

学校编码: 10384

密级_____

学 号: 33320130153903

厦 门 大 学

博 士 学 位 论 文

选择性多聚腺苷化在水稻发育过程中的基因表达调控作用以及适应性的研究

**The Role of Messenger RNA Alternative Polyadenylation
During Developmental Gene Expression Regulation and
Adapdation of Rice**

付海辉

指导教师姓名: 李庆顺 教授

专 业 名 称: 环 境 科 学

论文提交日期: 2016 年 6 月

论文答辩时间: 2016 年 7 月

2016 年 7 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下，独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为（ ）课题（组）的研究成果，获得（ ）课题（组）经费或实验室的资助，在（ ）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）：

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

() 1.经厦门大学保密委员会审查核定的保密学位论文，于 年 月 日解密，解密后适用上述授权。

() 2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

目录

摘要	I
Abstract.....	III
缩略语表	IV
第 1 章 绪论	1
1.1 水稻的发育	2
1.1.1 水稻的发育过程.....	2
1.1.2 水稻发育的转录组研究现状.....	3
1.2 多聚腺苷化的研究现状	4
1.2.1 什么是多聚腺苷化?	4
1.2.2 选择性多聚腺苷化.....	5
1.2.3 选择性多聚腺苷化影响基因表达的主要方式.....	7
1.2.4 选择性多聚腺苷化与植物发育.....	7
1.3 高通量测序检测 poly(A)位点的方法	8
1.4 研究内容和研究意义	9
1.4.1 研究内容.....	9
1.4.2 研究意义.....	11
1.4.3 技术路线.....	13
第 2 章 日本晴发育过程中选择性多聚腺苷化动态变化以及对基 因表达调控的影响	14
2.1 前言	14
2.2 材料与方法	15

2.2.1 样品采集.....	15
2.2.2 样本总 RNA 提取.....	15
2.2.3 样本总 RNA 纯化.....	16
2.2.4 样本 PAT-seq 文库的构建.....	16
2.2.5 样本文库质量检测.....	19
2.2.6 样本文库的 Illumina Hi-seq 2000 测序.....	19
2.2.7 基于 PAT-seq 检测水稻发育过程中不同组织的 poly(A)位点.....	20
2.2.8 基于表达序列标签的数据验证本实验日本晴的 poly(A)位点.....	21
2.2.9 日本晴发育过程中不同组织的 poly(A)位点表达模式分析.....	21
2.2.10 日本晴 14 个组织样本的 3'UTR 长度分析.....	22
2.2.11 日本晴 14 个组织样本中选择性多聚腺苷化位点变换基因的分析.....	22
2.2.12 日本晴 14 个组织样本的 poly(A)位点附近单核苷酸的轮廓分析.....	22
2.2.13 日本晴发育过程中不同时期 APA 位点变换基因的验证.....	23
2.2.14 基因功能富集分析.....	25
2.2.15 利用统计学的方法检验 APA 基因和 QTLs 的相关性.....	25
2.2.16 PAT-seq 和 RNA-seq 数据的相关性分析.....	26
2.3 结果.....	26
2.3.1 日本晴发育过程中发现大量基因有多个 poly(A)位点.....	26
2.3.2 PAT-seq 所取得的 poly(A)位点数据的验证.....	32
2.3.3 日本晴发育过程中 14 个组织的 PAC 表达模式特征.....	33
2.3.4 日本晴发育过程中组织特异性 PAC 分析.....	36
2.3.5 日本晴发育过程中相关基因的 3'UTR 长度的动态变化.....	38
2.3.6 日本晴发育过程中不同组织不同时期存在许多 APA 位点变换基因.....	40
2.3.7 日本晴发育早期基因倾向于远端 poly(A)位点, 后期倾向于近端 poly(A)位点.....	44
2.3.8 利用已经发表的 RNA-seq 数据证明 APA 变换基因表达水平发生显著的变化.....	47
2.3.9 日本晴 14 个组织样本 poly(A)位点上下游单核苷酸的特征.....	51
2.3.10 花粉 poly(A)位点的独特性可能与多聚腺苷化因子的表达丰度变化有	

