

黑曲霉木质纤维素降解酶基因的可变剪接分析及验证

徐轶凡, AJMAL Maria, 董飞宇, 张震, 林晖*, 陈红歌*

河南农业大学生命科学学院 农业农村部农业微生物酶工程重点实验室, 河南 郑州 450002

摘要: 黑曲霉 *Aspergillus niger* 因能够产生大量的木质纤维素降解酶而在木质纤维素资源利用中发挥重要作用。目前, 有关黑曲霉基因组中与木质纤维素降解相关的基因是否存在可变剪接的情况尚不清楚。本研究以黑曲霉 CBS513.88 菌株为研究对象, 采用 rMATS 和 ABLas 两种方法对黑曲霉在葡萄糖为唯一碳源(G 组)和小麦秸秆为唯一碳源(WS 组)下的 56 个木质纤维素降解酶基因的可变剪接事件进行分析, 并通过 RT-PCR 扩增和内含子特异性扩增对 3 个典型基因的可变剪接体进行了验证。结果表明, ABLas 可变剪接分析算法相较于 rMATS 分析算法更为准确, ABLas 分析算法显示 G 组和 WS 组共有 21 个木质纤维素降解酶基因出现了可变剪接, 可变剪接类型以内含子保留(IR)为主, 占有所有可变剪接事件的 82.85%。另外, G 组和 WS 组发生可变剪接的木质纤维素降解酶基因也有所不同: G 组发生可变剪接的基因为 13 个, WS 组发生可变剪接的基因为 14 个, 两组都发生可变剪接的基因为 6 个, 这表明黑曲霉木质纤维素降解酶基因的可变剪接在不同生长条件下存在差异, 另一方面, 黑曲霉中众多可变剪接体的存在也为开发新型的木质纤维素降解酶资源提供基础。

关键词: 黑曲霉; 木质纤维素降解酶; RNA-seq; 可变剪接; 内含子保留

[引用本文] 徐轶凡, AJMAL Maria, 董飞宇, 张震, 林晖, 陈红歌, 2022. 黑曲霉木质纤维素降解酶基因的可变剪接分析及验证. 菌物学报, 41(10): 1619-1633

Xu YF, Ajmal Maria, Dong FY, Zhang Z, Lin H, Chen HG, 2022. Analysis and verification of the alternative splicing events of the lignocellulose degrading enzyme genes from *Aspergillus niger*. Mycosystema, 41(10): 1619-1633

基金项目: 国家自然科学基金(32171476)

Supported by the National Natural Science Foundation of China (32171476).

*Corresponding authors. E-mail: huilin@henau.edu.cn, honggeyz@henau.edu.cn

ORCID: CHEN Hongge (0000-0002-7519-7341)

Received: 2022-01-23, accepted: 2022-02-24

Analysis and verification of the alternative splicing events of the lignocellulose degrading enzyme genes from *Aspergillus niger*

XU Yifan, AJMAL Maria, DONG Feiyu, ZHANG Zhen, LIN Hui*, CHEN Hongge*

Key Laboratory of Enzyme Engineering of Agricultural Microbiology, Ministry of Agriculture and Rural Affairs, College of Life Sciences, Henan Agricultural University, Zhengzhou 450002, Henan, China

Abstract: *Aspergillus niger* plays an important role in utilizing lignocellulosic biomass because of its ability to produce large amounts of lignocellulose degrading enzymes. Until now it is not clear whether there are alternative splicing events on genes related to lignocellulose degradation in *A. niger*. In this study, rMATS and ABLas analysis algorithms were used to analyze the alternative splicing events on a total of 56 lignocellulose degrading enzyme genes in *A. niger* strain CBS513.88 growing on media respectively using glucose (Group G) and wheat straw (Group WS) as the sole carbon source. The alternative splicing events occurred on the three typical genes were further verified by RT-PCR and intron-specific amplification. The results showed that ABLas analysis algorithm was more accurate than rMATS analysis algorithm. Based on ABLas analysis algorithm, a total of 21 lignocellulose degrading enzyme genes in both Group G and Group WS was found to have alternative splicing, and the type of alternative splicing occurred was mainly intron retention (IR) which accounted for 82.85% of all alternative splicing events. In addition, the lignocellulose degrading enzyme genes with alternative splicing in Group G and Group WS were different: 13 genes were found to have alternative splicing in Group G while 14 genes in Group WS, with 6 genes being the same in both groups. Our results indicates that the occurrence of alternative splicing on lignocellulose degrading enzyme genes in *A. niger* changes with different growth conditions. The existence of a large number of splicing variants in *A. niger* provides a basis for discovery of novel lignocellulose degrading enzymes.

Keywords: *Aspergillus niger*; lignocellulose degrading enzyme; RNA-seq; alternative splicing; intron retention

黑曲霉 *Aspergillus niger*, 因其能够大量分泌木质纤维素降解酶, 且被美国食品药品监督管理局(Food and Drug Administration, FDA)认定为GRAS (generally recognized as safe)级安全菌株, 因而成为目前工业上用于纤维素酶、半纤维素酶生产的最有潜力的微生物之一, 在木质纤维素转化利用中发挥重要作用(郭艳梅等2010)。不断阐明黑曲霉宽泛的木质纤维素降解酶系的产生机制对于更好地利用黑曲霉产酶细胞工厂具有实践意义。

可变剪接(alternative splicing, AS)可以使真核生物利用有限的基因产生不同的 mRNA, 从而产生不同的蛋白质, 该方式极大地扩展了蛋白质组的多样性(Black 2003)。可变剪接的基本类型包括外显子跳跃(exon skipping, ES)、可变 5'端位点(alternative 5' splice sites, A5SS)、可变 3'端位点(alternative 3' splice sites, A3SS)、外显子互斥(mutually exclusive exons, MXE)和内含子保留(intron retention, IR)等方式(Graveley 2001)。

可变剪接在很多高等生物中被发现与酶的不同功能相关, Christmas *et al.* (2001)在人类细胞色素 P450 基因 CYP4F3 中, 获得了 2 个发生外显子互斥的可变剪接体 CYP4F3A 和 CYP4F3B, 这 2 个可变剪接体相较于正常剪接体 CYP4F3 表现出了不同的组织分布, 且具有不同的底物特异性和生物学功能。在丝状真菌中也有证据表明可变剪接产生了来自同一基因的不同的酶, Boldo *et al.* (2010)在绿僵菌 *Metarhizium anisopliae* 中检测到了一个内含子保留的几丁质酶基因, 其产物推测是一个与细胞壁结合的几丁质酶, 参与细胞分裂过程中的细胞壁重建, 而正常剪接的几丁质酶则被分泌到胞外以降解胞外的几丁质提供营养。前期对黑曲霉 β -葡萄糖苷酶的研究中也发现了一型内含子保留的 β -葡萄糖苷酶剪接体, 它与正常剪接的 β -葡萄糖苷酶相比, 具有更高的最适温度和更好的热稳定性, 在水解京尼平苷制备京尼平的反应中, 内含子保留的 β -葡萄糖苷酶也具有更高的转化率(Zhu *et al.* 2019)。黑曲霉中 β -葡萄糖苷酶基因可变剪接体的发现促使我们考虑黑曲霉中其他的木质纤维素降解酶基因是否有发生可变剪接事件, 目前有关黑曲霉基因组以及其中与木质纤维素降解相关基因的可变剪接体存在情况尚不清楚。

在下一代测序(next generation sequencing, NGS)技术中, 真核生物包括真菌的可变剪接的检测主要通过对剪接位点(splice junction)的分析来进行, 即通过分析测序序列比对到参考基因组所出现的断层处的剪接位点, 可以初步检测到可变剪接事件(Xie *et al.* 2015)。另外, 通过概率建模进行基因模型预测的可变剪接分析方法同样也被用于分析真菌中的可变剪接(Zhao *et al.* 2013)。基于真菌基因组出现的高基因密度和转录本的多顺反子情况(Gordon *et al.* 2015), 本研究拟采用 rMATS 和 ABLas 两种可变剪接分

