

水稻第 6 染色体 S5 区重叠群构建、基因注释与 *OS-APH* 的克隆分析^{*}

王宏斌, 刘 兵, 黎 茵, 冯冬茹, 何炎明, 戚康标, 王金发

(中山大学有害生物控制与资源利用国家重点实验室 / 基因工程教育部重点实验室 // 生命科学学院, 广东 广州 510275)

摘 要: 以通过图位克隆所获得的 Cosmid 克隆 R2H9 序列信息为基础, 构建了一个长为 586 kb 的粳稻第 6 染色体 S5 座位的 BAC 重叠群 (JS5 BC)。通过生物信息学方法, 对 JS5 BC 进行了基因注释, 确定该区域含有 46 个基因位点。经功能预测, JS5 BC 重叠群中存在 3 个主要的基因家族, 分别是木葡聚糖岩藻糖基转移酶、脂酶和黄素氧化还原酶。JS5 BC 中注释基因的一个显著特点是功能相关基因密集存在, 如与植物细胞壁的合成代谢相关的 3 个基因、与呼吸链能量代谢的黄素氧化还原酶的 4 个基因等都是如此。为了验证基因注释的可靠性, 克隆了广亲和品种 Cp17 品种的 *OSAPH* 基因的 cDNA, 并检测了该基因的表达模式, 为注释基因提供了分子证据。

关键词: 水稻 *Oryza sativa* L; S5 座位; 基因注释; 酰胺水解酶; 重叠群

中图分类号: Q751 **文献标识码:** A **文章编号:** 0529-6579 (2006) 06-0067-05

栽培稻粳亚种间杂交子代具有较强的杂种优势: 产量高、适应性广、抗逆能力强等。但粳粳杂种普遍存在着不育或半不育现象。

Ikehashi 和 Araki 根据不同生态类型水稻品种间的杂交研究, 发现一些品种分别与粳、粳不同亚种杂交时, 都能产生正常可育的杂种 F₁ 的遗传现象, 从而提出了广亲和品种 (Wide compatibility varieties WCV) 的概念^[1]。他们把控制广亲和性的基因称为广亲和基因, 并将广亲和基因定位于水稻第 6 染色体上的 S5 座位。

为了研究和克隆 S5 座位的功能基因, 国内外许多研究者先后开展了 S5 座位的分子标记定位研究。Liu 等^[2]报道, S5 基因座与分子标记 R2349 紧密连锁, 二者之间相距仅 1.0 cM。最近, Qiu 等^[3]将 S₅^a 定位在一个 40 kb 的区域内, 而 Qing 等^[4]将 S₅^a 定位在一个 50 kb 的区域内。

我们为了通过图位克隆策略分离水稻广亲和基因, 构建了水稻广亲和品种的 Cosmid 文库, 以分子标记 23D12 的 R 末端为探针从 Cosmid 文库中筛选到一阳性克隆 R2H9^[5], 并进行了生物信息学分析^[6]。利用水稻基因组计划的信息资源, 构建了 S5-BAC 重叠群, 并分析了 S5 区的分子进化机制, 提出了不对称假说^[7], 同时对 S5-BAC 重叠群进行

了详细的基因注释, 在此基础上通过对水稻酰胺水解酶基因的克隆分析对注释基因进行了分子验证。本文报道相关研究结果。

1 材料和方法

1.1 材料

1.1.1 菌株和质粒 大肠杆菌菌株 BL21 (DE3) 和 DH5 α 由本实验室保存。pBluescript SK (M13 -), 由本实验室保存。pYLTAC17 载体由华南农业大学刘耀光教授惠赠。

1.1.2 试剂与材料 各种限制性内切核酸酶、T₄ DNA 连接酶、Taq DNA 聚合酶、高保真 Pfu DNA 聚合酶均购自 TaKaRa 公司; 质粒提取试剂盒、胶回收试剂盒、琼脂糖购自北京鼎国生物公司; PCR 扩增引物的合成和 DNA 测序由上海 Bioasia 公司完成。一步法 RT-PCR 试剂盒购自 CbneTech 公司; 其它常用试剂均为分子生物学级或分析纯。

1.2 实验方法

1.2.1 Cosmid 克隆 R2H9 的染色体电子定位 以根据图位克隆方法所获得的阳性 Cosmid 克隆 R2H9 序列信息为基础, 利用 BLASTN 程序, 针对 TIGR (The Institute of Genomic Research) 的 “All Rice BAC and PAC sequences in GenBank” 数据库

