CBM	Cellulase	6	AAA65585
<i>Coniophora puteana</i>	cel6B	Cellulase	6	BAH59083
<i>Albugo laibachii</i> Nc14	AlNc14C28G2684	Glucanase	6	CCA16954

<i>Melanocarpus albomyces</i>	cel7A	Endoglucanases	7	AJ515704
<i>Alternaria alternata</i>	C1	Exoglucanase	7	AAF05699
<i>Cochliobolus carbonum</i>	CEL1	Cellobiohydrolase	7	AAC49089
<i>Albugo laibachii</i> Nc14	AINc14C130 G6935	Exoglucanase	7	CCA21686
<i>Humicola grisea</i> var. <i>thermoidea</i>	exo1	Exoglucanase	7	BAA74517
<i>Gibberella zeae</i>	CBH-C	Cellobiohydrolase	7	AAO42612
<i>Xylaria polymorpha</i>	cbhI	exocellobiohydrolase	7	ABY61357
<i>Fusarium chlamyosporum</i>	cbh-c	Cellobiohydrolase	7	ABY16677
<i>Fusarium avenaceum</i>	Cbh-C	Glucanase	7	AAN07161
<i>Emericella nidulans</i>	AN0452-2	Endoglucanase	12	DQ490466
<i>Hypocrea jecorina</i>	egl	Endoglucanase	12	AB003694
<i>Cochliobolus carbonum</i>	MLG2	Glucanase	12	AF229447
<i>Fusarium equiseti</i>	cel12A	Endoglucanase	12	AF434182
<i>Bionectria ochroleuca</i>	cel12A	Endoglucanase	12	AF435063
<i>Rhizoctonia solani</i>	EG12	Endoglucanase	12	FJ550135
<i>Stachybotrys echinata</i>	cel12A	Endoglucanase	12	AF435067
<i>Humicola grisea</i>	cel12A	Endoglucanase	12	AF435071
<i>Stachybotrys chartarum</i>	cel12A	Endoglucanase	12	AM180511

<i>Gibberella zeae</i>		Endoglucanase	45	AY342397
<i>Fusarium oxysporum</i>	Cel45	Endoglucanase	45	L29381
<i>Rhizopus oryzae</i>	rce2	Endoglucanase	45	AB056667
<i>Penicillium decumbens</i>	Cel45	Endoglucanase	45	EU855738
<i>Humicola insolens</i>	egV	Endoglucanase	45	A21793
<i>Ustilago maydis</i>	egl1	Endoglucanase	45	S81598
<i>Melanocarpus albomyces</i>	Cel45A	Endoglucanase	45	AJ515703
<i>Humicola grisea</i>	egl3	Endoglucanase	45	AB003107
<i>Alternaria alternata</i>	K1	Endoglucanase	45	AF176572
<i>Gibberella zeae</i>		Endoglucanase	45	AY342397
<i>Volvariella volvacea</i>	egII	Endoglucanase	61	AY559101
<i>Trichoderma</i> sp. SSL		Endoglucanase	61	FJ040192