析算法对黑曲霉 CBS513.88 进行可变剪接分析, 其中 rMATS 分析算法是通过 rMATS 统计模型对样本中可变剪接事件进行表达定量, 主要用于分析人类等高等哺乳动物基因组的可变剪接(刘文荣等 2017; 郭家中等 2018; 鲍晶晶等 2019); ABLas 分析算法通过剪接位点进行分析, 主要用于分析人类基因组的可变剪接(Li *et al.* 2020; Liu *et al.* 2020), 对该算法进行修改, 也适用于分析真菌等低等真核生物的可变剪接(Jin *et al.* 2017)。

本研究对高产酶黑曲霉菌株 CBS513.88 在葡萄糖为唯一碳源和小麦秸秆为唯一碳源 2 种培养条件下进行 RNA-seq 分析, 基于 RNA-seq 结果采用 rMATS 和 ABLas 两种算法分别进行可变剪接分析, 从而鉴定出黑曲霉木质纤维素降解酶基因的可变剪接事件, 并通过 RT-PCR 和内含子的特异性扩增进行可变剪接事件的验证, 从而解析黑曲霉木质纤维素降解酶基因的可变剪接事件, 有利于加深可变剪接对黑曲霉产木质纤维素酶多样化的理解, 同时也为开发新型的木质纤维素降解酶资源提供理论基础。

1 材料与方法

1.1 菌株和试剂

实验所用的黑曲霉 CBS513.88 菌株购自广东省微生物研究所。本研究所用的 RNA 提取试剂 RNAiso plus 购自 TaKaRa 公司; 5×All-In-One RT Master Mix 购自爱必梦生物科技有限公司; Q5 超保真 DNA 聚合酶来自 NEB 公司; TA 零背景快速克隆试剂盒和大肠杆菌 DH5 α 来自北京庄盟国际生物基因科技有限公司。

1.2 不同碳源菌丝体获取

将黑曲霉 CBS513.88 的 4 代菌株接种于马铃薯葡萄糖琼脂(PDA)培养基(马铃薯 200 g、葡萄糖 20 g、琼脂 15–20 g、ddH₂O 1 000 mL),

在 28 °C 培养箱中培养 3–5 d 至黑曲霉孢子形成, 将菌株孢子悬液(10^7 个/mL)以每瓶 1 mL (10^7 个孢子)孢子悬液接入黑曲霉生长培养基 [2% 葡萄糖/小麦秸秆和 0.1% 蛋白胨加入到 Mandels' 基础培养基(KH_2PO_4 3 g/L、 $(\text{NH}_4)_2\text{SO}_4$ 2 g/L、 $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 0.5 g/L、 CaCl_2 0.5 g/L、 $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ 7.5 mg/L、 $\text{MnSO}_4 \cdot \text{H}_2\text{O}$ 2.5 mg/L、 $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ 3.6 mg/L、 $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$ 0.5 g/L)] 中, 生长培养基分装于 250 mL 三角瓶中, 每瓶 50 mL 在 28 °C、140 r/min 培养 72 h, 培养后获取菌丝体以液氮速冻保存于-80 °C。

1.3 RNA 提取和反转录

从-80 °C 冰箱中取出黑曲霉 CBS513.88 在葡萄糖和小麦秸秆两种碳源下生长的菌丝体, 使用液氮研磨充分, 按照 TaKaRa 公司的 RNAiso plus 试剂的 RNA 提取说明书提取葡萄糖组 G1、G2、G3 和小麦秸秆组 WS1、WS2、WS3 中的菌丝体总 RNA。将提取的 RNA 以 1% 的琼脂糖凝胶进行检测, 并利用 NanoDrop one 核酸浓度检测仪测量浓度, OD_{260}/OD_{280} 在 2.0 左右, 将符合条件的 RNA 按照爱必梦生物科技有限公司反转录试剂盒 5×All-In-One RT Master Mix (with AccuRT Genomic DNA Removal Kit) 说明进行反转录合成 cDNA 第一链, cDNA 用于后续实验或保存于-20 °C。

1.4 RNA-seq 文库构建和序列过滤

RNA-seq 文库构建和测序均由深圳华大基因有限公司完成, 分别取 3 个重复样品的总 RNA 后构建 2 个 cDNA 文库, 即 G 组(葡萄糖组)和 WS 组(小麦秸秆组), 在 BGISEQ500 平台进行测序。测序得到的原始数据, 里面含有带接头的、低质量的序列, 为确保测序数据的可靠性, 使用 SOAPnuke (Cock *et al.* 2010) 对测序数据进行过滤, 去除包含接头的序列、去除未知碱基 N 含量大于 5% 的序列、去除低质量的序列 (将质量值低于 10 的碱基占该序列总碱基数的

比例大于 20% 的序列定义为低质量序列), 得到有效数据(clean data)。

1.5 参考基因组的比对

黑曲霉 CBS513.88 的基因组序列和注释文件从美国国家生物技术信息中心 (National Center for Biotechnology Information, NCBI) 中获取, 参考基因组版本号为 GCF_000002855.3_ASM285v2, 使用 HISAT2 (Kim *et al.* 2015) 将有效测序数据(clean reads)与参考基因组进行比对, 再运用 Bowtie2 (Langmead & Salzberg 2012) 将测序数据比对到参考基因序列得到比对结果, 然后使用 RESM (Li & Dewey 2011) 计算各个样本的基因表达水平。

1.6 差异基因表达分析

使用 R 包 DESeq2 (Love *et al.* 2014) 筛选差异表达基因 (differentially expressed genes, DEG), 设置差异表达基因的筛选识别标准为 $P < 0.05$ 。

1.7 可变剪接分析

基于比对到参考基因组上的序列, 分别采用 rMATS 算法 (Shen *et al.* 2014) 和 ABLas 算法 (Jin *et al.* 2017; Xia *et al.* 2017; Liu *et al.* 2020) 分析黑曲霉木质纤维素降解酶基因的可变剪接。

rMATS 可识别的可变剪接事件有 5 种, 分别是外显子跳跃(exon skipping, ES)、可变 5' 端位点(alternative 5' splice site, A5SS)、可变 3' 端位点(alternative 3' splice site, A3SS)、内含子保留(intron retention, IR) 和外显子互斥(mutually exclusive exons, MXE)。

ABLas 算法使用 TOPHAT2 (Trapnell *et al.* 2009) 识别剪接位点, 可检测出 9 种可变剪接类型, 即外显子跳跃(ES)、可变 5' 端位点(A5SS)、可变 3' 端位点(A3SS)、外显子互斥(MXE)、可变启动子(mutually exclusive 5' UTRs, 5pMXE)、可变终止子(mutually exclusive 3' UTRs, 3pMXE)、

可变 5'端位点和外显子跳跃(A5SS&ES)、可变 3'端位点和外显子跳跃(A3SS&ES)、盒式外显子(cassette exon)。除此之外，内含子保留(IR)的检出是通过对候选剪接位点的序列深度减少为标准来计算的，其必须满足 4 个标准才能被判定为 IR 事件：(1) 候选内含子的平均碱基深度至少为侧翼外显子的 20%；(2) 内含子深度之和大于 100；(3) 候选内含子的 5'或 3'剪接位点上必须有边界序列存在；(4) 未发现其他类型的可变剪接事件。

1.8 黑曲霉典型的木质纤维素降解酶基因转录本的扩增和测序

本研究选取的 3 个典型的木质纤维素降解酶基因分别是内切木聚糖酶基因 *xynF1* (endo-1,4-beta-xylanase F1)、阿拉伯呋喃糖苷酶基因 *abnC* (alpha-L-arabinofuranosidase C)以及纤维二糖水解酶基因 *cbhC* (1,4-beta-D-glucan cellobiohydrolase C)。以反转录获得的 cDNA 为模板，利用 Q5 超保真 DNA 聚合酶和基因特异性引物(表 1)对 *xynF1* 基因(NCBI ID:4980082)、*abnC* 基因(NCBI ID: 4979546)和 *cbhC* 基因(NCBI ID: 4982491)进行转录本的扩增，PCR 产物经过 1%琼脂糖凝胶电泳检测后，使用 SanPrep 柱式 DNA 胶回收试剂盒进行产物纯

