DOI: 10.7671/j.issn.1001-411X.202006039

王朝欢, 宋博文, 余思佳, 等. 基于全基因组测序构建水稻 RIL 群体遗传图谱 [J]. 华南农业大学学报, 2021, 42(2): 44-50. WANG Chaohuan, SONG Bowen, YU Sijia, et al. Construction of a genetic map of rice RILs based on whole genome sequencing[J]. Journal of South China Agricultural University, 2021, 42(2): 44-50.

基于全基因组测序构建水稻 RIL 群体遗传图谱

王朝欢,宋博文,余思佳,肖武名,黄明(华南农业大学农学院/国家植物航天育种工程技术研究中心,广东广州510642)

摘要:【目的】以籼稻品种'MDS'和'R315'为亲本,构建一张高密度的遗传图谱,挖掘水稻 Oryza sativa L.重要农艺性状相关基因并加快水稻品种选育。【方法】对两亲本及其 192 个重组自交系 (RILs) 群体进行全基因组测序,筛选高质量单核苷酸多态性 (SNPs),划分 bin 标记,针对每个连锁群使用 JoinMap4.0 对 bin 标记进行排序,用 perl SVG 模块绘制连锁图,并对标记在基因组和遗传图谱上的位置进行共线性分析。【结果】两亲本间共筛选出 221 494 个高质量 SNPs,构建了一张高密度遗传图谱,包含 1 612 个 bin 标记,总图距为 1 327.82 cM,相邻标记间平均遗传图距为 0.82 cM。共线性分析显示各连锁群上的大部分标记顺序与基因组保持一致,共线性较好,图谱质量高。【结论】本研究构建的高密度遗传图谱质量较高,为后续功能基因的鉴定提供了基础。

关键词: 水稻; 重组自交系; 全基因组测序; 遗传图谱

中图分类号: Q37 文献标志码: A 文章编号: 1001-411X(2021)02-0044-07

Construction of a genetic map of rice RILs based on whole genome sequencing

WANG Chaohuan, SONG Bowen, YU Sijia, XIAO Wuming, HUANG Ming (College of Agriculture, South China Agricultural University/National Plant Breeding Engineering Technology Research center, Guangzhou 510642, China)

Abstract: 【Objective】 The *indica* cultivar 'MDS' and 'R315' were used as parents to construct a high-density genetic map, explore the important agronomic traits related genes of rice (*Oryza sativa* L.) and accelerate the breeding of rice cultivars. 【Method】 Whole genome sequencing of parents and their 192 recombinant inbred lines (RILs) were performed to screen high-quality single nucleotide polymorphisms (SNPs) and construct bin markers. The bin markers were sorted using JoinMap4.0 for each linkage group, and perl SVG was used to draw the linkage map. Collinearity analysis was conducted according to the position of the markers on the genome and genetic map. 【Result】 A total of 221 494 high-quality SNPs were screened between the two parents. The constructed high-density genetic map contained 1 612 bin markers. The total map distance was 1 327.82 cM, and the average genetic map distance between adjacent markers was 0.82 cM. The collinearity analysis showed that the order of most markers on each linkage group was consistent with that on the genome. The collinearity was good and the map was of high quality. 【Conclusion】 The constructed high-density genetic

收稿日期:2020-06-18 网络首发时间:2020-08-27 08:47:45

网络首发地址: https://kns.cnki.net/kcms/detail/44.1110.s.20200826.1521.004.html

作者简介: 王朝欢 (1995—),男,硕士研究生,E-mail: 461325829@qq.com; 通信作者: 肖武名 (1982—),男,副研究员,博士,E-mail: heredity24@126.com;黄 明 (1981—),男,助理研究员,博士,E-mail: mhuang@scau.edu.cn

基金项目:国家重点研发计划 (2017YFD0100100); 国家自然科学基金 (31872885); 广东省重点领域研发计划 (2020B020219004)

map is of high quality, providing a preliminary basis for the subsequent identification of functional genes.

Key words: Oryza sativa L.; recombinant inbred line; whole genome sequencing; genetic map

水稻 Oryza sativa L.是世界上超过一半人口的主要粮食作物,深入挖掘重要的功能基因对水稻主要农艺性状的遗传改良具有重要意义。高密度遗传图谱的构建为基因的精准定位和克隆创造了必要条件。