* 收稿日期: 2006 08 03

基金项目: 国家自然科学基金资助项目 (30370809); 中山大学青年教师科研启动基金资助项目 (2005-33000-1131226)

作者简介: 王宏斌 (1977 年生), 男, 博士, 讲师; 通讯联系人: 王金发, E-mail: lsswjf@mail.sysu.edu.cn

进行检索，确定 R2II9所对应的水稻 BAC PAC 克隆。进一步根据水稻物理图谱的信息，确定该 BAC PAC 克隆在水稻染色体的定位信息。

1.2.2 JS5-BC 重叠群 的电子构建^[7] 以所确定的 BAC PAC 克隆靠近 S5 座位的末端（选取长度为 5 kb 的序列）为检索源，采用上述策略进行下一轮的数据库检索，逐步向 S5 座位方向进行电子延伸。如此往复，共进行多轮检索，利用 Bioedit 软件将所获得的 BAC 克隆进行拼接，构建了约 586 kb 的水稻 S5 座位的 BAC 重叠群——JS5-BC（japonica S5 BAC contig JS5-BC）。

1.2.3 重叠群 JS5-BC 的基因注释 通过检索水稻 EST 数据库与全长 cDNA 数据库的策略，对 JS5-BC 进行了基因注释。利用 Bioedit 软件，寻找 JS5-BC 中各个 FL-cDNA 序列的中包含的 ORF，选取最长 ORF 所对应的氨基酸序列，分别检索 NCBI 中 NCBI Conserved Domain Search^[8]，根据检索结果，推测该 FL-cDNA 的功能。

1.2.4 注释基因 *OS-APH* 的克隆 根据对 JS5-BC 的基因注释结果，发现第 38 号注释基因（AK106301）所编码的蛋白质与拟南芥酰肽水解酶（一类脂酶）注释基因同源，可能是编码水稻酰肽水解酶候选基因，将其命名为 *OS-APH*。根据 AK106301 的序列信息，设计如下两条引物用来扩增 *aph* cDNA 全长：

OSAPH1F: CTCGCCTCCATCTCACCGCACGCC
OSAPH18R: GCCCCAGACTCCAACATIGATTGATTTT

以水稻广亲和品种 Cpsb17 叶总 RNA 为模板进行 RT-PCR（按 Clontech 公司试剂盒操作）。回收 PCR 产物，亚克隆后进行序列测定和分析。

1.2.5 *OS-APH* 表达谱分析 为了研究注释基因在水稻不同组织与发育时期的表达情况，我们采用 RT-PCR 方法，以水稻广亲和品种 Cpsb17 为材料进行检测。所选取的材料基本覆盖了水稻发育的不同时期和不同组织，包括：幼根、幼叶、老根、老叶、茎、小穗（<3 dM）、小穗（>3 dM）、花粉、开花期花序、成熟花序。

2 结果与分析

2.1 Cosmid 克隆 R2II9 的染色体电子定位

在我们前期的研究中，通过酶切图谱分析以及 Southern 杂交实验，初步确定了 R2II9 为接近 S5 区的阳性克隆。进一步对所构建亚克隆 TRW 1510 和 TRW 1517 全序列测定，然而，我们并没有能将 TRW 1510 和 TRW 1517 进行序列的拼接^[6]。本研究

中，我们利用水稻全序列数据库资源，对 R2II9 进行了重新的序列比较分析。发现 R2II9 是一个嵌合体，TRW 1510 和 TRW 1517 不是位于同一个染色体位置，可能是在 Cosmid 基因组文库构建时产生的两个独立基因组片段末端互补配对所致。所幸的是亚克隆 TRW 1517 对应于粳稻第 6 染色体的一个 BAC 克隆。将 23D12R 测序后进行检索分析，确定其序列完全位于亚克隆 TRW 1517 序列之中（图 1）。

