

· 热点评 ·

全基因组关联分析实现水稻粒型自然变异的分子解析

厉新民, 林鸿宣*

中国科学院上海生命科学研究院植物生理生态研究所, 植物分子遗传国家重点实验室/分子植物卓越中心, 上海 200032

摘要 全基因组关联分析(GWAS)近年来被广泛应用于解析生物自然变异的遗传基础。但限于其遗传定位精度, 在水稻(*Oryza sativa*)遗传学研究中, 该方法尚无法取代传统的图位克隆法在克隆复杂性状调控基因中的作用。近期, 中国科学家在应用GWAS等大数据来克隆控制水稻粒长和粒重等复杂性状的QTL方面取得了新突破。

关键词 全基因组关联分析, 水稻, 粒重, 演化, SPL

厉新民, 林鸿宣 (2016). 全基因组关联分析实现水稻粒型自然变异的分子解析. 植物学报 51, 411–415.

水稻(*Oryza sativa*)不仅是单子叶模式植物, 而且是我国重要的粮食作物。世界上种植面积最大的栽培水稻是亚洲栽培稻, 它是由亚洲野生稻(*O. rufipogon*)驯化而来, 在演化过程中经历了长期的自然选择和人工驯化, 形成了适宜在不同地域栽培的现代品种。亚洲栽培稻主要分为籼稻(*O. sativa* subsp. *indica*)与粳稻(*O. sativa* subsp. *japonica*) 2个亚种, 它们不仅在遗传上分化明显, 而且在生长环境和农艺性状上也有较大差异。籼稻适宜生长在低纬度的热带及亚热带地区, 一般粒型较细长; 而粳稻按其生长环境又可分为温带粳稻(Temperate *japonica*)和热带粳稻(Tropical *japonica*) (又称爪哇稻, *javanica*)。大部分粳稻属于温带粳稻, 适宜生长在较高纬度的温带地区, 籽粒为短圆型; 少部分品种属于热带粳稻, 适于种植在亚热带或热带地区, 籽粒通常是大粒型。水稻籽粒的大小是影响稻米产量和品质的重要因素, 在长期的水稻育种过程中一直备受关注。但籽粒大小是由多个遗传位点控制的复杂数量性状, 其遗传规律不甚清楚。直到水稻基因组测序完成后, 人们利用粳稻和籼稻的基因组信息, 对控制水稻籽粒大小的数量性状遗传位点(QTL)进行了研究。通过大量的遗传分析, 人们已经利用图位克隆的方法成功克隆到一批调控水稻粒型变异的QTL, 并解释了其分子机理。这些QTL主要通过调控水稻颖壳的大小来影响籽粒的形成, 按其调控方式可分为2类: 正向调控籽粒大小的有GS2/GL2 (Che et al., 2015; Duan et al., 2015; Hu

et al., 2015)、GS5 (Li et al., 2011)、GW6a (Song et al., 2015)和GW8 (Wang et al., 2012)等; 负向调控籽粒大小的则有GW2 (Song et al., 2007)、GS3 (Mao et al., 2010)、GL3.1 (Qi et al., 2012; Zhang et al., 2012)、GW5/qSW5 (Shomura et al., 2008; Weng et al., 2008)、TGW6 (Ishimaru et al., 2013)和GW7/GL7 (Wang et al., 2015a, 2015b)等。上述基因大多通过调控水稻颖壳细胞分裂过程来调节粒型大小, 有些则同时影响细胞大小(如GS2/GL2) (Che et al., 2015; Duan et al., 2015; Hu et al., 2015)。

随着基因组学的迅猛发展, 上千个水稻品种的基因组已被测序。这些数据为解析水稻自然变异和演化进程的分子机理提供了基础信息。通过考察各水稻品种的表型差异, 结合基因组信息, 利用全基因组关联分析(genome-wide association studies, GWAS)可以很快发现引起水稻不同品种生长发育表型或环境适应能力差异的遗传位点。但是, 由于水稻是自交作物, 在现有收集的水稻品种资源群体中, 其连锁不平衡衰退率不如理想的异交作物玉米那么高, 这造成在进行GWAS分析时, 最终的遗传定位区段只能精确到大约100–200 kb的范围内(Huang and Han, 2014)。而这个范围的水稻基因组中通常含有10余个功能基因。也就是说, 限于水稻的遗传特性, 现有的GWAS分析只能实现对目标基因的初步定位, 仍不能取代传统的图位克隆法实现对目标基因的精确定位。但是, 随着基因组数据的增加以及相应算法的改