本研究使用的重组自交系 (Recombinant inbred lines, RILs) 群体是由籼稻'MDS'和籼稻 'R315'构建的高代重组自交系,具有稳定的表型 性状,是一个良好的永久性遗传群体。自1988年 McCouch 等[1] 利用籼稻 'IR34583'与爪哇稻 'Bulu Dalam'衍生的 F₂ 群体构建了第 1 张含 135 个限制性内切酶片段长度多态性 (Restriction fragment length polymorphism, RFLP) 标记的水稻分 子连锁图谱以来,水稻的遗传图谱相继诞生。传统 的图谱构建通常利用 RFLP、简单重复序列 (Simple sequence repeat, SSR) 标记和随机扩增多态性 DNA(Random amplified polymorphic DNA, RAPD) 标记。结合多种分子标记技术进行遗传图 谱构建可进一步增加标记的密度。徐建龙等[2]利用 272 个均匀分布的标记 (141 个 RFLP 标记、99 个 SSR 标记、29个 RAPD 标记和 3个形态标记) 构建 了连锁图用于 QTL 分析,该连锁图总共覆盖基因 组 2 777.7 cM, 相邻标记间平均距离为 10.2 cM。早 期的分子标记,作图距离通常较大,后期精细定 位工作需要大量分离的 F2 代个体。随着 DNA 测序 技术的快速发展,基于高通量测序的新型分子标记 越来越多地应用于高密度遗传图谱的构建[3-4]。 Chen 等[5] 利用 148 个子代组成的 RIL 群体构建了 1 680.9 cM 的遗传图谱, 平均遗传距离为 1.16 cM, 共定位到 23 个耐盐相关 QTLs, 其中, 位于第 1、4、 12 号染色体上的 3 个 QTLs 表现出累加效应。单核 苷酸多态性 (Single nucleotide polymorphism, SNP) 标记作为第3代新型分子标记技术,具有数量 多、遗传稳定性高以及易于自动化分析等优点[6],已 成为近年来水稻高密度遗传图谱构建的优先选择。

本研究利用全基因组测序 (Whole genome sequencing, WGS) 技术对籼稻 'MDS'和籼稻 'R315'构建的1个RIL群体进行全基因组测序,先鉴定出两亲本间的SNP,再利用SNP构建 bin 标记,最后构建1张高密度遗传图谱,以期为进一步深入挖掘和研究重要农艺性状基因奠定有利基础。

1 材料与方法

1.1 试验材料

本研究使用的材料为国家植物航天育种工程技术研究中心前期收集的籼稻地方品种'MDS'和'R315'。父本'MDS'是大穗、抗倒、高产型的地方籼稻品种,具有高光效、高养分利用率等特点。母本'R315'是在'象牙香占'基础上改良的籼稻恢复系,具有抗病、优质、配合力好等特点,父母本杂交,F₂代通过单粒传法繁殖获得高世代 RILs。

1.2 CTAB 法提取植物叶片 DNA

2019 年晚季在华南农业大学校内教学科研试验基地的试验田中种植水稻亲本和 RILs 群体 (含192 个株系),每个材料按 6 行×6 株的规模种成小区,株行距均为 20 cm,单苗插植,常规栽培管理。F₅ 代单株收种。亲本及其 RILs(F₆代) 群体取20 粒饱满种子萌发,取各个株系的幼嫩叶片(每个株系随机取 5 株混样),按 CTAB 法[7] 提取 DNA 后送至北京诺禾致源生物信息科技有限公司进行测序分析。

1.3 DNA 文库构建及测序

检测合格的 DNA 样品通过酶切、加测序接头、 纯化、PCR 扩增等步骤完成整个文库制备。构建好 的文库通过 Illumina(测序仪) 进行测序。

1.4 测序数据质量评估及分析

检测合格的 DNA 文库进行 HiSeq 测序,产出Raw reads,结果以 fastq 文件格式存储。经过过滤,得到高质量的 Clean reads。对 192 个子代及亲本测序数据进行统计,包括测序 reads 数量、数据产量、测序错误率、Q20、Q30、GC 含量等。另外,将 Clean data 与 NCBI 的核苷酸数据库进行比对,以评估是否有其他来源的 DNA 污染。将 2 个亲本及 192 个子代测序数据与参考基因组 (http://www.mbkbase.org/R498/)进行比对,反映测序数据与参考基因组的相似性,覆盖深度和覆盖度能够反映测序数据的均一性及与参考序列的相似性。

1.5 单核苷酸多态性检测与标记开发

SNP 主要是指在基因组水平上由单个核苷酸的变异所引起的 DNA 序列多态性,包含单个碱基的转换、颠换等。基于 Burrows-Wheeler aligner (BWA) 比对结果,利用群体检测 SNP 的方式对亲本和 192 个子代进行 SNP 检测。对 BWA 比对结

果进行过滤:将比对到水稻参考基因组上唯一位置的 reads 挑选出来,采用 GATK(The genome analysis toolkit)^[8] 对过滤后的 bam 文件进行群体 SNP 的检测。为减少测序错误造成的假阳性 SNP,要求亲本 SNP 碱基支持数不少于 5,子代 SNP 碱基支持数不少于 3,统计杂合 SNP 数、纯合 SNP 数和杂合 SNP 比例 (杂合 SNP 数/总 SNP 数)。

