www. scichina. com

life, scichina, com



# 双胚苗水稻中单倍体与二倍体基因表达芯片分析

张红宇<sup>①3†</sup>、彭海<sup>2†</sup>、李平川<sup>①</sup>、邓其明<sup>①</sup>、徐培洲<sup>①</sup>、李云<sup>®</sup>、汪旭东<sup>①</sup>、吴先军<sup>①3\*</sup>

- ① 四川农业大学水稻研究所, 温江 611130;
- ② 江汉大学生命科学学院, 武汉 430056;
- ③ 西南作物基因资源与遗传改良教育部重点实验室, 雅安 625014;
- ④ 成都市农林科学院作物研究所, 温江 611130
- † 同等贡献
- \* 联系人, E-mail: wuxisan@126.com

收稿日期: 2008-01-21; 接受日期: 2008-01-21

国家自然科学基金(批准号: 30771157)与教育部长江学者和创新团队发展计划(批准号: IRT0453)资助项目

摘要 本研究运用基因芯片技术,对双胚苗水稻 SAR II-628 品系中的单倍体及其对应二倍体的 RNA 水平的表达变化情况进行对比分析,发现单倍化后基因表达迅速发生了不同程度的变异,主要结果如下: (i) 单倍化后,表达变化的序列占水稻总探针序列的 2.47%,随机分布在水稻的 12 条染色体上,激活的序列数量多于沉默的;(ii) 有 33 个序列在染色体上的分布具有成簇分布的特点;(iii) 对单倍化中已经预测功能的 575 条序列进行功能分类,发现都涉及了生物过程、细胞成份和分子功能这 3 个方面. SNF2 家族中控制转录调节的位点表达下降居多;与 DNA 重组修复相关的RAD54 家族中只有 2 个位点分别发生了表达上升和降低的变化;(iv) 采用 RT-PCR验证基因芯片的结果,总体上 83.78%序列表达结果与芯片结果相符;任选 7 个极端(激活和沉默)表达方式的序列,其表达结果与芯片相符率达到 91.86%.

#### 关键词

水稻(Oryza sativa L.)

双胚苗

二倍体

单倍体

基因芯片

基因表达

全基因组倍增或整体化过程被认为是植物尤其是禾本科作物物种形成和进化过程中非常普遍和重要的事件,50%~70%的开花植物在进化过程中均经历了一次或多次染色体加倍过程<sup>[1]</sup>.基因组加倍后,再经历所谓的二倍化过程(diploidization),进化成当代的二倍体物种.近年来植物多倍体基因表达研究是一个相当活跃的领域<sup>[2]</sup>.基因表达变化在众多异源多倍体植物中得到证实,包括蕨类植物<sup>[3,4]</sup>、拟南芥<sup>[5,6]</sup>、小麦<sup>[7]</sup>、棉花<sup>[8]</sup>和油菜<sup>[9]</sup>.仅有少量研究包含了同源多倍体如酵母<sup>[10]</sup>和拟南芥<sup>[5]</sup>.

单倍体仅有一套染色体,所有基因座位(不论是显性还是隐性)都能在RNA或表型水平上表现出来,因而是遗传学研究的重要工具.虽然对单倍体在甲

基化水平上的变化已有初步研究□□,但是至今对单倍体基因表达或基因表达调控机制还了解甚少.在异源多倍体基因表达研究中,难以将杂交与基因组加倍对基因表达的效应区分开□□.而双胚苗来源的多(单)倍体接近自然状态且都为DNA序列一致的同源多(单)倍体,可准确反映水稻倍性对基因表达的影响,弥补了前人研究工作的不足.通过水稻双胚苗可创造单倍体、二倍体、三倍体和四倍体等,其起源方式接近自然状态、DNA序列没有差异,实现了倍性系列的配套,可准确分析倍性对表达调控的影响.

本研究采用高密度基因芯片技术对来源于双胚苗SAR -628 品系的单倍体和其对应的二倍体检测基因表达的信息,比较不同倍性的分子水平差异与

表形上的差异之间的联系,将为研究水稻基因和基因组倍增后的二倍体化的进化机制提供非常好的素材.由于本实验的研究材料已经经过倍性验证和微卫星分析,并发现其一级结构没有发生变异[11],可以直接进行芯片分析和RT-PCR验证.

# 1 材料和方法

#### 1.1 材料

SAR -628来源于四川农业大学水稻研究所双胚 苗品系 9003 的自然群体,每年都能定期分离出单倍体与二倍体(n: 2n)的双苗植株.

# 1.2 RNA 的提取、纯化和定量

在叶枕距为 0 时的当天上午 8 时分别取该分蘖对应的剑叶,采用 Trizol 方法提取 RNA,并用紫外分光光度计检测 RNA 的产量,变性胶电泳质检后,作为反应的样本.上海晶泰生物技术有限公司则在预测的基因上选 11 个段并在原位合成探针后,利用Affymetrix 水稻基因芯片对检测样本进行杂交和扫描.