收稿日期: 2016-06-01; 接受日期: 2016-06-13

* 通讯作者。E-mail: hxlin@sibs.ac.cn

进,利用GWAS直接定位克隆作物的复杂性状QTL成为可能。最近,中国科学院上海生命科学研究院植物生态研究所韩斌研究组通过选择合适的群体材料,优化整合GWAS和表达分析、功能分析等方法,成功克隆了1个控制水稻籽粒大小的QTL (Si et al., 2016)。与传统的图位克隆方法相比,该研究效率高且速度快,有很大的优势,为利用大数据进行复杂性状的遗传解析提供了一个极好的范例。

在前期研究中,韩斌研究组曾重测序了517份水稻品种,并对其包括籽粒大小在内的14个农艺性状率先进行了GWAS分析(Huang et al., 2010)。该研究首次证明在水稻中进行GWAS分析的可行性和可靠性,可作为传统图位克隆方法的有效补充,大大提高了发掘水稻复杂性状相关QTL的效率。通过优化算法,他们又利用低倍覆盖的方案重测序了950多份水稻栽培种,对籽粒大小等重要性状进行了更高精度的GWAS分析(Huang et al., 2012b),表明优化整合后的GWAS分析有潜力直接定位到重要QTL。近期,他们利用GWAS结合其它功能分析,直接定位克隆了1个控制水稻籽粒大小的QTL (Si et al., 2016)。该研究取得了突破性的进展,把研究体系向前推进了一大步。

通过对381份粳稻品种(其中包括341份温带粳稻和40份热带粳稻)重新进行籽粒大小的表型分析,结合以前的基因型数据,韩斌研究组进行了严格的GWAS分析,在7号染色体上检测到了1个控制粳稻粒长和粒重变异的主效QTL,将其命名为GLW7 (GRAIN LENGTH AND WEIGHT ON CHROMOSOME 7)。他们将GWAS分析中关联程度最高的SNP所在的260 kb染色体区段作为候选基因所在位置,对其中含有的11个功能基因进行了表达模式与籽粒大小的关联分析。研究表明,11个基因在随机选取的10个大粒品种和10个小粒品种的根和叶中的表达情况无明显差异;而在穗中,第9个基因在大粒品种中的表达在mRNA和蛋白水平都比在小粒品种中明显更高。该基因较特异地在幼穗中表达,这与GLW7基因调控籽粒大小的功能也较为吻合。因此,该基因被认为是GLW7最可能的候选基因,其编码OsSPL13,属于植物特有的转录因子SPL (SQUAMOSA promoter-binding protein-like)家族。SPL被证明调控植物营养器官、花和果实的发育、发育时期的转变以及

育性等多种生物学进程,并且在次生代谢、激素响应和光信号转导等方面也发挥重要作用(Wang and Wang, 2015)。在水稻中,SPL转录因子参与调控很多重要的农艺性状。比如,OsSPL14是调控理想株型建成的重要QTL (Jiao et al., 2010; Miura et al., 2010);而OsSPL16则被证明是另一个调控水稻籽粒大小和品质的QTL-GW8 (Wang et al., 2012)。

研究人员通过对OsSPL13进行深入分析,发现有16个核苷酸差异与粒型差异连锁:其中6个在启动子区,3个在5'UTR区,1个是发生在外显子区的同义突变,2个在内含子区,还有1个单核苷酸差异(SNP)和3个插入缺失(indel)发生在3'UTR区。那么到底是哪些差异最终影响到籽粒大小呢?研究者进行了转基因功能验证实验。他们共设计了10种不同的转基因构建,分别包含上述不同的核苷酸差异,并用它们转化小粒的温带粳稻品种Dongjing。对转基因株系的分析表明,5'UTR区-146--135 bp的CCATTC重复序列是造成籽粒大小变异的遗传基础。在小粒品种的OsSPL13中该序列重复2次,而在大粒品种中则只有1次。这种重复次数的减少会造成OsSPL13在转录水平和蛋白水平积累量的显著增多,进而产生较大的籽粒。通过RNA干扰的方法将大粒品种GP579和小粒品种Dongjing中OsSPL13的表达量进行下调会使水稻籽粒的长度和粒重都显著降低。同时,他们筛选到了1个Dongjing来源的glw7突变体,其粒长和粒重比野生型均明显降低。用Dongjing的OsSPL13全长序列可以互补glw7突变体的表型;而用大粒品种的OsSPL13全长序列转化glw7突变体,则转基因株系的籽粒比Dongjing野生型还要大。这些结果表明,OsSPL13是控制粳稻中粒长的主效QTL-GLW7。