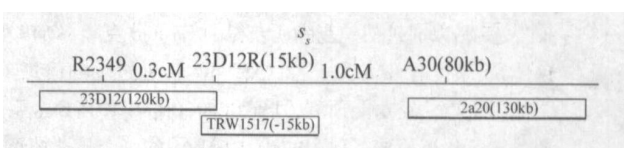


图 1 S5 区的物理图谱以及 R2349 与 TRW 1517 的相对位置关系

Fig. 1 Physical map of S5 region and the relative position of R2349 and TRW 1517

2.2 粳稻 S5 重叠群（JS5-BC）的电子构建

2.2.1 亚克隆 TRW 1517 向 S5 延伸方向的确定 考虑到我们所采用的水稻广亲和材料 Cpsb17 为偏粳型，可以结合 TGR 中的粳稻基因组精细图谱进行 S5 区的电子定位分析。根据 S5 位点区段高密度连锁图中分子标记的相对位置，我们从 NCBI 下载分子标记 R2349 序列，通过比较分子标记 R2349、23D12 和亚克隆 TRW 1517 的位置关系，可以确定 S5 座位位于分子标记 R2349 向亚克隆 TRW 1517 的延伸方向（图 2）。

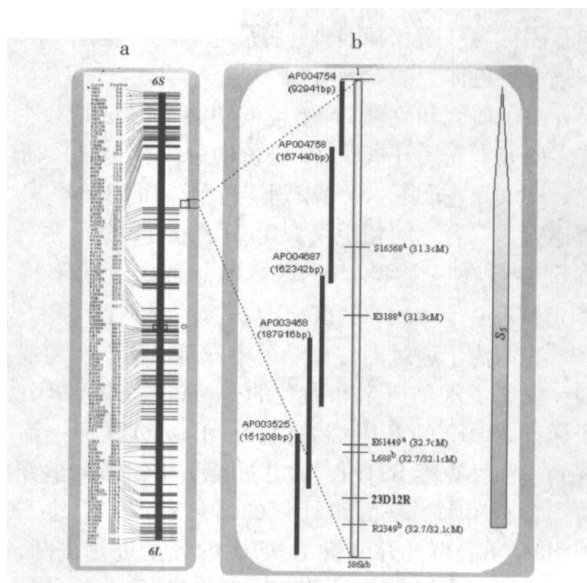


图 2 JS5-BC 的遗传图谱与物理图谱

Fig. 2 Genetic map and Physical map of JS5-BC

a 分子标记来自“BAC Ends Anchored to RGP

Chromosome 6M markers”

b 分子标记来自“YAC-Based Rice Transcript Map”

2.2.2 JS5-BC 的电子构建^[7] 进一步以 AP003525 为电子探针, 在 TGR 的粳稻基因组序列数据库中进行检索, 确定向 S5 座位方向延伸的 BAC 克隆。如此重复 4 次, 分别获得了 4 个 BAC 克隆: AP003458、AP004687、AP004758 和 AP004754。将这 5 个 BAC 克隆进行拼接, 组成了一个粳稻 BAC 重叠群 (BAC contig), 命名为 JS5-BC (Japonica S5 BAC contig), 并确定了位于该区段中的 5 个分子标记以及它们所对应的遗传图距 (图 2)。

2.3 JS5-BC 的基因注释

将 JS5-BC 检索水稻 FL cDNA 数据库, 确定了 43 个 FL cDNA 位点。为了做全面的基因注释, 又将 JS5-BC 在 NCBI 的中检索水稻 EST 数据库。基于 EST 检索以及 FL cDNA 检索及定位的基础上, 通过双向验证的检索策略, 排除了冗余的情况后, 一共确定 JS5-BC 中的 46 个基因位点, 并进行编号。将 46 个基因位点进行了功能预测、归类分析 (表 1)。

表 1 JS5-BC 中注释基因位点的功能分类
Tab.1 Functional classification of the annotated genes in JS5-BC

功能分类	JS5-BC 中的 基因位点	基因位 点总数
细胞代谢		13
脂酶	19 32 33 38 39	5
植物细胞壁合成	10 24 25 26	4
激酶	4 17	2
异构酶	46	1
脱酰酶	16	1
能量代谢		6
NADH	35 40 41 42 43	5
质体蓝素	45	1
转录因子	8 11 20	3
促进转运	15 18	2
分子伴侣	27	1
发动蛋白	2	1
未分类	1 3 5 6 29 31 36 44	8
非全长 ORF	9 13 21 22 30 37	6
未完全测序	7 12 14 23 28 34	6
合计		46