基于亲本基因型检测结果,进行亲本间多态性标记开发,并选择 RIL 群体的可用标记类型("aa×bb"型),即 2 个亲本的某个 SNP 基因型都为纯合且不相同。完成标记开发后,提取 192 个子代在亲本多态性标记位点的基因型,对分型后的标记进行筛选。首先进行异常碱基检查,子代分型结果中,可能会出现少数亲本中没有出现的碱基型,则认为该碱基是异常碱基,碱基缺失用"—"表示;基于上述结果继续进行异常基因型检查,本研究群体类型为 RILs,在子代分离群体中,主要以纯合为主,杂合类型占比很低,杂合标记也视为异常

基因型,转化为缺失。

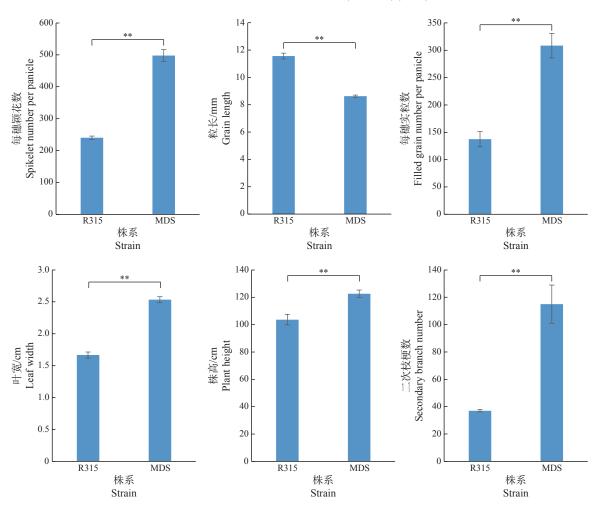
1.6 高密度图谱构建及质量评估

利用最终获得的高质量 SNP 标记,对每个个体使用 15 个 SNP 滑动窗口、步移长度为 1 的策略检测 RILs 间的重组断点,得到群体重组断点分布图。针对每个连锁群使用 JoinMap4.0 对每个连锁群的bin 标记进行排序 (连锁群使用回归算法排序,采用Kosambi 函数计算遗传距离)。根据获得的 bin 标记的遗传距离,使用 perl SVG 模块绘制连锁图,并依据文献 [9] 对标记在基因组上的位置和遗传图谱进行共线性分析。

2 结果与分析

2.1 亲本性状差异及测序数据信息分析

通过对两亲本的农艺性状数据进行差异显著性分析,发现两亲本在每穗颖花数、粒长、每穗实粒数、叶宽、株高和二级枝梗数上具有极显著差异(P<0.01)(图 1),每穗颖花数、粒长、每穗实粒



"**"表示亲本间差异极显著 (P<0.01, t 检验)

"**" indicates highly significant difference between parents (P < 0.01, t test)

图 1 亲本农艺性状分析

Fig. 1 Agronomic character analysis between parents

数、叶宽、株高、二级枝梗数是后续研究重点关注性状。

通过对两亲本和 RILs 群体 192 个子代检测合格的 DNA 文库进行 HiSeq 测序,总共获得549 496 399 500 bp 测序量。亲本'MDS'得到4 140 981 000 bp Raw reads,过滤后得到 Clean reads4 134 873 600 bp,平均测序深度为9.33×;亲本'R315'共得到4 395 794 100 bp Raw reads,过滤后得到4 390 405 500 bp Clean reads,平均测序深度9.66×。RILs 群体 192 个子代中平均每个个体的Raw reads约2 817 557 919 bp,平均测序深度6.33×。亲本总体测序质量高,Q20≥95%,Q30≥89%,GC分布正常,'MDS'和'R315'中GC含量分别为43.73%和44.12%。

将 2 个亲本及 192 个子代的测序数据与参考基因组比对。参考基因组大小为 390 983 850 bp, 192 个子代样本比对率均在 91% 以上, 1×覆盖度(至少有 1 个碱基的覆盖) 平均为 92.45%, 对参考基因组捕获区域的平均覆盖深度为 6.30×; 比对结果可用于后续的变异检测及相关分析。

2.2 SNP 识别与基因分型

基于 2 个亲本的基因型分析结果, 开发的标记类型及数量如图 2 所示, 父母亲本间共鉴定出具有多态性的位点 470 833 个, 可用标记类型为 "aa×bb"型, 共计 221 494 个。