化，将 PCR 产物与 One step ZTOPO-Blunt/TA 载体(北京庄盟国际生物基因科技有限公司)按照试剂盒要求连接，连接完成后转化到大肠杆菌 DH5 α ，筛选阳性克隆，获得的阳性克隆单菌落送往生工生物工程(上海)股份有限公司进行测序验证。

1.9 木质纤维素降解酶基因转录本中 A5SS 事件和 IR 事件的验证

对上述选定的 3 个木质纤维素降解酶基因 *xynF1*、*abnC* 和 *cbhC* 分别设计跨内含子引物进行 PCR，PCR 产物经 3%琼脂糖凝胶电泳检测。引物具体设置如下：(1) *xynF1* 基因的上游引物位于第 5 个外显子的第 11 bp 到 32 bp 处，下游引物位于第 7 个内含子的前 22 bp。(2) *abnC* 基因的上游引物位于第 1 个外显子的第 205 bp 到 225 bp 处，下游引物位于第 2 个外显子的第 48 bp 到 68 bp 处。(3) *cbhC* 基因针对第 1 个内含子和第 5 个内含子分别设置了 2 组引物，其中第 1 个上游引物位于第 1 个外显子的第 1 bp 到第 27 bp 处，第 1 个下游引物位于第 2 个外显子的第 82 bp 到第 102 bp 处。第 2 个上游引物位于第 5 个外显子的第 215 bp 到第 235 bp 处，第 1 个下游引物位于第 6 个外显子的第 40 bp 到第 60 bp 处。

表 1 黑曲霉木质纤维素降解酶基因转录本的 RT-PCR 引物和内含子特异性扩增引物
Table 1 RT-PCR primers and intron-specific amplification primers for transcripts of *Aspergillus niger* lignocellulose degrading enzymes

基因名称 Gene name	NCBI 编号 NCBI ID	RT-PCR 引物序列 RT-PCR primer sequence (5'→3')	
		上游 Forward	下游 Reverse
<i>xynF1</i>	4980082	ATGGTTCAGATCAAGGTAGCTGCA	CTAGAGAGCATTGCGATAGCAGTG
<i>abnC</i>	4979546	ATGCTTTCGTTTGTCTTGCTTCT	TCACGCCACAACAGGCCA
<i>cbhC</i>	4982491	ATGTTTAGCCGGATTACTTTGTCAC 内含子跨越引物序列(5'→3')	TCAAAGGGAAGGATTGGCGT
		Intron spanning primer sequence (5'→3')	
<i>xynF1</i>	4980082	ATGAAATCTTCAACGAAGACGG	CCAGAAATTCCAGCACCTCCAC
<i>abnC</i>	4979546	GCACTGCCTAACGGATCCATC	CGGCTGCCCAGGGTAGAGCTA
<i>cbhC</i>	4982491	ATGTTTAGCCGGATTACTTTGTCAC GCTGTCTAGTGCTGGGTTTGA	GAGAAACGAGGGGCCAGATGG CCTTGACATTGCACCAGTCAC

2 结果与分析

2.1 木质纤维素降解酶基因在不同碳源下的表达差异分析

木质纤维素降解酶包括纤维素酶系的 β -葡萄糖苷酶(β -glucosidase, EC3.2.1.21)、纤维二糖水解酶(cellobilhydrolase, EC3.2.1.91)、内切葡聚糖酶(endoglucanase, EC3.2.1.4); 木聚糖酶系的内切木聚糖酶(endoxylanase, EC3.2.1.8)、 β -木糖苷酶(β -xylosidase, EC3.2.1.37)、阿拉伯呋喃糖苷酶(α -L-arabinofuranosidase, EC3.2.1.55)、乙酰木聚糖酯酶(acetyl xylan esterase, EC3.1.1.72)等(Mach-Aigner *et al.* 2012); 以及多种辅助活性酶(auxiliary activities, AA)和一些含有碳水化合物结合结构域(carbohydrate-binding module, CBM)的酶等。检索黑曲霉 CBS513.88 基因组, 确定其中 56 个主要木质纤维素降解酶基因用于后续研究(表 2)。

将黑曲霉 CBS513.88 在以葡萄糖为唯一碳源培养的样本命名为 G 组, 以小麦秸秆为唯一碳源培养的样本命名为 WS 组, 对 2 个组共 6 个样本进行 RNA-seq 转录组测序。6 个测序样本有效数据(clean data)占比都达到 91%以上, 且有效数据比对到参考基因组的比例也达到 91%以上, 证明获得的转录组测序数据质量较高(表 S1, 表 S2)。RNA-seq 分析显示, 黑曲霉

CBS513.88 中的 56 个木质纤维素降解酶基因, 有 41 个基因都出现了差异表达, 其中 33 个基因在 WS 组上调, 8 个下调, 上调基因数量占比 80.48% (图 1)。值得关注的是, 在上调表达最高的前 5 个基因中有 4 个都是木聚糖酶系的内切木聚糖酶基因, 另一个为纤维素酶系的纤维二糖水解酶基因。多数差异表达的木质纤维素降解酶基因在小麦秸秆基质上均上调表达, 这与 Pullan *et al.* (2014)的研究结果一致。

2.2 不同算法对木质纤维素降解酶基因的可变剪接分析

2.2.1 不同算法检出的木质纤维素降解酶基因的可变剪接数量

本研究对上述选定的 56 个木质纤维素降解酶基因, 分别采用 rMATS 和 ABLas 两种可变剪接分析方法来分析黑曲霉的可变剪接事件。

rMATS 算法是基于 rMATS 统计模型对黑曲霉 CBS513.88 两种碳源条件下共 6 个样本进行可变剪接事件的表达定量, 然后以似然比检验(likelihood-ratio test)计算 *P* value 来表示两组样品在 IncLevel (inclusion level)水平上的差异, IncLevel 利用 Benjamini Hochberg 算法对 *P* value 进行校正得到 FDR 值, 从而鉴定黑曲霉中的可变剪接事件。rMATS 分析结果显示, 黑曲霉 CBS513.88 菌株木质纤维素降解酶的 56 个基因中, 共有 8 个基因出现了可变剪接, 其中 G 组

表 2 黑曲霉 CBS513.88 基因组中主要木质纤维素降解酶基因

Table 2 Main lignocellulose degrading enzyme genes in the genome of *Aspergillus niger* CBS513.88

NCBI 编号 NCBI ID	基因名称 Gene name	基因注释 Gene annotation	家族 Family	是否差异基因 DEG or no DEG
4977095	<i>xg74</i>	Xyloglucanase	GH74	Yes
4977345	<i>cbhB</i>	1,4-beta-D-glucan cellobiohydrolase B	GH7	Yes
4982202	<i>cbhB</i>	1,4-beta-D-glucan cellobiohydrolase B	GH7	Yes
4987438	<i>aguA</i>	Alpha-glucuronidase A	GH67	Yes
4982491	<i>cbhC</i>	1,4-beta-D-glucan cellobiohydrolase C	GH6	Yes
4985573	<i>cbhC</i>	1,4-beta-D-glucan cellobiohydrolase C	GH6	Yes
4987831	<i>abfB</i>	Alpha-N-arabinofuranosidase B	GH54	Yes
4978177	<i>abfA</i>	Alpha-N-arabinofuranosidase A	GH51	No