2.4 注释基因 *OsAPH* 的克隆及时空表达模式检测

2.4.1 注释基因 *OsAPH* 的克隆 以与 38 号注释基因同源的 cDNA 克隆 AK106301 的序列设计两端引物, 以水稻品种 Cpsb17 叶总 RNA 为模板, 通过 RT-PCR 方法克隆该基因。所克隆的 *OS-APH* 基因序列全长 2332bp, 与 AK106301 克隆相比, 只是第 10 外显子稍长, 其他部分基本相同。

通过 Bioedit 软件分析, 发现该基因内部有一个长 1 827 bp 的可读框。经 Blast 检索 Genbank 蛋白数据库, 发现该基因所编码的蛋白质与拟南芥酰肽水解酶样蛋白 (acylpeptide hydrolase like protein) 在氨基酸水平上具有 64% 的一致性。

2.4.2 *OsAPH* 的时空表达 为了研究 *OsAPH* 基因在水稻广亲和品种 Cpsb17 不同组织与发育时期的表达情况, 我们采用 RT-PCR 方法, 检测了该基因的表达模式, 结果表明: *OsAPH* 基因在水稻的各个组织及不同的发育时期均有表达, 但在在花粉组织中的表达异常, 比其余时期的表达产物为小 (图 3)。

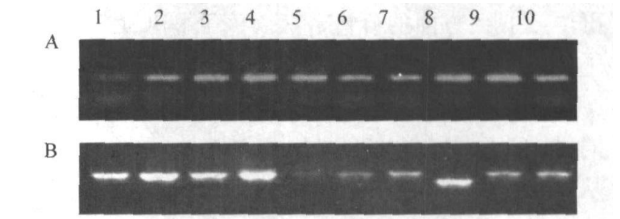


图 3 RT-PCR 检测 *OsAPH* 在水稻广亲和品种 Cpsb17 中的时空表达特模式检测

Fig. 3 Analysis of the expression pattern of *OsAPH* gene in Cpsb17 by RT-PCR
A: *Osact1* 引物扩增结果; B: *OS-APH* 引物扩增结果
1-10 幼根、幼叶、老根、老叶、茎、小穗 (< 3 dM)、小穗 (> 3 dM)、花粉、开花期花序、成熟花序

我们将花粉期差异表达的 cDNA 亚克隆及测序后表明, *OS-APH* 基因在花粉组织表达出现了可变剪接, 第 8 外显子发生了缺失, 导致原 ORF 提前终止 (图 4)。通过分析 *OsAPH* 各外显子-内含子剪接位点及其附近的序列, 并与一致保守的单子叶植物的外显子-内含子剪接位点比较, 发现所有剪接位点均符合通用的 GTAG 规则, 并且部分符合单子叶植物的一致序列^[9], 同时, 在外显子 8 的剪接位点附近并未发现序列异常。这一结果提示, *OS-APH* 基因的表达模式可能与水稻的花粉发育有关, 其变剪接的机制, 值得进一步研究。



图 4 *OS-APH* 基因外显子结构图示
Fig. 4 Exon organization of the *OS-APH* gene
E1-E18 外显子, E8 在花粉中发生可变剪接

3 讨论

3.1 关于 JS5-BC 的跨度问题

由于水稻第 6 染色体的 S5 区与水稻广亲和基因相关、分子标记稀少、图距离单位特别等特点,引起育种学家和分子生物学家的极大兴趣,采取了不同的研究策略进行研究。我们采取的策略是先研究整体,尔后研究细节,即先研究 S5 区,再研究 S5 区内的相关基因及位点。

根据 Liu 等^[2]对 S5 座位精细定位的遗传图谱,与 S5 座位最接近的分子标记为 R2349 相距 S5 座位约 1.0 cM。我们构建了一个由 5 个 BAC 克隆组成的总长度达 586 kb 的粳稻 BAC 重叠群——JS5-BC。根据 3 方面的分析,认为该 JS5-BC 已经覆盖了 S5 座位。

首先,粳稻基因组序列中遗传图距与物理图距的对应关系在 214 ~ 288 kb cM⁻¹^[10],而 JS5-BC 中从 R2349 向 S5 座位延伸了 540 kb,至少跨越了 1.5 cM。其次,根据现有的水稻高密度连锁图,JS5-BC 中包含了 5 个分子标记,其中间距最远的是 S16568 (31.3 cM) 与 R2349 (32.7 / 32.1 cM),在 JS5-BC 中它们之间物理距离 342 kb,相对的遗传距离为 1.4 / 0.8 cM,则推测该区域遗传图距与物理图距的对应关系为 244 / 427 kb cM⁻¹。以此为标准,JS5-BC 从 R2349 向 S5 座位延伸了 2.2 / 1.3 cM。若以 E6119 和 E3188 之间的图距对应关系换算,则 110 kb cM⁻¹。最后, R2349 在第 6 染色体的遗传图谱定位为 32.7 / 32.1 cM,而组成 JS5-BC 中最远的 BAC 克隆 AP004754 在 TGR 中定位为 19.1 cM,因此 JS5-BC 对应的遗传距离也大于 1.0 cM。