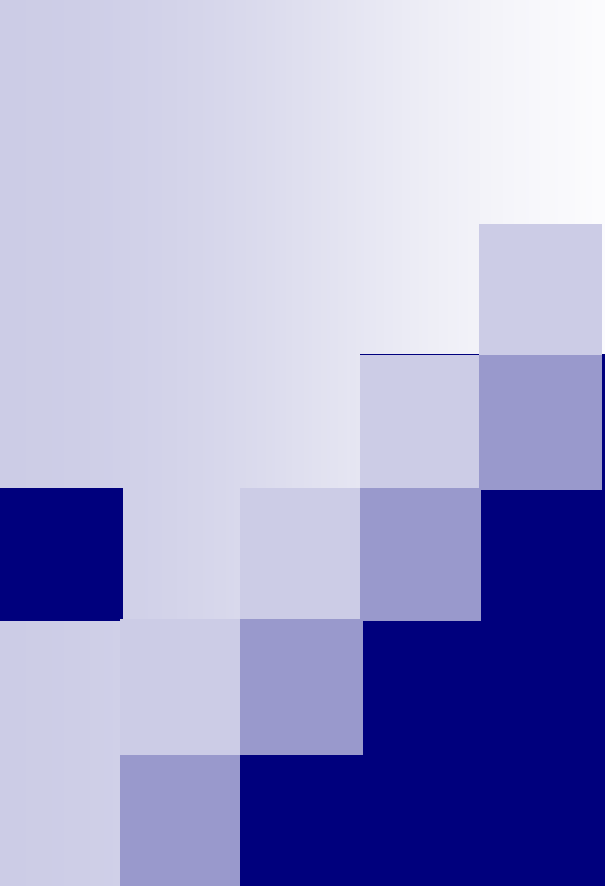
研究材料

- **cDNA:**

- **ATGAAGCTCGTCATTTCTCTCGCTACCACCGTCTTTGGTGCTTTGACTAT
TGTGAATGCTCAGTCCGGTACGGGCGCTACGACGCGCTACTGGGACTG
CTGCAAAGAAAGCTGCGGTTGGTCTGGAAAAGCGAGTGTGAGCGCTCC
AGTGCGCAGCTGCGACAAGAACAACAACCCGCTCTCGGACTACAACGCC
CAGTCAGGGTGCGGTAGTGGTGGAGCATATGCGTGTGCTAACCATTAC
CGTGGGCTGTAAACGATAGTCTTTCCTACGGCTTTGCAGCTGTAAAAC
TCAGGGTGGCGCCGAGTCTACATGGTGCTGCCAATGCTACGAGCTGAC
ATTCACGAGTGGCCCTGTTTCTGGAAAGAAGATGATTGTCCAGACCACC
AATACAGGTGGGGATCTCGGAAACAACCACTTTGACCTTATGATCCCGG
GCGGTGGAGTAGGTATTTTCGGGCAAGGGTGCATCTCAGTACGGGG
CACCTTCTACAGGGTGGGGTGCACAGTATGGAGGCGTTTCGAGCCGAT
CGCAGTGCTCTCAACTTCCTTCGGCACTCCAAGCCGGCTGTAACCTGGCG
CTTTGACTGGTTCCAGAACGCGGACAACCCATCGGTCTCGTTCAAGCAA
GTTACCTGCCCGAGCCAGCTCACTAGCATCACCGGATGCTCTCGCAAGT
AA**

- **DNA:**

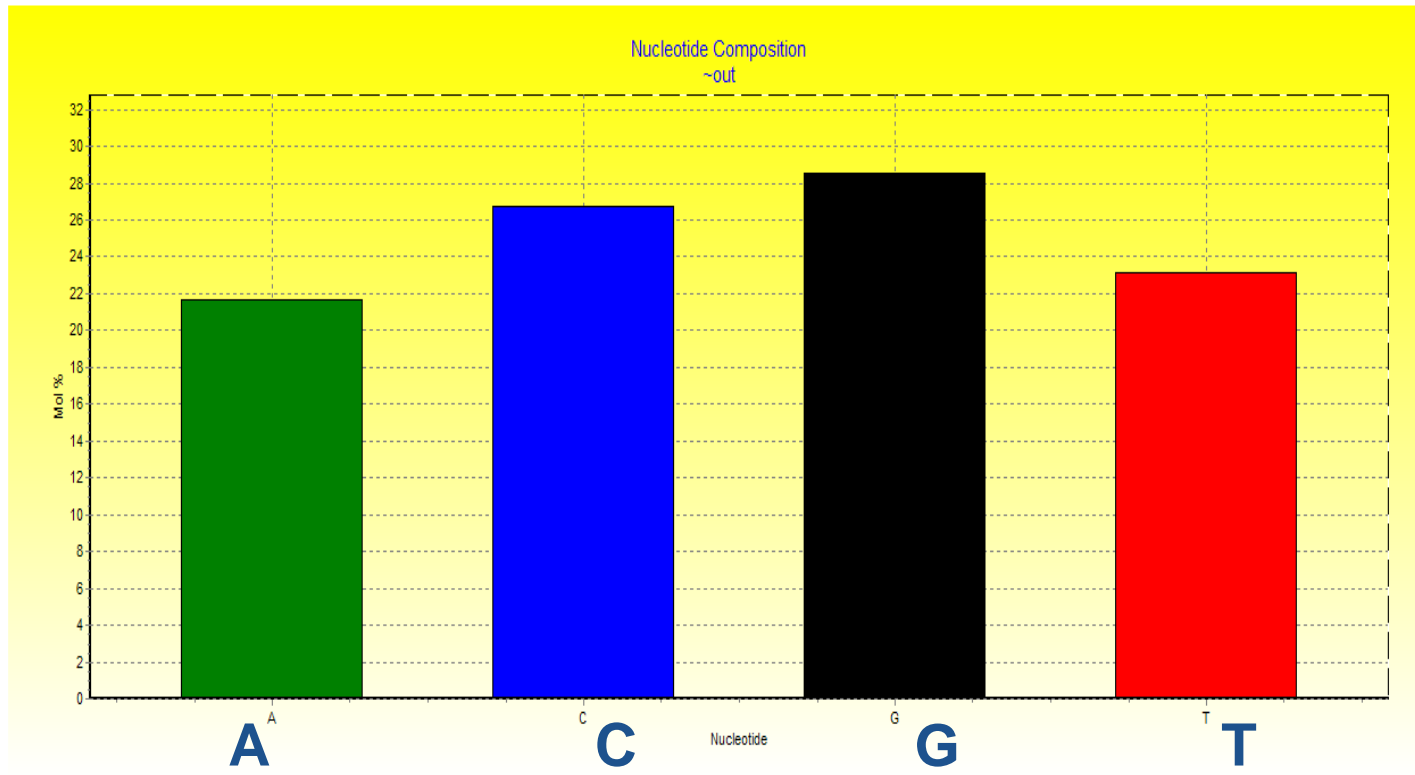
- **ATGAAGCTCGTCATTTCTCTCGCTACCACCGTCTTTGGTGCTTTGACTAT
TGTGAATGCTCAGTCCGGTACGGGCGCTACGACGCGCTACTGGTGAGTT
CAAACCTGAATGCCAATTTCTTCTCCCTCATATCTTGCCTTTAATAGGG
ACTGCTGCAAAGAAAGCTGCGGTTGGTCTGGAAAAGCGAGTGTGAGCG
CTCCAGTGCGCAGCTGCGACAAGAACAACCCGCTCTCGGACTACAA
CGCCCAGTCAGGGTGCAGTAGTGGTGGAGCATATGCGTGTGCTAACCA
TGTTTCGTACGCATGTTTCGCTGTGCCCTGCGTCTTACTAATATCAAGTGTA
TTATAGTCACCGTGGGCTGTTAACGATAGTCTTTCCTACGGCTTTGCAG
CTGTA AAACTTCAGGGTGGCGCCGAGTCTACATGGTGCTGCCAATGCTA
CGAGCTGACGTAAGTTTTCCCGCATTGTAACGCATATATTCGCTATAGA
TAACACGTGCTATATAGATTCACGAGTGGCCCTGTTTCTGGAAAAGAAGA
TGATTGTCCAGACCACCAATACAGGTGGGGATCTCGGAAACGTAGGTAT
TCTCACTTTGTAATCGAAGCACCCTCTAACCACCCTTTAGAACCACTT
TGACCTTATGATCCCGGGCGGTGGAGTAGGGTGCATATTGAGTAGATTG
AGATCATGATTAGTACTTATACTCGAAAAAGTATTTTCGGGCAAGGGTG
CGCATCTCAGTACGGGGCACCTTCTACAGGGTGGGGTGCACAGTATGG
AGGCGTTTCGAGGTACGCAAACGGACCAGGCTCGATTCAATAACATTTT
CTGACATCATATCTCAAGCCGATCGCAGTGCTCTCAACTTCCTTCGGCA
CTCCAAGCCGGCTGTA ACTGGCGCTTTGACTGGTTCCAGAACGCGGACA
ACCCATCGGTAGGCCCGAGTTTACTCGGAGATTTGATCTCGTCCATTGA
CCAAA ACTTCTCAGGTCTCGTTCAAGCAAGTTACCTGCCCGAGCCAGCT
CACTAGCATCACCGGATGCTCTCGCAAGTAA**



