

水稻基因组 MSAP 指纹图谱构建及 DNA 甲基化修饰位点分离与鉴定^①

张 勇^{②*} 邓科君^{③*} 张 韬* 杨足君* 彭金华* 周建平* 任正隆^{④*}

(* 电子科技大学生命科学与技术学院 成都 610054)

(** 四川农业大学植物遗传育种省级重点实验室 雅安 625014)

摘 要 采用 *EcoR* I 和 *Hpa* II/*Msp* I 双酶切建立了适合于水稻基因组的甲基化敏感扩增多态性(MSAP)分析体系,在全基因组水平检测了水稻 DNA 甲基化修饰位点。以 12 对 MSAP 引物进行选择性地扩增,共检测到甲基化修饰位点 120 个,“CCGG/GGCC”位点甲基化修饰比例为 20.17%。对部分水稻基因组甲基化修饰位点进行回收,最终分离了 55 条存在甲基化位点变异的 DNA 序列,通过 BLAST 比对分析将其联配到水稻基因组序列上。分析表明,这些甲基化修饰位点主要集中于基因启动子区(47%)和第一外显子区(22%),在其侧翼序列中存在类似“CpG island”典型序列特征的“CpG”二核苷酸成簇富集区。在此基础上,对应用 MSAP 技术分离水稻基因组 DNA 甲基化修饰位点的有效性以及水稻基因组序列中“CpG island”类似序列分布特征和生物意义进行了讨论。

关键词 水稻,表观遗传, DNA 甲基化,甲基化敏感扩增多态性(MSAP), CpG 岛

0 引 言

表观遗传学(epigenetics)是指在 DNA 序列没有发生变异的情况下,基因表达的可遗传改变,这种改变又可随细胞的有丝分裂和(或)减数分裂遗传下去的现象^[1]。作为表观遗传学研究的主要内容,DNA 甲基化是指在 DNA 甲基转移酶催化下,将甲基转移到胞嘧啶 5' 位置上的过程。DNA 的甲基化现象涉及许多生命科学的基础研究和应用领域,与基因表达调控、胚胎发育、细胞分化、基因组印迹、性染色体失活等诸多生物学过程密切相关^[1-4],是当今表观遗传学研究的活跃领域。

水稻是半数世界人口赖以生存的主要食物,随着其全基因组序列草图及精细测序的完成,可以解码大量与水稻生长、发育相关的遗传信息。但到目前为止,水稻基因组中蕴藏的表观遗传信息却未得到研究者的足够重视,对于水稻基因组 DNA 甲基化修饰状况及模式、基因表达表观调控等知之甚少。本研究选用粳稻日本晴(*Oryza sativa* ssp. *japonica*)为研究材料,采用 *EcoR* I 和 *Hpa* II/*Msp* I 双酶切建立适合于水稻基因组的甲基化敏感扩增多态性

(methylation-sensitive amplification polymorphism, MSAP)技术体系,在全基因组水平检测 DNA 甲基化修饰位点。针对目标位点,预测、鉴定功能性甲基化修饰位点,解析 DNA 甲基化对相关基因表达的潜在调控机制,为水稻表观遗传学研究积累理论依据。

1 材料与方法

1.1 材料

粳稻日本晴(*Oryza sativa* ssp. *japonica* cultivar Nipponbare)由四川农业大学水稻研究所提供。

1.2 DNA 提取

选新鲜、饱满籽粒,以 70% 乙醇浸种消毒,置于湿润培养皿中培养,10 天后取幼叶用于 DNA 提取, DNA 提取详见文献[5]报道。

1.3 MSAP 分析

EcoR I 接头及引物序列以 Vos 等^[6]序列为参照,*Hpa* II(*Msp* I)接头及引物序列以 Xiong 等^[7]序列为参照。MSAP 分析流程基本参照 Vos 等^[6]的扩增片段长度多态性(amplified fragment length polymorphism, AFLP)方法并经修改,详见文献[5]报道。

① 国家自然科学基金(30671136、30730065),国家博士后科学基金(20070411158)和电子科技大学青年基金资助项目。

② 男,1977 年生,博士,副教授;研究方向:植物分子细胞遗传学及表观遗传学;E-mail: zhangyong916@uestc.edu.cn

③ 同等贡献。

④ 通讯作者,E-mail: renzl@uestc.edu.cn
(收稿日期:2008-09-09)