待续

续表 2

4982486	<i>abfC</i>	Alpha-N-arabinofuranosidase C	GH51	Yes
4983621	<i>abfA</i>	Alpha-N-arabinofuranosidase A	GH51	No
4977614	<i>eglA</i>	Endo-beta-1,4-glucanase A	GH5	Yes
4982163	<i>eglB</i>	Endo-beta-1,4-glucanase B	GH5	Yes
4989072	<i>eglA</i>	Endoglucanase A	GH5	No
4979546	<i>abnC</i>	Alpha-L-arabinosidase C	GH43	Yes
4980522	<i>bglJ</i>	Beta-glucosidase J	GH3	Yes
4981203	<i>bglB</i>	Beta-glucosidase B	GH3	No
4982032	<i>bglG</i>	Beta-glucosidase G	GH3	Yes
4982244	<i>bglD</i>	Beta-glucosidase D	GH3	Yes
4983142	<i>bglD</i>	Beta-glucosidase D	GH3	No
4984238	<i>bglM</i>	Beta-glucosidase M	GH3	Yes
4987033	<i>bglM</i>	Beta-glucosidase M	GH3	No
4988081	<i>bglG</i>	Beta-glucosidase G	GH3	Yes
4989339	<i>bgl</i>	Beta-glucosidase	GH3	Yes
4989375	<i>bglJ</i>	Beta-glucosidase J	GH3	No
4989921	<i>bglA</i>	Beta-glucosidase A	GH3	Yes
4977682	<i>xlnD</i>	Exo-1,4-beta-xylosidase xlnD	GH3	Yes
4977958	<i>xynA</i>	Endo-1,4-beta-xylanase A	GH11	Yes
4978152	<i>xynA</i>	Endo-1,4-beta-xylanase A	GH11	Yes
4987596	<i>xyn5</i>	Endo-1,4-beta-xylanase 5	GH11	Yes
4988056	<i>xynB</i>	Endo-1,4-beta-xylanase B	GH11	Yes
4980082	<i>xynF1</i>	Endo-1,4-beta-xylanase F1	GH10	Yes
4980363	<i>bgl1B</i>	Beta-glucosidase 1B	GH1	Yes
4983612	-	Acetyl xylan esterase	CE5	No
4983631	-	Acetyl xylan esterase	CE5	Yes
4985851	<i>axeA</i>	Acetyl xylan esterase A	CE1	Yes
4980408	-	Hypothetical protein	CBM63	Yes
4978139	-	Hypothetical protein	CBM50	No
4988766	-	Hypothetical protein	CBM48	Yes
4979015	-	Hypothetical protein	CBM21	Yes
4988091	<i>eglD</i>	Probable endo-beta-1,4-glucanase D	CBM1/AA9	Yes
4987123	<i>gun4</i>	Endoglucanase-4	CBM1/AA9	Yes
4988058	-	Hypothetical protein	CBM1	Yes
4982839	<i>gun4</i>	Endoglucanase-4	AA9	No
4985809	<i>gun4</i>	Endoglucanase-4	AA9	Yes
4991259	<i>gun4</i>	Endoglucanase-4	AA9	No
4979417	<i>cdh</i>	Cellobiose dehydrogenase	AA2	Yes
4990292	<i>cdh</i>	Cellobiose dehydrogenase	AA2	Yes
4978017	<i>abr2</i>	Laccase abr2	AA1	Yes
4977636	<i>abr2</i>	Laccase abr2	AA1	Yes
4980364	<i>abr2</i>	Laccase abr2	AA1	No
4980991	<i>abr2</i>	Laccase abr2	AA1	Yes
4991441	<i>abr2</i>	Laccase abr2	AA1	No
4985931	<i>lac1</i>	Laccase ARB	AA1	Yes
4988595	<i>lac1</i>	Laccase ARB	AA1	Yes
4980971	<i>lac2</i>	Laccase-2	AA1	No
4984589	<i>lac2</i>	Laccase-2	AA1	No

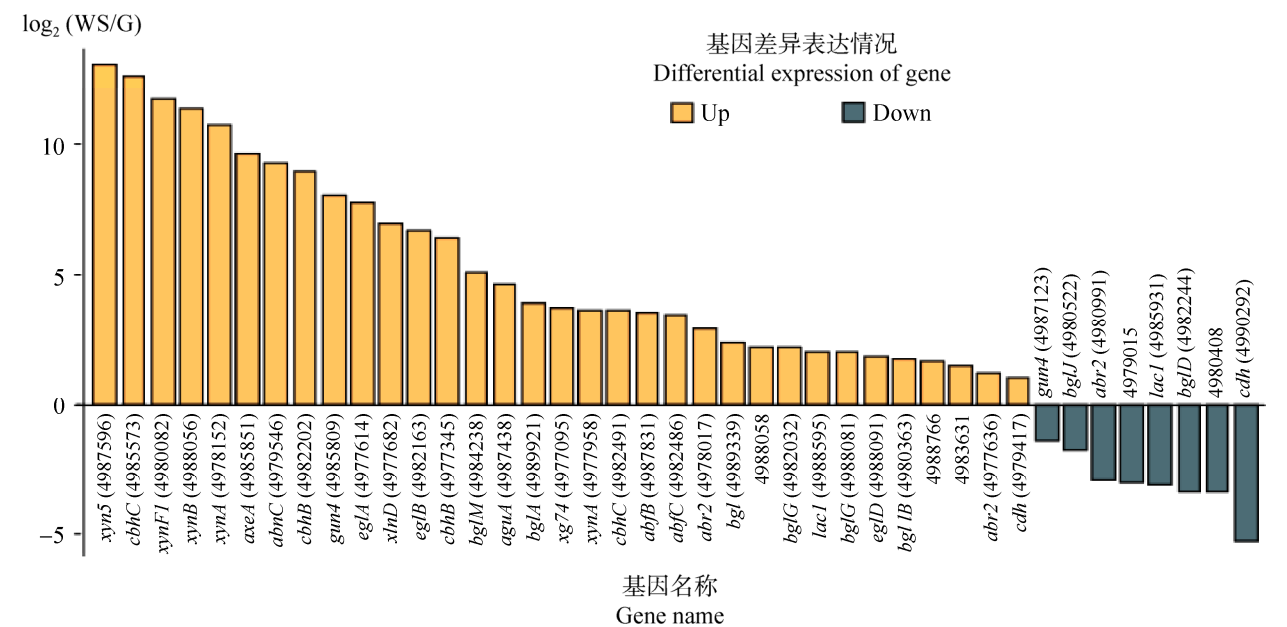


图 1 黑曲霉 CBS513.88 的木质纤维降解酶基因在不同碳源下的差异表达情况 $\log_2(WS/G)$:反映差异倍数, 即 $\log_2(WS \text{ 组基因表达量}/G \text{ 组基因表达量})$, 大于 0 表示相对于 G 组, WS 组基因表达量上调, 小于 0 表示下调. 上图所有基因的差异显著值(P 值)均小于 0.05, 都具有显著性差异. 黄色表示在 WS 组中上调的木质纤维素降解酶基因, 灰色表示在 WS 组中下调的木质纤维素降解酶基因

Fig. 1 Differential expression of lignocellulose degrading enzyme genes in *Aspergillus niger* CBS513.88 growing on media with glucose as the sole carbon source (G group) and wheat straw as the sole carbon source (WS group). $\log_2(WS/G)$: Reflects the multiple of difference, $\log_2(\text{Expression of WS group}/\text{Expression of G group})$, greater than 0 means that the gene expression of WS group is up-regulated relative to group G, and less than 0 means down-regulation. The significant differences (P values) of all genes in the above figure are all less than 0.05, and all have significant differences. Yellow indicates the up-regulated lignocellulose degrading enzyme genes in the WS group, and gray indicates the down-regulated lignocellulose degrading enzyme genes in the WS group.