二. β -1, 4-内切纤维素酶 EG-1基因的生物信息学分析

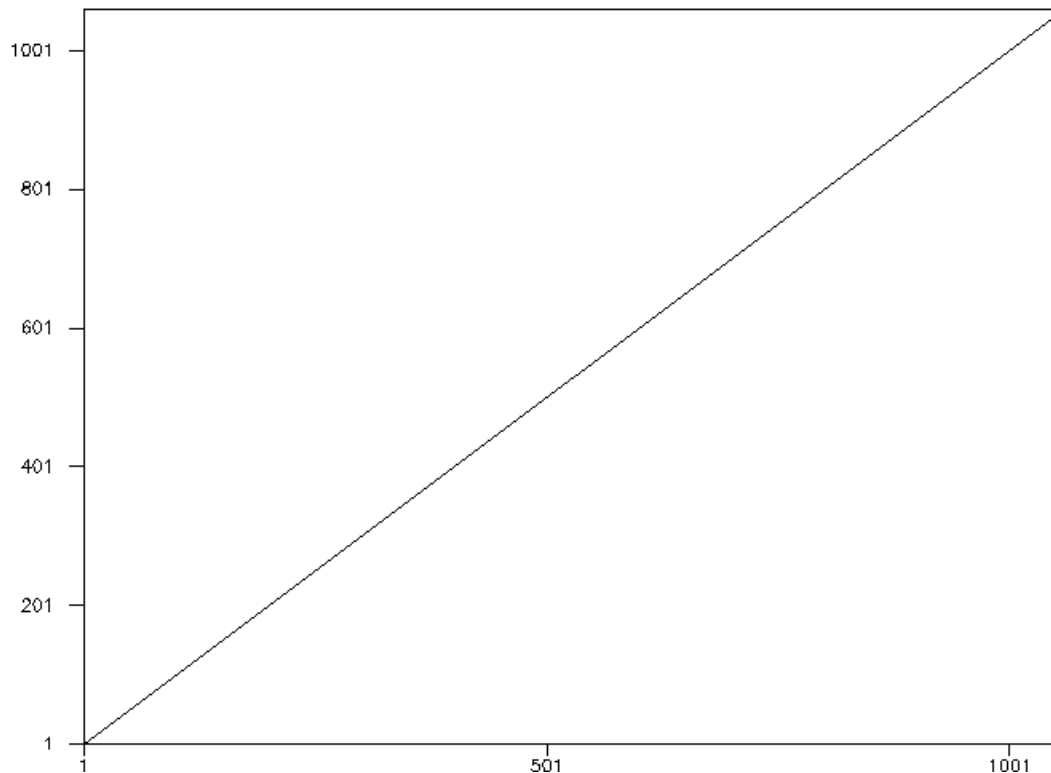
1. 通过BioEdit软件对EG-1基因cDNA序列进行分析:

CDNA长度: 684个核苷酸, GC含量为55.26%, AT含量44.74%。

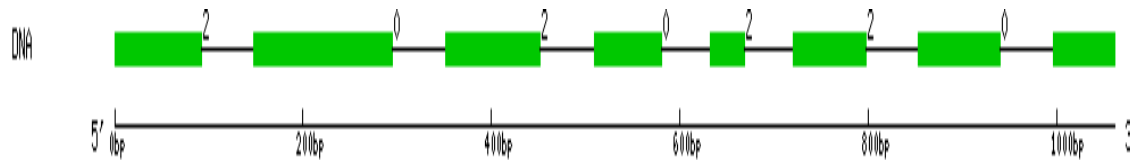


2. 通过abc weblab dotup对EG-1基因DNA序列进行分析,表明没有重复基因。

Dottup: raw::493433 vs raw::493433
Fri 23 Dec 2011 22:22:52



3. 通过GSDS软件，构建EG-1基因的序列结构。

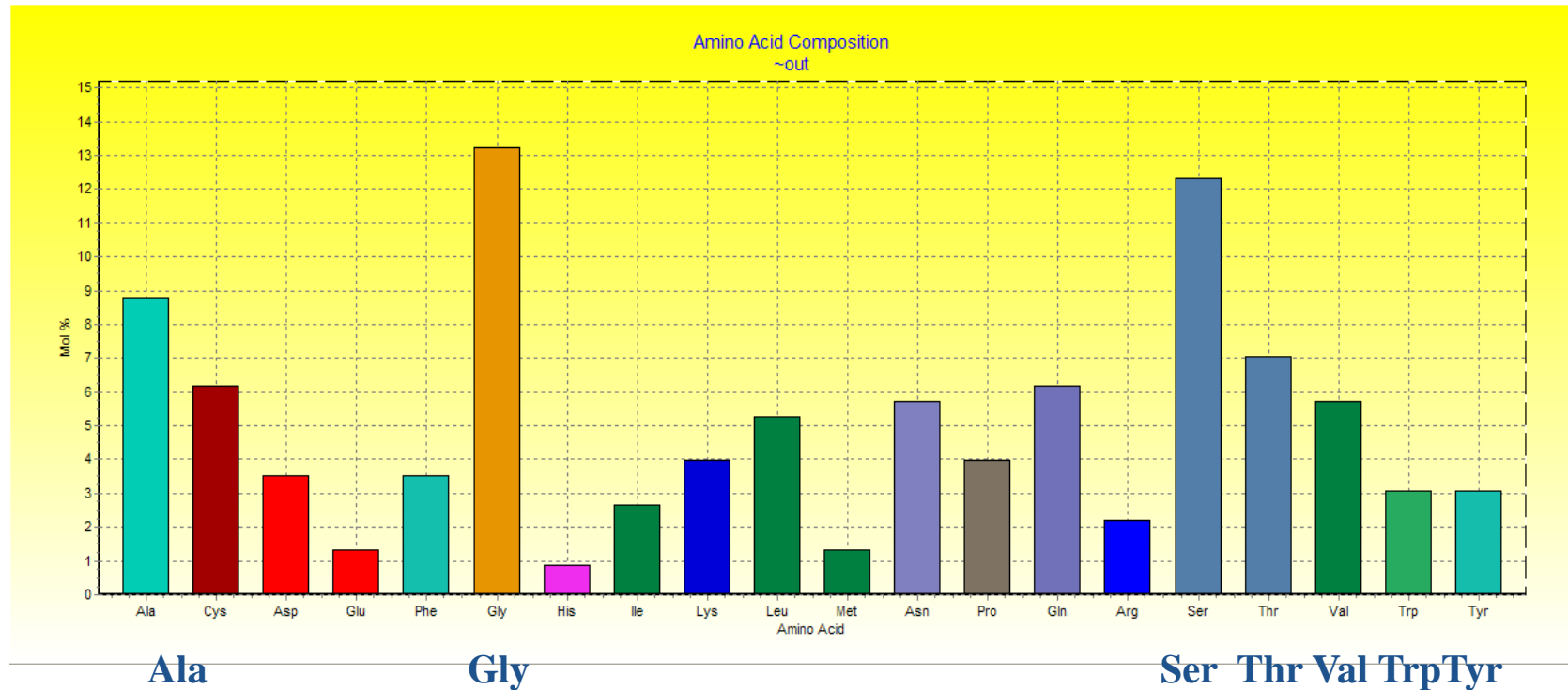


Legend:

■ exon ■ marked region — intron ■ UTR 0 1 2: intron phase

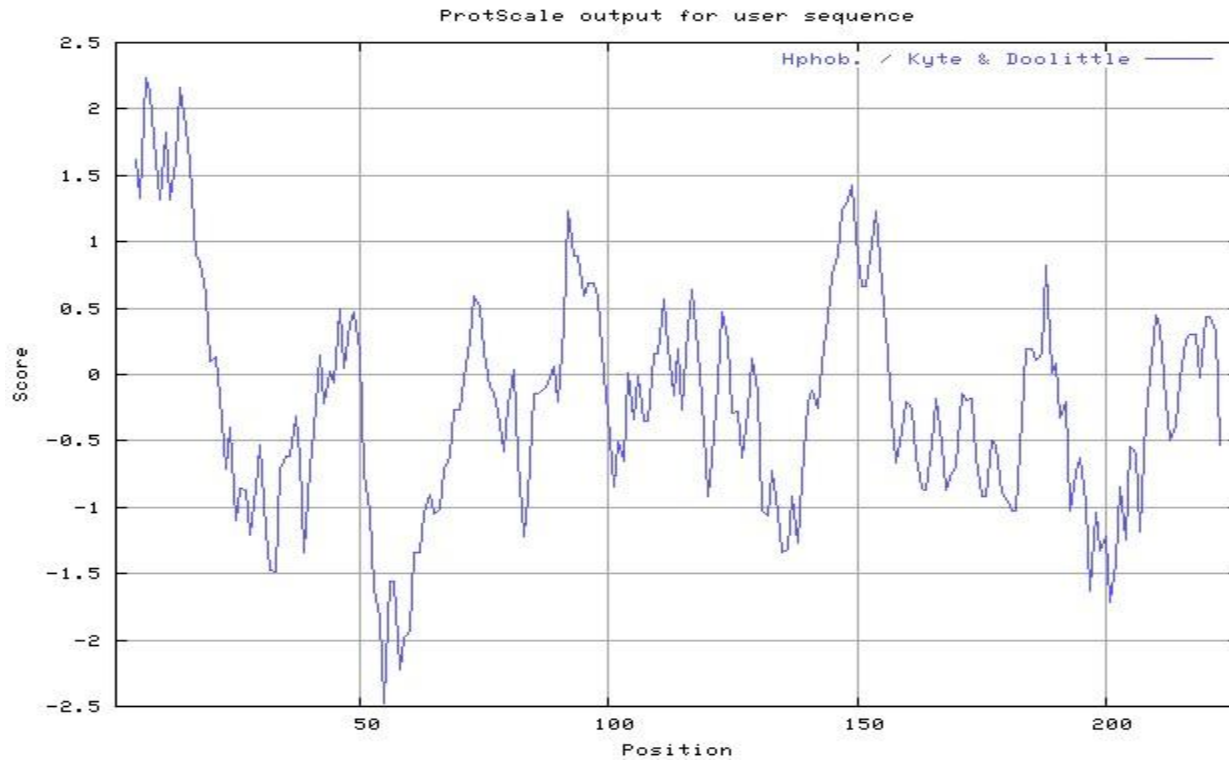
β -1,4-内切纤维素酶基因全长DNA序列为1061bp，包含一个684bp的开放阅读框ORF和7段内含子。

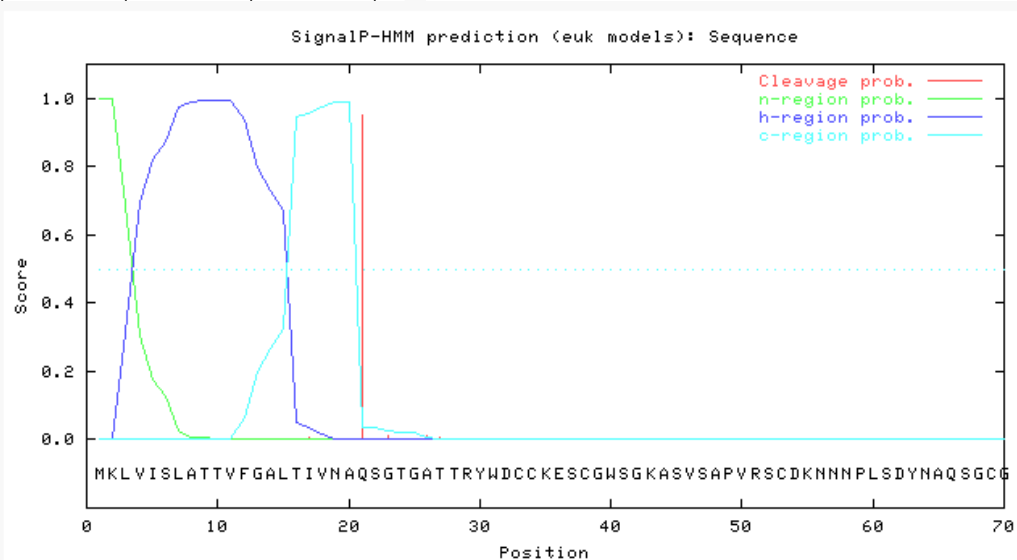
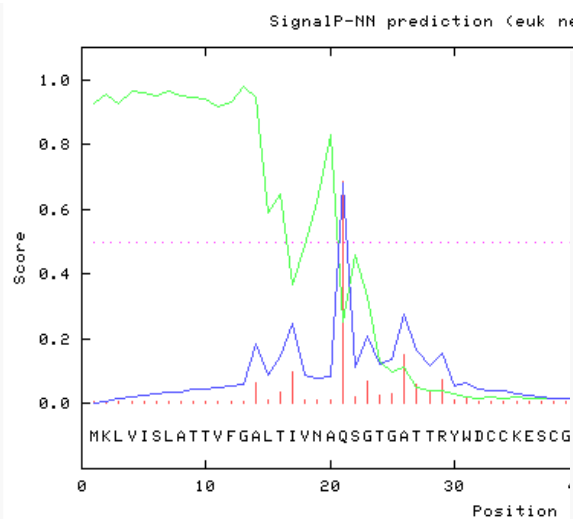
4. 通过BioEdit软件对EG-1基因推测氨基酸序列进行分析： 发现推测的氨基酸序列中，疏水性Gly的含量最高。