1.4 甲基化修饰位点回收、克隆及测序

针对甲基化修饰位点进行回收,以相应选扩引物及扩增条件进行扩增。PCR 产物进行回收(Sangon, China)、连接(TaKaRa, Japan),转化大肠杆菌感受态细胞,挑阳性单克隆。序列测定委托 Invitrogen 公司进行(Invitrogen, America),序列信息登录 GenBank 注册。

1.5 甲基化修饰位点序列分析

从国际水稻基因组测序计划(IGSP)下载水稻基因组全序列(Build 4.0),通过本地 BLAST 比对分析,将克隆的水稻甲基化修饰位点联配到基因组序列上,记录定位信息及所在或相邻基因 ID,分别对 5'-非翻译区、启动子区、外显子区、内含子区、3'-非翻译区以及基因间隔区进行分布分析。依据 Gene Ontology (GO)系统,对 DNA 甲基化修饰序列进行功能描述。

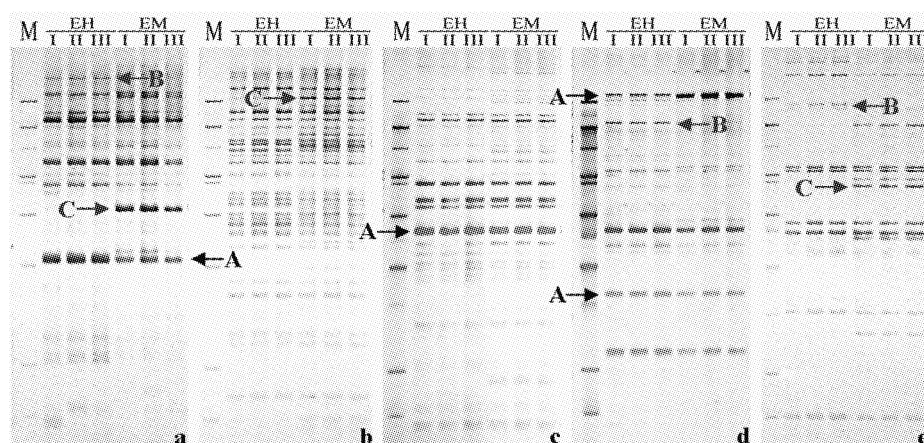
1.6 甲基化修饰位点侧翼序列“CpG island”分布及检测

以 MSAP 检出的水稻 DNA 甲基化修饰位点为中点,提取上、下游各 1000bp 序列,依据文献[8]对“CpG island”的定义,分析“CpG island”的分布特征。针对部分序列设计 PCR 扩增引物,参照文献[9]的方法,对预测的“CpG island”区域进行检测。

2 结果与分析

2.1 水稻 MSAP 指纹图谱构建

同裂酶 *Hpa* II 和 *Msp* I 的酶切位点均为“CCGG/GGCC”,但二者对酶切位点 DNA 序列的甲基化敏感性却有不同^[10],当以此同裂酶分别与 *EcoR* I 组合进行 AFLP 反应时,每一扩增位点均代表一个(CCGG/GGCC)位点,对比 *Hpa* II 与 *Msp* I 不同的扩增模式就可反映出该位点的甲基化状态(A:非甲基化位点;B:半甲基化位点;C:全甲基化位点)(图 1)。



(EH: *EcoR* I + *Hpa* II 双酶切处理;EM: *EcoR* I + *Msp* I 双酶切处理;M:分子量标准;I-III:重复)

图 1 水稻基因组 DNA 甲基化位点 MSAP 指纹图谱

根据扩增所得指纹图谱清晰、多态性丰富、重复性好的要求,选出 12 对 MSAP 引物组合,共扩增产生 595 条清晰可辨且可重复的 DNA 条带,平均每对引物扩增获得 49.58 条带。在全部扩增位点中,检测到甲基化位点 120 个(20.17%)。其中,全甲基化位点为 63 个(10.59%);半甲基化位点为 57 个(9.58%)。

2.2 DNA 甲基化修饰位点序列分析

对分离的 55 条甲基化修饰位点进行本地 BLAST 比对分析,将其联配到水稻基因组序列上。检索结果发现,55 条甲基化修饰序列随机分布于水

稻 12 条染色体上,表明水稻基因组中存在广泛的 DNA 甲基化修饰现象。进一步的分析发现,DNA 甲基化修饰位点分布区域并不是随机的,绝大多数甲基化修饰位点存在于编码基因附近(51/55)(表 1)。

将克隆序列的定位信息输入 GFF 数据库进行查询,得到其在基因序列中相对位置的分布信息。分析结果表明,甲基化修饰位点在水稻编码基因不同区域间的分布差异较大,其中启动子区分布频率最高(47%),其次为第一外显子区(22%),而在基因上(2%)、下(9%)游非翻译区及内含子区(7%)出现频率较低,另外有 13% 的甲基化位点分布与远离编

