

· 研究报告 ·

水稻糯性相关基因的全基因组关联分析

杨行海^{1,2}, 农保选^{1,2}, 夏秀忠^{1,2}, 张宗琼^{1,2}, 曾宇^{1,2}, 刘开强³, 邓国富³, 李丹婷^{1,2*}

¹广西农业科学院水稻研究所, 南宁 530007; ²广西水稻遗传育种重点实验室, 南宁 530007; ³广西农业科学院, 南宁 530007

摘要 利用1 090 071个SNP标记, 对419份广西地方稻种资源核心种质的糯性进行全基因组关联分析。结果表明, 运用混合线性模型, 在 $P < 4.72 \times 10^{-8}$ ($4.72E-8$)水平下, 检测到45个与糯性显著关联的SNP位点, 均位于第6号染色体上。通过筛选显著关联位点上、下游各150 kb区域内的基因, 共找到305个候选基因, 其中包含2个与淀粉合成酶相关的*Wx*和*SSIIa*基因; 同时第6号染色体5.69–5.89 Mb区域发现4个与糯性显著关联的SNP位点, 该区域可能是影响水稻(*Oryza sativa*)糯性的重要候选区域。

关键词 水稻, 全基因组关联分析, 糯性, 候选基因

杨行海, 农保选, 夏秀忠, 张宗琼, 曾宇, 刘开强, 邓国富, 李丹婷 (2016). 水稻糯性相关基因的全基因组关联分析. 植物学报 51, 737–742.

水稻(*Oryza sativa*)是世界上最重要的粮食作物之一, 全世界超过一半人口以稻米为主食。随着生产的发展和人民生活水平的提高, 人们对稻米的品质提出了更高的要求。稻米品质改良已成为水稻研究的热点(毛孝强等, 2003)。目前国内外有关稻米品质性状的评价体系基本相同, 通常包括加工品质、外观品质、蒸煮与食味品质、营养品质及特异品质5个方面(Tan et al., 2000)。

水稻籽粒中除了含少量蛋白质外, 淀粉含量一般为糙米重的90%, 是稻米的主要贮存物质(FAO, 1993)。淀粉由直链淀粉和支链淀粉组成。稻米的淀粉品质与直链淀粉和支链淀粉的含量密切相关(Vandeputte et al., 2003; Nakamura et al., 2005), 其中直链淀粉含量是决定稻米蒸煮与食味品质及加工品质最主要的因素(蔡一霞等, 2005; 何秀英等, 2015)。水稻胚乳中的直链淀粉合成由水稻蜡质基因(*Wx*)编码的颗粒结合淀粉合成酶(GBSS)来控制(Hirano and Sano, 2000)。因此, *Wx*基因是控制水稻直链淀粉含量的主要基因(Wang et al., 1990), 对稻米的蒸煮与食味品质及加工品质起着决定性作用。水稻糯与非糯主要是由该位点中的1对等位基因, 即*Wx*和*wx*的差异引起的。目前, 已在水稻中鉴定了*Wx*的多个等位基因, 如*Wx^a*、*Wx^b*、*Wx^{mq}*、*Wx^{mp}*、*Wxⁱⁿ*和*wx*等(Sato et al., 2002; Mikami et al., 2008; Liu et

al., 2009; Yang et al., 2013)。同时, Tian等(2009)的研究结果表明, *Wx*基因与其它基因间还存在互作, 并发现存在水稻的直链淀粉含量微效应基因*SSII-3*、*SSIII-2*、*AGPlar*、*PUL*和*SSI*。Takemoto-Kuno等(2015)定位了1个能够与*Wx*互作并且控制水稻低直链淀粉含量的QTL——*qAC2*。

核心种质是由现存的种质资源收集品种中的1套数量有限的记录材料组成, 它代表整个收集品种中遗传资源的遗传类型, 包含尽可能多的遗传多样性(Hodgkin, 2012)。目前, 在水稻中利用核心种质开展全基因组关联分析, 鉴定了许多影响水稻重要农艺性状的QTL (Huang et al., 2010, 2012; Zhao et al., 2011; 邱先进等, 2015)。

