

李青青,王旭歌,钟甲丽,等. 纤维素酶系基因的克隆与序列分析[J]. 江苏农业科学,2016,44(3):40-43.
doi:10.15889/j.issn.1002-1302.2016.03.010

纤维素酶系基因的克隆与序列分析

李青青¹, 王旭歌¹, 钟甲丽¹, 李旺^{1,2}

(1. 河南科技大学动物科技学院,河南洛阳 471003; 2. 河南宏翔生物科技有限公司,河南汝州 467500)

摘要:根据 GenBank 已报道的枯草芽孢杆菌的基因序列,设计特异性引物,经 PCR 分别从 *B. subtilis* L、*B. subtilis* K 基因组中扩增得到 3 个纤维素酶基因序列,基因片段分别长约 700、1 500、1 700 bp。连接到 T 载体上进行测序,获得该 3 个纤维素酶基因片段的 DNA 序列,在线比对分析表明 3 个基因的部分碱基序列与已报道的纤维素酶基因的序列同源性达 99%。这 3 个基因的全序列在 GenBank 中未见报道,且通过软件对比,三者没有任何相似性。根据 3 个纤维素酶基因的来源分别命名为 *Lcen*、*Ken*、*Kkg*。借助软件分析确定 *Lcen* 基因全长 699 bp,基因序列为 1 个完整的阅读框,连续编码 232 个氨基酸。*Ken* 基因全长 1 500 bp,连续编码 499 个氨基酸。*Kkg* 基因全长 1 686 bp,基因序列为 1 个完整的阅读框,连续编码 561 个氨基酸。3 个基因均属于纤维素酶家族大类。

关键词:纤维素酶;基因;克隆;序列分析

中图分类号: Q785 **文献标志码:** A **文章编号:** 1002-1302(2016)03-0040-04

纤维素是地球上最丰富的可再生自然资源^[1]。全球每年纤维素产量达 2 000 亿 t^[2],仅我国秸秆产量就达 5 亿~7 亿 t,大部分被焚烧、丢弃,不仅浪费资源,而且会对环境造成污染^[3]。纤维素是一种无色、无味的白色丝状物,难溶于一般的有机溶剂及水,是植物细胞壁的主要组成部分。植物纤维素结构基本由结晶区域、无定型区域组成,纤维素结晶区域比无定型区域难降解^[4]。酶解法是目前降解纤维素最有效的方法^[5]。纤维素酶是能够分解纤维素、最终将其降解成葡萄糖的一类酶的总称。根据纤维素酶催化反应功能的不同可将其分为:(1)内切葡聚糖酶,这类酶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;(2)外切葡聚糖酶,这类酶作用于纤维素线状分子末端,水解 β -1,4 糖苷键;(3) β -葡萄糖苷酶,这类酶将纤维二糖水解成葡萄糖分子^[6]。纤维素酶分子多数由球状的催化结构域(catalytic domains, CD)、连接桥(linker)、纤维素结合结构域(cellulose-binding domains, CBD)3 个部分构成^[7-9]。纤维素酶分布广泛,目前饲料用纤维素酶主要从微生物中获得。随着分子生物学、基因工程技术的发展,对纤维素酶分子层面研究也随之展开,纤维素酶基因的克隆与表达成了研究焦点。细菌、真菌的纤维素酶系不断被人们发现、分离,大量纤维素酶基因得到克隆、表达,丰富了纤维素酶的研究材料^[10]。结构功能完整的纤维素酶基因克隆到高效表达载体上再进行异源表达,能使纤维素酶的产量成倍提高^[11]。本试验对纤维素酶系内的 3 个纤维素酶基因进行克隆和序列分析,旨在为后续高效联合

表达纤维素酶系基因进行研究准备。

1 材料与与方法

1.1 材料

1.1.1 菌株与质粒 *Bacillus subtilis* K、*Bacillus subtilis* L 均由实验室筛选并保存,大肠杆菌 DH5 α 购于北京天根生化科技有限公司, pGEM-T Easy Vector System 购于 Promega 公司。

