

研究报告

Research Report

基于高通量测序的出芽短梗霉转录组学研究

马赛箭 安超 薛文娇* 上官亦卿

陕西省微生物研究所, 西安, 710043

* 通讯作者, xwenjiao@163.com

摘要 为获得出芽短梗霉合成普鲁兰多糖的转录组信息, 揭示出芽短梗霉生长发育过程中转录组的整体表达特征, 本研究采用 Illumina HiSeq 2500 平台对出芽短梗霉 swp 35 进行高通量测序, 结合生物信息学方法开展基因表达谱研究和功能基因预测。通过测序、组装共得到 29 919 条 Transcript 和 20 312 条 UniGene。所有 UniGene 与 NCBI 的非冗余蛋白质数据库(Nr)进行相似搜索(E 值 $<10^{-5}$) 结果一共有 13 465 个 UniGene 与数据库中的已知基因同源。此外, 还对 UniGene 进行了 GO、COG 和 KEGG 的功能注释、分类及通路分析, 获得一批与普鲁兰多糖合成相关的基因。通过高通量测序, 获得了丰富的出芽短梗霉合成普鲁兰多糖的转录组数据, 为深入研究出芽短梗霉合成普鲁兰多糖的代谢调控提供了有价值的信息。

关键词 高通量测序, 出芽短梗霉, 转录组, 基因注释

Deep Sequencing-based Transcriptome Analysis of *Aureobasidium pullulans*

Ma Saijian An Chao Xue Wenjiao* Shangguan Yiqing

Microbiology Institute of Shaanxi, Xi'an, 710043

* Corresponding author, xwenjiao@163.com

DOI: 10.13417/j.gab.035.000931

Abstract To obtain transcriptome information of *Aureobasidium pullulans*, and reveal the overall expression characteristics of *A. pullulans* transcriptome during growth process. The transcriptome of *A. pullulans* was sequenced by Illumina HiSeq 2500 platform, meanwhile, gene expression profiling and functional gene were predicted by using bioinformatics method. De novo assembly of ESTs data generated 29 919 Transcript and 20 312 UniGene. Similarity searches against the NCBI-Nr database revealed that 13 465 of these UniGenes have significant matches (E -value $<10^{-5}$). In addition, the UniGenes were similarity searched against the GO, COG and KEGG databases, and annotated with gene functional descriptions, gene ontology terms or pathways. Then a number of genes associated with pullulan synthesis were obtained. By high-throughput deeply studying the theoretical basis of sequencing, we obtained abundant transcriptome information, which is very important databases for pullulan biosynthesis pathway and regulation.

Keywords High-throughput sequence, *A. pullulans*, Transcriptome, Gene annotation

出芽短梗霉(*Aureobasidium pullulans*) 是半知菌门、丝孢菌纲、丛梗孢目、暗丛梗孢科、短梗霉属的一种, 其发育过程中有酵母状、菌丝体状、厚垣孢子、芽生孢子等多种细胞形态(Ronen et al., 2002; 付湘晋等, 2005)。出芽短梗霉能以淀粉水解物、蔗糖和葡萄糖等糖类发酵生产胞外多聚糖—普鲁兰多糖, 它具有极佳的成膜性、成纤维性、阻氧性、可塑性、粘结性

和易自然降解等许多独特的理化和生物学特性, 无毒无害, 对人体无任何副作用, 可广泛地应用于农业(朱桂兰等, 2012; 卢星达, 2012)、食品工业(相茂功等, 2008; 周美丽, 2015)、制药工业(刘国军, 2014)和化妆品工业中。除此之外, 经过化学改良的各种普鲁兰衍生物还是很重要的抗病毒、抗血栓、抗血凝固、抗肿瘤药物的工业原料(Suginoshita et al., 2002; 张超等,

2014; Singh et al., 2015)。但由于现有工业化生产菌株产量低、提取成本高、价格昂贵使得实际应用受到了很大的限制。目前报道的研究多集中在普鲁兰多糖的发酵生产(包括发酵条件的研究和突变株的筛选)和生产应用上(Singh et al., 2008; Rajeeva et al., 2010),有关分子结构、产糖机理、分子修饰以及该多糖生产菌株分子遗传背景的研究较少(Cheng et al., 2011)。另外出芽短梗霉的合成途径尚不是完全清楚,而关于出芽短梗霉合成普鲁兰多糖转录组数据也鲜有报道(Kang et al., 2010; 盛龙, 2015)。

近些年来,有关基因组、转录组、蛋白质组、代谢组和代谢通量组等各种组学技术的研究越来越多,也越来越成熟,它们在揭示细胞生理活动规律和生物代谢机理的研究中也起着更加重要的作用(方翔等, 2015; 王松等, 2015),其中转录组学是率先发展起来以及应用最广泛的技术(Lockhart and Winzeler, 2000)。细胞功能是从基因表达开始的,转录组则是某一特定时间和条件下细胞内所有基因转录的RNA总称,通过转录组分析,能够高通量的获得基因表达的RNA信息,并揭示基因表达与某些生命现象之间的内在联系(史硕博等, 2010)。转录组测序也可用于研究基因结构与功能、预测可变剪接和新转录本等,与传统的芯片杂交平台技术相比,具有更广泛的检测范围。对于缺乏基因组信息的物种而言,转录组研究也得到了更广泛的应用(周国艳等, 2013; 贾新平等, 2014a)。

为了获得出芽短梗霉合成普鲁兰多糖的转录组信息,本研究采用Illumina HiSeq 2500平台进行高通量测序,结合生物信息学方法开展基因表达谱研究和功能基因预测,以期能够更多的了解普鲁兰的合成途径及代谢调控机理。