# 1.3 芯片数据分析

表达芯片扫描数据统计按 MAS5.0 操作手册进行 (http://www.affymetrix.com/support/technical/whitepapers/sadd\_whitepaper.pdf),分析过程中包括 cut-offs 在内的所有参数都采用 MAS5.0 使用手册推荐的值.

#### 1.4 功能分类分析

提取Affiymetrix芯片检测中表达变化位点的探针序列,利用Swissprot 和Tremble数据库(http://www.ebi.ac.uk/)进行Blastx 比对,比对结果利用GoPipe软件[13]进行功能归类,参数选用GoPipe操作手册推荐值.

#### 1.5 RT-PCR 验证

选取 SAR -628 来源的 1N-2N 双胚苗的 8 个水稻单株,来自同一株系的命名为 A,其他 7 个单倍体单株则命名为 A,B,C,D,E,F和 G.取它们在同一发育时期的叶的总 RNA 逆转录合成 ss-cDNA,作为RT-PCR 反应的模板.

根据基因芯片的扫描结果,任选出单倍体中的 3 个激活位点和 4 个沉默位点作为目标序列,重新编号 为 1~7(表 1).

# 2 结果和分析

## 2.1 单倍化敏感序列的表达及分布

通过水稻芯片扫描,发现当SAR -628 发生单倍化后,很多序列也随之发生表达上的上调和下调变化.根据表达的不同情况,只筛选出表达变化且变化2 倍以上这些序列.其中序列在单倍体中表达而在二倍体中不表达的情况定义为激活,相反的情况则定义为沉默,在单倍体中的表达量高于二倍体2倍以上的定义为表达上升,相反则定义为下降.单倍化后的共有1475条序列发生表达变化,占整个水稻探针序列的2.47%.其中包括激活、上升、沉默和下降共4种类型(表2),称这些发生变化的序列为倍性敏感序列.它们表达变化的倍数也不同,表达下降的倍数在2~9.8之间,表达上升的倍数在2~10.5之间,但是总体变化倍数集中在2~3倍之间.

这 1475 条单倍性特征序列在水稻染色体上明确定位的共有 1358 条,它们在水稻基因组的 1~12 条染色体上均有分布(表 3). 除了第 3 条染色体上表达下调的序列数目多于表达上调的外,其他 11 条染色体上的特征序列表达变化情况均是表达上调所占的比例大于 50%.

表 1	RT-PCR	反应的 8	对引物序列

编号	上游引物 5′~3′方向	下游引物 5′~3′方向
1	CAACGCGCAGACATACAACC	ATGGACACCACACGGCACTAT
2	CGCAGTTCATGGGTGCTGGC	AGGAGCCATGGTGATGGTGC
3	GAACATCGCCGCAGGAAAAGCTC	ATTAGTGGCGATTTGTTTATGTG
4	GAATCTCCCTACTAGCTACT	CATGCTTATACGTACATCCA
5	CTTCGTCGAGGTAGATCACAT	GCTACATAAACAAGTCAGCGT
6	TTGTAACAGTTGGCCTTGTAGT	GGAGAGGCATTCAAGATGTGCT
7	GCTGTCATACATCACTTCATGG	CTCATGATTTCTCTCAGGCT
内参	GAACTGGTATGGTCAAGGCTG	ACACGGAGCTCGTTGTAGAAG

表 2 SAR -628 单倍化后倍性敏感序列的表达情况

单倍化	沉默	激活	上升	下降	变化 合计
数量	160	206	643	466	1475
比例%	10.85	13.97	43.59	31.59	
在芯片中的比例%	0.27	0.34	1.08	0.78	2.47

#### 2.2 单倍化敏感序列变化成簇分布

已有研究发现在四倍体拟南芥中,具有新表达模式的基因不但编码各种功能类别的蛋白质,而且在染色体上存在一定的成簇分布<sup>[14]</sup>.通过对单倍化后水稻的基因芯片扫描分析发现,有一些代表不同基因功能的序列不但位置相邻,而且表达量也随着倍性的变化同时发生变化,就是具有成簇分布的特点.筛选的这些序列虽然相邻但没有重叠,而且单个序列与序列间的长度之和大于6kb,保证这些序列分别指向不同的基因.共有16组序列(共33条)成簇分布,其中有1组中共有3条序列相邻的且同时发生表达变化的,其他15组均是2条序列相邻.这16组序列成簇分布在不同的染色体上,第1染色体上数目最多,有7组,占43.75%.同时发生表达上调反应的共有8组,远远多于发生其他两种表达情况的序列组数(表4).这些成簇分布序列的长度范围在0.312~12.41

kb之间,组内间距大小在 0.69~54.01 kb之间,与表达情况没有一定的联系.