表 1 水稻基因组 DNA 甲基化修饰位点测序及比对结果

序列	基因组定位			E 值	基因位点	功能注释
	染色体	起点	终点			
RMS-01	12	21703137	21703337	1.0E-110	Os12g0539333	conserved hypothetical protein.
RMS-02	3	18215515	18215795	1.0E-157	Os03g0422200	conserved hypothetical protein.
RMS-03	3	36641351	36641535	1.0E-100	Os03g0850400	similar to aspartate kinase precursor
RMS-04	12	15118871	15118970	5.0E-50	Os12g0444500	conserved hypothetical protein.
RMS-05	10	18415513	18416001	0.0	Os10g0478100	heat shock protein DnaJ, N-terminal domain containing protein.
RMS-06	6	23715437	23715734	1.0E-163	Os06g0585200	conserved hypothetical protein.
RMS-07	7	19424604	19424843	1.0E-133	Os07g0499300	hypothetical protein.
RMS-08	3	2376499	2376624	2.0E-65	Os03g0143300	similar to Cysteine-tRNA ligase.
RMS-09	2	32032208	32032414	1.0E-113	Os02g0743200	conserved hypothetical protein.
RMS-10	2	6998623	6998832	1.0E-115	Os02g0224200	similar to Ser/Thr specific protein phosphatase 2A B regulatory subunit beta isoform.
RMS-11	4	18479920	18480035	4.0E-59	Os04g0377600	similar to TBC1 domain family member 13.
RMS-12	5	13289426	13289905	0.0	Os05g0298200	ankyrin repeat containing protein.
RMS-13	4	14995264	14995503	1.0E-131	Os04g0323500	conserved hypothetical protein.
RMS-14	11	22666284	22666424	6.0E-72	Os11g0559533	non-protein coding transcript, putative npRNA
RMS-15	12	12698681	12698854	6.0E-94	Os12g0411700	ABC transporter related domain containing protein.
RMS-16	2	3707424	3707633	1.0E-115	Os02g0168600	ovarian tumour, otubain domain containing protein.
RMS-17	8	16668755	16668893	4.0E-73	Os08g0359300	esterase/lipase/thioesterase domain containing protein
RMS-18	3	17765977	17766287	1.0E-170	Os03g0413400	glycosyl transferase, family 8 protein.
RMS-19	9	5262950	5263186	1.0E-131		No significant hit available
RMS-20	9	17354426	17354583	2.0E-84	Os09g0447100	conserved hypothetical protein.
RMS-21	2	33284743	33284936	1.0E-103	Os02g0768800	non-protein coding transcript.
RMS-22	5	3320880	3321087	1.0E-112	Os05g0157100	RNA polymerase Rpb7, N-terminal domain containing protein.
RMS-23	8	16668755	16668893	4.0E-73	Os08g0359300	esterase/lipase/thioesterase domain containing protein.
RMS-24	5	26179387	26179763	0.0	Os05g0524500	protein kinase-like domain containing protein.
RMS-25	4	3972364	3972695	0.0		No significant hit available
RMS-26	1	4415041	4415295	1.0E-142	Os01g0183300	endonuclease/exonuclease/phosphatase domain containing protein.
RMS-27	9	15665806	15666151	0.0	Os09g0418000	protein kinase-like domain containing protein.
RMS-28	6	8234427	8234708	1.0E-158	Os06g0257450	ribonucleotide reductase family protein.
RMS-29	1	44745783	44746241	0.0	Os01g0972900	similar to Clone ZZD405 mRNA sequence.
RMS-30	4	24466567	24466590	0.098	Os04g0481100	similar to Seed imbibition protein.
RMS-31	5	17967071	17967340	1.0E-151	Os05g0372800	cyclin-like F-box domain containing protein.
RMS-32	4	18479920	18480035	9.0E-57	Os04g0377600	similar to TBC1 domain family member 13.
RMS-33	6	6481860	6482310	0.0	Os06g0225200	mitochondrial transcription termination factor-related family
RMS-34	1	12264422	12264726	1.0E-164	Os01g0319000	similar to Pectin acetyltransferase.
RMS-35	7	27739707	27739973	1.0E-149	Os07g0648100	non-protein coding transcript.
RMS-36	8	27677104	27677593	0.0	Os08g0550100	similar to 26S proteasome subunit RPN3a.