1 结果与分析

1.1 出芽短梗霉转录组数据组装

应用Illumina HiSeq 2500高通量测序技术,对出芽短梗霉 swp35 菌体的转录组进行了测序,共获得15 525 942 reads 片段,包含了4.54 Gb的序列信息,GC%含量平均值是51.57%,将获得的reads片段进行聚类组装得到7 173 509个Contig序列。其中长度为200~300 bp的Contig序列有7 160 014个,占总体的99.81%。对Contig片段组装后共获得29 919个Transcript,序列信息达到39 402 845 bp (39.4 Mb),平均长度为1 317 bp,N50为2 227 bp。其中长度在200~500 bp的Transcript有11 448个,占总体的38.26%;500~1 000 bp的Transcript有4 563个,占总体的15.65%;≥1 000 bp的Transcript有13 908个,占总体的46.48%。对Transcripts序列进行进一步组装,共获得20 312个UniGene,序列信息达到21 109 737 bp,平均长度为1 039.27,N50为1 961 bp。200~500 bp的UniGene有10 175个,占总体的50.09%;500~1 000 bp的UniGene有2 849个,占总体的14.03%;≥1 000 bp的UniGene有7 288个,占总体的35.88%(表1)。

1.2 Unigene 的功能注释

1.2.1 Unigene 的序列比对及功能注释

使用BLAST软件将Unigene序列与NR、Swiss-Port、COG、GO等数据库进行比对,预测Unigene的氨基酸序列,然后使用HMMER软件与Pfam数据库比对,获得UniGene的注释信息。结果表明,在Nr数据库中含有的UniGene相似序列最多,为13 445个,在其他数据库中找到UniGene的相似序列相对较少。本项目通过选择BLAST参数E-value不大于

表1 出芽短梗霉转录组 Contig, Transcript 及 UniGene 数据组装质量统计

Table 1 Quality statistics of assembled Contigs, Transcripts, and UniGene

长度范围 Length range	重叠群 Contig	转录体 Transcript	非冗余的基因数据库 Unigene
200~300	7 160 014 (99.81%)*	7 481 (25.00%)	6 943 (34.18%)
300~500	3 384 (0.05%)	3 967 (13.26%)	3 232 (15.91%)
500~1 000	2 816 (0.04%)	4 563 (15.25%)	2 849 (14.03%)
1 000~2 000	3 934 (0.05%)	6 964 (23.28%)	4 006 (19.72%)
2 000+	3 361 (0.05%)	6 944 (23.21%)	3 282 (16.16%)
Total number	7 173 509	29 919	20 312
Total length	275 926 330	39 402 845	21 109 737
N50 length	40	2 227	1 961
Mean length	38.46	1 316.98	1 039.27

10⁻⁵ 和 HMMER 参数 E-value 不大于 10⁻¹⁰, 最终获得 13 465 个有注释信息的 UniGene, 占比 66.3%, 还有很大比例的未知序列。基因注释的统计结果如下(表 2)。

1.2.2 unigene 的 GO 分类

GO 是一个采用动态更新的标准词汇表来描述基因和其产物功能的数据库, 目前被广泛应用于生物的转录组数据分析研究中, GO 总共分为 3 大功能类, 分别为基因的分子功能(molecular function)、所处的细胞位置(cellular component)和所参与的生物过程(biological process)(曾地刚等, 2013)。应用 GO 数据库对出芽短梗霉的 Unigene 进行功能分类, 初步了解出芽短梗霉表达基因的功能分布特征。研究结果表明, 出芽短梗霉的 Unigene 可划分为 49 个功能组, 其中细胞组分 16 个、分子功能 15 个、代谢进程 18 个(表 3)。其中代谢进程(3 614 个)、催化活性(3 061 个)、细胞进程(2 613 个)、单一有机体进程(2 188 个)和结合活性(2 177 个)功能组中涉及的 UniGene 较多, 而细胞外基质(2 个)、蛋白标记(1 个)和营养库活性类(1 个)功能组中涉及的 UniGene 很少。

1.2.3 COG 功能分类

COG (clusters of orthologous groups)是对基因产物进行同源分类的数据库, 是一个较早的识别直系同源基因的数据库, 通过对多种生物的蛋白质序列大量比较而来的。将所测出芽短梗霉 UniGene 与 COG 数据库进行比对, 结果显示, 出芽短梗霉 UniGene 可分为 25 类, 并对每类的 UniGene 进行了统计分析(图 1)。从图 1 可以看出, 出芽短梗霉 UniGene 所涉及的 COG 功能较为全面涵盖了大部分的生命活动。其中最多的是一般功能预测类基因(R), 然后依次是氨基酸运输和代谢类基因(E)、碳水化合物运输和代谢类基因(G)、矿脂运输和代谢类基因(P)、复制,

表 2 基因注释统计

Table 2 The statistical analysis of annotated gene

注释数据库	基因数据库	≥ 300 nt	≥ 1 000 nt
Annotated databases	Unigene		
COG	4 632	4 203	3 255
GO	5 134	4 179	2 855
KEGG	2 853	2 652	2 102
KOG	6 239	5 590	4 420
Pfam	7 315	6 850	5 474
Swiss-Prot	7 528	7 021	5 556
nr	13 445	10 946	7 113
All	13 465	10 952	7 114