关.....	52
2.3.11 RNA-seq 数据验证 PAT-seq 数据能够很好的反映基因的表达水平 ...	54
2.3.12 日本晴发育过程中差异表达的 PAC 与差异表达的基因是相互独立的生物学事件.....	55
2.3.13 日本晴发育过程中几个重要时期相同组织 PAC 表达模式特征.....	57
2.3.14 日本晴发育过程中花器官的不同组织 PAC 表达模式特征.....	61
2.3.15 日本晴发育过程中 APA 基因高度地富集于 QTLs	65
2.3.16 低温耐受因子 OsDREB1F 的 APA 可能赋予日本晴对低温环境的适应性.....	69
2.4 讨论	70
第3章 93-11 发育过程中选择性多聚腺苷化动态变化以及对基因表达调控的影响	76
3.1 前言	76
3.2 材料与方法	77
3.2.1 样品采集.....	77
3.2.2 93-11 样本处理方法与数据分析	77
3.2.3 93-11 发育过程中不同时期 APA 位点变换基因的验证	78
3.3 结果	80
3.3.1 93-11 发育过程中也有大量基因有多个 poly(A)位点.....	80
3.3.2 93-11 发育过程中 PAC 表达模式特征	85
3.3.3 93-11 发育过程中组织特异性 PAC 很少	87
3.3.4 93-11 发育过程中相关基因的 3'UTR 长度动态变化	90
3.3.5 93-11 发育过程中也存在许多 APA 位点变换基因	92
3.3.6 93-11 发育过程中早期基因倾向于远端 poly(A)位点, 晚期倾向于使用近端 poly(A)位点	94
3.3.7 93-11 中 14 个组织样本 poly(A)位点上下游单核苷酸的特征.....	98
3.3.8 93-11 发育过程中 PAC 的表达与基因的表达是相互独立的生物学事件	

.....	99
3.3.9 93-11 发育过程中几个重要时期相同组织 PAC 表达模式特征	101
3.3.10 93-11 发育过程中花器官的不同组织 PAC 表达模式特征	105
3.4 讨论	109
第 4 章 结论与展望	113
4.1 结论	113
4.2 特色与创新点	114
4.3 不足与展望	115
参考文献	117
附录	124
致谢	125

厦门大学博硕士学位论文摘要库

Content

Abstract(in Chinese).....	I
Abstract(in English)	III
Abbreviation	IV
Chapter 1 Introduction.....	1
1.1 The development of rice.....	2
1.1.1 The development process of rice	2
1.1.2 Research progress of transcriptome in the development of rice	3
1.2 Research progress of polyadenylation	4
1.2.1 What is polyadenylation?	4
1.2.2 Alternative polyadenylation.....	5
1.2.2 The main model that APA impact on gene expression regulation	7
1.2.3 Alternative polyadenylation and the growth of plant	7
1.3 Method for detecting poly(A) site using high throughput sequencing	8
1.4 Contents and significant.....	9
1.4.1 Contents	9
1.4.2 Significance.....	11
1.4.3 Technical routes	13
Chapter 2 The dynamic change of APA in the development of <i>Japonica</i> Nipponbare and APA affect gene expression regulation ...	14
2.1 Introduction.....	14

2.2 Material and method	15
2.2.1 Sample collected	15
2.2.2 Total RNA isolation	15
2.2.3 Total RNA purified.....	16
2.2.4 PAT-seq library constructed	16
2.2.5 PAT-seq library test	19
2.2.6 Illumina Hi-seq 2000 sequencing	19
2.2.7 The poly(A) site of the different tissues in the development of rice was detected using PAT-seq.....	20
2.2.8 Poly(A) site of japonica was identified by EST data.....	21
2.2.9 The expression model of poly(A) site in the development of rice.....	21
2.2.10 Analyses of 3'UTR length across 14 tissues	22
2.2.11 Analyses of APA switching gene in the development of rice	22
2.2.12 Single nucleotide profile across 14 samples	22
2.2.13 APA switching gene confirmation	23
2.2.14 Analyses of gene functional enrichment.....	25
2.2.15 The analysis of APA gene and QTLs using stastical method.....	25
2.2.16 Correlation between PAT-seq and RNA-seq	26
2.3 Results.....	26
2.3.1 Many genes had several poly(A) sites in the development of rice	26
2.3.2 PAT-seq data identifiedlots of poly(A) sites of EST data	32
2.3.3 PAC expression model across 14 samples.....	33
2.3.4 Analysis of tissue specific PAC in the development of rice.....	36
2.3.5 The dynamic change of 3'UTR length in the development of rice.....	38
2.3.6 Many APA switching gene occurred in the development of rice	40
2.3.7 Gene tends to use distalpoly(A) site in the early development stage of rice while gene tends to use proximal poly(A) site in the later stage	44
2.3.8 APA switching that changes gene expression level was identified by published RNA-seq	47
2.3.9 Characteristic of single nucleotide profile around PAC in the growth of rice....	51