#### 2.3 单倍化敏感序列的功能分类

在单倍体的 1475 条倍性敏感位点的特征序列中,有 575 条已经预测了功能,占总变化序列的 38.98%.利用 GoPipe (版本 1.21)软件对这 575 条序列进行已知功能分类,发现这 575 条序列的功能分类主要涉及生物过程(biological process)、细胞成份(cell component)和分子功能(molecular function)这 3 个方面(表5~7).由于这 575 个序列中,有的序列不只具有一个功能,还具有其他多个功能,因此在分类的时候,会多次出现在不同的功能中,因此所有的功能分类对应的序列位点数之和会大于 575.

由表 5 得出, 575 条预测功能的序列中, 与生物过程相关的有 472 条. 其中变化最多的是新陈代谢过程, 所占的比例达到了 65.04%; 其次是蛋白代谢方面, 比例为 25.91%; 还有有关细胞生长和维持、生物合成、蛋白修饰和运输相关的功能也占了一定的比例, 分别为 16.87%, 16.7%, 15.65%和 11.65%. 新陈代谢变化中值得注意的是 DNA 新陈代谢与 RNA 转录代谢的变化, 因为前者是基因组增减的必然要求, 后者

表 3 单倍化的敏感序列在染色体上的分布

染色体(chr.)	chr01	chr02	chr03	chr04	chr05	chr06	chr07	chr08	chr09	chr10	chr11	chr12	合计
扫描序列	8940	5017	5538	4035	3673	3854	4635	3329	2593	3358	2920	2643	50535
单倍化后下调	103	58	86	40	50	50	58	38	24	33	28	20	588
单倍化后上调	140	81	76	85	56	56	67	43	45	49	36	36	770
合计	243	139	162	125	106	106	125	81	69	82	64	56	1358
比例%	2.72	2.77	2.93	3.1	2.89	2.75	2.7	2.43	2.66	2.44	2.19	2.12	2.69

表 4 表达同时变化的相邻序列在染色体的分布及数目

相邻序列条数	染色体编号		数目合计(组)		
相邻方列余数	采巴仰绸专	同时上调	同时下调	上调和下调	数日言 (组)
3	1	0	0	1	1
2	1	2	1	3	6
2	2	1	1	0	2
2	4	2	0	0	2
2	5	1	1	0	2
2	6	1	0	0	1
2	7	1	0	0	1
2	12	0	1	0	1
合计		8	4	4	16

表 5 单倍化中预测功能序列位点按生物过程的功能分类

GO 号	功能描述	位点数	百分率(%)
GO:0008150	生物进程	472	82.09
GO:0007154	细胞通讯	18	3.13
GO:0009719	内源刺激的响应	4	0.70
GO:0009605	外源刺激的响应	31	5.39
GO:0009628	非生物胁迫响应	4	0.70
GO:0009607	生物胁迫响应	29	5.04
GO:0007165	信号转导	14	2.43
GO:0008151	细胞生长/维持	97	16.87
GO:0007049	细胞周期	21	3.65
GO:0016043	细胞结构与发生	10	1.74
GO:0007028	胞质结构与发生	2	0.35
GO:0006996	细胞器结构与发生	1	0.17
GO:0007010	细胞骨架与发生	1	0.17
GO:0008283	细胞增殖	22	3.83
GO:0019725	细胞内环境稳定	1	0.17
GO:0008152	新陈代谢	374	65.04
GO:0006519	氨基酸和衍生物代谢	27	4.70
GO:0009058	生物合成	96	16.7
GO:0005975	碳代谢	38	6.61
GO:0009056	分解代谢	34	5.91
GO:0006731	辅酶和辅基代谢	10	1.74
GO:0006118	电子转动	51	8.87
GO:0006091	能量通路	19	3.30
GO:0006629	脂代谢	6	1.04
GO:0006139	碱基、核苷酸和核酸代谢	69	12
GO:0006259	DNA 代谢	26	4.52
GO:0006350	转录	31	5.39
GO:0019538	蛋白质代谢	149	25.91
GO:0006412	蛋白质合成	35	6.09
GO:0006464	蛋白质修饰	90	15.65
GO:0019748	二级代谢	10	1.74
GO:0006950	逆境反应	34	5.91
GO:0006810	转运	67	11.65
GO:0006811	离子转运	27	4.70
GO:0015031	蛋白质转运	5	0.87
GO:0016265	死亡	15	2.61
GO:0008219	细胞死亡	15	2.61
GO:0007275	发育	10	1.74
GO:0030154	细胞分化	1	0.17
GO:0009653	形态发生	1	0.17

注: 百分率=(位点数/575)×100%

则直接与将要分析到的表达变化相关;细胞大小的改变等也是倍性变化的典型特征,因而细胞生长与维持相关基因表达也有较大变化.