发生可变剪接的基因有 4 个, WS 组中出现可变剪接的基因有 7 个, 两组共同出现可变剪接的基因数目为 3 个(图 2)。

ABLas 算法是基于黑曲霉中剪接位点(splice junction)进行分析, 检出了黑曲霉 CBS513.88 基因组中所有注释外显子和内含子的边界区域的剪切位点, 一共产生了 163 234 个剪切位点, 基于剪切位点处序列的数量, 对黑曲霉可变剪接事件进行鉴定。ABLas 分析结果显示 56 个木质纤维素降解酶基因中, 共有 21 个基因出现了可变剪接事件, 其中 G 组发生可变剪接的基因数目为 13 个, WS 组发生可变剪接的基因数目为 14 个, 两组都发生可变剪接的基因数目为 6 个(图 2)。

两种分析方法的结果对比显示, 使用 ABLas 分析方法检出的可变剪接基因数量较多, 其检出的木质纤维素降解酶的可变剪接基因数量是 rMATS 分析方法检出数量的 2.6 倍以上。

2.2.2 不同算法检出的木质纤维素降解酶基因的可变剪接类型

rMATS 算法和 ABLas 两种算法检测出的木质纤维素降解酶基因的可变剪接也显示出差异, 其发生类型见表 3。对上述选定的 56 个木质纤维素降解酶基因, 分别用 2 种算法分析了这 56 个基因的可变剪接情况。分析发现其中 23 个基因(任一种算法检测出可变剪接即统计在内)发生了至少 1 种可变剪接事件。

rMATS 可变剪接分析结果显示在 WS 组和 G 组只有 1 种可变剪接类型，为外显子跳跃 (exon skipping, ES)，其中 WS 组检出 7 个 ES 事件，占 2 组全部可变剪接事件的 63.64%，G 组检出 4 个 ES 事件，占比 36.36% (表 3)。

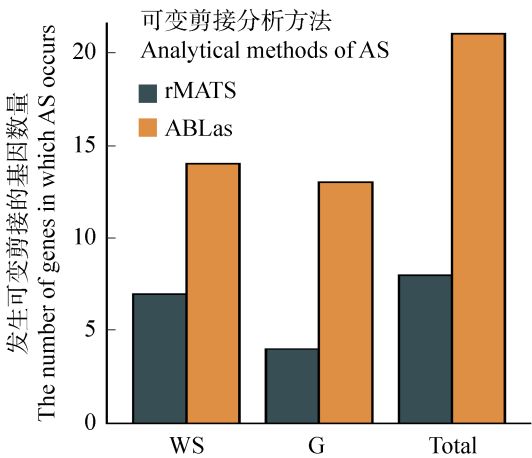


图 2 两种算法检出的不同碳源下培养的黑曲霉木质纤维素降解酶的可变剪接基因数量
WS：以小麦秸秆为唯一碳源培养的黑曲霉 CBS513.88 菌株；G：以葡萄糖为唯一碳源培养的黑曲霉 CBS513.88 菌株。灰色表示采用 rMATS 算法分析的可变剪接结果，黄色表示采用 ABLas 算法分析的可变剪接结果

Fig. 2 The number of alternative splicing (AS) genes of *Aspergillus niger* lignocellulose degrading enzymes cultured under different carbon sources detected by the two algorithms. WS: *A. niger* CBS513.88 strain cultured on medium with wheat straw as the sole carbon source; G: *A. niger* CBS513.88 strain cultured on medium with glucose as the sole carbon source. Gray represents the AS results analyzed by rMATS algorithm, and yellow represents the AS results analyzed by ABLas algorithm.

ABLas 可变剪接分析结果显示在 WS 组和 G 组共出现了 3 种可变剪接类型，分别为 ES、A5SS 和 IR。WS 组发生的主要可变剪接类型为 IR，IR 事件数量为 13 个，占 WS 组所有可变剪接事件的 92.85%，除去 IR 事件的 1 个可变剪接事件类型为 A5SS，G 组发生的可变剪接情况与

WS 组相似，G 组中可变剪接事件主要为 IR，有 13 个，占 G 组所有可变剪接事件的 92.85%，剩余 1 个可变剪接事件的类型为 ES (表 3)。

使用 rMATS 算法的可变剪接分析结果显示可变剪接类型以 ES 为主，值得一提的是，IR 在结果中并未检出，而在 ABLas 的可变剪接分析中检出了占比高于 92% 的 IR 事件，且检出的 IR 数量要远远高于 rMATS 算法检出的 ES 数量，2 种算法的可变剪接分析结果在数量和类型上差异巨大，基于 2 种算法的差异，因此黑曲霉 CBS513.88 的木质纤维素降解酶基因在 2 种碳源条件下发生的可变剪接事件的数量和类型并不确定，需要验证 2 种算法可变剪接分析结果的可靠性。

2.3 典型木质纤维素降解酶基因可变剪接事件的验证

2.3.1 *xynF1*、*abnC* 和 *cbhC* 基因可变剪接体的克隆

为了验证 2 种算法可变剪接分析结果的准确性，选取了半纤维素酶系的内切木聚糖酶基因 *xynF1* (NCBI ID: 4980082) 和阿拉伯呋喃糖苷酶基因 *abnC* (NCBI ID: 4979546)，以及纤维素酶系的纤维二糖水解酶基因 *cbhC* (NCBI ID: 4982491) 进行 RT-PCR 扩增。这 2 种算法的可变剪接分析都检测到这 3 个基因出现了可变剪接事件，但 2 种分析结果的可变剪接类型存在差异，其中 rMATS 分析方法检出 *xynF1*、*abnC* 和 *cbhC* 发生的可变剪接事件均为 ES 类型，而 ABLas 分析方法检出 *xynF1* 发生的可变剪接事件为 A5SS 类型，*abnC* 和 *cbhC* 发生的可变剪接事件均为 IR 类型 (表 3)。

对黑曲霉 CBS513.88 的 WS 组和 G 组的菌丝体分别提取了总 RNA，使用 1% 的琼脂糖凝胶电泳进行 RNA 质量的检测，采用 RT-PCR 分别扩增了目的基因 *xynF1*、*abnC* 和 *cbhC*，PCR 产物在 1% 的琼脂糖凝胶电泳检测下，分别获得了 3 条大小约为 984、1 019 和 1 468 bp 的特异

性单一条带，与 ABLas 算法分析的目的基因的可变剪接体大小一致(图 3)。对 PCR 产物进行纯化和 TA 克隆，并对获得的阳性转化子进行测序验证，结果显示除正常剪接的转录产物外，*xynF1* 转录物中检测到了第 7 个外显子的可变 5' 端位点，*abnC* 转录物中检测到了第 1 个内含子保留的转录产物，这都与 ABLas 算法预测的可变剪接事件一致，且都不同于 rMATS 算法预测的 ES 类型。在 *cbhC* 基因的转录产物中，也检测到了 ABLas 算法预测的 IR 事件，而未出现 rMATS 算法预测的 ES 类型，不过测序结果显示

其 IR 事件并未发生在 ABLas 算法检出的第 5 个内含子，而是发生在第 1 个内含子处。

2.3.2 *xynF1*、*abnC* 和 *cbhC* 基因的特异性内含子扩增

为了进一步验证这 3 个基因的可变剪接情况，对 *xynF1*、*abnC* 和 *cbhC* 发生可变剪接的区域设计特异内含子跨越引物进行 PCR 扩增，分别获得了大小约为 271、181 和 217 bp 的特异性单一条带(图 4)。对于 *xynF1* 基因，扩增产物显示出第 5 个内含子和第 6 个内含子的正常剪接和第 7 个内含子的部分保留，即从内含子

表 3 两种算法检出的黑曲霉 CBS513.88 木质纤维素降解酶基因的可变剪接类型

Table 3 AS types of *Aspergillus niger* CBS513.88 lignocellulose degrading enzyme genes detected by two algorithms