本研究以419份广西地方稻种质资源核心种质为实验材料, 在2014年早稻和晚稻中对该材料的糯性进行鉴定, 结合全基因组基因型数据解析水稻的糯性遗传基础, 挖掘影响水稻糯性的相关基因, 旨在为解析稻米品质变异的途径和改良稻米品质提供基因及材料。

1 材料与方法

1.1 实验材料

实验材料为本课题组构建的414份广西地方稻(*Oryza sativa* L.)种质资源核心种质(李丹婷等, 2012), 另5份

收稿日期: 2015-11-11; 接受日期: 2016-03-25

基金项目: 广西自然科学基金(No.2015GXNSFBA139054, No.2015GXNSFBA139080)、广西科学研究与技术开发计划(No.桂科重14121001-1-6)、广西八桂学者专项经费(No.桂人才通字[2011]4号)和广西农业科学院优势团队(No.2015YT15)

* 通讯作者。E-mail: lidanting@gxaas.net

为增补材料。将2014年早、晚稻种植在广西农业科学院水稻研究所试验田，待成熟后，当谷粒水分降到15%以下时，观测胚乳。随机挑选饱满的稻谷10粒，去掉颖壳，将胚乳横切后，用1% I₂-KI溶液在断口处染色，呈现棕红色反应为糯，呈蓝色反应为非糯。

1.2 DNA提取、基因组修饰、PCR扩增及测序

采用CTAB法(Rogers and Bendich, 1988)提取样品DNA，利用事先选好的酶切组合对DNA进行酶切，将酶切产物进行5'末端修复，同时对5'末端进行磷酸化修饰，3'末端加A，使之与solexa接头5'端T互补，提高接头连接效率，阻止solexa接头自连。3'和5'端分别完成修饰之后，继续连接solexa测序接头，便于将连接产物锚定在flowcell上，进行桥式扩增。用琼脂糖凝胶电泳进行片段大小选择，通过PCR扩增增大文库量，建好的文库用 Illumina HiSeq TM 2000 进行测序。

1.3 数据分析

对测序得到的原始reads (双端序列)进行数据评估，得到各个样品的原始reads，将reads比对到参考基因组上，基因组为从Ensembl数据库上下载的最新水稻基因组。获得SLAF片段后进行SNPs检测。

基于SNP，利用TASSEL (Zhang et al., 2010)软件的一般线性模型(general linear model)和混合线

性模型(compressed mixed linear model)得到关联值。计算公式如下：

$$Y=X\alpha+Q\beta+K\mu+e$$

其中，样品群体结构Q使用admixture软件计算，样品间亲缘关系K通过SPAGeDi (Hardy and Veke-mans, 2002)软件计算，一般线性模型使用Q群体结构信息，而混合线性模型使用Q+K，即群体结构和亲缘关系的信息。X为基因型，Y为表型。最终每个SNP位点都能得到1个关联值。

2 结果与讨论

2.1 表型统计分析

研究表明，在419份广西地方稻种质资源核心种质中，非糯性材料有311份，占74.2%；糯性材料有108份，占25.8%。

2.2 全基因组关联分析

利用419份水稻品种的简化基因组测序序列，经质控后，获得1 090 071个单核苷酸多态性位点(SNPs)。使用混合线性模型检测关联的SNPs，在 $P<4.72\times 10^{-8}$ (4.72E-8)水平下，共检测到45个与水稻糯性相关联的SNP位点，均分布在第6号染色体上；有2个显著关联区域(图1)，其中有34个点集中分布在1.46-1.96 Mb区域内(表1)。