细胞学观察显示,相比于Dongjing,glw7突变体以及OsSPL13转基因水稻颖壳中的细胞数目并没有显著变化,而细胞长度则差异较大。glw7的颖壳细胞长度明显变短,而OsSPL13转基因水稻的颖壳细胞长度则显著增加。这些证据表明,OsSPL13通过调控细胞的延伸来控制水稻籽粒的大小。对籽粒灌浆速率的检测表明,glw7突变体的灌浆速率和最终灌浆水平都远低于野生型;而OsSPL13转基因植株则正好相反。除了对籽粒大小进行调控,OsSPL13对水稻穗发育也有调节作用。在Dongjing中转入OsSPL13会明显增加水稻的每穗粒数;而glw7突变体和OsSPL13

RNAi株系中, 穗长、穗枝梗数以及每穗粒数都显著降低。综上所述, *OsSPL13*可以显著增加水稻的粒重和每穗粒数, 因而可以使水稻增产, 其中大粒品种中的*OsSPL13*增产效果更为明显。

原位杂交实验表明, *OsSPL13*主要在幼穗一级枝梗和二级枝梗起始时期的花原基中表达, 并且在1 mm的幼嫩颖壳中表达较高, 随着颖壳逐渐发育成熟, 其表达量也从颖壳基部开始慢慢降低。这种表达模式与*OsSPL13*调控水稻穗部发育以及籽粒大小的功能相吻合。研究表明, *SPL*类基因通常在3'UTR区有一个miR156识别位点, 其表达受miR156的调控(Wang and Wang, 2015)。将*OsSPL13*的miR156识别位点突变以后, *OsSPL13*的表达不仅仅局限在幼穗和颖壳, 在叶片中也有大量表达。在*OsmiR156*过表达的水稻中, *OsSPL13*的表达量急剧降低, 而通过转基因手段降低*OsmiR156*的表达量, 则可以检测到*OsSPL13*的表达量明显升高。这表明*OsSPL13*的表达模式受到miR156的精确调控。

*SPL*类转录因子能识别并结合DNA上的1个核心元件GTAC, 进而发挥其功能(Wang and Wang, 2015)。编码 α -tubulin的基因*SRS5* (*SMALL AND ROUND SEED 5*)在启动子区有4个GTAC核心元件, 已被证明是*OsSPL13*的靶基因。研究表明, *OsSPL13*能够结合在*SRS5*的启动子区, 并且*SRS5*在*OsSPL13*转基因植物和Dongjing中的表达量要明显高于*glw7*突变体。同时, *OsSPL13*还被证明可以直接转录调控水稻穗发育的重要调节基因*DEP1* (Huang et al., 2009)。为了更全面地解析*OsSPL13*的下游基因, 研究人员利用RNA-seq技术比较了Dongjing和*glw7*的转录组。结果表明, 一些expansin相关基因在*glw7*突变体中的表达受到明显抑制。这进一步证明*OsSPL13*通过调节细胞延展相关的基因来影响细胞大小, 进而调控水稻籽粒的大小。

*OsSPL13*可以分为大粒型*OsSPL13^{LGH}*和小粒型*OsSPL13^{SGH}*两种单倍体型。群体遗传学分析表明, *OsSPL13^{SGH}*只存在于温带粳稻中; 而*OsSPL13^{LGH}*则存在于绝大部分的热带粳稻和籼稻中, 只有少部分的温带粳稻含有*OsSPL13^{LGH}*单倍体型。在水稻演化过程中, 粳稻是最早从野生稻中驯化出来的, 然后在长期的地理扩张和驯化过程中, 一些粳稻向高纬度地区发展, 成为温带粳稻, 还有一些又和低纬度地区的

野生稻杂交融合, 产生了籼稻, 同时也有一些粳稻发展到热带亚热带地区(Huang et al., 2012a)。通过比较*OsSPL13*遗传位点在不同群体中的遗传分化系数可以发现, 热带粳稻中的*OsSPL13^{LGH}*很可能是从籼稻中渗入的。