1.1.2 主要试剂 DL 2000 Marker、溴酚蓝、琼脂糖购自北京天根生化科技有限公司; *Taq* 酶、DNTP、各种限制性内切酶购自 Promega 公司;溴化乙锭(EB)、抗生素(Amp)购于北京索莱宝科技有限公司;RNA 酶、蛋白酶 K 购于北京中科瑞泰生物科技有限公司;琼脂粉、蛋白胨购于 Oxid 公司。基因组提取试剂盒、质粒提取试剂盒、DNA 胶纯化回收试剂盒均购自于中科瑞泰(北京)生物科技有限公司。

1.1.3 引物 试验中的 3 对引物序列:

ENF: 5' - ATGAAACGGTCAATCTCTATT - 3';

ENR: 5' - CTAATTTGGTTCTGTTCCTCCA - 3';

CENF: 5' - ATGAAAAAGATCATGAGTGCAT - 3';

CENR: 5' - TTATTCAGGAACTGAACATGG - 3';

KGF: 5' - ATGAGTGAATGGTGGAAAGAAG - 3';

KGR: 5' - TCATATACTAATGCCCATCACAG - 3'。

1.2 方法

LB 培养基的配制:蛋白胨 10 g,酵母提取物 5 g,氯化钠 10 g,蒸馏水 1 L,固体 LB 培养基加入 20 g 琼脂粉,120 $^{\circ}$ C 高压灭菌 30 min。细菌基因组 DNA 的提取、质粒 DNA 的提取、DNA 的凝胶回收参照试剂盒提供的说明进行。纤维素酶基因的 PCR 扩增、纤维素酶基因片段与 T 载体的连接、重组质粒导入感受态细胞、含纤维素酶基因的重组质粒的酶切鉴定参照《分子克隆试验指南》规定的方法进行。纤维素酶系基因的序列测定由北京擎科新业公司完成。用 VectorNT 软件和 NCBI Blast 2.0 在线软件对该纤维素酶基因的序列进行分析。

收稿日期:2015-03-06

基金项目:国家自然科学基金(编号:31101744);河南省重大科技专项(编号:131100110300)。

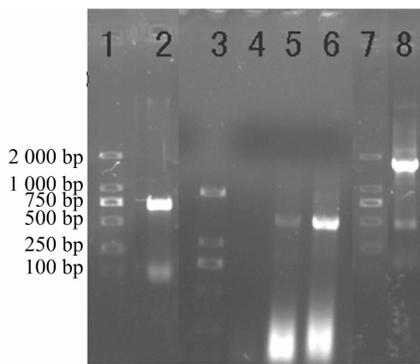
作者简介:李青青(1990—),女,从事饲料酶制剂研究。E-mail: liwang@haust.edu.cn。

通信作者:李旺,博士,副教授,主要从事分子营养与饲料生物技术研究。E-mail:liwang@haust.edu.cn。

2 结果与分析

2.1 纤维素酶基因的克隆

根据 GenBank 公布的纤维素酶系的内切葡聚糖酶基因、外切葡聚糖酶基因、葡萄糖苷酶基因序列设计特异性引物: ENF(R)、CENF(R)、KGF(R),以提取得到的 *B. subtilis* K、*B. subtilis* L 基因组 DNA 为模板进行 PCR 扩增,扩增结果如图 1 所示。扩增得到的 DNA 片段大小分别为 1 500、700、1 700 bp 左右。根据来源不同分别将 3 个基因命名为 *Ken*、*Lcen*、*Kkg*。