重组和修复类基因(L)较多, 而胞外结构类基因(W)和核结构类基因(Y)数量最少, 其他种类的基因表达量都各有不同。

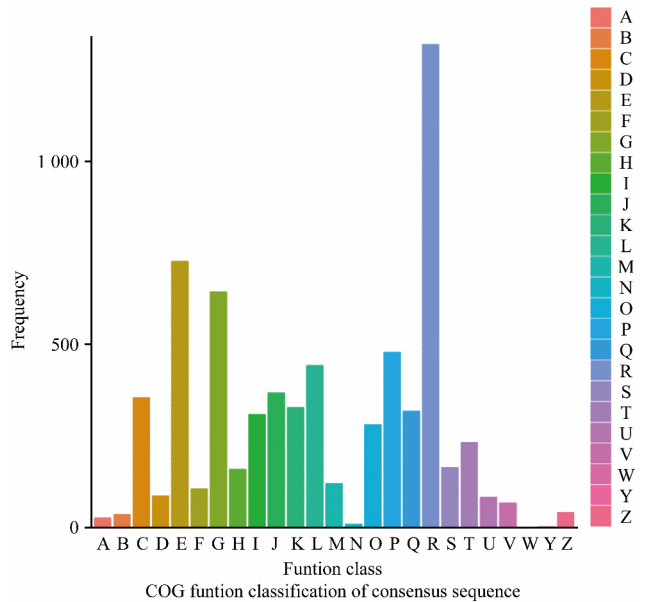


图 1 出芽短梗霉 Unigene 的 COG 功能分类

注: A: RNA 加工和修饰; B: 染色质结构和活力; C: 能量生成和转换; D: 细胞周期控制, 细胞分裂, 染色体分区; E: 氨基酸运输和代谢; F: 核苷酸运输和代谢; G: 碳水化合物运输和代谢; H: 辅酶运输和代谢; I: 脂质运输和代谢; J: 翻译, 核糖体结构和生物发生; K: 转录; L: 复制, 重组和修复; M: 细胞壁 / 膜发生; N: 细胞运动; O: 翻译后修饰, 蛋白质折叠和分子伴侣; P: 矿脂运输和代谢; Q: 次生代谢物合成, 运输和代谢; R: 一般功能预测; S: 未知功能预测; T: 信号传导机制; U: 防卫机制; V: 细胞内转运, 分泌和小泡运输; W: 胞外结构; Y: 核结构; Z: 细胞构架

Figure 1 COG function classification of *Aureobasidium pullulans* Unigenes

Note: A: RNA processing and modification; B: Chromatin structure and vitality; C: Energy generation and transformation; D: Cell cycle control, cell division and chromosome partition; E: Amino acid transport and metabolism; F: Nucleotide transport and metabolism; G: Carbohydrate transport and metabolism; H: Coen zyme transport and metabolism; I: Lipid transport and metabolism; J: Translation, the ribosome structure and biological occurred; K: Transcription; L: Copy, restructuring and repair; M: Cell/membrane; N: Cell movement; O: Modification after translation, protein folding and molecular partner; P: Petrolatum transport and metabolism; Q: Secondary metabolites synthesis, transport and metabolism; R: General function prediction; S: The unknown function prediction; T: Signal transduction mechanism; U: Defense mechanism; V: Transported within cells, secretion and vesicle transport; W: Extracellular structure; Y: Nuclear structure; Z: Cellular architecture

表 3 出芽短梗霉 *Unigene* 的 GO 分类
Table 3 GO functional categories of *Aureobasidium pullulans* Unigenes

细胞组分		分子功能		生物学进程	
Cellular component		Molecular function		Biological process	
GO 分类	基因数据库	GO 分类	基因数据库	GO 分类	基因数据库
GO_classify	A.P.Unigene	GO_classify	A.P.Unigene	GO_classify	A.P.Unigene
胞外成分	88	结合蛋白转录活性	6	繁殖	27
Extracellular region		Protein binding		Reproduction	
		transcription factor activity			
细胞	1 480	结合核酸转录活性	27	免疫系统进程	5
Cell		Protein binding		Immune system process	
		transcription factor activity			
拟核	5	催化活性	3 061	代谢过程	3 614
Nucleoid		Catalytic activity		Metabolic process	
细胞隔膜	1 080	受体活性	8	细胞进程	2 613
Membrane		Catalytic activity		Cellular process	
病毒粒子	8	核苷交换因子活性	5	繁殖过程	36
Virion		Catalytic activity		Reproductive process	
细胞连接	3	结构分子活性	122	生物黏附	3
Cell junction		Structural molecule activity		Biological adhesion	
细胞外基质	2	转运活性	431	信号传导	54
Extracellular matrix		Transporter activity		Signaling	
膜关闭内腔	55	结合活性	2 177	有机体多细胞进程	43
Membrane-enclosed lumen		Binding		Multicellular organismal process	
复杂大分子	533	电子载体活性	83	发育进程	98
Macromolecular complex		Electron carrier activity		Developmental process	
细胞器	1 110	抗氧化活性	33	生长	20
Organelle		Antioxidant activity		Growth	
细胞外基质成分	2	金属伴侣蛋白活性	3	细胞活动	5
Extracellular matrix part		Metallochaperone activity		Locomotion	
胞外成分	7	酶调节活性	28	单一有机体进程	2 188
Extracellular region part		Enzyme regulator activity		Single-organism process	
细胞器成分	469	蛋白标签	1	节律进程	3
Organelle part		Protein tag		Rhythmic process	
病毒体成分	8	营养储藏活性	1	应激反应	319
Virion part		Nutrient reservoir activity		Response to stimulus	
细胞膜成分	710	分子转导活性	24	细胞活动	853
Membrane part		Molecular transducer activity		Localization	
细胞成分	1 483			有机体进程	38
Cell part				Multi-organism process	
				生物调节	377
				Biological regulation	
				细胞成分组织	278
				Cellular component	
				organization or biogenesis	

1.2.4 KEGG 代谢通路分析

KEGG (kyoto encyclopedia of genes and genomes) 是系统分析基因产物在细胞中的代谢途径以及该基

因产物功能的数据库, 利用该数据库有助于把基因及表达信息作为一个整体的网络进行研究。将本实验获得的数据以 KEGG pathway 数据库作为参考, 最

终获得 13 465 个有注释信息的 UniGene。并将检测到的 UniGene 定位到 105 个具体的代谢途径分支,每个代谢通路所涉及到的基因数量如下(表 4)所示。

1.2.5 普鲁兰多糖合成途径关键基因的鉴定

根据目前已有的文献报道,参与普鲁兰多糖合成的关键酶有葡萄糖-6 磷酸变位酶、尿苷二磷酸葡萄糖焦磷酸化酶和葡糖基转移酶。通过转录组测序,并结合生物信息学分析获得了三个酶系的基因序列(表 5),首次从分子生物学水平证实并完善了普鲁兰多糖生物合成途径。

1.2.6 SSR 分析

MISA (MicroSatellite identification tool)是鉴定简单重复序列(simple sequence repeat, SSR)的软件,它可以通过对 UniGene 序列的分析,鉴定出 6 种类型的 SSR。利用 MISA 软件对筛选得到的 7 288 条 1 kb 以上的 UniGene 做 SSR 分析,共检测到 670 个 SSR 位点(表 6)。其中单核苷酸重复为 369 个,达到 55.07%,其次是三核苷酸重复为 226 个,占比 33.73%,在筛选过程中,为确保 SSR 位点的潜在多态性,我们将四核苷酸、五核苷酸和六核苷酸的最小重复次数均设定为 5,一定程度上影响了这几类核苷酸重复在总 SSR 位点中的比例。SSR 的特征分析,有助于开展出芽短梗霉基因组差异分析、通用性标记开发和遗传图谱构建等研究。