韩斌研究团队首次在水稻中通过GWAS对籽粒大小这类复杂的数量性状进行了分子解析, 结合功能分析成功克隆到了1个控制粳稻粒型分化的重要QTL。这是一项基因组学时代水稻重要复杂性状研究的标志性成果, 为利用GWAS进行QTL克隆打开了一扇大门。*OsSPL13^{LGH}*是从热带粳稻中发掘的一个新的正向调控水稻粒长和粒重的主效QTL, 可以进一步应用到我国粳稻品种改良中。我国常规的粳稻品种多是短圆粒型, 从外观品质上有增加粒长的需求。虽然籼稻粒型较长, 但是籼粳遗传背景差异较大, 属于不同亚种, 杂交育种难度相对较大。因此粳稻来源的*OsSPL13^{LGH}*提供了一个宝贵的粳稻育种基因资源。籼稻中的粒长主效QTL是*GS3*, 其可以解释大部分籼稻群体内的粒长变异(Mao et al., 2010)。研究表明, *GLW7*和*GS3*在遗传上处于相互独立的调控通路。*GLW7*与*GS3*协调整合应用, 可为水稻育种过程中优化粒型和品质提供重要的基因工具。

对*OsSPL13*调控水稻籽粒大小的分子机理进行深入研究, 探究它与其它粒型基因间的调控网络, 将会加深我们对水稻粒型性状分子基础的认识, 也可以增加我们对*SPL*家族转录因子调控植物发育的系统认识。这方面研究既有重要的理论意义, 又有广泛的应用价值。

参考文献

- Che R, Tong H, Shi B, Liu Y, Fang S, Liu D, Xiao Y, Hu B, Liu L, Wang H, Zhao M, Chu C (2015). Control of grain size and rice yield by *GL2*-mediated brassinosteroid responses. *Nat Plants* 2, 15195.
- Duan P, Ni S, Wang J, Zhang B, Xu R, Wang Y, Chen H, Zhu X, Li Y (2015). Regulation of *OsGRF4* by *OsmiR396* controls grain size and yield in rice. *Nat Plants* 2, 15203.
- Hu J, Wang Y, Fang Y, Zeng L, Xu J, Yu H, Shi Z, Pan J, Zhang D, Kang S, Zhu L, Dong G, Guo L, Zeng D, Zhang G, Xie L, Xiong G, Li J, Qian Q (2015). A rare allele of *GS2* enhances grain size and grain yield in rice. *Mol Plant* 8, 1455–1465.
- Huang X, Han B (2014). Natural variations and genome-wide