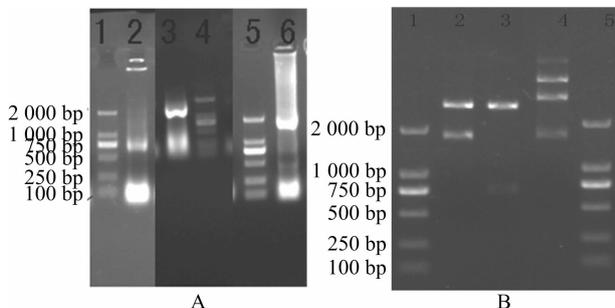
1、3、7号均为DL 2000 Marker(梯度为100、250、500、750、1 000、2 000 bp),2号为*Lcen*基因扩增的产物,5、6号分别为*Ken*基因扩增的产物,8号为*Kkg*基因扩增的产物

图1 PCR产物电泳图

2.2 纤维素酶基因的鉴定

分别将片段回收,连接于 T 载体上转化到大肠杆菌培养,挑取单菌落,进行重组质粒 PCR 鉴定(图 2 - A),凝胶上有明

显的 DNA 条带,与 PCR 扩增的结果一样。进行酶切鉴定(图 2 - B),凝胶上分别出现 2 条明显的 DNA 条带,一条大小约为 3 kb 与 T-easy 载体的片段大小一致,另一条带大小分别为 1 500、700、1 700 bp,与 PCR 扩增的结果一致。由此确定 PCR 扩增产物顺利连接到 T-easy 载体上,构建了 3 个基因与 T-easy 连接的重组质粒,分别命名为 *LcenT*、*KenT*、*KkgT*。



A为重组质粒PCR产物电泳图,其中M为DL 2000 Marker,泳道2为*LcenT*质粒的PCR鉴定,泳道3为*KenT*质粒的PCR鉴定,泳道6为*KkgT*质粒的PCR鉴定。B为重组质粒的酶切鉴定电泳,泳道2为*KkgT*酶切鉴定,泳道3为*LcenT*酶切鉴定,泳道4为*KenT*酶切鉴定

图2 重组质粒 PCR 鉴定和酶切鉴定

2.3 纤维素酶基因的序列分析

将 3 个纤维素酶基因重组质粒 *LcenT*、*KenT*、*KkgT* 分别进行测序,并通过 VectorNT 软件和 GenBank 数据库对 3 个基因的序列进行分析。

2.3.1 *Lcen* 基因的序列分析 对 *LcenT* 重组质粒进行测序,得到的基因序列如图 3 所示。

```

Met Lys Lys Ile Met Ser Ala Phe Val Gly Met Val Leu Leu Thr Ile Phe Cys Phe Ser
ATG AAA AAG ATC ATG AGT GCA TTT GTT GGT ATG GTT TTG TTG ACG ATC TTC TGT TTT TCC
Pro Gln Ala Ser Ala Ala Tyr Asp Asn Leu His Glu Gly Tyr Ala Thr Tyr Thr Gly Ser
CCG CAA GCT TCG GCA GCA TAT GAC AAC CTG CAT GAA GGT TAT GCA ACG TAT ACA GGG TCA
Gly Tyr Ser Gly Gly Ala Phe Leu Leu Asp Pro Ile Pro Ser Asp Met Glu Ile Thr Ala
GGT TAT TCA GGA GGA GCT TTC CTG CTG GAT CCC ATT CCT TCC GAT ATG GAG ATT ACT GCA
Ile Asn Pro Ala Asp Leu Asn Tyr Gly Gly Val Lys Ala Ala Leu Ala Gly Ser Tyr Leu
ATA AAT CCG GCG GAT CTC AAT TAC GGA GGA GTA AAA GCA GCA CTT GCC GGC TCT TAT TTG
Glu Val Glu Gly Pro Lys Gly Lys Thr Thr Val Tyr Val Thr Asp Leu Tyr Pro Glu Gly
GAA GTT GAA GGG CCA AAA GGG AAA ACA ACC GTA TAT GTT ACT GAT CTT TAC CCC GAA GGC
Ala Arg Gly Ala Leu Asp Leu Ser Pro Asn Ala Phe Arg Lys Ile Gly Asn Met Lys Asp
GCT CGG GGA GCT CTT GAT CTG TCA CCT AAT GCC TTC CGT AAA ATC GGC AAT ATG AAA GAC
Gly Lys Ile Asn Ile Lys Trp Arg Val Val Lys Ala Pro Ile Thr Gly Asn Phe Thr Tyr
GGA AAA ATC AAT ATT AAA TGG CGT GTT GTC AAA GCT CCA ATC ACC GGC AAT TTC ACG TAC
Arg Ile Lys Glu Gly Ser Ser Arg Trp Trp Ala Ala Ile Gln Val Arg Asn His Lys Tyr
CGG ATC AAA GAA GGC AGC AGC AGG TGG TGG GCA GCA ATC CAA GTC AGA AAT CAC AAG TAT
Pro Val Met Lys Met Glu Tyr Glu Lys Asp Gly Lys Trp Ile Asn Met Glu Lys Met Asp
CCT GTT ATG AAA ATG GAA TAT GAA AAG GAT GGT AAG TGG ATC AAC ATG GAG AAA ATG GAC
Tyr Asn His Phe Val Ser Thr Asn Leu Gly Thr Gly Ser Leu Lys Val Arg Met Thr Asp
TAT AAC CAT TTT GTG AGT ACG AAT TTA GGG ACA GGT TCT CTC AAG GTC AGA ATG ACT GAC
Ile Arg Gly Lys Val Val Lys Asp Thr Ile Pro Lys Leu Pro Glu Ser Gly Thr Ser Lys
ATC CGC GGA AAA GTT GTG AAA GAC ACC ATT CCA AAG CTG CCT GAA AGC GGA ACG TCC AAA
Ala Tyr Thr Val Pro Gly His Val Gln Phe Pro Glu ***
CGC TAT ACA GTA CCG GGC CAT GTT CAG TTT CCT GAA TAA