2 讨论

随着二代高通量测序技术的成熟和广泛应用,不仅完全改变了过去的研究模式,给人类和动植物基因组学、转录组学、宏基因组学研究方面带来了全新的变化,并且逐步深入到微生物学的研究领域(秦楠等, 2011)。但是目前关于出芽短梗霉基因组的研究还相对较少,转录组研究还处于空白状态。Illumina 高通量测序的特点是数据量大、速度快、成本低、效率高,适合于没有基因组信息的样本开展转录组测序研究(祁云霞等, 2011)。基于转录组学在功能基因研究中的重要地位,本研究通过 Illumina 高通量测序技术对出芽短梗霉转录组进行测序,研究其基因表达谱,挖掘不同代谢产物中的差异表达基因。通过对出芽短梗霉转录组测序共获得 15 525 942 个 reads 序列,对 reads 序列拼接后获得了 7 173 509 个 Contig 序列。将 Contig 序列进行组装得到了 20312 个 UniGene,它们的平均长度为 1039.27 N50 为 1961 bp。N50 值越大,反映组装得到的长片段越多,组装效果

就越好。测序数据产量和数据组装质量是转录组测序完成情况的重要指标(贾新平等, 2014b)。上述结果表明,此次序列组装的质量和长度均可满足转录组分析要求,进一步说明 Illumina HiSeq 2500 是转录组测序的可靠平台。

将出芽短梗霉 UniGene 与 Nr、SwissPort 等数据库进行序列比对,出芽短梗霉的 20 312 条 Unigene 中有 13 465 条 UniGene 与数据库中的其他已知基因具有不同程度的同源性,另外还获得了 6 867 条 UniGene 与数据库中的已知基因同源性较低,占总体的 33.81%,有可能是未知基因。

COG 和 GO 的功能分类对初步了解基因的功能起着重要作用,而 KEGG 数据库中的参考 pathway 不仅可以推测基因的功能,而且可以研究基因在不同代谢通路中所在位置及作用,三者相辅相成,成为新物种中发掘功能基因的重要手段(Wang et al., 2013; 贾新平等, 2014a),基于 KEGG 通路的分析有助于我们进一步了解基因的生物学功能(Kanehisa and Goto, 2000)。通过这些注释,我们能够了解基因的分子功能、所处细胞位置、所参与的生物过程、所处代谢途径或信号通路等等(曾地刚等, 2013),同时我们也找到了普鲁兰多糖合成过程中的 3 个关键酶的基因序列,这为今后进一步挖掘出芽短梗霉发酵生产普鲁兰多糖过程中的重要表达调控基因,开展通过基因或代谢途径调整控制普鲁兰多糖合成的产量及分子量等方面的研究提供了基础数据。

SSR 分子标记具有重复性好、耗时短、成本低、遗传信息量大、共显性遗传等优点,因此在基因定位、基因克隆、功能基因发掘、遗传图谱构建、分子标记辅助育种等研究中得到了广泛应用(Yang et al., 2009; 刘峰等, 2012)。目前,出芽短梗霉可用的 SSR 数量非常有限,转录组数据为开发 SSR 分子标记提供了丰富和极有价值的可利用资源。本研究通过查找发现了 670 个 SSR 位点,SSR 出现频率高且类型丰富。SSR 分子标记的开发有助于出芽短梗霉功能基因的挖掘和遗传多样性分析的研究,并有助于促进真菌类分子生物学的发展。

3 材料与方法

3.1 实验材料

出芽短梗霉 *Aureobasidium pullulans* swp35,由本课题组选育获得,20%甘油管中保藏,存于 -80℃ 低温冰箱。采用 PDA (potato D-glucose agar)培养基

表 4 出芽短梗霉合成普鲁兰多糖的 Unigene 的代谢途径分析
Table 4 Analysis of pullulan synthetic metabolic pathways of *A. pullulans*

代谢通路	通路编码	基因数量	代谢通路	通路编码	基因数量
Pathway	Patway ID	Gene number	Pathway	Pathway ID	Gene number
氧化磷酸化	ko00190	104	甘氨酸, 丝氨酸, 苏氨酸代谢	ko00260	36
Oxidative phosphorylation			Glycine, serine and threonine metabolism		
细胞周期	ko04111	100	半胱氨酸和蛋氨酸代谢	ko00270	36
Cell cycle-yeast			Cysteine and methionine metabolism		
RNA 转运	ko03013	90	甘油磷脂代谢	ko00564	36
RNA transport			Glycerophospholipid metabolism		
剪接体	ko03040	87	丙氨酸, 天门冬氨酸和谷氨酸的代谢	ko00250	35
Spliceosome			Alanine, aspartate and glutamate metabolism		
蛋白质内质网加工	ko04141	85	甲烷代谢	ko00680	34
Protein processing in endoplasmic reticulum			Methane metabolism		
嘌呤代谢	ko00230	84	MAPK 信号转导通路	ko04011	34
Purine metabolism			MAPK signaling pathway-yeast		
核糖体	ko03010	81	谷胱甘肽代谢	ko00480	33
Ribosome			Glutathione metabolism		
淀粉和蔗糖代谢	ko00500	75	蛋白酶体	ko03050	33
Starch and sucrose metabolism			Proteasome		
真核细胞核糖体合成	ko03008	69	基底转录因子	ko03022	32
Ribosome biogenesis in eukaryotes			Basal transcription factors		
减数分裂	ko04113	69	错配修复	ko03430	32
Meiosis-yeast			Mismatch repair		
嘧啶代谢	ko00240	66	吞噬体	ko04145	32
Pyrimidine metabolism			Phagosome		
氨基糖和核苷酸糖代谢	ko00520	61	N-糖链的合成	ko00510	31
Amino sugar and nucleotide sugar metabolism			N-Glycan biosynthesis		
泛素介导的蛋白水解	ko04120	58	TCA 循环	ko00020	30
Ubiquitin mediated proteolysis			Citrate cycle (TCA cycle)		
糖酵解和糖异生途径	ko00010	51	丁酸代谢	ko00650	30
Glycolysis / Gluconeogenesis			Butanoate metabolism		
RNA 降解	ko03018	51	果糖和甘露糖代谢	ko00051	29
RNA degradation			Fructose and mannose metabolism		
DNA 修复	ko03030	49	丙酮酸代谢	ko00620	29
DNA replication			Pyruvate metabolism		
过氧化物酶体	ko04146	47	丙酸代谢	ko00640	28
Peroxisome			Propanoate metabolism		
核苷酸切除修复	ko03420	45	戊糖磷酸途径	ko00030	27
Nucleotide excision repair			Pentose phosphate pathway		
基因监测途径	ko03015	44	各种类型的 N-糖链的合成	ko00513	27
mRNA surveillance pathway			Various types of N-glycan biosynthesis		
精氨酸和脯氨酸代谢	ko00330	42	甘油脂代谢	ko00561	27
Arginine and proline metabolism			Glycerolipid metabolism		
酪氨酸代谢	ko00350	41	核糖核酸聚合酶	ko03020	27
Tyrosine metabolism			RNA polymerase		
			脂肪酸代谢	ko00071	25
			Fatty acid metabolism		