由表 6 得出, 575 条预测功能的序列中, 与分子功能相关的有 523 条. 其中变化最多的是具有催化激活功能的序列, 比例为 59.83%, 表明倍性变化中各

表 6 单倍化中预测功能序列位点按分子功能的分类

GO 号	功能描述	位点数	百分率(%)
GO:0003674	分子功能	523	90.96
GO:0016209	抗氧化活性	8	1.39
GO:0005488	绑定	271	47.13
GO:0005509	钙离子绑定	22	3.83
GO:0030246	糖绑定	18	3.13
GO:0008289	脂绑定	6	1.04
GO:0003676	核酸绑定	91	15.83
GO:0003677	DNA 绑定	58	10.09
GO:0003682	染色体绑定	2	0.35
GO:0003700	转录因子活性	10	1.74
GO:0004518	核酸酶活性	8	1.39
GO:0003723	RNA 绑定	27	4.70
GO:0008135	转录因子活性、核酸绑定	7	1.22
GO:0000166	核苷酸绑定	136	23.65
GO:0005515	蛋白质绑定	6	1.04
GO:0003754	分子伴侣活性	5	0.87
GO:0003824	催化活性	344	59.83
GO:0016301	激酶	100	17.39
GO:0004672	蛋白激酶活性	84	14.61
GO:0016787	水解酶	85	14.78
GO:0008233	肽酶活性	17	2.96
GO:0004721	蛋白磷酸酶活性	3	0.52
GO:0016740	转移酶	156	27.13
GO:0030234	酶调节活性	6	1.04
GO:0005554	未知分子功能	7	1.22
GO:0003774	活动能力	1	0.17
GO:0004871	信号转导活性	42	7.30
GO:0005102	受体绑定	3	0.52
GO:0004872	受体活性	36	6.26
GO:0045735	营养储存	2	0.35
GO:0005198	结构分子	26	4.52
GO:0030528	转录调节活性	11	1.91
GO:0045182	转录调控子	7	1.22
GO:0005215	转运子活性	71	12.35
GO:0005489	电子转运子活性	16	2.78
GO:0005216	离子通道活性	1	0.17

注: 百分率=(位点数/575)×100%

种酶的表达出现了较大调整; 其次是具有绑定功能的序列, 比例为 47.13%, 显示了 DNA 和 RNA 复合物的改变; 其他的就是转移酶、核苷酸绑定、激酶、核酸绑定和水解酶等也占了一定的比例, 分别为27.13%, 23.65%, 17.39%, 15.83%和 14.78%. 激酶与信号转导在基因表达调控网络中位置特殊, 它们的改变必然对倍性变化中基因表达产生全面影响.这些改变应该都是直接与倍性变化或与倍性变化后

表达变化相关. 虽然酶类变化位点很多, 但是因为酶 在生物体内种类多, 不能说明酶类对倍性更敏感.

由表 7 得出, 575 条预测功能的序列中, 与细胞结构相关的有 242 条. 其中直接与细胞相关的序列变化比例就达到了 40.7%, 其次就是与细胞内的相关, 比例达到 27.3%, 还有就是与细胞质、质粒和细胞核等相关. 从细胞结构的角度显示了倍性改变后在酶、能量代谢、蛋白质代谢和 DNA 活动等方面都发生的改变.

表 7 单倍化中预测功能序列位点按细胞成份的功能分类

GO 号	功能描述	位点数	百分率(%)
GO:0005575	细胞成份	242	42.09
GO:0005623	细胞	234	40.70
GO:0005622	胞间	157	27.30
GO:0005694	染色体	7	1.22
GO:0005929	纤毛	1	0.17
GO:0005737	胞质	103	17.91
GO:0016023	细胞质囊	1	0.17
GO:0005856	细胞骨架	2	0.35
GO:0005829	胞浆	2	0.35
GO:0005783	内质网	1	0.17
GO:0005794	高尔基体	3	0.52
GO:0005739	线粒体	13	2.26
GO:0005777	过氧化物酶体	2	0.35
GO:0009536	质粒	62	10.78
GO:0005840	核糖体	22	3.83
GO:0005634	细胞核	49	8.52
GO:0000228	核染色体	3	0.52
GO:0005635	核膜	1	0.17
GO:0005730	核仁	1	0.17
GO:0005654	核质	1	0.17
GO:0005886	质膜	2	0.35
GO:0009579	类囊体	21	3.65
GO:0030312	外膜结构	1	0.17
GO:0005618	细胞壁	1	0.17
GO:0005576	细胞外	5	0.87
GO:0005941	未定位	3	0.52

注: 百分率=(位点数/575)×100%

#### 2.4 单倍化中重要功能位点的表达变化

(1) DNA重组修复相关位点. DNA重组过程中最重要的是DNF家族中RAD54,它负责从单链DNA尾部替下RPA并组成前联会复合体,最终完成断裂DNA遗传修复[15]. RAD54 在四倍体A.suecica及其远缘杂交种中高度表达[16]. 可推测: RAD54 的表达模式与新形成的同源或异源多倍体中的染色体行为不稳定性相一致[17]. 单倍体水稻中只具有一套基因组,其减数分裂中染色体行为异常和不稳定也将与RAD54相关.