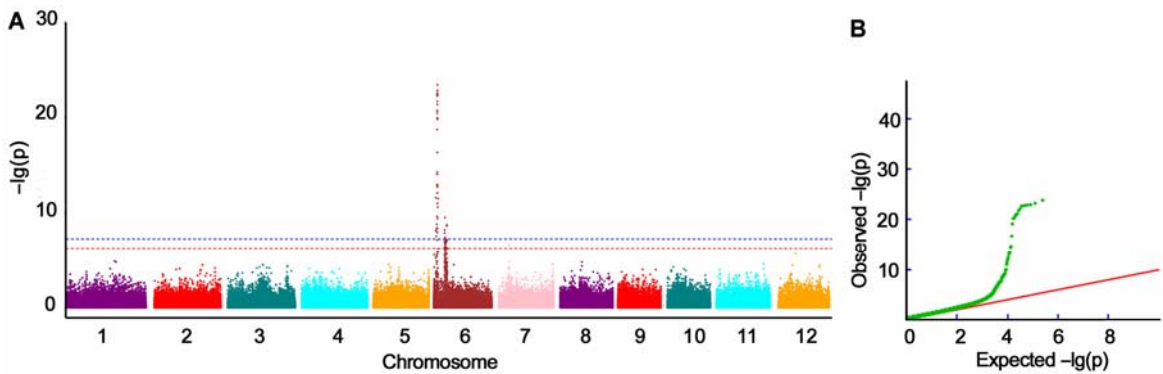


图1 水稻糯性的全基因组关联分析
(A) Manhattan图; (B) QQ图

Figure 1 Genome-wide association studies of waxy traits in rice
(A) Manhattan plots of compresses mixed linear model for waxy trait; (B) Quantitle-quantitle plot of compresses mixed linear model for waxy trait

表1 水稻中与糯性显著关联的SNP位点

Table 1 Significantly associated SNPs of waxiness in rice

Chromosome	Position	Allele	Major allele	Minor allele	P-value	Known loci
6	921039	A, T	455	75	2.32E-12	
6	945908	G, T	455	73	6.02E-09	
6	945924	G, A	455	73	6.02E-09	
6	1179280	C, A	516	32	2.60E-08	
6	1457896	C, A	465	339	1.42E-09	
6	1512757	G, C	381	197	3.57E-09	
6	1512819	A, G	384	198	3.38E-09	
6	1592445	G, C	593	231	1.20E-13	
6	1597045	G, A	470	236	9.37E-09	
6	1607061	T, A	559	247	1.61E-19	
6	1654904	G, T	297	161	2.12E-10	
6	1661829	A, C	587	241	1.69E-21	
6	1684538	A, G	407	171	3.63E-09	
6	1684543	C, T	444	136	4.13E-17	
6	1684784	T, C	431	155	7.26E-13	
6	1694761	G, C	481	231	3.78E-22	
6	1694808	A, G	482	230	9.55E-21	
6	1694818	T, C	455	251	8.14E-15	
6	1703886	T, G	574	264	3.20E-21	
6	1720764	A, G	533	269	4.73E-15	
6	1720799	G, A	561	241	3.70E-23	
6	1720821	C, T	560	240	4.80E-23	
6	1745339	C, T	338	134	1.00E-08	
6	1754453	A, T	561	253	1.75E-22	
6	1756202	T, C	552	264	1.79E-21	
6	1756454	A, G	556	262	2.45E-23	
6	1780293	T, C	482	248	4.08E-23	
6	1780336	C, A	483	249	2.81E-23	
6	1780519	C, T	479	251	3.96E-09	
6	1792217	C, T	557	241	3.29E-24	Wx
6	1797551	C, T	571	253	1.32E-23	
6	1807558	A, G	295	159	9.55E-14	
6	1807737	T, C	272	166	8.98E-12	
6	1807778	G, T	279	159	1.67E-13	
6	1807797	A, G	534	250	1.38E-20	
6	1822395	T, C	525	219	1.12E-20	
6	1865861	G, A	543	137	3.62E-10	
6	1958767	T, C	405	211	2.23E-11	
6	2256752	G, A	578	34	2.85E-12	
6	5688919	T, G	581	211	5.68E-09	
6	5711650	C, T	664	34	1.17E-08	
6	5843510	G, A	763	49	2.79E-10	
6	5881999	T, C	784	48	3.34E-08	
6	6879531	A, G	488	272	1.62E-09	SSI/a
6	6879731	C, G	486	274	2.07E-09	