(续表 1)

RMS-37	11	16577653	16578036	0.0	Os11g0445300	protein kinase-like domain containing protein.
RMS-38	4	15654812	15655168	0.0	Os04g0334700	peptidase A1, pepsin family protein.
RMS-39	7	29917530	29917865	0.0	Os07g0688300	similar to Importin alpha 1.
RMS-40	4	11535266	11535580	1.0E-173	Os04g0273600	non-protein coding transcript.
RMS-41	4	6592945	6593355	0.0	Os04g0196600	unknown function DUF563 family protein.
RMS-42	9	5290450	5290726	1.0E-148		No significant hit available
RMS-43	7	8673022	8673287	1.0E-148		No significant hit available
RMS-44	9	4088119	4088413	1.0E-166	Os09g0246700	conserved hypothetical protein.
RMS-45	10	11657429	11657790	0.0	Os10g0365200	non-protein coding transcript.
RMS-46	5	26179387	26179763	0.0	Os05g0524500	protein kinase-like domain containing protein.
RMS-47	11	26776382	26776538	8.0E-84	Os11g0633800	cyclin-like F-box domain containing protein.
RMS-48	8	6355827	6356021	1.0E-106	Os08g0208700	similar to Transposase.
RMS-49	2	34356324	34356521	1.0E-103	Os02g0787600	ionotropic glutamate receptor family protein.
RMS-50	2	4696337	4696624	1.0E-159	Os02g0184200	inorganic H + pyrophosphatase family protein.
RMS-51	4	1317697	1318004	1.0E-173	Os04g0122000	leucine rich repeat, N-terminal domain containing protein.
RMS-52	7	23351250	23351720	0.0	Os07g0565600	similar to Peptidyl-prolyl cis-trans isomerase TLP38, chloroplast precursor.
RMS-53	12	22232991	22233348	0.0	Os12g0545600	non-protein coding transcript, putative npRNA.
RMS-54	1	37542825	37543223	0.0	Os01g0835000	conserved hypothetical protein.
RMS-55	10	18806783	18806991	1.0E-114	Os10g0485300	conserved hypothetical protein.

码基因的区域(图 2)。由此可见,水稻基因组中甲基化修饰位点主要集中于启动子区和第一外显子区,这与哺乳动物中的分布相一致。

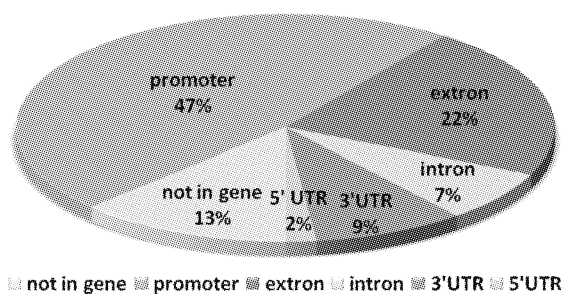


图 2 水稻基因组 DNA 甲基化修饰位点分布

根据 Rice Annotation Project 的注释, 51 条与 MSAP 检出甲基化修饰位点相邻的水稻基因被注释到包括物质能量转运、合成代谢、功能定位、信号传导、基因表达在内的 9 个 GO 节点(图 3)。注释结果说明, 水稻基因组中 DNA 甲基化修饰发生在功能多样的编码序列中, 可能参与相关基因的表达调控。

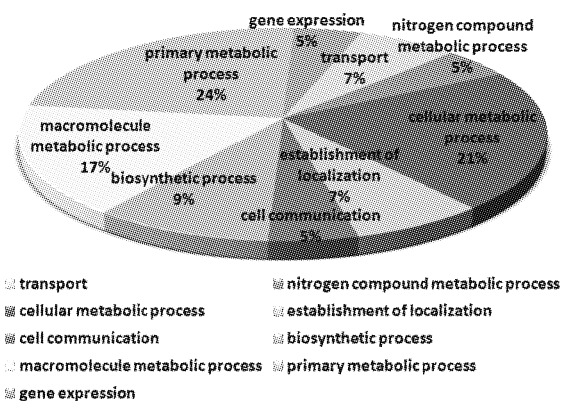


图 3 水稻基因组 DNA 甲基化修饰位点相邻基因 GO 注释节点分布

2.3 “CpG”二核苷酸分布特征分析及 MS-PCR 验证

依据 Gardiner-Garden 和 Frommer^[8] 对 “CpG island” 的定义, 设定 “GC” 含量 $\geq 50\%$, “CpG” 实际/期望值 ≥ 0.60 , “CpG island” 长度 $\geq 200\text{bp}$, 间隔 $\geq 100\text{bp}$ 为阈值, 在 42 个水稻 DNA 甲基化修饰位点中共检出 58 个具典型 “CpG island” 特征的区域, 其中 “GC” 含量、“CpG” 实际/期望值、“CpG island” 长度最高值分别为 66.3%、1.529、1758bp(表 2)。