续表 4

Continuing table 4

代谢通路	通路编码	基因数量	代谢通路	通路编码	基因数量
Pathway	Patway ID	Gene number	Pathway	Patway ID	Gene number
氨酰 tRNA 生物合成	ko00970	39	苯丙氨酸代谢	ko00360	25
Aminoacyl-tRNA biosynthesis			Phenylalanine metabolism		
半乳糖代谢	ko00052	38	氰基氨基酸代谢	ko00460	25
Galactose metabolism			Cyanoamino acid metabolism		
缬氨酸, 亮氨酸和异亮氨酸降解	ko00280	38	磷酸肌醇代谢	ko00562	24
Valine, leucine and isoleucine degradation			Inositol phosphate metabolism		
色氨酸代谢	ko00380	37	缬氨酸, 亮氨酸和异亮氨酸的生物合成	ko00290	23
Tryptophan metabolism			Valine, leucine and isoleucine biosynthesis		
内吞作用; 内噬作用	ko04144	37	硒化合物的代谢	ko00450	9
Endocytosis			Selenocompound metabolism		
苯丙氨酸 酪氨酸和色氨酸生物合成	ko00400	23	一个由叶酸碳池	ko00670	9
Phenylalanine, tyrosine and tryptophan biosynthesis			One carbon pool by folate		
核黄素代谢	ko00740	21	花生四烯酸代谢	ko00590	8
Riboflavin metabolism			Arachidonic acid metabolism		
氮代谢	ko00910	21	非同源末端连接	ko03450	8
Nitrogen metabolism			Non-homologous end-joining		
泛酸和辅酶 A 的生物合成	ko00770	20	硫接力系统	ko04122	8
Pantothenate and CoA biosynthesis			Sulfur relay system		
碱基切除修复	ko03410	20	酮体的合成与降解	ko00072	7
Base excision repair			Synthesis and degradation of ketone bodies		
β - 丙氨酸代谢	ko00410	19	昼夜节律	ko04710	7
beta-Alanine metabolism			Circadian rhythm-mammal		
鞘脂类代谢	ko00600	19	抗坏血酸和 aldarate 代谢	ko00053	6
Sphingolipid metabolism			Ascorbate and aldarate metabolism		
乙醛酸盐代谢	ko00630	19	其他多糖降解	ko00511	6
Glyoxylate and dicarboxylate metabolism			Other glycan degradation		
类固醇合成	ko00100	18	维生素 B6 代谢	ko00750	6
Steroid biosynthesis			Vitamin B6 metabolism		
卟啉与叶绿素代谢	ko00860	18	牛磺酸和牛磺酸代谢	ko00430	5
Porphyrin and chlorophyll metabolism			Taurine and hypotaurine metabolism		
赖氨酸降解	ko00310	17	亚油酸代谢	ko00591	5
Lysine degradation			Linoleic acid metabolism		
糖基磷脂酰肌醇(GPI)锚生物合成	ko00563	17	α - 亚麻酸的代谢	ko00592	5
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis			alpha-Linolenic acid metabolism		
硫代谢	ko00920	17	鞘糖脂生物合成系列	ko00603	5
Sulfur metabolism			Glycosphingolipid biosynthesis-globo series		
组氨酸代谢	ko00340	16	烟酸和烟酰胺代谢	ko00760	5
Histidine metabolism			Nicotinate and nicotinamide metabolism		
蛋白输出	ko03060	16	其他类型的 O- 聚糖的生物合成	ko00514	4
Protein export			Other types of O-glycan biosynthesis		
			生物素代谢	ko00780	4
			Biotin metabolism		

续表 4
Continuing table 4

代谢通路	通路编码	基因数量	代谢通路	通路编码	基因数量
Pathway	Pathway ID	Gene_number	Pathway	Pathway ID	Gene number
不饱和脂肪酸的生物合成	ko01040	15	脂肪酸生物合成	ko00061	10
Biosynthesis of unsaturated fatty acids			Fatty acid biosynthesis		
同源重组	ko03440	15	醚酯代谢	ko00565	10
Homologous recombination			Ether lipid metabolism		
磷脂酰肌醇信号系统	ko04070	14	ABC 转运	ko02010	4
Phosphatidylinositol signaling system			ABC transporters		
戊糖, 葡萄糖醛酸转换	ko00040	13	自然杀伤细胞介导细胞毒	ko04650	4
Pentose and glucuronate interconversions			Natural killer cell mediated cytotoxicity		
赖氨酸生物合成	ko00300	13	线粒体脂肪酸延伸率	ko00062	3
Lysine biosynthesis			Fatty acid elongation in mitochondria		
萜类化合物骨架合成	ko00900	13	硫胺素代谢	ko00730	3
Terpenoid backbone biosynthesis			Thiamine metabolism		
囊泡运输中的陷阱相互作用	ko04130	13	硫辛酸代谢	ko00785	3
SNARE interactions in vesicular transport			Lipoic acid metabolism		
泛醌等萜醌的生物合成	ko00130	12	咖啡因的代谢	ko00232	2
Ubiquinone and other terpenoid-quinone biosynthesis			Caffeine metabolism		
叶酸的合成	ko00790	11	青霉素和头孢菌素合成	ko00311	2
Folate biosynthesis			Caffeine metabolism		
自噬调控	ko04140	11	精氨酸和 D- 鸟氨酸代谢	ko00472	1
Regulation of autophagy			D-arginine and D-ornithine metabolism		
			粘多糖的降解	ko00531	1
			Glycosaminoglycan degradation		