```

图3 *Lcen* 基因序列

用 VectorNT 软件对 *Lcen* 基因序列进行分析发现,该基因全长 699 个碱基,组成 1 个完整的开放阅读框(open read

frame, ORF), 连续编码 232 个氨基酸,起始密码 ATG,终止密码 TAA。

利用 NCBI 网站上的 Blast 功能对该基因序列进行比对发现,基因 *Lcen* 与已报道的枯草芽孢杆菌纤维素酶基因 GU327817.1、CP003695.1、CP003329.1、AP012496.1、AP012495.1 十分相似,相似度为 99%。对其编码的氨基酸进行比对分析发现,该基因编码的氨基酸序列与已报道的枯草芽孢杆菌胞外的葡萄糖苷酶基因 NP389744.1、YP007427045.1、YP004203797.1、YP006231826.1、ZP06873100.1 的氨基酸序列十分相似,相似度达到 95%。故可初步确定该基因来源于枯草芽孢杆菌,属于

```
Met Lys Arg Ser Ile Ser Ile Phe Ile Thr Cys Leu Leu Ile Thr Leu Leu Thr Met Gly
ATG AAA CCG TCA ATC TCT ATT TTT ATT ACG TGC TTA CTG ATT ACG TTA TTG ACA ATG GGC
Gly Met Pro Ala Ser Pro Ala Ser Ala Ala Gly Thr Lys Thr Pro Val Ala Lys Asn Gly
GGC ATG CCG GCT TCG CCG GCA TCA GCA GCA GGG ACA AAA ACG CCA GTA GCC AAG AAT GGC
Gln Leu Ser Ile Lys Gly Thr Gln Leu Val Asn Arg Asp Gly Lys Ala Val Gln Leu Lys
CAG CTT AGC ATA AAA GGT ACA CAG CTC GTT AAC CGA GAC GGT AAA GCG GTA CAG CTG AAG
.....
.....
Gly Thr Leu Ala Pro Gly Ala Ser Ala Gly Asn Ile Gln Leu Arg Leu His Asn Asp Asp
GGA ACG TTG GCA CCG GGA GCA AGC GCA GGG AAT ATT CAG CTC CGT CTT CAC AAT GAT GAC
Trp Ser Asn Tyr Ala Gln Ser Gly Asp Tyr Ser Phe Phe Lys Ser Asn Thr Phe Lys Thr
TGG AGC AAT TAT GCA CAA AGC GGC GAT TAT TCC TTT TTC AAA TCA AAT ACG TTT AAA ACA
Thr Lys Lys Ile Thr Leu Tyr Asp Gln Gly Lys Leu Ile Trp Gly Thr Glu Pro Asn ***
ACG AAA AAA ATC ACA TTA TAT GAT CAA GGA AAA CTG ATT TGG GGA ACA GAA CCA AAT TAG
```