表 2 水稻基因组 DNA 甲基化修饰位点侧翼序列 CpG island 预测结果

序列	起点	终点	GC(%)	CpG 实际/期望值	长度	序列	起点	终点	GC(%)	CpG 实际/期望值	长度
RMS-01	1007	1256	50.0	1.359	250	RMS-29	1	492	60.8	0.977	492
RMS-02	-	-	-	-	-		1191	1446	50.0	0.638	256
RMS-03	23	1006	63.1	1.042	984	RMS-30	-	-	-	-	-
RMS-04	19	1076	46.6	0.783	1058		26	914	52.2	0.89	889
	1628	2001	66.3	1.223	374	RMS-31	915	1263	50.7	0.656	349
RMS-05	-	-	-	-	-		1733	2001	52.4	0.932	269
RMS-06	1	1186	53.9	1.104	1186	RMS-32	1327	1756	50.7	0.618	430
RMS-07	833	1072	50.0	1.149	240	RMS-33	335	1726	59.4	0.994	1392
RMS-08	-	-	-	-	-	RMS-34	1	1050	59.1	1.141	1050
RMS-09	596	2001	54.2	1.417	1409		1263	1806	45.8	0.877	544
RMS-10	-	-	-	-	-	RMS-35	-	-	-	-	-
RMS-11	1327	1756	50.7	0.618	430	RMS-36	617	876	50.80	0.722	260
RMS-12	237	456	50.0	0.804	220	RMS-37	1565	2001	56.30	0.838	437
	694	1527	52.8	0.758	834	RMS-38	374	579	50.00	0.729	206
RMS-13	840	1866	50.1	1.05	1027		1118	1902	51.30	0.756	785
RMS-14	960	1149	50.0	0.95	190	RMS-39	-	-	-	-	-
RMS-15	1	1758	52.6	1.065	1758	RMS-40	1230	1674	51.50	0.982	445
RMS-16	-	-	-	-	-	RMS-41	1317	2001	51.10	0.783	685
RMS-17	15	275	51.0	0.726	261	RMS-42	-	-	-	-	-
	1469	2001	62.9	1.231	533	RMS-43	-	-	-	-	-
RMS-18	-	-	-	-	-	RMS-44	828	1515	44.3	0.931	688
RMS-19	468	715	50.0	0.774	248		1740	1947	50.0	0.854	208
	1091	1396	50.0	1.03	306	RMS-45	1	1309	57.1	0.846	1309
RMS-20	465	995	41.2	1.529	531	RMS-46	6	377	50.3	0.64	372
	1376	2001	62.3	1.037	626		1079	1796	56.7	0.671	718
RMS-21	1513	2001	55.6	1.22	489	RMS-47	1	840	63.3	1.104	840
RMS-22	192	515	50.0	0.865	324	RMS-48	-	-	-	-	-
RMS-23	15	275	51.0	0.726	261	RMS-49	1	263	51.7	1.195	263
	1469	2001	62.9	1.231	533		533	1166	58.4	1.196	634
RMS-24	6	377	50.3	0.64	372	RMS-50	-	-	-	-	-
	1079	1796	56.7	0.671	718	RMS-51	-	-	-	-	-
RMS-25	1	616	50.0	0.827	616	RMS-52	854	1133	50.0	0.753	280
	911	1493	44.3	0.946	583	RMS-53	864	1165	50.0	1.062	302
RMS-26	1	1210	65.2	1.153	1210	RMS-54	-	-	-	-	-
RMS-27	1	352	51.4	0.695	352	RMS-55	-	-	-	-	-
RMS-28	756	1490	52.4	0.803	735						

针对部分序列,依据预测的“CpG island”区域设计 PCR 引物,以 *McrBC* 酶切水稻基因 DNA,进行 PCR 扩增验证预测结果。结果显示,*McrBC* 酶切后无扩增产物检出,而未酶切反应中有扩增产物检出,

这表明在预测区域内“CpG”二核苷酸存在明显的 DNA 甲基化修饰现象,即所检测甲基化修饰位点侧翼序列中确实存在大量甲基化修饰的“CpG”二核苷酸富集分布的“CpG island”区域(图 4)。