5. β -1,4-内切纤维素酶疏水性及信号肽分析

ExPASy Proteomics Server





data

```
>Sequence          length = 70
# Measure  Position  Value  Cutoff  signal pep
max. C      21      0.686  0.32   YES
max. Y      21      0.682  0.33   YES
max. S      13      0.980  0.87   YES
mean S      1-20     0.841  0.48   YES
D           1-20     0.762  0.43   YES
```

Most likely cleavage site between pos. 20 and

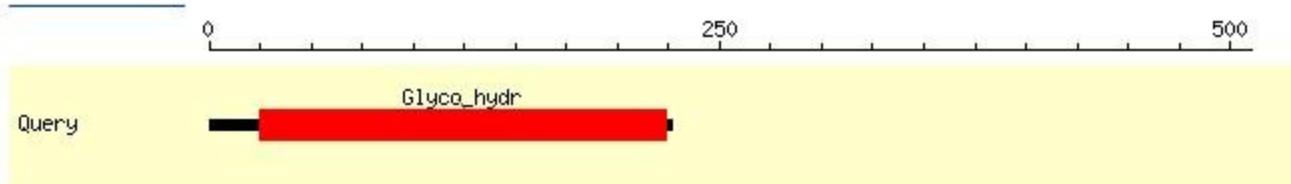
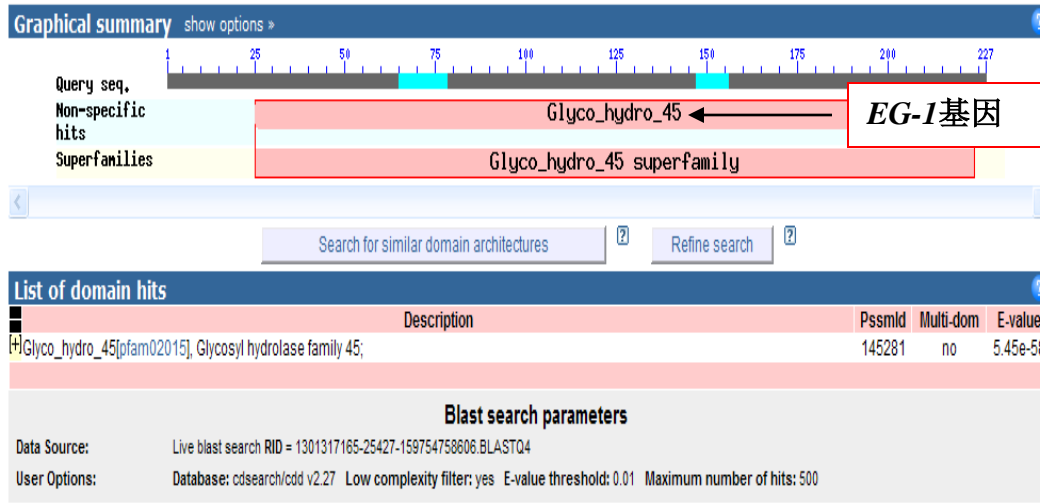
data

```
>Sequence
Prediction: Signal peptide
Signal peptide probability: 0.998
Signal anchor probability: 0.000
```

Max cleavage site probability: 0.953 between pos. 20 and 21

通过Signal.P服务器分析，发现β-1,4-内切纤维素酶基因 N端存在信号肽，推测其一级信号肽切割位点在A²⁰和Q²¹之间，信号肽序列MKLVISLATTVFGALTIVNA。

6. 推导的氨基酸序列提交到NCBI后的分析结果



将该基因的氨基酸序列在NCBI上进行protein blast，结果显示该 β -1,4-内切纤维素酶EG-1基因属于Glycosyl hydrolase family 45，含有一个催化区(Catalytic Domain, CD)

7. 通过ProtParam软件分析, 发现推测的蛋白质分子量为 23661.3, 理论等电点为8.18, 分子式为 $C_{1020}H_{1557}N_{289}O_{328}S_{17}$ 。

Number of amino acids: 227

Molecular weight: 23661.3

Theoretical pI: 8.18

Formula: $C_{1020}H_{1557}N_{289}O_{328}S_{17}$

Total number of atoms: 3211

Amino Acid	Number	Mol%
Ala	20	8.81
Cys	14	6.17
Asp	8	3.52
Glu	3	1.32
Phe	8	3.52
Gly	30	13.22
His	2	0.88
Ile	6	2.64
Lys	9	3.96
Leu	12	5.29
Met	3	1.32
Asn	13	5.73
Pro	9	3.96
Gln	14	6.17
Arg	5	2.20
Ser	28	12.33
Thr	16	7.05
Val	13	5.73
Trp	7	3.08
Tyr	7	3.08