图4 *Ken* 基因头尾序列

用 VectorNT 软件对 *Ken* 基因序列分析,*KenT* 的基因全长 1 500 个碱基,组成 1 个完整的开放阅读框,连续编码 499 个氨基酸。起始密码 ATG,终止密码 TAG。利用 NCBI 网站上的 Blast 功能对该基因序列进行比对发现,基因 *Ken* 的序列与已报道的枯草芽孢杆菌纤维素酶基因 FJ800366.1、EF070194.1、KC477685.1、CP002468.1、HM543165.1 基因序列十分相似,相似度为 99%。对其编码的氨基酸进行比对,该基因编码的氨基酸序列与已报道的枯草芽孢杆菌的内切葡萄糖苷酶基因 NP389695.2、AFX88666.1、YP007209477.1、ACK38261.1 的氨基酸序列十分相似,相似度达到 99%。故可确定该基因来源于枯草芽孢杆菌,属于纤维素酶基因。

对该基因编码的氨基酸结构进行分析预测发现,该基因编码的氨基酸属于水解酶超级家族,整体由 2 个主要部分组成(图 5),一部分为纤维素酶区域(cellulase,图 5 - A),一部分为纤维素酶的绑定区域(CBM - 3,图 5 - B),其中的纤维素酶区域结构符合糖基水解酶家族 5 的结构特征。

2.3.3 *Kkg* 基因的序列分析 对 *KkgT* 重组质粒进行测序,得到的基因序列如图 6 所示。

用 VectorNT 软件对 *Kkg* 基因序列分析,*Kkg* 基因全长 1 686 个碱基,组成 1 个完整的开放阅读框,连续编码 561 个氨基酸。起始密码 ATG,终止密码 TGA。

利用 NCBI 网站上的 Blast 功能对该基因序列进行比对发现,*Kkg* 的基因序列与已报道的枯草芽孢杆菌纤维素酶基因 cp002468.1、CP003695.1、CP003329.1、AP012496.1、AP012495.1 序列十分相似,相似度为 99%。对其编码的氨基酸进行比对分析发现,该基因编码的氨基酸序列与枯草芽孢杆菌 1,4 - 6 - α - 葡糖苷酶基因 YP004205293.1、YP007208039.1、YP007428400.1、YP007664110.1、ZP12669813.1

纤维素酶基因。

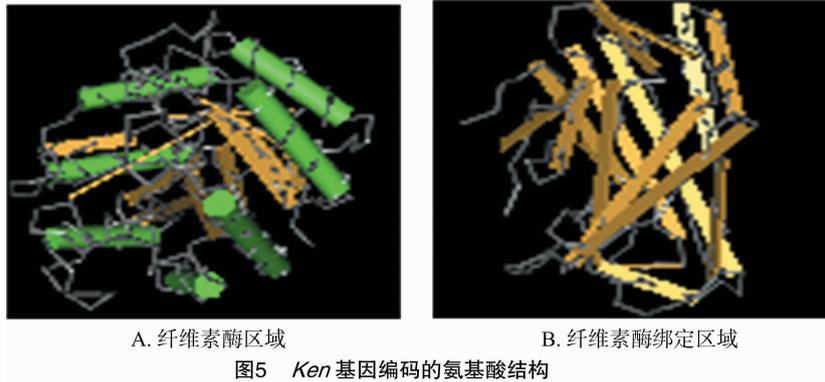
对基因编码的氨基酸序列进行分析和预测得知,该基因编码的氨基酸结构属于罕见的脂蛋白(RlpA)超家族,基因序列部分与罕见的脂蛋白、假定的 EG45-like 域包含蛋白 1、内切葡聚糖酶 c 末端域/亚单位和相关蛋白质相似性高。