表2 3个显著关联位点候选基因

Table 2 Summary of 3 significantly associated loci and candidate genes

Position	Candidate gene	Annotation
Chr06_1792217	<i>Os06g0133000</i>	<i>Wx</i>
Chr06_5843510	<i>Os06g0214900</i>	FAD-dependent pyridine nucleotide-disulphide oxidoreductase domain containing protein
Chr06_5843510	<i>Os06g0215000</i>	Hypothetical protein
Chr06_5843510	<i>Os06g0215100</i>	Protein of unknown function DUF1645 family protein
Chr06_5843510	<i>Os06g0214850</i>	Similar to gibberellin receptor GID1L2
Chr06_5843510	<i>Os06g0214800</i>	Alpha/beta hydrolase fold-3 domain containing protein
Chr06_5843510	<i>Os06g0214300</i>	Alpha/beta hydrolase fold-3 domain containing protein
Chr06_6879531	<i>Os06g0229800</i>	<i>SSIa</i>

2.3 候选基因分析

基于水稻基因组注释(RAP-DB, <http://rapdb.dna.affrc.go.jp/>), 依据LD衰退水平, 本研究选取关联SNP位点的上下游各150 kb的区域, 共检测到305个基因, 对其中8个与本研究相关的基因进行注释(表2)。利用所检测到的SNPs候选基因, 我们在P值最高的关联位点Chr06_1792217附近的21.6 kb处找到1个编码颗粒结合淀粉合成酶的*Wx*基因, 并鉴定其是控制水稻籽粒直链淀粉含量的主效基因。在关联位点Chr06_6879531找到*Os06g0229800*, 即*SSIa*基因。*SSIa*基因是控制水稻糊化温度的主效基因, 编码可溶性淀粉合成酶II。在关联位点Chr06_5843510附近注释了6个候选基因, 其中*Os06g0214850*基因与赤霉素受体GID1L2有关。

2.4 讨论

稻米直链淀粉含量是其品质的一个重要性状, 水稻胚乳中的直链淀粉主要由*Wx*基因控制。因此, 挖掘影响糯性的相关基因是解析稻米品质变异的重要途径, 也为改良稻米直链淀粉含量提供了基因和材料。本研究在水稻第6号染色体上检测到45个与水稻直链淀粉含量显著关联的SNP位点, 并成功在显著关联的SNP区域内找到2个与淀粉合成酶相关的*Wx*基因和*SSIa*基因, 这与前人的研究结果一致(Wang et al., 1990; Huang et al., 2010), 进一步说明了本研究结果的可靠性。水稻*Wx*基因通过编码颗粒结合淀粉合成酶来控制稻米的直链淀粉合成(Hirano and Sano, 2000)。Tian等(2009)研究发现, *Wx*对直链淀粉含量和胶稠度具有主效作用, 而对糊化温度具有微效性; *SSI-3*是唯一对糊化温度具有主效作用而对直链

淀粉含量和胶稠度具有微效作用的基因; *SSI-2*、*AGPlar*、*PUL*和*SSI*这4个基因只对直链淀粉含量有微效作用。对*SSIa*基因的研究表明, 其编码可溶性淀粉合成酶II (Nakamura, 1996)。通过对水稻*SSIa*基因的克隆, 发现其只有1个不完整的开放阅读框(高振宇等, 2003)。进而通过对不同水稻品种间该基因的序列进行比较, 测出*SSIa*基因编码区内的碱基替换可能引起了支链淀粉晶体层结构的改变, 从而导致糊化温度的变化(Umemoto et al., 2004; Nakamura et al., 2005)。Zhang等(2011)研究显示, *SSIa*和*SSIla*在水稻籽粒淀粉合成中存在基因间互作。Takemoto-Kuno等(2015)研究表明, *qAC2*是能够与*Wx*互作并且控制水稻低直链淀粉含量的QTL, 可能在*Wx^b*的mRNA剪接过程中发挥功能。

本研究中的另一显著关联SNP位点在水稻第6号染色体5 843 510 Mb上, 该区域*Os06g0214850*基因编码赤霉素(GA)受体GID1L2。已有研究表明, 在小麦(*Triticum sativum*)灌浆期GA能显著提高籽粒中总淀粉和支链淀粉的含量, 降低籽粒的直连淀粉含量(于安玲, 2011)。该基因在水稻基因组中的具体功能尚不清楚, 需要进一步检测和验证。

本研究结果表明, 全基因组关联分析是挖掘水稻农艺性状候选基因的高效手段。基于高密度的SNPs, 全基因组关联分析能挖掘到与性状相关的候选基因, 同时有利于对候选基因进行遗传机制分析。