8. 通过PredictProtein软件预测氨基酸的基本性质，结果如下。

Predicted secondary structure composition

sec str type	H	E	L
% in protein	4.85	33.04	62.11

Non-Ordinary Secondary Structure

Sequence length	227
Secondary structure	Helix=4.8%, Strand=33.0%, Loop=6
Window size	70
Structure content cutoff	12%
Minimum consecutive exposed residues	10

NORS	N=NORS region
Secondary structure	H=helix, E=strand, '='=loop

Pattern-ID:	AMIDATION PS00009 PDOC00009
Pattern-DE:	Amidation site
Pattern:	.G[RK][RK]
121	SGKK

Pattern-ID:	GLYCOSYL_HYDROL_F45 PS01140 PDOC00877
Pattern-DE:	Glycosyl hydrolases family 45 active site
Pattern:	[STA] TRY [FYW] D . {5} [CA]
27	TTRYWDCKESC

Pro site

Pattern-ID:	ASH_GLYCOSYLATION PS00001 PDOC00001
Pattern-DE:	N-glycosylation site
Pattern:	N[^P][ST][^P]
86	NDSL

Pattern-ID:	PKC_PHOSPHO_SITE PS00005 PDOC00005
Pattern-DE:	Protein kinase C phosphorylation site
Pattern:	[ST].[RK]
27	TTR
41	SGK
121	SGK
176	SSR
208	SFK
225	SRK

Pattern-ID:	CK2_PHOSPHO_SITE PS00006 PDOC00006
Pattern-DE:	Casein kinase II phosphorylation site
Pattern:	[ST].[2][DE]
132	TGGD

Pattern-ID:	MYRISTYL PS00008 PDOC00008
Pattern-DE:	N-myristoylation site
Pattern:	G[^EDRKHPFYW].[2][STAGCN][^P]
23	GTGATT
68	GCGSGG
100	GGAEST
133	GGDLGN
147	GGGVGI
154	GQGCAS
162	GAPSTG
169	GAQYGG

9. 通过Prosite预测氨基酸序列保守位点，发现该序列属于糖苷水解酶45家族，活性位点序列为**TTRYWDCCKESC**。

Hits by PS01140 GLYCOSYL_HYDROL_F45 *Glycosyl hydrolases family 45 active site* :

USERSEQ1



(227 aa)

27 - 38:

[level tag: (0)] TTRYWDCCKESC

10. 通过Swiss Model件进行同源建模，结果如下。



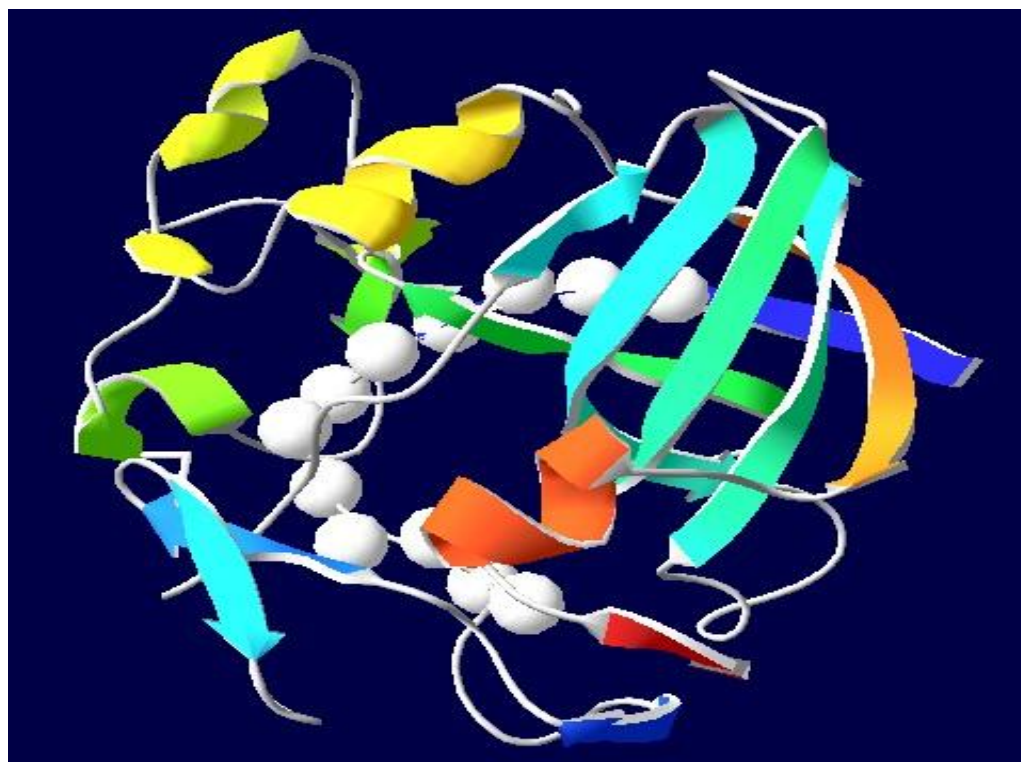
Model Details: Batch.1

model pic

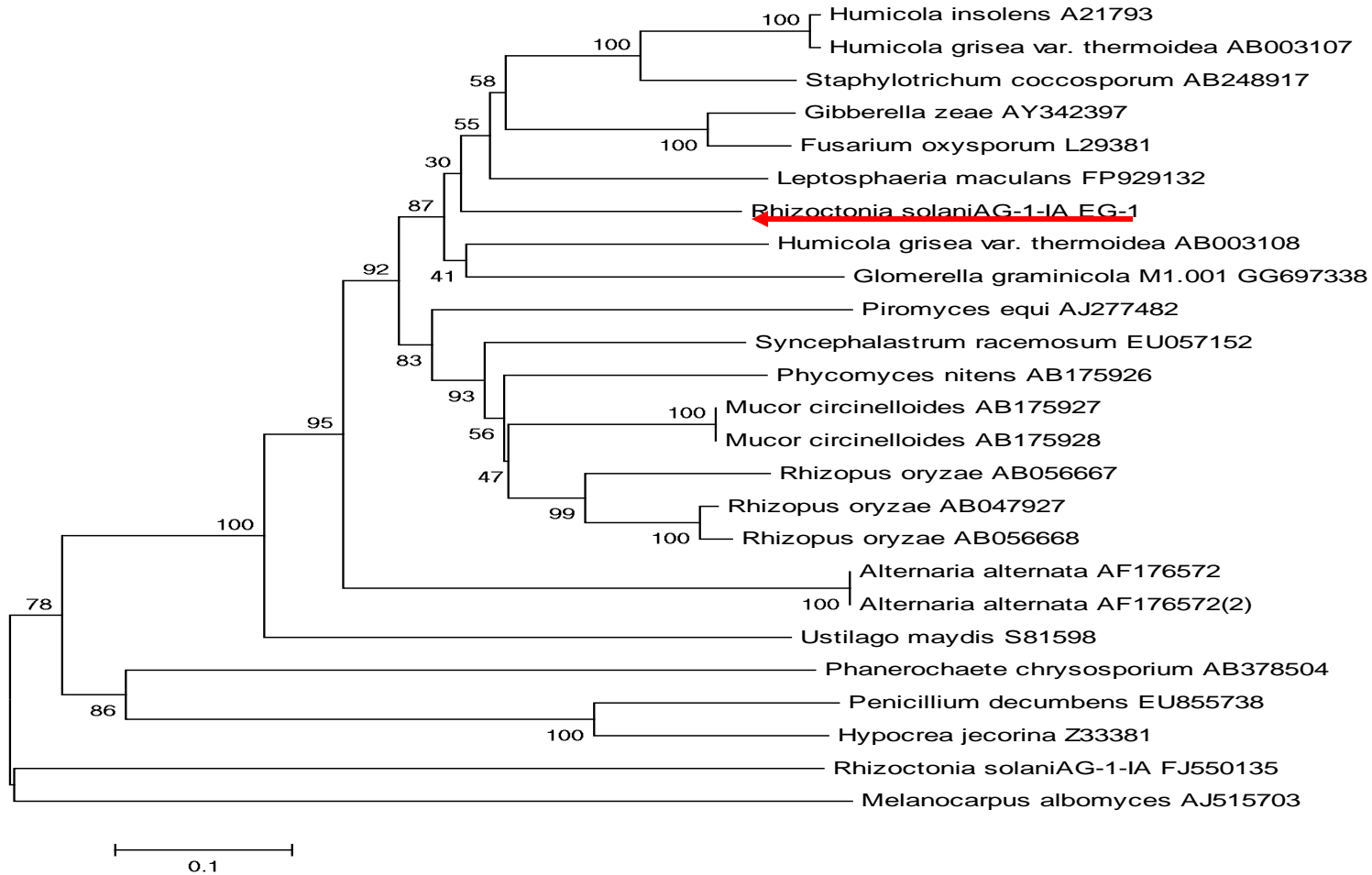


Target:	
modelled residue range:	23 to 226
based on template	1I8fA (1.80 A)
Sequence Identity [%]:	61.275
Evalue:	2.23423e-66

11. 通过Swiss-pdbviewer进行分析，并标记出糖苷水解酶45家族活性位点，如下图：



12. 根据同源序列构建系统发育树。





三、结论

- β -1,4-内切纤维素酶EG-1基因是根据真菌第45家族糖苷水解酶氨基酸同源保守序列，设计兼并引物，通过RT-PCR、3'-RACE PCR和5'-TAIL PCR分离得到的，cDNA和DNA序列在GenBank中的登录号分别为GU372731、GU372728（未释放）。
- 本文对EG-1基因序列和推测的编码蛋白进行了结构和功能的预测分析,分析结果表明GC含量为55.26%，富含丝氨酸和甘氨酸，具有信号肽序列，引导蛋白质跨膜运输。

- 通过prosite预测氨基酸序列保守位点，发现该序列属于糖苷水解酶45家族，活性位点序列为**TTRYWDCC****KESC**，同源建模分析表明保守位点位于结构内部。
- 通过上述分析，能促进人们对 β -1,4-内切纤维素酶基因的进一步了解，在后续的实验中，可以通过实验进一步对**保守位点**进行突变分析验证其酶活性中心及生物学功能，以期为防治玉米纹枯病提供新的策略。

b&t
welcome to our home
© B&T Studio. All Rights Reserved.



谢谢!