2.3.10 Specificity of intron in mature pollen may be associated with the change of expression level of poly(A) site factor	52
2.3.11 PAT-seq data was able to reflect gene expression level	54
2.3.12 DE PAC and DE gene were two independent biological events	55
2.3.13 PAC expression pattern in the developmental important stage of rice	57
2.3.14 PAC expression pattern in the flower of rice	61
2.3.15 QTLs tends to use APA gene in the development of rice	65
2.3.16 Cold stress response factor (<i>OsDREB1F</i>) that occurred APA in <i>Japonica</i> Nipponbare may increase cold tolerance	69
2.4 Discussion	70

Chapter 3 The dynamic change of APA in the development of <i>indica</i> 93-11 and APA affect gene expression regulation.....	76
3.1 Introduction	76
3.2 Material and method	77
3.2.1 Sample collected	77
3.2.2 Sample treatment and bioinformatic analyses	77
3.2.3 APA switching gene confirmation.....	78
3.3 Results	80
3.3.1 Many genes had several poly(A) sites in the development of 93-11	80
3.3.2 PAC expression model across 14 samples of 93-11	85
3.3.3 Tissue specific PAC in the development of rice was a few	87
3.3.4 The dynamic change of 3'UTR length in the development of 93-11	90
3.3.5 Many APA switching gene occurred in the development of 93-11	92
3.3.6 Gene tends to use distal poly(A) site in the early development stage of <i>indica</i> while gene tends to use proximal poly(A) site in the later stage	94
3.3.7 Characteristic of single nucleotide profile around PAC in the growth of 93-11	98
3.3.8 DE PAC and DE gene were two independent biological events	99

3.3.9 PAC expression pattern in the developmental important stage of 93-11.....	101
3.3.10 PAC expression pattern in the flower of 93-11	105
3.4 Discussion	109
Chapter 4 Conclusion and prospect.....	113
4.1 Conclusion	113
4.2 Innovation of the paper	114
4.3 Defect and prospect.....	115
Reference	117
Appendix	124
Acknowledgements	125

厦门大学博士论文摘要库

摘要

水稻 (*Oryza sativa*) 不仅是世界上最重要的粮食作物之一, 也是单子叶模式植物。水稻品种 japonica Nipponbare 和 indica 93-11 为两个水稻亚种代表, 广泛的种植于我国的南北区域, 对环境与气候形成一定的适应性, 在我国形成北粳南籼的生产格局。它们各自具有独自的生理特性和生长周期。转录组为特定生理条件下, 细胞中全部转录产物的集合。通过对水稻发育过程转录组的研究, 可以全面地了解基因表达调控在水稻发育过程中的调节作用以及对环境适应性。多聚腺苷化是基因在转录形成成熟的转录本之前的加工过程之一, 即在转录本的 3' 末端适当位点进行切割并加上 50-250 个腺苷酸残基, 即 poly(A) 尾巴, 通常也是产生一个成熟的转录本的最后一步。Poly(A) 在基因的转录后调控过程中具有重要的生物学作用, 比如决定转录子细胞核外运、参与蛋白质翻译起始复合物的形成、以及防止 mRNA 降解。水稻发育过程的转录组一直都是国内外学者研究的热点, 但是利用高通量测序技术全基因组范围内对水稻发育过程中相关基因的 3' 末端的信息研究的非常少。