通过检索和比对, 共 9 个位点与 DNA 重组修复有关(表 8), 仅有探针 Os.7297.1.S1\_at 和 Os.8966.1. S1\_at 所对应的位点表达发生了变化. 位点注释显示其中与 RAD50 同源负责 DNA 重组修复功能的蛋白

表达量下降,而负责不精确转录具有修复功能的 RAD23 的同源蛋白表达量上升.推测这可能与单倍 体只有一套染色体相关,并且单倍体在自交时不结 实,而与正常的二倍体杂交可正常结实.说明单倍体 同源重组的错误率低,修复工作量减少.

(2) SNF2 家族中与基因调节相关的位点. 不同物种 SNF2 家族蛋白执行以下功能: 转录调控如 M0T1, SNF2 和 BRM; 维持减数分裂时染色体稳定如 lodestar; DNA 损伤修复,包括核酸切除修复,如 RAD16 和 ERCC6、复制后子链缺口修复如 RAD5、重组修复如 RAD54.单倍体中共检索到 7 个具有 SNF2 特征结构域的位点(表 9),其中 4 个位点表达量降低,一个中度降低,另外 2 个不变.这些蛋白表达量的降低也说明单倍体中修复重组的几率降低.

表 8 单倍化后 DNA 重组修复相关位点的表达变化

探针	二倍	体	单倍	体	二VS单		
	信号值	判定	信号值	判定	改变量	判定	<b>为能抽</b> 处
Os.24886.1.S1_at	1700.5	P	1585.1	P	-0.2	NC	DNA 修复与重组蛋白 (推定)
Os.30497.1.S1_at	23.3	A	26.4	A	-0.9	NC	DNA 修复蛋白 rhp54 (推定, 与 RAD54 同源)
Os.30497.1.S2_at	59.7	A	5.8	A	-2.6	NC	DNA 修复蛋白 rhp54 (推定,与 RAD54 同源)
Os.7297.1.S1_at	1462	P	825.9	P	-0.6	D	DNA 修复重组蛋白(rad50)
Os.8966.1.S1_at	2142.2	P	2912.7	P	0.9	I	转录序列与 DNA 修复蛋白 RAD23 (拟南芥) 存在弱同源性
OsAffx.15402.1.S1_at	43.1	A	7.9	A	-2.3	NC	DNA 切除修复家族 HhH-GPD (推定)
OsAffx.27392.1.S1_at	122.4	A	135.5	Α	0.1	NC	DNA 切除修复家族 HhH-GPD (推定)
OsAffx.2831.1.S1_at	32.8	A	30.6	Α	-0.1	NC	类似于 DNA 修复蛋白 RAD5
OsAffx.2831.1.S1_x_at	469.3	P	524.2	P	-0.3	NC	类似于 DNA 修复蛋白 RAD5

注: 1, 信号值是芯片扫描值, 不等于表达量; 2, 判定: 判断表达与否或增减与否是根据每条序列 11 对探针信号值作出的统计推断, 各符号意义为, A: 不表达, P: 表达, NC: 没有变化, I: 表达上升, D: 表达下降; 3, 改变量=二倍体的表达量/单倍体的表达量; 4, 功能是通过 interpro 比对得出的

表 9 单倍体中 SNF2 家族中与基因调节相关的位点表达变化

探针	=-	二倍体		单倍体		/S 单	- 功能描述	
1本口	信号值	判定	信号值	判定	改变量	判定	- 切形畑处	
Os.10905.1.S1_x_at	337.4	P	216.9	P	-0.8	D	SNF2 相关结构域	
Os.10925.1.S1_at	3636.7	P	2374.1	P	-0.6	D	SNF2 相关结构域	
Os.11138.2.S1_x_at	1199.7	P	639	P	-0.4	D	带有 SNF2 相关结构域的染色质重构 DNA 结合 蛋白(推定)	
Os.16806.1.S1_at	1153.3	P	906.6	P	-0.3	NC	SNF2 相关结构域	
Os.28431.1.S1_at	3942.5	P	2789	P	-0.5	D	SNF2 相关结构域	
Os.30497.1.S2_at	59.7	A	5.8	A	-2.6	NC	DNA 修复蛋白 rhp54(推定,与 RAD54 同源)	
Os.52584.1.S1_at	2604	P	1737.3	P	-0.6	MD	SNF2 相关结构域	