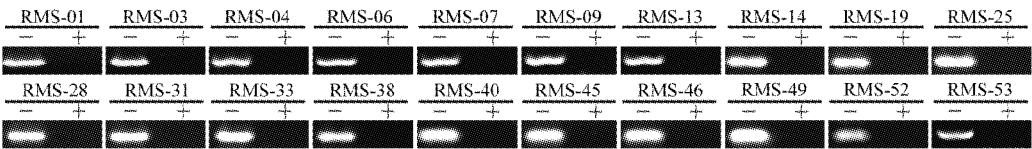


图 4 DNA 甲基化修饰位点侧翼序列“CpG island”检测(-:无 *McrBC* 酶切;+:*McrBC* 酶切)

3 讨论

3.1 MSAP 分离水稻基因组 DNA 甲基化修饰位点的有效性

MSAP 是以 AFLP 技术为蓝本,利用对同一酶切识别位点(“CCGG/GGCC”)中胞嘧啶甲基化修饰敏感性不同的一对同裂酶(*Hpa* II、*Msp* I)代替 AFLP 的 *Mse* I,经扩增后产生甲基化敏感多态性谱带。MSAP 所需 DNA 量少,敏感性强,检出位点多,稳定可靠、重复性强,适用于全基因组水平的甲基化位点扫描分析^[10]。目前,MSAP 被广泛应用于小麦^[5]、棉花^[11]、大米草^[12]、水稻-菰渐渗系^[13]、高粱^[14]等多种植物表观遗传学研究中。在植物研究中 MSAP 对于基因组 DNA 甲基化检出率在 4.7%~45% 之间,这反应了不同物种基因组间甲基化修饰水平的差异性^[5,13,14]。本文通过优化实验体系、筛选引物组合,获得了清晰、稳定的水稻 MSAP 指纹图(图 1),检测到甲基化修饰位点 120 个,甲基化修饰比例为 20.17%,这较之前对水稻亚种 *Oryza sativa* ssp. *indica* 检出率有所提高,这同样反映了基因组在 DNA 修饰层面的物种差异性^[7,13,15]。由于 MSAP 是基于对甲基化敏感程度不同的一对等裂酶(普遍使用 *Hpa* II 和 *Msp* I)进行甲基化位点检出,因此只能检出发生于“CCGG/GGCC”位点的甲基化修饰现象,而不能对该位点以外的甲基化修饰进行检出,从而造成绝对检出率的一定偏低。但因为该位点多位于“CpG”二核苷酸重复区内,而植物甲基化修饰主要发生在此区域内,因此多数研究者认为“CCGG/GGCC”位点的甲基化修饰率能够客观反映植物基因组甲基化修饰水平^[5,10,12]。基于目前的研究水平,具有高效、重复性强、适用范围广等特性的 MSAP 技术依然是进行植物基因组甲基化修饰位点分析的有效方案。

3.2 水稻基因组 DNA 甲基化修饰的生物学意义

DNA 甲基化修饰在调节基因表达、控制细胞分化及系统发育等过程中起着重要的生物学功能^[1,4,16],高等植物基因组中,大约 20%~50% 的胞嘧啶处于甲基化状态,这是植物维持正常生长发育所必需的^[2,3]。如果 DNA 甲基化水平非正常变化,将会导致发育异常、生活力下降等诸多表型变异^[17]。Xiao 等^[18]的研究证明,拟南芥幼胚发育过程中 DNA 甲基化水平与发育所需特异基因表达量成反比,其中胚发育特异基因表达缺失,限制了幼胚中激素的正常积累,造成幼胚凋亡及种子失活。Zil-

berman 等^[3]发现,拟南芥基因组中,大部分编码基因及转座子存在相当比例的 DNA 甲基化,甲基化程度的不同与其转录活性密切相关,进而影响植株的表型。例如在拟南芥 *ibm1* 突变体中,细胞周期调控相关基因甲基化水平表现异常,突变体表现植株矮化、叶片变小、开花延迟、种子变小、结实率降低以及育性下降等表型变异^[16]。