NCBI 登录号 NCBI ID	基因名称 Gene name	采用 rMATS 分析的可变剪接种类 AS types analysed by rMATS		采用 ABLas 分析的可变剪接种类 AS types analysed by ABLas	
		WS	G	WS	G
4980363	<i>bglIB</i>	-	-	-	IR
4980522	<i>bglJ</i>	-	-	IR	-
4982032	<i>bglG</i>	-	-	IR	-
4983142	<i>bglD</i>	-	-	IR	-
4984238	<i>bglM</i>	-	-	IR	IR
4987033	<i>bglM</i>	-	-	-	IR
4989339	<i>bgl</i>	-	-	-	IR
4982202	<i>cbhB</i>	-	-	-	IR
4982491	<i>cbhC</i>	ES	-	IR	-
4985573	<i>cbhC</i>	ES	-	-	-
4989072	<i>eglA</i>	-	ES	-	-
4977958	<i>xynA</i>	-	-	IR	-
4980082	<i>xynF1</i>	ES	-	A5SS	-
4978177	<i>abfA</i>	ES	ES	IR	IR
4982486	<i>abfC</i>	ES	ES	-	IR
4979546	<i>abnC</i>	ES	-	IR	-
4983631	-	-	-	-	IR
4991259	<i>gun4</i>	-	-	IR	IR
4980991	<i>abr2</i>	-	-	-	IR
4978139	-	-	-	IR	-
4988091	<i>eglD</i>	-	-	IR	IR
4988766	-	ES	ES	IR	ES/IR
4987123	<i>gun4</i>	-	-	IR	IR

注：“-”表示 3 个重复样本均未检出，对于标注可变剪接类型的基因是指 3 个样本中至少有 1 个检出，当 3 个重复样本中检出不同的可变剪接类型时，用“/”隔开

Note: “-” indicates that none of the AS type was detected in all 3 replicate samples. Genes marked with AS types mean that the AS types occurred in at least 1 of 3 replicate samples. When different AS types were detected in 3 replicate samples, separate them with “/”.

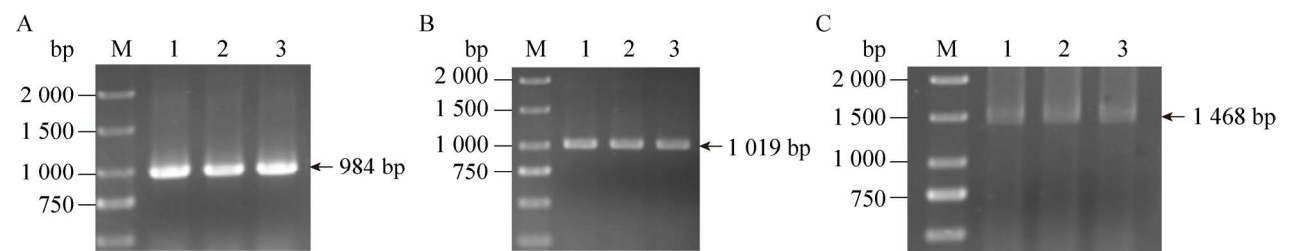


图 3 典型木质纤维素降解酶基因的 RT-PCR 扩增情况 M: DL2000 marker; 1, 2, 3: 分别为 WS 组的 3 个样本; A: *xynF1* 基因的 RT-PCR 电泳图; B: *abnC* 基因的 RT-PCR 电泳图; C: *cbhC* 基因的 RT-PCR 电泳图

Fig. 3 RT-PCR amplification of typical lignocellulose degrading enzyme genes. M: DL2000 marker; 1, 2, 3: Three samples in the WS group; A: The RT-PCR electrophoresis image of the *xynF1* gene; B: The RT-PCR electrophoresis image of the *abnC* gene; C: The RT-PCR electrophoresis image of the *cbhC* gene.

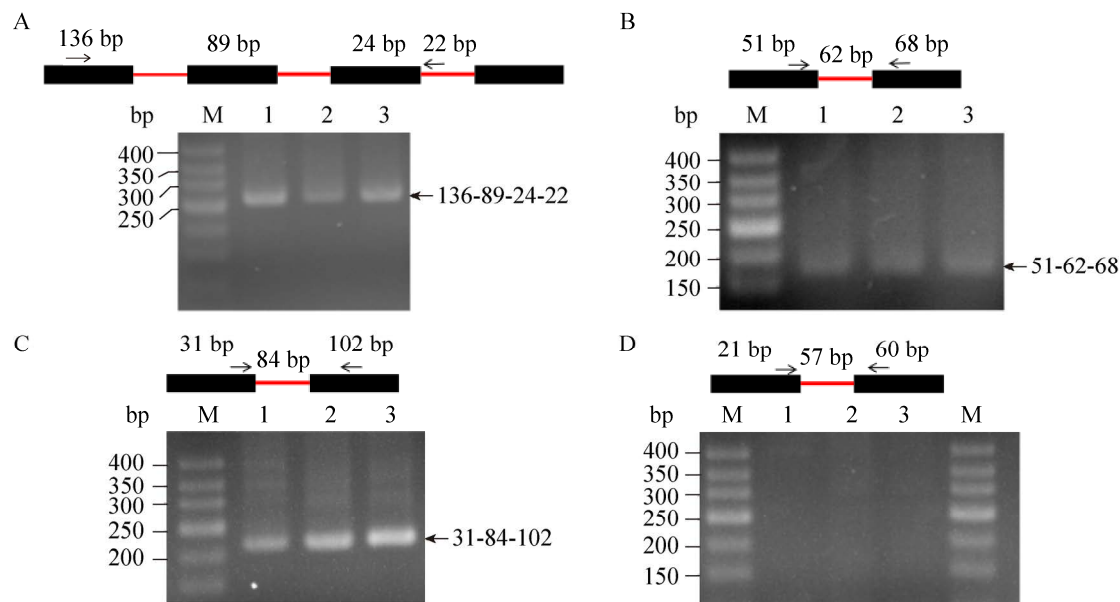


图 4 典型木质纤维素降解酶基因的内含子特异性扩增情况 M: 50 bp DNA ladder; 1, 2, 3: 分别为 WS 组的 3 个样本; A: *xynF1* 基因的 A5SS 事件验证电泳图; B: *abnC* 基因的 IR 事件验证电泳图; C: *cbhC* 基因的 IR 事件(第 1 个内含子保留)验证电泳图; D: *cbhC* 基因的 IR 事件(第 5 个内含子保留)验证电泳图. 示意图中黑色方块代表外显子, 红色细线代表内含子, 箭头代表内含子跨越引物

Fig. 4 Intron-specific amplification of typical lignocellulose degrading enzyme genes. M: 50 bp DNA ladder; 1, 2, 3: Three samples in the WS group; A: The validation electropherogram of the A5SS event of the *xynF1* gene; B: The validation electropherogram of the IR event of the *abnC* gene; C: The validation electropherogram of the IR event (retention of the first intron) of the *cbhC* gene; D: The validation electropherogram of the IR event (5th intron retention) of the *cbhC* gene. In the schematic diagram, black squares represent exons, red thin lines represent introns, and arrows represent intron-spanning primers.

特异序列扩增出目的条带, 即证明了其存在 A5SS 类型的可变剪接事件(图 5A); 对于 *abnC* 基因, 第 1 个内含子的特异性扩增结果证明其在第 1 个内含子处存在内含子保留(图 5B); 在

cbhC 基因中, ABLas 算法显示其在第 5 个内含子处存在内含子保留, 但扩增结果显示其在第 1 个内含子处存在内含子保留, 在第 5 个内含子处没有出现内含子保留(图 5C)。