最近, Qiu^[3] 和 Qing 等^[4] 分别将 S5n 定位在一个 40 kb 和 50 kb 的区域内,有趣的是,这一区域也在我们构建的重叠群中 (360 ~ 410 kb),距离 23D12R 约 150 kb,也就是说该区域 1.0 cM 约相当于 150 kb,在我们根据已有的分子标记所进行的遗传图距的分析范围内。

3.2 JS5-BC 基因注释的一个分子证明——OS-APH 基因的克隆分析

虽然通过基因注释,初步确定了 S5 区的 46 个基因或位点,但是,注释基因的可靠性有多大,这需要进行分子水平的验证。为此,我们选择了 38 号注释基因进行了克隆分析。克隆该基因,主要基于两方面的考虑。

第一,在经典遗传图上 S5 座位与一个同工酶标记 Est 2 (一种脂酶) 紧密连锁^[11],而 JS5-BC

重叠群则是根据分子遗传图谱与 S5 连锁的分子标记为基础构建的,因此,如果能够证明该同工酶是 JS5-BC 中的一个注释基因,则从分子水平上为注释基因提供了证据。有趣的是,根据我们对 JS5-BC 的基因注释和功能预测的结果,发现第 38 号基因位点可能是编码水稻酰肽水解酶 (也是一类脂酶) 候选基因。

第二,酰肽水解酶 (acylpeptide hydrolase APH) 属于一类新的丝氨酸蛋白酶家族。该酶能够催化 N 末端酰基化的肽段水解以释放出 N 末端酰化的氨基酸。在高等真核生物细胞中,细胞内合成蛋白质的 N 乙酰化被认为是一种重要的保护机制^[12]。据报道,哺乳动物细胞中被 N 酰基化修饰的蛋白质高达 80% ~ 90%,酵母中为 50%,原核细胞中则很少,植物中的情况还不清楚^[13]。

目前的研究表明, N -酰基的去除是蛋白质活性调控过程的重要步骤,酰肽水解酶通过其对蛋白质的去酰基化作用从而调控某些重要的生理生化过程。因此,克隆分析水稻酰肽水解酶基因,不仅可为 JS5-BC 注释基因提供证据,同时可通过该基因的研究进一步探讨 S5 区的功能。

3.3 JS5-BC 重叠群构建及基因的意义

通过构建 JS5-BC 重叠群、基因注释和功能预测,获得了水稻第 6 染色体 S5 区的一些重要信息。

首先,我们发现 JS5-BC 中存在 3 个主要的基因家族,分别是木聚糖岩藻糖基转移酶 (XG_Ftase, Xyloglucan fucosyltransferase)、脂酶 (Esterase / lipase) 和黄素氧化还原酶 (flavin oxidoreductases),前两个与细胞代谢过程相关 (其中 XG_Ftase 与植物细胞壁的合成代谢有关),第三个则参与能量代谢。特别是参与细胞壁合成的基因家族的存在,引起我们极大的兴趣,因为植物的育性与细胞壁的合成有很大的关系。

第二,JS5-BC 中的基因位点的显著特点是基因位点分布不均匀,功能相关基因密集存在,例如参与呼吸链能量代谢的黄素氧化还原酶的 4 个基因就位于 528 kb ~ 570 kb 的 42 kb 的范围内,提示基因在基因组中的位置与其功能有一定的相关性。

第三,在 FL cDNA 数据库中参与能量代谢的基因位点所占比例仅为 1.1%,而在 JS5-BC 中却高达 13%,这与 Cui 等在对某些植物不育现象的研究中发现线粒体呼吸链组分的异常破坏,可能是造成败育的主要原因之一^[14]的说法是否有某种联系?