注: 1, 信号值是芯片扫描值, 不等于表达量; 2, 判定是判断表达与否或增减与否是根据每条序列 11 对探针信号值作出的统计推断, 各符号意义为, A: 不表达, P: 表达, NC: 没有变化, D: 表达下降, MD: 中度降低; 3, 改变量=二倍体的表达量/单倍体的表达量; 4, 功能是通过 interpro 比对得出的

#### 2.5 RT-PCR 验证

由于单倍化后基因芯片中发生表达变化的序列较多,因此根据沉默和激活的不同、序列的功能、片段的大小和表达量的多少初步选取 37 个序列(17 个沉默的,20 个激活的)来进行 RT-PCR 的表达验证,发现有 83.78%的验证结果和芯片的数据表达情况一致.这很可能是由于芯片分析中不存在 PCR 扩增,因而对低丰度表达的 mRNA 检出率不高.为了减少误差,特只选择极端的表达方式及沉默和激活进行分析.这里只初步选择与芯片结果一致的 7 个位点,1~3 为激活序列,4~7 为沉默的序列,具体扩增结果见表 10和图 1,2. 运用 RT-PCR 方法验证这些位点在芯片中的表达情况、发现在叶片中一致率达到 91.86%.

表 10 不同引物在单倍体与二倍体的叶中的 RT-PCR 扩增 结果

引物	A	A	В	С	D	Е	F	G
1	0	0	1	1	1	1	1	1
2	0	1	1	1	1	1	1	1
3	0	0	1	1	1	1	0	0
4	1	0	0	0	0	0	0	0
5	1	0	0	1	0	0	0	0
6	1	0	0	0	0	0	0	0
7	1	0	0	0	0	0	0	0
内参	1	1	1	1	1	1	1	1

注: 1: 有带, 0: 无带. 1~3 为单倍体特异激活序列的引物, 4~7 为单倍体特异沉默序列的引物

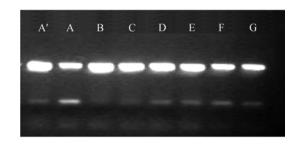


图 1 内参在 A 和 A~G 中的扩增情况

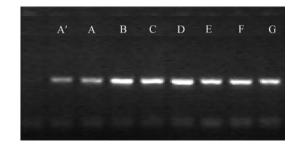


图 2 2 号序列在 A 和 A~G 的扩增结果

# 3 讨论

# 3.1 单倍化后植株在 RNA 水平发生的激活和沉默

本实验中,发现SAR -628 在单倍化后确实发生了一定比例的激活与沉默表达变化,且激活位点数多于沉默位点数(表 2). 激活代表新的生物性状或功能的出现或加强,沉默则代表新的生物性状或功能的消失或减弱,激活多于沉默对于新性状与物种的形成具有重要的意义.

但本研究与其他同源多倍体如酵母[10]和拟南芥<sup>[5]</sup>的表达差异变化比较存在两点明显的不同:一是 SAR -628 沉默与激活变化位点比例更大,二是激活位点数多于沉默位点数.

首先,从基因数量上来看,水稻共有 59712 个预计基因,而作为双子叶植物代表的拟南芥,估计有 20000 个基因,酵母是最简单的真核生物,共有 6400 条可读框,因此水稻在基因数目上的优势表明一旦倍性发生变化,其可供调节改变的基因数量就多.

其次,水稻与拟南芥、酵母基因组的 GC 含量分布差异明显. GC 含量高的区域,基因密度高,重组频率也高.由于基因的高密度分布,会形成局域化的基因簇,从而更多的分布在染色体臂上.遗传学研究已

有结论, 染色体臂的重组频率比染色体的中间区域 高的多, 这样, 染色体臂比中间区域有更快的演化速 度。因此基因在染色体的不同区域的安排就有重要 的生物学意义, 研究发现, 保守的真核生物基因, 如 在酵母中, 那些结构上相似的所谓核心基因, 多分布 在常染色体的中间区域: 这就保证了这些基因的相 对稳定性, 从而也保证了这些生物生命的相对稳定 性. 水稻基因内GC含量具有以下特点: (i) 大部分GC 含量变异在基因内; (ii) 水稻基因至少有一个高 GC 含量外显子; (iii) GC 含量梯度明显. 在蛋白质编码区, 沿着转录方向, 从 5 端开始, 水稻基因组 GC 含量有 一个负梯度, 5 端 GC 含量比 3 端高. GC 含量梯度在 基因内差异明显, 虽然也有少数表现为零梯度, 但没 有基因表现为正梯度, 相反, 拟南芥没有可比的 GC 含量差异. 检测成百上千个其同源序列, 水稻 GC 在 编码区所有位置上均等同或超过拟南芥. 所以水稻 相对于拟南芥和酵母, 面对外界的共同影响, 会更易 发生基因的表达变化.