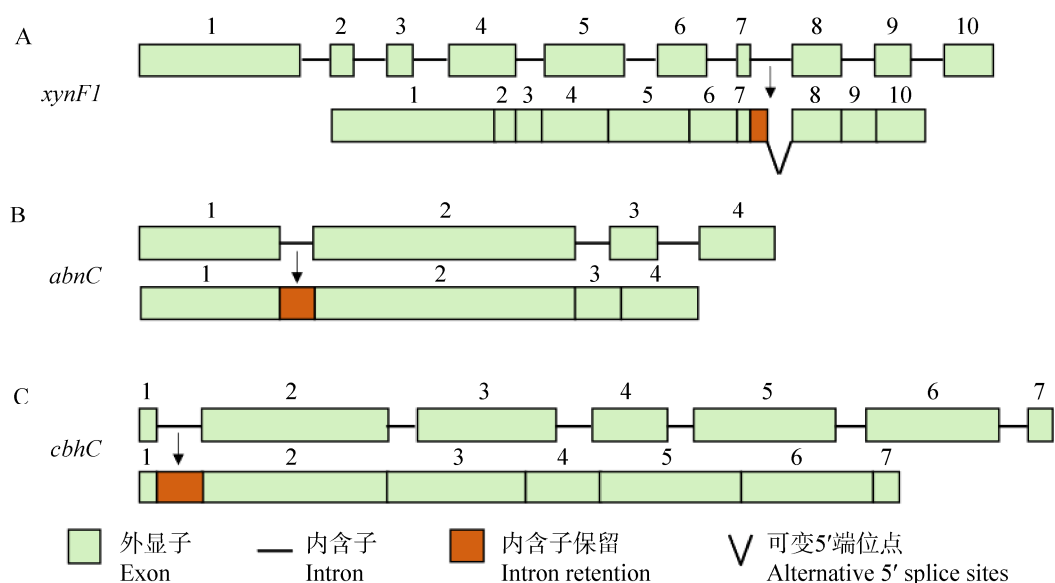


图 5 典型木质纤维素降解酶基因的可变剪接模式示意图 上排：基因的基因组 DNA 结构；下排：可变剪接体的结构

Fig. 5 Schematic of alternative splicing patterns of typical lignocellulose degrading enzyme genes. A: *xynFI*; B: *abnC*; C: *cbhC*. Upper: The genomic DNA of each gene; Below: Splicing variant.

综合 RT-PCR 和内含子特异性扩增结果，可知 *xynFI* 基因存在的 A5SS 事件与第 2 种算法 ABLas 分析结果一致，*abnC* 基因存在的 IR 事件也与 ABLas 分析结果一致，*cbhC* 基因虽也出现了内含子保留情况，但其出现位置与 ABLas 分析结果不同，且 3 个基因均没有出现 rMATS 分析结果中的 ES 事件，由此证实算法 ABLas 分析的可变剪接事件较为准确。

3 讨论

真菌基因组的特异性和复杂性以及算法间的差异导致了可变剪接检出结果的巨大差异。本研究采用 rMATS 和 ABLas 两种算法分析黑曲霉木质纤维素降解酶基因的可变剪接情况无论是发生可变剪接的基因数目还是可变剪接的类型都不相同，在黑曲霉 CBS513.88 的 56 个木质纤维素降解酶基因中，通过 rMATS 分析方法共得到以 ES 为主要类型的 8 个可变剪接基因，通过 ABLas 分析方法共得到以 IR 为主要类型的 21 个可变剪接基因。为了确定可变剪接分析的

准确性，本研究选择典型的 3 个木质纤维素降解酶基因进行 RT-PCR 验证，结果显示，ABLas 算法检出的基因 *xynFI* 的 A5SS 事件和基因 *abnC* 的 IR 事件都被 RT-PCR 所验证，而 rMATS 算法检出的可变剪接事件并没有在 RT-PCR 结果中被验证，这显示基于 ABLas 算法分析黑曲霉的可变剪接事件准确度高。另外，从总体上 2 种方法所检出的可变剪接事件的类型看，rMATS 检出的可变剪接事件中 ES 为主要类型，占比 91.81% 以上，而 ABLas 检出的可变剪接事件中 IR 占比 71% 以上，而 ES 很少见，仅占比约 1% (表 S3，表 S4)。多项研究已表明真菌中 IR 是其最常见的可变剪接事件(McGuire *et al.* 2008; Zhao *et al.* 2013; Xie *et al.* 2015; Jin *et al.* 2017)，在真菌中 IR 的流行与哺乳动物中 ES 的流行(Xing & Lee 2006; Sammeth *et al.* 2008)形成明显的对比，本研究中 ABLas 以 IR 事件为主的分析结果与已有的真菌可变剪接研究结果相符。因此，ABLas 算法对黑曲霉的可变剪接分析结果可信度更高，后续讨论也将基

于 ABLas 的可变剪接分析结果开展。

可变剪接在不同物种中的发生频率不同。在人类中有 95% 以上的基因可能经历可变剪接事件(Pan *et al.* 2008); 在拟南芥 *Arabidopsis thaliana* 中约有 42% 的基因存在可变剪接(Marquez *et al.* 2012); 而在真菌中发生可变剪接的基因比例要低得多, 例如新生隐球菌 *Cryptococcus neoformans*、黄曲霉 *Aspergillus flavus*、稻瘟病菌灰巨座壳 *Magnaporthe grisea* 和禾谷镰刀菌 *Fusarium graminearum*, 它们的可变剪接基因占比分别为 4.3% (277 个基因)、1.3% (162 个基因)、1.4% (151 个基因)和 1.7% (231 个基因)(Loftus *et al.* 2005; McGuire *et al.* 2008; Ramani *et al.* 2011; Zhao *et al.* 2013)。本研究可变剪接分析结果显示黑曲霉 CBS513.88 (WS 组)中发生可变剪接的基因有 3 486 个, 占总基因数的 23.55% (表 S4), 与之前的研究相比, 可变剪接的发生率要高得多。这可能意味着发育越复杂且基因组越复杂的真菌发生可变剪接的频率就越高, 这与 Gehrmann *et al.* (2016)的观点一致。

本研究分别分析了黑曲霉在木质纤维素降解酶诱导条件(WS 组)和木质纤维素降解酶阻遏条件(G 组)下的可变剪接事件, 2 种条件下木质纤维素降解酶基因表现出了不同的可变剪接基因数目。只在 G 组出现而不在 WS 组出现可变剪接的木质纤维素降解酶基因有 7 个, 只在 WS 组出现而不在 G 组出现可变剪接的木质纤维素降解酶基因有 8 个, 在 2 组中均出现可变剪接的木质纤维素降解酶基因有 6 个。考察本研究选定的 3 个典型木质纤维素降解酶基因 *xynF1*、*abnC* 和 *cbhC* 在 2 组中出现可变剪接的情况, 差异更为显著, 3 个基因在 WS 组均出现可变剪接, 而在 G 组中却没有出现, 这也与本研究的 RT-PCR 结果、内含子跨越 PCR 结果和测序结果相一致。这表明黑曲霉木质纤维素降解酶基因的可变剪接在不同生长条件下存在差异, 推

测可变剪接可能在黑曲霉木质纤维素降解酶的表达调控中也具有一定的调节作用。一些研究表明, 可变剪接是受到调节的, 或者与不同环境条件有关。例如在拟南芥中, 富含丝氨酸/精氨酸的蛋白编码基因的剪接会因激素或者非生物胁迫而改变(Palusa *et al.* 2007)。在人类中一些涉及胚胎干细胞凋亡和分化的基因通过可变剪接进行调节(Salomonis *et al.* 2010)。本研究显示的黑曲霉基因组总体发生可变剪接的比例伴随生长条件的改变而变化, 也证实了 Zhao *et al.* (2013)有关禾谷镰刀菌等真菌基因可能会在不同的生长条件下出现不同的可变剪接比例的推测。

可变剪接是导致生物蛋白质组多样性的重要途径, 但是可变剪接也会产生转录本中终止密码子提前或移码突变等情况, 从而使这些转录本不具有相应的生物学功能(Vilardell *et al.* 2000), 本研究中检出的 *abnC* 基因和 *cbhC* 基因的 IR 类型可变剪接体即产生了终止密码子的提前或移码突变情况, 因而此 2 种可变剪接体不会产生功能产物。其他的木质纤维素降解酶基因的可变剪接体是否具有蛋白产物以及是否具有相应酶的功能, 还需要进行蛋白质水平及酶功能的分析检测, 对这些可变剪接体的深入探讨将可能发现黑曲霉中新型的木质纤维素降解酶资源。

[REFERENCES]