不同亚种水稻基因组中 15.90%~20.17% 的胞嘧啶处于甲基化修饰状态^[7,13,15],这种甲基化修饰的物种差异性被认为有着重要的系统学意义^[19]。最近的研究证实,在外源种质导入、组织培养继代及高静水压等因素诱导下,水稻基因组甲基化修饰程度与模式会发生显著改变,进而引起转录组发生广泛变异,从而表现不同的性状^[13,20,21],这说明正常 DAN 甲基化修饰水平的维持是水稻生命过程的必须要求。对甲基化修饰位点进行定位分析,发现绝大多数位点位于水稻编码基因附近,其中大部分修饰位点发生于启动子区(47%)及第一外显子编码区(22%)(图 2),这与 Li 等^[22]对水稻 4、10 号染色体的序列分析结果相一致。对存在甲基化修饰的基因位点进行功能分析,发现其编码包括蛋白激酶、热激蛋白、磷酸化酶、糖基转移酶在内的多种调控及结构蛋白(表 1),在功能上包括物质能量转运、合成代谢、功能定位、信号传导、基因表达调控在内的 9 个 GO 生物注释节点(图 3)。这种集中于主要功能基因位点的甲基化修饰现象,与研究者们对拟南芥全基因组^[2]及水稻两条染色体^[22]分析得出的结论相一致,表明通过 DNA 甲基化修饰对功能基因表达进行调控是包括水稻在内的植物表观遗传修饰的重要途径。

3.3 水稻基因组序列中“CpG”岛分布及潜在功能

以“CpG”二核苷酸中胞嘧啶 5' 甲基化为特征的 DNA 甲基化修饰,是表观遗传修饰最主要的方式之一,具有重要的生物学功能。在进化过程中,以人为代表的哺乳动物基因组中绝大多数“CpG”发生了丢失,但在基因组的某些区域,簇集了过量的“CpG”二核苷酸^[23]。Gardiner-Garden 和 Frommer^[8]首先提出“CpG 岛”(CpG island)的定义,随后的研究表明这些“CpG 岛”通常位于基因启动子及第一外显子区域,极少部分靠近基因 3' 端^[24]。最近的研究证明,高等植物基因组中同样存在丰富的胞嘧啶甲基化修饰现象,其中“CpG”二核苷酸是最主要的修饰位点,但对“CpG”二核苷酸富集区在植物基因组中相对于基因的分布位置却有着不同的结论^[2,3,22]。对分离的水

稻甲基化修饰位点侧翼序列进行筛选,检出 58 个具典型“CpG island”特征的“CpG”二核苷酸成簇富集分布区(表 2,图 4)。结合水稻基因组的定位信息(图 2),可知在水稻基因编码序列 5'端(启动子及第一外显子编码区)存在类似哺乳动物“CpG island”的“CpG”二核苷酸成簇富集区。这一实验证据与之前对水稻基因组“CpG”二核苷酸分布特征的预测分析相一致^[22],但与拟南芥中“CpG”二核苷酸成簇富集区主要分布于基因中部^[2,3]的结论有所差异。这种“CpG”二核苷酸成簇富集区在基因序列中分布不一致的结果在一定程度上反映了单子叶与双子叶植物在基因组组成与进化上的差异性,代表了两种不同的甲基化修饰系统,可能是不同的 DNA 甲基化表观调控机制的一种体现。