参考文献

蔡一霞, 朱智伟, 王维, 张祖健, 朱庆森 (2005). 直链淀粉含量与稻米品质主要性状及米饭质地关系的研究. 扬州大学学报(农业与生命科学版) 26, 52–55.
高振宇, 曾大力, 崔霞, 周奕华, 颜美仙, 黄大年, 李家洋, 钱

- 前 (2003). 水稻稻米糊化温度控制基因 ALK 的图位克隆及其序列分析. 中国科学(C辑) **33**, 481–487.
- 何秀英, 程永盛, 刘志霞, 陈钊明, 刘维, 卢东柏, 陈粤汉, 廖耀平 (2015). 国标优质籼稻的稻米品质与淀粉RVA谱特征研究. 华南农业大学学报 **36**, 37–44.
- 李丹婷, 夏秀忠, 农保选, 刘开强, 张宗琼, 梁耀懋 (2012). 广西地方稻种资源核心种质构建和遗传多样性分析. 广西植物 **32**, 94–100.
- 毛孝强, 余腾琼, 林谦, 徐绍忠, 杨德 (2003). 我国水稻品质性状数量遗传研究进展. 云南农业大学学报 **18**, 203–207.
- 邱先进, 袁志华, 陈凯, 杜斌, 何文静, 杨隆维, 徐建龙, 邢丹英, 吕文恺 (2015). 用全基因组关联分析解析籼稻垩白的遗传基础. 作物学报 **41**, 1007–1016.
- 于安玲 (2011). 芸藁素(BR)和赤霉素(GA)对小麦籽粒淀粉积累、粒度分布及加工特性的影响. 博士论文. 泰安: 山东农业大学. pp. 31–90.
- Hodgkin T (李自超等译) (2012). 植物遗传资源核心种质. 北京: 中国农业出版社. pp. 4.
- FAO (1993). Rice in Human Nutrition. Laguna: International Rice Research Institute. pp. 35–60.
- Hardy OJ, Vekemans X (2002). SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* **2**, 618–620.
- Hirano HY, Sano Y (2000). Comparison of $WAXY$ gene regulation in the endosperm and pollen in *Oryza sativa* L. *Genes Genet Syst* **75**, 245–249.
- Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang QF, Li J, Han B (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet* **42**, 961–967.
- Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, Li W, Guo Y, Deng L, Zhu C, Fan D, Lu Y, Weng Q, Liu K, Zhou T, Jing Y, Si L, Dong G, Huang T, Lu T, Feng Q, Qian Q, Li J, Han B (2012). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat Genet* **44**, 32–39.
- Liu L, Ma X, Liu S, Zhu C, Jiang L, Wang Y, Shen Y, Ren Y, Dong H, Chen L, Liu X, Zhao Z, Zhai H, Wan J (2009). Identification and characterization of a novel $Waxy$ allele from a Yunnan rice landrace. *Plant Mol Biol* **71**, 609–626.
- Mikami I, Uwatoko N, Ikeda Y, Yamaguchi J, Hirano HY, Suzuki Y, Sano Y (2008). Allelic diversification at the wx locus in landraces of Asian rice. *Theor Appl Genet* **116**, 979–989.
- Nakamura Y (1996). Some properties of starch debranching enzymes and their possible role in amylopectin biosynthesis. *Plant Sci* **121**, 1–18.
- Nakamura Y, Francisco PB, Hosaka Y, Sato A, Sawada T, Kubo A, Fujita N (2005). Essential amino acids of starch synthase IIa differentiate amylopectin structure and starch quality between japonica and indica rice varieties. *Plant Mol Biol* **58**, 213–227.
- Rogers SO, Bendich AJ (1988). Extraction of DNA from plant tissues. *Plant Mol Biol Manual* **A6**, 1–10.
- Sato H, Suzuki Y, Sakai M, Imbe T (2002). Molecular characterization of Wx -mq, a novel mutant gene for low-amylose content in endosperm of rice (*Oryza sativa* L.). *Breed Sci* **52**, 131–135.
- Takemoto-Kuno Y, Mitsueda H, Suzuki K, Hirabayashi H, Ideta O, Aoki N, Umemoto T, Ishii T, Ando I, Kato H, Nemoto H, Imbe T, Takeuchi Y (2015). $qAC2$, a novel QTL that interacts with Wx and controls the low amylose content in rice (*Oryza sativa* L.). *Theor Appl Genet* **128**, 563–573.
- Tan YF, Xing YZ, Li JX, Yu SB, Xu CG, Zhang QF (2000). Genetic bases of appearance quality of rice grains in Shanyou 63, an elite rice hybrid. *Theor Appl Genet* **101**, 823–829.
- Tian ZX, Qian Q, Liu QQ, Yan MX, Liu XF, Yan CJ, Liu GF, Gao ZY, Tang SZ, Zeng DL, Wang YH, Yu JM, Gu MH, Li JY (2009). Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities. *Proc Natl Acad Sci USA* **106**, 21760–21765.
- Umemoto T, Aoki N, Lin HX, Nakamura Y, Inouchi N, Sato Y, Yano M, Hirabayashi H, Maruyama S (2004). Natural variation in rice starch synthase IIa affects enzyme and starch properties. *Funct Plant Biol* **31**, 671–684.
- Vandeputte GE, Vermeylen R, Geroms J, Delcour JA (2003). Rice starches. I. Structural aspects provide insight into crystallinity characteristics and gelatinisation behaviour of granular starch. *J Cereal Sci* **38**, 43–52.
- Wang ZY, Wu ZL, Xing YY, Zheng FG, Guo XL, Zhang WG, Hong MM (1990). Nucleotide sequence of rice $WAXY$ gene. *Nucleic Acids Res* **18**, 5898.