本论文利用 PAT-seq (Poly(A) tag sequence) 高通量测序技术研究了水稻两个模式品系日本晴和 93-11 发育过程中不同时期不同组织中基因的选择性多聚腺苷化 (Alternative polyadenylation, 或 APA) 动态变化以及多聚腺苷化信号使用情况, 同时利用统计学方法对日本晴数量性状基因 (Quantitative trait loci, 或 QTL) 的表达和 APA 的关系进行阐述, 最后对 93-11 和日本晴形成南北适应性的分子机制进行探讨。通过全面而系统的了解两个水稻亚种发育过程中选择性多聚腺苷化动态变化以及对基因表达调控的影响, 并深入探索 APA 在水稻发育过程中的生物学作用, 为理解水稻发育过程中基因表达转录后调控提供新的认识, 同时为育种工作提供参考。

主要结果如下:

(1) 利用 PAT-seq 测序技术对水稻日本晴发育过程中不同时期不同组织共 14 个样本的转录本进行 3' 末端测序, 共得到 136,955,352 个 PAT, 聚成 68,220 个 PAC (Poly(A) site cluster), 分布在 28,032 个基因。其中, 有 13,419 个基因 (47.9%) 有 2 个或者 2 个以上的 PAC, 说明在水稻日本晴发育过程中广泛发生 APA 现象。

在日本晴水稻花粉中发现最多组织特异性的 PAC, 其中有大约一半来源于表达基因的内含子区域。进一步分析发现, 这些组织特异性的 PAC 富集在能量代谢、重要氨基酸代谢以及防御等通路之中。主成分分析 (Principal component analysis, 或 PCA) 发现日本晴花粉中具有明显不同的 PAC 表达模式, 并且发现花粉表达的基因内含子区域的 PAC 单核苷酸轮廓和 PAS (Poly(A) signal) 与其他组织明显不同, 进一步研究发现日本晴花粉的多聚腺苷化因子的表达水平与其他组织 (比如 20 天的叶) 也不同。这暗示水稻花粉的多个生物学事件与其他组织的不同可能与花粉的多聚腺苷化因子的表达水平发生变化有关。在水稻日本晴发育过程中发现许多 APA 位点变换基因, 日本晴发育的早期阶段基因倾向于使用远端的 PAC, 后期倾向使用近端的 PAC, 同时发现 APA 变换现象改变了基因的表达水平。我们还发现差异表达的 PAC 和差异表达的基因是两个独立的生物学事件, 两类基因富集的生物学过程也基本不同。通过统计学方法发现, QTL 倾向于使用 APA 基因。QTL 可以在不增加基因数量的前提下通过使用 APA 基因从而达到对 QTL 的微调。同时发现, 与 QTL 关联的高表达基因的 PAC 主要分布在植株高度、抽穗日数、穗数、千粒重、分蘖数、根干重、籽粒产量、穗长和穗数等, 暗示 APA 在调控水稻农艺性状方面的潜能。通过比较两个水稻亚种的 PAT-seq 数据, 发现冷胁迫关键应答因子 *OsDREB1F* 在日本晴中发生了 APA 现象, 而在 93-11 中没有发生 APA 现象, 可能与日本晴耐受相对低温环境而适合北方种植有关。

(2) 利用同日本晴相同的方法, 对籼稻品系 93-11 的发育过程中不同时期不同组织共 14 个样本的转录本进行 3' 末端测序, 共得到 137,398,368 个 PAT, 聚成 63,502 个 PAC, 分布在 24,786 个基因中。其中, 有 10,724 (42%) 个基因有 2 个或者 2 个以上的 PAC, 说明在 93-11 发育过程中也广泛发生 APA 现象。在 93-11 花粉中也发现最多组织特异性的 PAC, 其中有大约一半来源于表达基因的内含子区域。这些组织特异性的 PAC 富集在一些杂环化合物分解代谢相关的生物学过程, 暗示组织特异性的 PAC 参与了花粉中的一些防御性物质的生成。聚类分析发现 93-11 花粉中也具有明显不同的 PAC 表达模式, 但是花粉内含子区域的 PAC 单核苷酸轮廓和 PAS 与其他组织相似, 这一点不同于日本晴。在 93-11 发育过程中也发现许多 APA 位点变换基因, 早期阶段基因也倾向于使用远端的

PAC，后期阶段也倾向使用近端的 PAC，也发现 APA 变换现象改变了基因的表达水平。研究发现，差异表达的 PAC 和差异表达的基因是两个独立的生物学事件，两类基因富集的生物学过程也基本不同。