上述的信息,引起我们思考这样一个问题:水稻的广亲和性是由单个基因决定的?还是由一个社

区基因 (block genes) 共同作用的结果? 根据基因注释的信息, 我们正在对其中一些重要的基因位点进行较深入的研究。

参考文献:

[1] IKEHASHI H. ARAKI A. Genetics of sterility in recom cross of rice[J]. In Rice Genetics (IRR I), 1986 119 - 130

[2] LIU K D. WANG J. LIU H B. et al. A genome wide analysis of wide compatibility in rice and the precise location of the S5 locus in the molecular map[J]. Theor Appl Genet 1997 95 809 - 814

[3] QIU S Q. LIU K D. JIANG J X. et al. Delimitation of the rice wide compatibility gene S5n to a 40 kb DNA fragment[J]. Theor Appl Genet, 2005 111 1080 - 1086

[4] QING J. LU J F. CHAO Q. et al. Delimiting a rice wide compatibility gene S5n to a 50 kb region[J]. Theor Appl Genet 2005 111 1495 - 1503

[5] 易厚富, 刘兵, 范云, 等. 水稻广亲和品种核 DNA Cosmid文库的构建和鉴定 [J]. 热带亚热带植物学报, 2001, 9 185 - 189

[6] 王宏斌, 刘兵, 易厚富, 等. 水稻 S5 区侯选克隆 R2119 的筛选及序列信息学分析 [J]. 中山大学学报: 自然科学版, 2002 41(6): 78 - 82

[7] WANG H B. YU L J. LAI F. et al. Molecular evidence for asymmetric evolution of duplicated sister regions after rice polypoidy [J]. Plant Molecular Biology 2005 59

63 - 74

[8] MARCHLER B A. ANDERSON J B. DEWEESE S. et al. CDD: a curated Entrez database of conserved domain alignments [J]. Nucleic Acids Res 2003 31 383 - 387

[9] SM PSON G G. FILIPOW ICZ W. Splicing of precursors to mRNA in higher plants: mechanism, regulation and sub nuclear organisation of the spliceosomal machinery [J]. Plant Molecular Biology 1996 32 1 - 41.

[10] GOFF S A. REKE D. LAN T H. et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp *japonica*) [J]. Science 2002 296 92 - 100

[11] CAUSSE M A. FULTON T M. CHO Y G. et al. Saturated molecular map of the rice genome based on an interspecific backcross population [J]. Genetics 1984 138 1251 - 1274.

[12] SENTHILKUMAR R. SHARMA K K. Effect of chaotropic agents on the structure function of recombinant acylpeptide hydrolase [J]. J Protein Chem 2002 21 323 - 332

[13] PERRIER J. GIARDNA T. DURAND A. et al. Specific enhancement of acylase I and acylpeptide hydrolase activities by the corresponding N- acetylated substrates in primary rat hepatocyte cultures [J]. Biol Cell 2002 94 45 - 54.

[14] CUI X. WISE R P. SCHNABLE P S. The r2 nuclear restorer gene of male sterile T- cytoplasm maize [J]. Science 1996 272 1334 - 1336

Construction and Annotation of the JS5-BC Contig of Rice Sixth Chromosome S5 Locus and OS-APH Gene Analysis

WANG Hong-bin, LU Bing, Li Yin, Feng Dong-nu, HE Yan-ming, QI Kang-biao, Wang Jin-fu
(The State Key Laboratory of Biocontrol and The Key Laboratory of Gene Engineering of Ministry of Education, School of Life Science, Sun Yat-sen University, Guangzhou 510275, China)

Abstract Based on the sequences information of the rice Cosmid clone R2119, the JS5-BC contig (BAC contig of Japonica rice S5 locus) had been constructed. The length of JS5-BC was 586 kb and covering the S5 locus, as determined from the relative map of molecular marker around the S5 locus. After annotation of JS5-BC, 46 gene loci were predicted. Among the 26 genes with predicted function, there were three gene families mainly: xyloglucan fructosyltransferase gene family, esterase/lipase gene family and flavin oxidoreductases gene family, respectively. A remarkable characteristic of annotation genes in the JS5-BC was the crowded existence of function correlation genes, for example, three genes related to plant cell wall synthesis metabolism, four genes related to breath chain energy metabolism and so on. To verify the annotation of JS5-BC, one of the annotated genes, OS-APH, was isolated from Cps10 17 and analyzed.

Key words rice (*Oryza sativa* L.); S5 locus; gene annotation; acylpeptide hydrolase (APH); contig