Yang J, Wang J, Fan FJ, Zhu J, Chen T, Wang CL, Zheng TQ, Zhang J, Zhong WG, Xu JL (2013). Development of AS-PCR marker based on a key mutation confirmed by resequencing of *Wx*-mp in Milky Princess and its application in japonica soft rice (*Oryza sativa* L.) breeding. *Plant Breed* **132**, 595–603.

Zhang GY, Cheng ZJ, Zhang X, Guo XP, Su N, Jiang L, Mao L, Wan JM (2011). Double repression of soluble starch synthase genes *SSIIa* and *SSIIIa* in rice (*Oryza sativa* L.) uncovers interactive effects on the physico-chemical properties of starch. *Genome* **54**, 448–459.

Zhang ZW, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat Genet* **42**, 355–360.

Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH, Gij N, Islam MR, Reynolds A, Mezey J, McClung AM, Bustamante CD, Mccouch SR (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat Commun* **2**, 467.

Genome-wide Association Study of Genes Related to Waxiness in *Oryza sativa*

Xinghai Yang^{1,2}, Baoxuan Nong^{1,2}, Xiuzhong Xia^{1,2}, Zongqiong Zhang^{1,2}, Yu Zeng^{1,2}
Kaiqiang Liu³, Guofu Deng³, Danting Li^{1,2*}

¹Rice Research Institute, Guangxi Academy of Agricultural Science, Nanning 530007, China; ²Guangxi Key Laboratory of Rice Research Genetics and Breeding, Nanning 530007, China; ³Guangxi Academy of Agricultural Science, Nanning 530007, China

Abstract We examined rice waxiness by genome-wide association study based on genotyping 1 090 071 single nucleotide polymorphism (SNP) variants across landraces in the core collection of 419 rice germplasm derived from Guangxi. We identified 45 significant associated loci located on chromosome 6 at $P < 4.72 \times 10^{-8}$ ($4.72E-8$) in the compressed mixed linear model. Screening the genes in the candidate regions about 150 kb upstream and downstream of each significant loci revealed 305 candidate genes, including *Wx* and *SSIIa* related to starch synthase. Furthermore, 4 significant SNPs associated with waxiness were intensively located in the region of 5.69–5.89 Mb on chromosome 6. This region may be an important candidate region related to waxiness in rice.

Key words *Oryza sativa*, genome-wide association study, waxiness, candidate genes

Yang XH, Nong BX, Xia XZ, Zhang ZQ, Zeng Y, Liu KQ, Deng GF, Li DT (2016). Genome-wide association study of genes related to waxiness in *Oryza sativa*. *Chin Bull Bot* **51**, 737–742.

* Author for correspondence. E-mail: lidanting@gxaas.net