表 5 参与普鲁兰多糖合成的关键酶基因
Table 5 The key function gene involved the pullulan synthesis

位置	基因注释	数据来源
Locus	Annoation	Database
Pullulan synthetase		
c3489.graph_c0	UDP-glucose pyrophosphorylase	nr_annotation
c3455.graph_c0	Glucose-6-phosphate isomerase	Swissprot_annotation
c7006.graph_c0	Glycosyltransferase family	Pfam_annotation
c3976.graph_c0	glycosyltransferase family 34 protein	nr_annotation
c9289.graph_c0	Glycosyltransferase like family 2	Pfam_annotation
c12086.graph_c0	Glycosyltransferase family group 2	Pfam_annotation
c13480.graph_c0	Dolichyl-diphosphooligosaccharide-protein glycosyltransferase	nr_annotation
c8944.graph_c0	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase	Swissprot_annotation
c4365.graph_c0	Glycosyltransferase family	Pfam_annotation
c11392.graph_c0	glycosyltransferase family 2 protein	nr_annotation
c11394.graph_c0	Glycosyltransferase family 25 (LPS biosynthesis protein)	Pfam_annotation
c10751.graph_c0	Glycosyltransferase sugar-binding region containing DXD motif	Pfam_annotation
c10493.graph_c0	Glycosyltransferase sugar-binding region containing DXD motif	Pfam_annotation
Degrading enzyme		
c12983.graph_c0	Isopullulanase (Precursor)	Swissprot_annotation

表 6 出芽短梗霉 SSR 不同重复基序分布及优势碱基组成

Table 6 Distribution and compositions of the dominant repeat of the different repeat motifs for SSR

重复基元长度	数量	优势重复基序
Searching item	Number	Advantage repeat motif
单碱基重复	369	A/T
SSR mono nucleotide		
双碱基重复	54	AG/TA
SSR di nucleotide		
三碱基重复	226	TGC/CTT/TCA/CAG
SSR tri nucleotide		
四碱基重复	12	--
SSR tetra nucleotide		
五碱基重复	4	--
SSR penta nucleotide		
六碱基重复	5	--
SSR hexa nucleotide		

活化, 28℃, 培养 5 d。液体培养基(g/L): 蔗糖 50, MgSO₄·7H₂O 0.2, K₂HPO₄ 5.0, (NH₄)₂SO₄ 0.6, 酵母粉 1.7, NaCl 1.0, pH 6.5。通过液体培养基扩大培养, 72 h 后通过离心收集菌体, 使用 50 mmol/L Tris-HCl (pH 7.8) 洗涤菌体两次, 菌体储存到 -80℃ 备用。

3.2 实验方法

3.2.1 RNA 提取

使用 TaKaRa RNA Plus* 试剂盒提取出芽短梗霉菌体的总 RNA。将菌体放入研钵, 加入适量液氮速冻研磨, 将菌体干粉转移到 DEPC 处理的 1.5 mL 离心管中, 加入 1 mL RNAiso Plus 振荡混匀, 静置 5 min; 12 000 r/min, 4℃ 离心 5 min, 将上清转移至新的 1.5 mL 离心管, 加入 200 μL 的氯仿振荡混匀, 室温静置 5 min; 12 000 r/min, 4℃ 离心 15 min, 将上清液移至新的离心管, 加入 1 mL 的异丙醇, 室温静置 10 min; 12 000 r/min, 4℃ 离心 10 min, 除上清, 用 1 mL RNase-free 水配制的 75% 乙醇清洗沉淀; 10 000 r/min, 4℃ 离心 5 min, 弃上清保留沉淀; 自然干燥, 溶解于适量的 DEPC 处理水中, 保存于 -70℃ 备用。用 Agilent 2100 Bioanalyzer 检测 RNA 提取质量, RIN 值 ≥ 7.0。

3.2.2 RNA 文库构建及数据组装

- (1) 用带有 Oligo (dT) 的磁珠富集 mRNA;
- (2) 加入 Fragmentation Buffer 将 mRNA 随机打断;
- (3) 以 mRNA 为模板, 用六碱基随机引物 (random hexamers) 合成第一条 cDNA 链, 然后加入缓冲液、dNTPs、RNase H 和 DNA polymerase 合成第二

条 cDNA 链, 利用 AMPure XP beads 纯化 cDNA;

(4) 纯化后的双链 cDNA 再进行末端修复、加 A 尾并连接测序接头, 然后用 AMPure XP beads 进行片段大小选择;

(5) 最后通过 PCR 富集得到 cDNA 文库;

(6) 文库构建完成后, 分别使用 Qubit2.0 和 Agilent 2100 对文库的浓度和插入片段大小 (insert size) 进行检测, 使用 Q-PCR 方法对文库的有效浓度进行准确定量, 以保证文库质量;

(7) 库检合格后, 用 HiSeq2500 进行高通量测序, 测序读长为 PE125;

(8) 对 Raw Data 进行数据过滤, 去除其中的接头序列及低质量 Reads, 获得高质量的 Clean Data。将 Clean Data 进行序列组装, 获得该物种的 UniGene 库。

3.2.3 功能注释、分类和代谢途径分析

使用 BLAST 软件将 UniGene 序列与 NR、Swiss-Prot、GO、COG、KOG、KEGG 数据库比对 (BLAST 参数 E-value 不大于 10⁻⁵), 预测完 UniGene 的氨基酸序列之后使用 HMMER 软件与 Pfam 数据库比对 (HMMER 参数 E-value 不大于 10⁻¹⁰), 获得 UniGene 的注释信息、分类及 KEGG 代谢网络分析。