本研究的激活位点多于沉默位点的现象与其他 同源多倍体的表达结果正好相反, 说明水稻植株单 倍化后只具有一套基因组,为了生存,必然要在基因 表达上做出更大的调整. 因而植株在倍性变化后当 代就已经为性状选择提供了原料, 如果这些新出现 或消失的性状在以后选择中能成功适应环境而保留 下来, 就能形成具有与祖先种不同性状的新物种, 水 稻单倍化"积极主动"的调整自身基因表达变化对新 物种形成以及适应环境与自身剧烈变化(如染色体组 的减少)具有重要意义. 为了加强对自身生存有利的 性状和功能的表达,减弱对生存不利因素的影响,单 倍化后出现激活位点多于沉默的位点是可以理解的. 但是, 对于单倍体在整个生长过程中, 这些发生激活 和沉默的位点的比例是否随之发生不同变化, 发生 激活和沉默变化的基因具体控制哪些功能和性状, 单倍体植株又是如何通过控制这些激活和沉默的位 点的表达来适应生存的, 目前还不清楚, 需要在以后 的研究中继续探索.

#### 3.2 倍性敏感序列与成簇分布

具有相关功能的基因成簇分布的现象在植物中是十分普遍的. Cai和Morishima<sup>[18]</sup>在普通野生稻和栽培稻中均发现与植物驯化相关的QTLs性状聚集在染

色体特定的区域.在许多玉米基因定位资料中也发现QTLs性状成簇现象.本研究中通过基因芯片扫描得到的在染色体上有明确定位的 1358 条倍性敏感序列,在水稻的 12 条染色体上均有分布,并且差异不大,但是发现有 16 组序列成簇分布在 7 条染色体上,虽然对这 33 条序列的功能还没有定论,但可以推断它们之间有着功能上的联系.这些功能相关的基因定位在一个基因簇的现象可解释功能基因网络的适应意义.它是在漫长的进化过程中,有机体发展的将遗传信息进行有效组织的机制.一方面,基因与基因之间通过各种特异的相互作用形成有序的调节网络,另一方面,一些基因成簇排列或位于同一染色质环境,协调基因间的表达.这种基因成簇排列的现象有利于结构功能相关基因的行为协调一致,是遗传信息组织的一种高级形式.

# 3.3 小 RNA 对 SAR -628 倍性特殊表达变化的 解释

多倍体基因表达的活跃研究激发了科学家去了解控制多倍体基因表达的调控机制.在众多综述中,不同作者对多倍体基因表达调控机制提出了自己的构想[19.20],但这些设想缺乏实验证据.同源多倍体中基因表达变化可以从多个角度进行解释,如根据基因表达与染色体或染色体臂的相关关系,一些学者提出用基因组剂量效应与剂量补偿来解释倍性效应对基因表达的影响[21.22].另外,多(单)倍化中,细胞体积变化也被认为是酵母基因表达在倍性间变化的

原因[23]. 不同于其他多倍体的表达变化, SAR -628 单倍体和三倍体(结果待发表)在基因芯片和RT-PCR 研究中, 都表现出较大的特殊性, 如激活多于沉默以及与基因组剂量效应的背离等. 这些令人费解的现象似乎可以通过小RNA干扰进行较合理的解释.

在基因表达倍性调控机制中,小RNA参与表达调控的实验证据较为明确. RNA介导的基因表达调控的证据来自于小麦<sup>[24]</sup>. 逆转座子的激活可以引起基因沉默, 其作用方式是通过转录逆转座子时发生"读过"现象, 因而转录了下游部分邻近基因的反义序列(小RNA)后, 通过RNA干涉而沉默相应基因<sup>[24]</sup>. 在某些情况下, 逆转座子也可激活下游一些不表达基因, 其作用方式与基因沉默相反, 即"读过"后转录了与下游基因的同义RNA序列而激活相应基因. 拟南芥转录组分析也表明其中含有大量未知功能的正义或反义RNA<sup>[25]</sup>. 同源四倍体拟南芥中转基因表达被倍性水平所调控, 转入的基因在二倍体和同源四倍体中沉默, 而在三倍体中被重新激活了<sup>[26]</sup>. 内源基因的表达水平也在玉米偶数和奇数倍性间变化<sup>[22]</sup>.