2.3.2 *Ken* 基因的序列分析 对 *KenT* 重组质粒进行测序,得到的基因序列如图 4 所示。

的氨基酸序列十分相似,相似度达到 99%。故可初步确定该基因来源于枯草芽孢杆菌,属于纤维素酶基因的类别。对该基因编码的氨基酸结构进行分析和预测发现,该基因编码的氨基酸属于 α - 淀粉酶的超家族,基因编码的氨基酸序列上包含纤维素酶活性部位、Ca 绑定结构域、催化部位(图 7)。

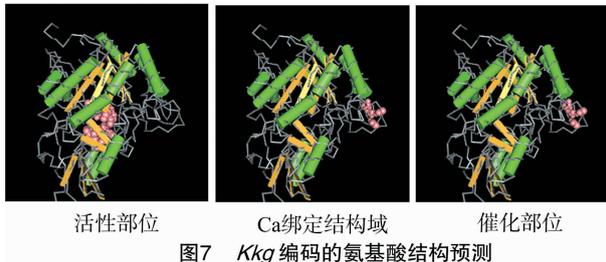
3 结论与讨论

纤维素酶的应用非常广泛,然而天然纤维素酶由于酶活较低以及成本高等因素的限制,对于大规模的工业化应用有一定困难。近年来,随着分子生物学和基因工程技术的发展,对纤维素酶分子层面的研究也随之展开,纤维素酶基因的克隆与表达成了研究焦点。目前常用的基因克隆方法有人工合成法、PCR 扩增法以及构建基因文库等^[12]。本试验通过 GenBank 检索枯草芽孢杆菌纤维素酶基因,根据基因的编码序列设计 3 对引物,从 *B. subtilis* K、*B. subtilis* L 基因组 DNA 扩增出了 3 个基因片段,通过测序得到基因序列。枯草芽孢杆菌是目前细菌纤维素酶基因的主要来源,已报道的大部分纤维素酶基因均从枯草芽孢杆菌中获得^[13-14]。已报道的枯草芽孢杆菌纤维素酶基因大多数属于内切葡聚糖酶(endo β - 1,4 glucanase)基因^[14]。在纤维素酶基因克隆上,不论是从何种微生物上进行克隆,大部分报道均是克隆获得 1 个基因。本试验同时从 *Bucillus subtilis* K 基因组 DNA 中克隆得到 2 个纤维素酶基因 *Ken*、*Kkg*,其中 *Ken* 大小为 1 500 bp,与已报道的大多数枯草芽孢杆菌内切葡聚糖酶基因一致,属于纤维素水解酶家族;*Kkg* 基因全长 1 686 bp,与已报道的枯草芽孢杆菌 1,4 - 6 - α - 葡糖苷酶基因一致,在结构分析上,该基因属于 α - 淀粉酶的超家族。



Met Ser Glu Trp Trp Lys Glu Ala Val Val Tyr Gln Ile Tyr Pro Arg Ser Phe Tyr Asp
ATG AGT GAA TGG TGG AAA GAA GCT GTC GTT TAT CAA ATT TAC CCG CGC AGT TTT TAT GAT
 Ala Asn Gly Asp Gly Phe Gly Asp Leu Gln Gly Val Ile Gln Lys Leu Asp Tyr Ile Lys
GCC AAT GGA GAT GGA TTC GGG GAT TTG CAA GGT GTG ATT CAA AAG CTG GAT TAC ATC AAA
 Asn Leu Gly Ala Asp Val Ile Trp Leu Ser Pro Val Phe Asp Ser Pro Gln Asp Asp Asn
AAT CTC GGG GCG GAT GTC ATC TGG CTC TCC CCG GTA TTT GAT TCC CCG CAG GAT GAC AAC