3.2.4 SSR 位点搜索及分析

对出芽短梗霉转录组中的 UniGene 序列进行 SSR 位点搜索, 标准为: 单、二、三、四、五、六核苷酸基序最少重复次数分别为 10、8、5、5、5、5, 并对查找到的 SSR 类型进行特征分析。

作者贡献

马赛箭和安超负责数据分析、论文撰写和部分实验工作; 上官亦卿负责出芽短梗霉的发酵及样品处理; 薛文娇负责总体设计和研究指导。

致谢

本研究由西部之光 (2013DF01)、陕西省自然科学基金 (2014JM2-2015)、陕西省科学院科技计划项目 (No. 2013K-06) 和科技计划重点项目 (No. 2014K-01) 共同资助。

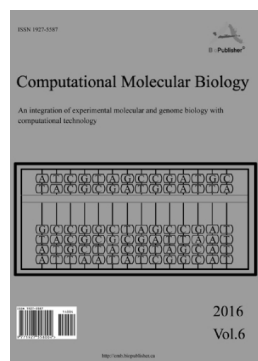
参考文献

- Cheng K.C., Demirci A., and Catchmark J.M., 2011, Pullulan: biosynthesis, production, and applications, *Applied Microbiology and Biotechnology*, 92(1): 29-44
- Fang X., Zeng W.W., Wang Q., Shi C.B., and Wu S.Q., 2015,

- Application of genomics, transcriptomics, proteomics and structural biology in aquatic science, *Dongwu Yuxue Jinzhan (Progress in Veterinary Medicine)*, 36(7): 108-112 (方翔, 曾伟伟, 王庆, 石存斌, 吴淑勤, 2015, 基因组学, 转录组学, 蛋白质组学和结构生物学在水产科学中的应用, *动物医学进展*, 36(7): 108-112)
- Fu X.J., and Tong Q.Y., 2005, A research review of pullulan, *Shipin Yanjiu Yu Kaifa (Food Research and Development)*, 26(6): 16-22 (付湘晋, 童群义, 2005, 短梗霉多糖的研究, *食品研究与开发*, 26(6): 16-22)
- Jia X.P., Ye X.Q., Liang J.L., Deng Y.M., Sun X.B., and She J.M., 2014a, Transcriptome characteristics of *Paspalum vaginatum* analyzed with illumina sequencing technology, *Caoye Xuebao (Acta Prataculturae Sinica)*, 23(6): 242-252 (贾新平, 叶晓青, 梁丽建, 邓衍明, 孙晓波, 余建明, 2014, 基于高通量测序的海滨雀稗转录组学研究, *草业学报*, 23(6): 242-252)
- Jia X.P., Sun X.B., Deng Y.M., Liang L.J., and Ye X.Q., 2014b, Sequencing and Analysis of the of *Asplenium nidus*, *Yuanyi Xuebao (Acta Horticulturae Sinica)*, 41(11): 2329-2341 (贾新平, 孙晓波, 邓衍明, 梁丽建, 叶晓青, 2014, 鸟巢蕨转录组高通量测序及分析, *园艺学报*, 41(11): 2329-2341)
- Kanehisa M., and Goto S., 2000, KEGG: kyoto encyclopedia of genes and genomes, *Nucleic Acids Research*, 28(1): 29-34
- Kang B.K., Yang H.J., Choi N.S., Ahn K.H., Park C.S., Yoon B.D., and Kim M.S., 2010, Production of pure beta-glucan by *Aureobasidium pullulans* after pullulan synthetase gene disruption, *Biotechnology Letters*, 32(1): 137-142
- Liu F., Wang Y.S., Tian X.L., Mao Z.C., Zou X.X., and Xie B.Y., 2012, SSR mining in pepper (*Capsicum annum* L.) transcriptome and the polymorphism analysis, *Yuanyi Xuebao (Acta Horticulturas Sinica)*, 39(1): 168-174 (刘峰, 王运生, 田雪亮, 茆振川, 邹学校, 谢丙炎, 2012, 辣椒转录组 SSR 挖掘及其多态性分析, *园艺学报*, 39(1): 168-174)
- Liu G.J., 2014, Studies on hard capsules material based on pullulan-carrageenan, Thesis for M.S., Jiangnan University, Supervisor: Tong Q.Y., pp.34-45 (刘国军, 2014, 普鲁兰多糖—卡拉胶基硬胶囊材的研究, 硕士学位论文, 江南大学, 导师: 童群义, pp.34-45)
- Lockhart D.J., and Winzeler E.A., 2000, Genomics, gene expression and DNA arrays, *Nature*, 405(6788): 827-836
- Lu X.D., 2012, Screening of pullulanstrains by UV mutagenesis and optimization of the fermentation conditions, Thesis for M.S., Tianjin University of Science & Technology, Supervisor: Wang J.M., pp.28-31 (卢星达, 2012, 产普鲁兰多糖出芽短梗霉的菌种筛选及发酵条件优化, 硕士学位论文, 天津科技大学, 导师: 汪建明, pp.28-31)
- Qi Y.X., Liu Y.B., and Rong W.H., 2011, RNA-seq and its applications: a new technology for transcriptomics, *Yichuan (Hereditas)*, 33(11): 1191-1202 (祁云霞, 刘永斌, 荣威恒, 2011, 转录组研究新技术: RNA-seq 及其应用, *遗传*, 33(11): 1191-1202)
- Qin N., Li D.F., and Yang R.F., 2011, Next-generation sequencing technologies and the application in microbiology—a review, *Weishengwuxue Tongbao (Acta Microbiologica Sinica)*, 51(4): 445-457 (秦楠, 栗东芳, 杨瑞馥, 2011, 高通量测序技术及其在微生物学研究中的应用, *微生物通报*, 51(4): 445-457)
- Rajeeva G., Ranjan S., Monika G., and Manogya K.G., 2010, *Aureobasidium pullulans*, an economically important polymorphic yeast with special reference to pullulan, *African Journal of Biotechnology*, 47(9): 7989-7997
- Ronen M., Guterman H., and Shabtai Y., 2002, Monitoring and control of pullulan production using vision sensor, *Journal of Biochemical and Biophysical Methods*, 51(3): 243-249
- Sheng L., 2015, The study of key factors and mechanism of pullulan biosynthesis, Dissertation for Ph.D., Jiangnan University, Supervisor: Tong Q.Y., pp.15-17 (盛龙, 2015, 普鲁兰多糖生物合成的关键影响因素及其机理研究, 博士学位论文, 江南大学, 导师: 童群义, pp.15-17)
- Shi S.B., Chen T., and Zhao X.M., 2010, Transcriptome platforms and applications to metabolic engineering, *Shengwu Gongcheng Xuebao (Journal of Biotechnology)*, 26(9): 1187-1198 (史硕博, 陈涛, 赵学明, 转录组平台技术及其在代谢工程中的应用, *生物工程学报*, 26(9): 1187-1198)
- Singh R.S., Kaur N., and Kennedy J.F., 2015, Pullulan and pullulan derivatives as promising biomolecules for drug and gene targeting, *Carbohydrate Polymers.*, 123(6): 190-207
- Singh R.S., Saini G.K., and Kennedy J.F., 2008, Pullulan: microbial sources, production and applications, *Carbohydrate Polymers.*, 73(4): 515-531
- Suginoshita Y., Tabata Y., Mataumura T., Toda Y., Nabeshima M., and Moriyasu F., 2002, Liver targeting of human interferon-beta with pullulan based on metal Coordination, *Journal of Controlled Release Official*, 83(1): 75-88
- Wang R., Xu S., Jiang Y., Jiang J.W., Li X.D., He J., Peng F., and Xia B., 2013, De novo sequence assembly and characterization of *Lycoris aurea* transcriptome using GS FLX Titanium platform of 454 pyrosequencing, *PLoS One*, 8(4): 60449-60459
- Wang S., Liang H.Z., Jia C.F., and Zhang B.L., 2015, Comparative genomics analysis of carbon metabolism of *Lactobacillus casei* 12A, *Weishengwuxue Tongbao (Microbiology)*, 42(2): 315-324 (王松, 梁宏彰, 贾春凤, 张柏林, 2015, 干酪乳杆菌 12A 碳源代谢的比较基因组学分析, *微生物学通报*, 42(2): 315-324)
- Xiang M.G., Deng D.J., Guo X.P., Wang F.S., and Zhu X.Q., 2008, Application of Pullulan in food industry, *Shipin Yu*