由以上的分析可以设想,在同源三倍体与单倍体中由转座子"读出"所形成的干扰 RNA 都是奇数倍性的,由于双链 RNA 才对目标基因进行有效的抑制,因而所形成的小RNA 很可能起正向调节作用.这可以解释为什么单倍体与三倍体中激活多于沉默,也可以解释为什么增加与减少一组染色体对 SAR -628 同一目标基因调节方向大致相同.

## 参考文献 \_\_\_\_\_

- 1 Wendel J F. Genome evolution in polyploids. Plant Mol Biol, 2000, 42: 225—249 DOI
- 2 Adams, K.L. Evolution of duplicate gene expression in polyploid and hybrid plants, J. Heredity, 2007, 98(2): 136—141[DOI]
- Pichersky E, Soltis D, Soltis P. Defective chlorophyll a/b-binding protein genes in the genome of a homosporous fern. Proc Natl Acad Sci USA, 1990, 87(1): 195—199[DOI]
- 4 Gastony G J. Gene silencing in a polyploid homosporous fern: paleopolyploidy revisited. Proc Natl Acad Sci USA, 1991, 88(5): 1602—1605[DOI]
- Wang J, Tian L, Madlung A, et al. Stochastic and epigenetic changes of gene expression in Arabidopsis polyploids. Genetics, 2004, 167 (4): 1961—1973[DOI]
- 6 Lee H S, Chen Z J. Protein-coding genes are epigenetically regulated in Arabidopsis polyploids. Proc Natl Acad Sci USA, 2001, 98(12): 6753—6758[DOI]
- 7 Kashkush K, Feldman M, Levy A A. Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. Genetics, 2002, 160(4): 1651—1659

- 8 Adams K L, Cronn R, Percifield R, et al. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc Natl Acad Sci USA, 2003, 100(8): 4649—4654[DOI]
- 9 Albertin W, Balliau T, Brabant P, et al. Numerous and rapid nonstochastic modifications of gene products in newly synthesized Brassica napus allotetraploids. Genetics, 2006, 173(2): 1101—1113[DOI]
- 10 Galitski T, Saldanha A J, Styles C A, et al. Ploidy regulation of gene expression. Science, 1999, 285(5425): 251—254[DOI]
- 11 张红宇, 彭海, 李云, 等. 水稻基因组 DNA 胞嘧啶甲基化在单倍体和对应二倍体间的差异研究. 科学通报, 2006, 51(13): 1529—1535
- 12 Chen Z J, Ni Z. Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. Bioessays, 2006, 28(3): 240—252[DOI]
- 13 Chen Z Z, Xue C H, Zhou F F, et al., GoPipe: streamlined gene ontology annotation for batch anonymous sequences with statistics. Prog Biochem Biophys, 2005, 32(2): 187—191
- Wang J, Tian L, Madlung A, et al. Stochastic and epigenetic changes of gene expression in *Arabidopsis polyploids*. Genetics, 2004, 167(4): 1961—1973[DOI]
- Pques F, Haber J E. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. Microbiol Mol Biol Rev, 1999, 63: 349—404
- Wang J, Tian L, Madlung A, et al. Stochastic and epigenetic changes of gene expression in *Arabidopsis polyploids*. Genetics, 2004, 167: 1961—1973[DOI]
- 17 Comai L, Tyagi A P, Winter K, et al. Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis allotetraploids*. Plant Cell, 2000, 12: 1551—1568[DOI]
- 18 Cai H W, Morishima H. QTL clusters reflect character associations in wild and cultivated rice. Thero Appl Genet, 2002, 104: 1217—1228[DOI]
- 19 Chen Z J, Ni Z. Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. Bioessays, 2006, 28(3): 240—252[DOI]
- Osborn T C, Pires J C, Birchler J A, et al. Understanding mechanisms of novel gene expression in polyploids. Trends Genet, 2003, 19(3): 141—147[DOI]
- 21 Guo M, Davis D, Birchler J A. Dosage effects on gene expression in a maize ploidy series. Genetics, 1996, 142(4): 1349—1355
- Guo M, Birchler J A. Trans-acting dosage effects on the expression of model gene systems in maize aneuploids. Science, 1994, 266(5193): 1999—2002[DOI]
- 23 Galitski T, Saldanha A J, Styles C A, et al. Ploidy regulation of gene expression. Science, 1999, 285(5425): 251—254[DOI]
- 24 Kashkush K, Feldman M, Levy A A. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nat Genet, 2003, 33(1): 102—106[DOI]
- Yamada K, Lim J, Dale J M, et al. Empirical analysis of transcriptional activity in the *Arabidopsis* genome. Science, 2003, 302(5646): 842—846[DOI]
- 26 Mittelsten Scheid O, Jakovleva L, Afsar K, et al. A change of ploidy can modify epigenetic silencing. Proc Natl Acad Sci USA, 1996, 93: 7114—7119[DOI]