- Bao JJ, Pu YB, Ma YH, Zhao QJ, 2019. Identification and analysis of alternative splicing in *Longissimus dorsi* of sheep at different development stages. *Biotechnology Bulletin*, 35(7): 32-38 (in Chinese)
- Black DL, 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry*, 72(1): 291-336
- Boldo JT, do Amaral KB, Junges A, Pinto PM, Staats CC, Vainstein MH, Schrank A, 2010. Evidence of alternative splicing of the *chi2* chitinase gene from *Metarhizium anisopliae*. *Gene*, 462(1-2): 1-7
- Christmas P, Jones JP, Patten CJ, Rock DA, Zheng Y, Cheng

- SM, Weber BM, Carlesso N, Scadden DT, Rettie AE, Soberman RJ, 2001. Alternative splicing determines the function of CYP4F3 by switching substrate specificity. *Journal of Biological Chemistry*, 276(41): 38166-38172
- Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM, 2010. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, 38(6): 1767-1771
- Gehrmann T, Pelkmans JF, Lugones LG, Wösten HA, Abeel T, Reinders MJ, 2016. *Schizophyllum commune* has an extensive and functional alternative splicing repertoire. *Scientific Reports*, 6(1): 1-11
- Gordon SP, Tseng E, Salamov A, Zhang J, Meng X, Zhao Z, Wang Z, 2015. Widespread polycistronic transcripts in fungi revealed by single-molecule mRNA sequencing. *PLoS One*, 10(7): e0132628
- Graveley BR, 2001. Alternative splicing: increasing diversity in the proteomic world. *Trends in Genetics*, 17(2): 100-107
- Guo JZ, Tao HX, Li PF, Li L, Zhang HP, 2018. Identification and analysis of alternative splicing in *Longissimus dorsi* of Jianzhou Da'er goat from an early embryo stage to a neonatal stage. *Acta Agriculturae Boreali-occidentalis Sinica*, 27(3): 316-325 (in Chinese)
- Guo YM, Zheng P, Sun JB, 2010. *Aspergillus niger* as a potential cellular factory: prior knowledge and key technology. *Chinese Journal of Biotechnology*, 26(10): 1410-1418 (in Chinese)
- Jin L, Li G, Yu D, Huang W, Cheng C, Liao S, Zhang Y, 2017. Transcriptome analysis reveals the complexity of alternative splicing regulation in the fungus *Verticillium dahliae*. *BMC Genomics*, 18(1): 1-14
- Kim D, Langmead B, Salzberg SL, 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, 12(4): 357-360
- Langmead B, Salzberg SL, 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4): 357-359
- Li B, Dewey CN, 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12(1): 1-16
- Li SQ, Liu J, Zhang J, Wang XL, Chen D, Wang Y, Xu YM, Huang B, Lin J, Li J, Wang XZ, 2020. Transcriptome profiling reveals the high incidence of *hnRNPA1* exon 8 inclusion in chronic myeloid leukemia. *Journal of Advanced Research*, 24: 301-310
- Liu J, Li C, Wang J, Xu D, Wang H, Wang T, Li L, Li H, Nan P, Zhang J, Wang Y, Huang C, Chen D, Zhang Y, Wen T, Zhan Q, Ma F, Qian H, 2020. Chromatin modifier MTA1 regulates mitotic transition and tumorigenesis by orchestrating mitotic mRNA processing. *Nature Communications*, 11(1): 1-17
- Liu WR, Ding RF, Zhang YM, Li YP, Li L, Guo ZY, 2017. Transcriptome sequencing analysis of hepatocellular carcinoma HepG2 cells induced by antitumor peptide 9R-P201. *Biotechnology Bulletin*, 33(7): 210-215 (in Chinese)
- Loftus BJ, Fung E, Roncaglia P, Rowley D, Amedeo P, Bruno D, Hyman RW, 2005. The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science*, 307(5713): 1321-1324
- Love MI, Huber W, Anders S, 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12): 550
- Mach-Aigner AR, Omony J, Jovanovic B, van Boxtel AJ, de Graaff LH, 2012. D-xylose concentration-dependent hydrolase expression profiles and the function of *CreA* and *XlnR* in *Aspergillus niger*. *Applied and Environmental Microbiology*, 78(9): 3145-3155
- Marquez Y, Brown JWS, Simpson C, Barta A, Kalyna M, 2012. Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. *Genome Research*, 22(6): 1184-1195
- McGuire AM, Pearson MD, Neafsey DE, Galagan JE, 2008. Cross-kingdom patterns of alternative splicing and splice recognition. *Genome Biology*, 9(3): 1-19
- Palusa SG, Ali GS, Reddy AS, 2007. Alternative splicing of pre-mRNAs of *Arabidopsis* serine/arginine-rich proteins: regulation by hormones and stresses. *The Plant Journal*, 49(6): 1091-1107
- Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ, 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, 40(12): 1413-1415
- Pullan ST, Daly P, Delmas S, Ibbett R, Kokolski M, Neiteler A, Archer DB, 2014. RNA-sequencing reveals the complexities of the transcriptional response to lignocellulosic biofuel substrates in *Aspergillus niger*. *Fungal Biology and Biotechnology*, 1(1): 1-14
- Ramani AK, Calarco JA, Pan Q, Mavandadi S, Wang Y, Nelson AC, Fraser AG, 2011. Genome-wide analysis of alternative splicing in *Caenorhabditis elegans*. *Genome Research*, 21(2): 342-348
- Salomonis N, Schlieve CR, Pereira L, Wahlquist C, Colas A, Zambon AC, Conklin BR, 2010. Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. *Proceedings of the National Academy of Sciences*, 107(23): 10514-10519
- Sammeth M, Foissac S, Guigó R, 2008. A general definition and nomenclature for alternative splicing events. *PLoS Computational Biology*, 4(8): e1000147

- Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, Xing Y, 2014. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proceedings of the National Academy of Sciences*, 111(51): E5593-E5601
 - Trapnell C, Pachter L, Salzberg SL, 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 25(9): 1105-1111
 - Vilardell J, Chartrand P, Singer RH, Warner JR, 2000. The odyssey of a regulated transcript. *RNA*, 6(12): 1773-1780
 - Xia H, Chen D, Wu Q, Wu G, Zhou Y, Zhang Y, Zhang L, 2017. *CELF1* preferentially binds to exon-intron boundary and regulates alternative splicing in HeLa cells. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1860(9): 911-921
 - Xie BB, Li D, Shi WL, Qin QL, Wang XW, Rong JC, Song XY, 2015. Deep RNA sequencing reveals a high frequency of alternative splicing events in the fungus *Trichoderma longibrachiatum*. *BMC Genomics*, 16(1): 1-15
 - Xing Y, Lee C, 2006. Alternative splicing and RNA selection pressure—evolutionary consequences for eukaryotic genomes. *Nature Reviews Genetics*, 7(7): 499-509
 - Zhao C, Waalwijk C, de Wit PJ, Tang D, van der Lee T, 2013. RNA-Seq analysis reveals new gene models and alternative splicing in the fungal pathogen *Fusarium graminearum*. *BMC Genomics*, 14(1): 1-16
 - Zhu H, Lu W, Lin H, Ju Z, Liu X, Chen H, 2019. Effects of intron retention on properties of β -glucosidase in *Aspergillus niger*. *Fungal Biology*, 123(6): 465-470
- [附中文参考文献]
- 鲍晶晶, 浦亚斌, 马月辉, 赵倩君, 2019. 绵羊不同发育阶段背最长肌组织中可变剪接的鉴定与分析. *生物技术通报*, 35(7): 32-38
 - 郭家中, 陶海溪, 李鹏飞, 李利, 张红平, 2018. 简州大耳羊早期胎儿到新出生阶段背最长肌中可变剪接的鉴定与分析. *西北农业学报*, 27(3): 316-325
 - 郭艳梅, 郑平, 孙际宾, 2010. 黑曲霉作为细胞工厂: 知识准备与技术基础. *生物工程学报*, 26(10): 1410-1418
 - 刘文荣, 丁若凡, 张一鸣, 李宇鹏, 李玲, 郭志云, 2017. 抗肿瘤多肽 9R-P201 诱导下的肝癌 HepG2 细胞转录组测序分析. *生物技术通报*, 33(7): 210-215