- association studies in crop plants. *Annu Rev Plant Biol* **65**, 531–551.
- Huang X, Kurata N, Wei X, Wang ZX, Wang A, Zhao Q, Zhao Y, Liu K, Lu H, Li W, Guo Y, Lu Y, Zhou C, Fan D, Weng Q, Zhu C, Huang T, Zhang L, Wang Y, Feng L, Furuumi H, Kubo T, Miyabayashi T, Yuan X, Xu Q, Dong G, Zhan Q, Li C, Fujiyama A, Toyoda A, Lu T, Feng Q, Qian Q, Li J, Han B (2012a). A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**, 497–501.
- Huang X, Qian Q, Liu Z, Sun H, He S, Luo D, Xia G, Chu C, Li J, Fu X (2009). Natural variation at the *DEP1* locus enhances grain yield in rice. *Nat Genet* **41**, 494–497.
- 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 (2012b). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat Genet* **44**, 32–39.
- Ishimaru K, Hirotsu N, Madoka Y, Murakami N, Hara N, Onodera H, Kashiwagi T, Ujiie K, Shimizu B, Onishi A, Miyagawa H, Katoh E (2013). Loss of function of the IAA-glucose hydrolase gene *TGW6* enhances rice grain weight and increases yield. *Nat Genet* **45**, 707–711.
- Jiao Y, Wang Y, Xue D, Wang J, Yan M, Liu G, Dong G, Zeng D, Lu Z, Zhu X, Qian Q, Li J (2010). Regulation of *OsSPL14* by *OsmiR156* defines ideal plant architecture in rice. *Nat Genet* **42**, 541–544.
- Li Y, Fan C, Xing Y, Jiang Y, Luo L, Sun L, Shao D, Xu C, Li X, Xiao J, He Y, Zhang Q (2011). Natural variation in *GS5* plays an important role in regulating grain size and yield in rice. *Nat Genet* **43**, 1266–1269.
- Mao H, Sun S, Yao J, Wang C, Yu S, Xu C, Li X, Zhang Q (2010). Linking differential domain functions of the *GS3* protein to natural variation of grain size in rice. *Proc Natl Acad Sci USA* **107**, 19579–19584.
- Miura K, Ikeda M, Matsubara A, Song XJ, Ito M, Asano K, Matsuoka M, Kitano H, Ashikari M (2010). *OsSPL14* promotes panicle branching and higher grain productivity in rice. *Nat Genet* **42**, 545–549.
- Qi P, Lin YS, Song XJ, Shen JB, Huang W, Shan JX, Zhu MZ, Jiang L, Gao JP, Lin HX (2012). The novel quantitative trait locus *GL3.1* controls rice grain size and yield by regulating Cyclin-T1;3. *Cell Res* **22**, 1666–1680.
- Shomura A, Izawa T, Ebana K, Ebitani T, Kanegae H, Konishi S, Yano M (2008). Deletion in a gene associated with grain size increased yields during rice domestication. *Nat Genet* **40**, 1023–1028.
- Si LZ, Chen JY, Huang XH, Gong H, Luo JH, Hou QQ, Zhou TY, Lu TT, Zhu JJ, Shangguan YY, Chen EW, Gong CX, Zhao Q, Jing YF, Zhao Y, Li Y, Cui LL, Fan DL, Lu YQ, Weng QJ, Wang YC, Zhan QL, Liu KY, Wei XH, An K, An G, Han B (2016). *OsSPL13* controls grain size in cultivated rice. *Nat Genet* **48**, 447–456.
- Song XJ, Huang W, Shi M, Zhu MZ, Lin HX (2007). A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat Genet* **39**, 623–630.
- Song XJ, Kuroha T, Ayano M, Furuta T, Nagai K, Komeda N, Segami S, Miura K, Ogawa D, Kamura T, Su-zuki T, Higashiyama T, Yamasaki M, Mori H, Inukai Y, Wu J, Kitano H, Sakakibara H, Jacobsen SE, Ashikari M (2015). Rare allele of a previously unidentified histone H4 acetyltransferase enhances grain weight, yield, and plant biomass in rice. *Proc Natl Acad Sci USA* **112**, 76–81.
- Wang H, Wang H (2015). The miR156/SPL module, a regulatory hub and versatile toolbox, gears up crops for enhanced agronomic traits. *Mol Plant* **8**, 677–688.
- Wang S, Wu K, Yuan Q, Liu X, Liu Z, Lin X, Zeng R, Zhu H, Dong G, Qian Q, Zhang G, Fu X (2012). Control of grain size, shape and quality by *OsSPL16* in rice. *Nat Genet* **44**, 950–954.
- Wang S, Li S, Liu Q, Wu K, Zhang J, Wang S, Wang Y, Chen X, Zhang Y, Gao C, Wang F, Huang H, Fu X (2015a). The *OsSPL16-GW7* regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nat Genet* **47**, 949–954.
- Wang Y, Xiong G, Hu J, Jiang L, Yu H, Xu J, Fang Y, Zeng L, Xu E, Xu J, Ye W, Meng X, Liu R, Chen H, Jing Y, Wang Y, Zhu X, Li J, Qian Q (2015b). Copy number variation at the *GL7* locus contributes to grain size diversity in rice. *Nat Genet* **47**, 944–948.
- Weng J, Gu S, Wan X, Gao H, Guo T, Su N, Lei C, Zhang X, Cheng Z, Guo X, Wang J, Jiang L, Zhai H, Wan J (2008). Isolation and initial characterization of *GW5*, a major QTL associated with rice grain width and weight. *Cell Res* **18**, 1199–1209.
- Zhang X, Wang J, Huang J, Lan H, Wang C, Yin C, Wu Y,

Tang H, Qian Q, Li J, Zhang H (2012). Rare allele of *OsPPKL1* associated with grain length causes extra-large grain and a significant yield increase in rice. *Proc Natl Acad Sci USA* **109**, 21534–21539.

Genome-wide Association Study Opens a Window to Molecular Dissection of Rice Grain Size

Xinmin Li, Hongxuan Lin*

National Key Laboratory of Plant Molecular Genetics, Centre for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China

Abstract High-throughput sequencing technologies bring us the genomics age, consequently facilitates genome-wide association studies (GWAS) of complex traits in crops. But GWAS has not yet been successful in detecting the genetic basis of phenotypic variations in rice due to limited mapping resolution. Recently, chinese scientists have cloned a QTL for rice grain length and weight using GWAS combining with functional investigations and propelled the molecular dissection of rice QTL from genetics to genomics. Their study provided us not only a model for investigating rice complex traits and evolutionary changes using “omics” resources but also a valuable gene for rice breeding.

Key words genome-wide association studies, rice, grain weight, evolution, SPL

Li XM, Lin HX (2016). Genome-wide association study opens a window to molecular dissection of rice grain size. *Chin Bull Bot* **51**, 411–415.

* Author for correspondence. E-mail: hxlin@sibs.ac.cn