 Glu Tyr Arg Gly Glu Lys Leu Leu Val Val Val Asn Leu Ser Glu Glu Lys Ala Leu Phe
GAA TAT CGA GGG GAA AAG CTT CTT GTC GTT GTG AAT TTA TCG GAA GAA AAG GCT CTG TTC
 Glu Ala Pro Pro Glu Leu Ile His Glu Arg Trp Lys Val Leu Ile Ser Asn Tyr Pro Gln
GAA GCG CCT CCA GAA CTG ATT CAT GAG CGT TGG AAA GTG CTG ATT TCA AAC TAT CCG CAG
 Glu Arg Ala Asp Leu Lys Ser Ile Ser Leu Lys Pro Tyr Glu Ala Val Met Gly Ile Ser
GAG CGG GCT GAC TTA AAG AGT ATT AGC CTC AAA CCT TAT GAA GCT GTG ATG GGC ATT AGT
 Ile ***
 ATA TGA

图6 *Kkg* 基因头尾序列

参考文献:

- [1] 杨家华, 郭志宏, 杜海祖. 纤维素酶的研究与应用[J]. 中兽医医药杂志, 2007, 26(5): 30-32.
- [2] Lynd L R, Weimer P J, van Zyl W H, et al. Microbial cellulose utilization: fundamentals and biotechnology [J]. Microbiology and Molecular Biology Reviews, 2002, 66(3): 506-577.
- [3] 张平平, 刘宪华. 纤维素生物降解的研究现状与进展[J]. 天津农学院学报, 2004, 11(3): 48-54.
- [4] Juy M, Amit G, Alzati M, et al. Crystal structure of a thermostable bacterial cellulose degrading enzyme[J]. Nature, 1992, 357(6373): 89-91.
- [5] 刘萌, 战利, 马红霞, 等. 纤维素酶及纤维素酶基因工程学研究进展[J]. 安徽农业科学, 2011, 39(16): 9515-9517.
- [6] 张杰, 张晓东, 孟祥梅, 等. 纤维素酶研究进展[J]. 可再生能源, 2007, 25(5): 57-60.
- [7] Klyosov A A. Trends in biochemistry and enzymology of cellulose gradation[J]. Biochemistry, 1990, 29(47): 10577-10585.
- [8] Tilbeurg H, Tomme P, Claeysens M. Limited proteolysis of the celD lobiohydrolase I from Treesei [J]. FEBS Letters, 1986, 204(2): 223-227.
- [9] 李旺, 杨明明, 陈玉林. 产纤维素酶枯草杆菌 *B. subtilis* DR 的鉴定与酶特性研究[J]. 饲料研究, 2012(1): 33-35.
- [10] 李雪峰, 侯红萍. 选育高产纤维素酶菌种的研究进展[J]. 酿酒科技, 2010(5): 92-94.
- [11] Li W, Huan X, Zhou Y, et al. Simultaneous cloning and expression of two cellulase genes from *Bacillus subtilis* newly isolated from Golden Takin (*Budorcas taxicolor bedfordi*) [J]. Biochemical and Biophysical Research Communications, 2009, 383(4): 397-400.
- [12] Shelomi M, Watanabe H, Arakawa G. Endogenous cellulase enzymes in the stick insect (Phasmatodea) gut [J]. Journal of Insect Physiology, 2014, 60: 25-30.
- [13] 卢敏, 王帅豪, 狄元冉, 等. 纤维素酶基因克隆与表达[J]. 动物营养学报, 2012, 24(6): 1013-1018.
- [14] Li W, Zhang W W, Yang M M, et al. Cloning of the thermostable cellulase gene from newly isolated *Bacillus subtilis* and its expression in *Escherichia coli* [J]. Molecular Biotechnology, 2008, 40(2): 195-201.