参考文献

- [1] Henderson I R, Jacobsen S E. Epigenetic inheritance in plants. *Nature*, 2007, 447 (7143): 418-424
- [2] Zhang X Y, Yazaki J, Sundaresan A, et al. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell*, 2006, 126 (6): 1189-1201
- [3] Zilberman D, Gehring M, Tran R K, et al. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genetics*, 2007, 39 (1): 61-69
- [4] Jullien P E, Kinoshita T, Ohad N, et al. Maintenance of DNA methylation during the *Arabidopsis* life cycle is essential for parental imprinting. *Plant Cell*, 2006, 18 (6): 1360-1372
- [5] Zhang Y, Liu Z H, Liu C, et al. Analysis of DNA methylation variation in wheat genetic background after alien chromatin introduction based on methylation-sensitive amplification polymorphism. *Chinese Science Bulletin*, 2008, 53 (1): 58-69
- [6] Vos P, Hogers R, Bleeker M, et al. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, 1995, 23: 4407-4414
- [7] Xiong L Z, Xu C G, Maroof M A S, et al. Patterns of cytosine methylation in an elite hybrid and its parental lines, detected by a methylation-sensitive amplification polymorphism technique. *Molecular Genetics and Genomics*, 1999, 26: 439-446
- [8] Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes. *Journal of Molecular Biology*, 1987, 196 (2): 261-282
- [9] Rabinowicz P D, Palmer L E, May B P, et al. Genes and transposons are differentially methylated in plants, but not in mammals. *Genome Research*, 2003, 13: 2658-2664
- [10] Reyna-Lopez G E, Simpson J, Ruiz-Herrera J. Differences in DNA methylation patterns are detectable during the dimorphic transition of fungi by amplification of restriction polymorphisms. *Molecular Genetics and Genomics*, 1997, 253: 703-710
- [11] Adams K L, Percifield R, Wendel J F. Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. *Genetics*, 2004, 168: 2217-2226
- [12] Salmon A, Ainouche M L, Wendel J F. Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). *Molecular Ecology*, 2005, 14: 1163-1175
- [13] Dong Z Y, Wang Y M, Zhang Z J, et al. Extent and pattern of DNA methylation alteration in rice lines derived from introgressive hybridization of rice and *Zizania latifolia* Griseb. *Theoretical and Applied Genetics*, 2006, 113: 196-205
- [14] Zhang M S, Yan H Y, Zhao N, et al. Endosperm-specific hypomethylation, and meiotic inheritance and variation of DNA methylation level and pattern in sorghum (*Sorghum bicolor* L.) inter-strain hybrids. *Theoretical and Applied Genetics*, 2007, 115 (2): 195-207
- [15] Zhang H Y, Peng H, Li Y, et al. Patterns of DNA cytosine methylation between haploids and corresponding diploids in rice. *Chinese Science Bulletin*, 2006, 51 (14): 1721-1728
- [16] Saze H, Shiraishi A, Miura A, et al. Control of genic DNA methylation by a jmjC domain-containing protein in *Arabidopsis thaliana*. *Science*, 2008, 319 (5862): 462-465
- [17] Kinoshita Y, Saze H, Kinoshita T, et al. Control of FWA gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. *The Plant Journal*, 2007, 49 (1): 38-45
- [18] Xiao W, Custard K D, Brown R C, et al. DNA methylation is critical for *Arabidopsis* embryogenesis and seed viability. *Plant Cell*, 2006, 18 (4): 805-814
- [19] Vanyushin B F. DNA methylation in plants. *Current Topics in Microbiology and Immunology*, 2006, 301: 67-122
- [20] Liu Z, Wang Y, Shen Y, et al. Extensive alterations in DNA methylation and transcription in rice caused by introgression from *Zizania latifolia*. *Plant Molecular Biology*, 2004, 54 (4): 571-582
- [21] Ding Y, Wang X, Su L, et al. SDG714, a histone H3K9 methyltransferase, is involved in *Tos17* DNA methylation and transposition in rice. *Plant Cell*, 2007, 19 (1): 9-22
- [22] Li X, Wang X, He K, et al. High-resolution mapping of epigenetic modifications of the rice genome uncovers interplay between DNA methylation, histone methylation, and gene expression. *Plant Cell*, 2008, 20 (2): 259-276

- [23] Issa J P. CpG island methylator phenotype in cancer. *Nature Reviews Cancer*, 2004, 4 (12): 988-993
- [24] Jones P A, Takai D. The role of DNA methylation in mammalian epigenetics. *Science*, 2001, 293 (5532): 1068-1070

Construction of rice genomic methylation-sensitive amplification polymorphism fingerprint and characterization of DNA methylated sites

Zhang Yong^{* **}, Deng Kejun^{*}, Zhang Tao^{*}, Yang Zujun^{*}, Peng Jinhua^{*}, Zhou Jianping^{*}, Ren Zhenglong^{* **}
(^{*} School of Life Sciences and Technology, University of Electronic Science and Technology of China, Chengdu 610054)
(^{**} Sichuan Provincial Key Laboratory of Plant Breeding and Genetics, Sichuan Agriculture University, Ya'an 625014)

Abstract

In this study, the double digestion of *EcoR* I and *Hpa* II/*Msp* I was used to construct the high-density rice genomic methylation-sensitive amplification polymorphism (MSAP) fingerprint, and the methylated sites were detected at the whole genome level. By using twelve pairs selective primers, a total of 120 methylated sites were detected, which represented that the ratio of methylation modification at the 'CCGG/GGCC' site in rice genome was 20.17%. Some of methylated sites were extracted and sequenced. Finally, 55 fragments with methylation modification were characterized and located to rice genome sequence data through the basic local alignment search tool (BLAST). The further sequence analysis showed that the methylated sites were mainly distributed in the gene promoter region (47%) and the first exon region (22%). The region of CpG rich cluster with typical CpG island sequence features was also detected in the flanking sequence of rice methylated sites. Based on these experimental evidences, the effectiveness to isolate rice methylated site by MSAP and the distribution features of the CpG island-like sequence in rice genome were discussed deeply in the present paper.

Key words: *Oryza sativa*, epigenetic, DNA methylation, methylation-sensitive amplification polymorphism (MSAP), CpG island