- Yaopin (Food and Drug), 10(1): 67-69 (相茂功, 邓长江, 郭学平, 王凤山, 朱希强, 2008, 普鲁兰糖在食品工业中的应用进展, 食品与药品, 10(1): 67-69)
- Yang H.B., Liu M.Y., Kang W.H., John M., and Kang B.C., 2009, Development of SNP markers linked to the L locus in *Capsicum* spp. by a comparative genetic analysis, *Molecular Breeding*, 24(4): 433-446
- Zeng D.G., Cheng X.L., Xie D.X., Zhao Y.Z., Yang C.L., Ma N., Li Y.M., and Chen X.H., 2013, Deep sequencing-based transcriptome analysis of *lipopenaeus vannamei*, *Jiyinzuxue Yu Yingyong Shengwuxue (Genomics and Applied Biology)*, 32(3): 308-313 (曾地刚, 陈秀荔, 谢达祥, 赵永贞, 杨春玲, 马宁, 李咏梅, 陈晓汉, 2013, 基于高通量测序的凡纳滨对虾的转录组分析, 基因组学与应用生物学, 32(3): 308-313)
- Zhang C., Yang J., Xiao B., Zhang L.H., Wang H., Sun H.F., Kong D.L., and Ma G.L., 2014, Study on cellular uptake mechanism and subcellular distribution of cholesterol-modified pullulan nanoparticles, *Guoji Shengwu Yixue Gongcheng Zazhi (International Journal of Biomedical Engineering)*, 37(4): 204-209 (张超, 杨菁, 肖宝, 张琳华, 王海, 孙洪范, 孔德领, 马桂蕾, 2014, 胆固醇基修饰的普鲁兰纳米粒的摄取机制及亚细胞分布研究, 国际生物医学工程杂志, 37(4): 204-209)
- Zhou G.Y., Hu W.X., Xu J.H., and Xue Q.Z., 2013, Integrating multiple "omics" analysis for plant metabolites and their functions, *Zhejiang Daxue Xuebao (Nongye Yu Shengming Kexueban) (Journal of Zhejiang University (Agric.&Life Sci.))*, 39(3): 237-245 (周国艳, 胡望雄, 徐建红, 薛庆中, 2013, 整合多个组学(omics)分析植物代谢产物及其功能, 浙江大学学报(农业与生命科学版), 39(3): 237-245)
- Zhou M.L., 2015, Cellulose enhanced chitosan/pullulan packaging materials, *Baozhuang Gongcheng (Packaging Engineering)*, 36(11): 61-64 (周美丽, 2015, 纤维素增强壳聚糖/普鲁兰多糖包装材料的研究, 包装工程, 36(11): 61-64)
- Zhu G.L., and Tong Q.Y., 2012, Research progress on microbial polysaccharide, *Shipin Gongye Keji (Science and Technology of Food Industry)*, 33(6): 444-448 (朱桂兰, 童群义, 2012, 微生物多糖的研究进展, 食品工业科技, 33(6): 444-448)

Computational Molecular Biology (CMB)



Computational Molecular Biology (ISSN 1927-5587) is an open access, peer reviewed journal published online by BioPublisher. The Journal is publishing all the latest and outstanding research articles, letters, methods, and reviews in all areas of Computational Molecular Biology, covering new discoveries in molecular biology, from genes to genomes, using statistical, mathematical, and computational methods as well as new development of computational methods and databases in molecular and genome biology. The papers published in the journal are expected to be of interests to computational scientists, biologists, and teachers / students / researchers engaged in biology, as well as are appropriate for R & D personnel and general readers interested in computational technology and biology.

Email: edit@cmb.biopublisher.ca

Web: <http://cmb.biopublisher.ca>