关键词：水稻；日本晴；93-11；信使 RNA；发育；多聚腺苷化；选择性的多聚腺苷化；3'末端形成；基因表达调控；转录后调控

厦门大学博硕士论文摘要库

Abstract

Rice is not only one of the most important food crops in the world, but also is the model plant of monocots. Two subspecies of rice (represented by cultivars *japonica* Nipponbare and *indica* 93-11) were widely planted throughout China's north and south regions, and it had certain adaptability on the environment and climates. Transcriptome is a collection of all transcripts in cells under specific physiological conditions. Through transcriptome analysis during developmental process of rice, we could understand the role of gene expression regulation. Polyadenylation is an important biological event, where pre-mRNA is cleaved and was added to a poly(A) tail in the 3'UTR post-transcriptionally. Poly(A) plays its role in mRNA nuclear export to cytoplasm, mRNA stability, and the formation of a translational initiation complex. However, the 3'-end information of genes about rice is scarce. Most previous studies did not reveal the poly(A) sites at the nucleotide level. Hence, we used a poly(A) tag sequencing approach (PAT-seq) to investigate the genome-wide landscape of alternative polyadenylation (APA) and poly(A) signals (PAS) in 14 different tissues and developmental stages of both rice subspecies, Japonica and Indica, and revealed the relationship of APA and QTL (Quantitative trait loci) using statistical methods. We comprehensively and systematically studied the dynamic changes of the APA in the development of rice and the effect of APA in the regulation of gene expression. We further investigated the role of APA in the growth of rice, which will provide for new view for understanding the post-transcriptional regulation of the development of rice and provide potential basis for future breeding work.

(1) To study the 3'-end of genes during the different developmental stages and tissues of rice, we used a poly(A) tags sequencing (PAT-seq) protocol and identified 136,955,352 PATs and 68,220 poly(A) site clusters (PACs), which were dispersed in 28,032 genes. Of these genes, 13,419 genes (47.9%) had more than 2 PACs. It was suggested that extensive APA event occurred in the growth of rice. PACs in the mature pollen were the most and nearly half PACs were from the introns, which were

enriched in the following pathways, including energy production, defense and core amino acid of pollen. Principal component analysis (PCA) and cluster analyses found that the expression model of PACs in mature pollen was different from other tissues and the single nucleotide profile around PAS in mature pollen were also distinguished from other tissues. To probe into the reason why of such differences, the expression levels of several poly(A) factors in mature pollen were found very different from other tissue. This is indicative of the change poly(A) site selection is associated with the change in the expression level of these poly(A) factors. We further found that many APA switching happened during development of rice, and many genes tend to use distal PAC in the early developmental stages of rice while tend to use proximal PAC in later stages, and APA site switching was able to change the expression level of genes. Meanwhile, we found that differentially expressed (DE) PAC and DE genes were two independent biological events in the growth of rice and they were enriched in different biological processes. Through statistical analysis, we found that QTL tend to use APA genes. This may be associated with QTL controlled by many genes, and thus QTL can be finely regulated using APA. Highly expressed PACs were mainly enriched in the following QTLs, including plant height, days to heading, spikelet number, 1000-seed weight, root dry weight, tiller number, panicle number, panicle length and grain yield, indicated that the potential role of APA in the determination of important agricultural traits. By comparing the PAT-seq data of two sub-species of rice, we found that a cold stress response factor, *OsDREB1F*, use APA in *Japonica* Nipponbare but not in *Indica* 93-11, indicating that APA may be involved in post-transcriptional regulation of the genes and potentially enhance low temperature tolerance.

(2) To study the 3'-end of genes during the different developmental stages and tissues of 93-11 using PAT-seq, we obtained 137,398,368 PATs and 63,502 PACs, dispersed in 28,032 genes. Of these genes, 10,724 genes (42%) had more than 2 PACs. This result demonstrated that extensive APA also occurred in the development of indica rice. Tissue specific PAC in mature pollen was the most and half PAC was from intron of gene the same as *japonica* Nipponbare, which was mainly enriched in the

Degree papers are in the “[Xiamen University Electronic Theses and Dissertations Database](#)”.

Fulltexts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.