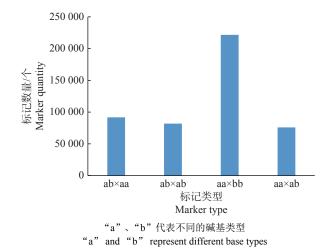


图 2 开发的单核苷酸多态性标记的类型及数量 Fig. 2 Type and quantity of developed single nucleotide

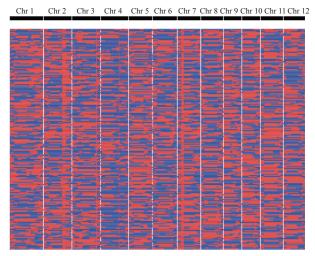
polymorphism markers

对分型后的遗传标记进行筛选,未发现异常碱基,说明基因分型准确性较好。基于筛选的结果对子代分型结果进行异常基因型检查,未发现

杂合类型,说明 RILs 子代群体中,杂合类型占比较低。最终获得 221 494 个有效标记用于连锁性分析。

2.3 高密度遗传图谱构建及质量评估

基于获得的 221 494 个 SNP 标记,对每个个体使用 15 个 SNP 滑动窗口、步移长度为 1 的策略^[9] 检测 RILs 间的重组断点,得到群体重组断点图 (图 3)。利用图 2 的标记信息,可追踪到 192 个子代中每个子代的每条染色体的重组事件发生的位置。



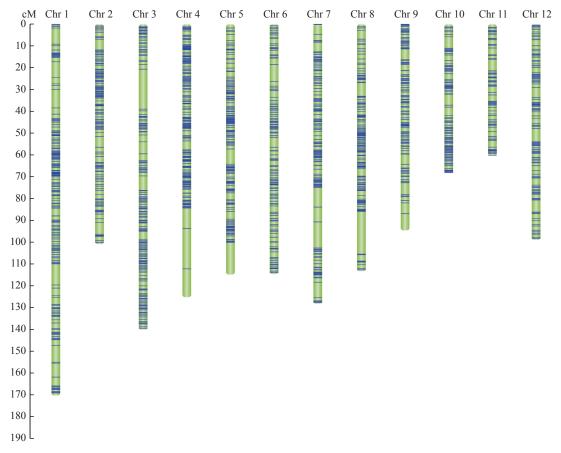
Chr 1~12: 染色体编号: 红色表示与亲本'R315'一致的区域, 蓝色表示与亲本'MDS'一致的区域

Chr 1–12: Chromosome number; Red area is consistent with parent 'R315', blue area is consistent with parent 'MDS'

图 3 群体重组断点图

Fig. 3 Group recombination breakpoint diagram

使用 JoinMap 4.0 对每个连锁群的 bin 标记进行排序 (连锁群使用回归算法排序,使用 Kosambi 函数计算遗传距离),最终获得 bin 标记 1612 个,均匀分布在各染色体上 (图 4)。用 perl SVG 模块绘制连锁图,总图距 1 327.82 cM,覆盖了 12 个连锁群 (分布在水稻的 12 条染色体上),标记间的平均图距为 0.82 cM。各个连锁群 bin 标记数量、总图距、平均图距等基本信息如表 1 所示。其中第 1 号染色体 (Chr 1)连锁群图距最长,为 170.19 cM,包含 bin 标记最多,为 194 个,标记间平均图距是 0.88 cM;第 11 号染色体 (Chr 11)连锁群图距最短,为 60.18 cM,包含 bin 标记最少,为 68 个,标记间的平均图距是 0.89 cM;连锁群中最大间隔 (Maximum gap)长度小于 5 cM 的比例高达 98.2%。



Chr 1~12: 染色体编号; 图上横线表示标记位置

Chr 1-12: Chromosome number; The horizontal line on the map indicates marker location

图 4 连锁群标记分布图

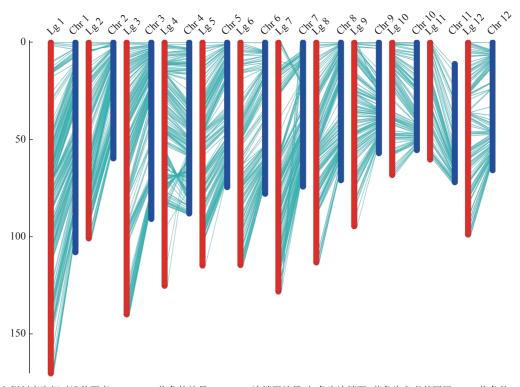
Fig. 4 Distribution map of linkage group marker

表 1 遗传连锁群信息 Table 1 Characteristics of genetic linkage group

染色体编号	标记数量	总图距/cM	平均图距/cM	最大间隔/cM
Chromosome number	Marker number	Total map distance	Average map distance	Maximum gap
1	194	170.19	0.88	9.73
2	137	100.75	0.74	5.39
3	164	139.83	0.85	18.29
4	174	125.12	0.72	18.53
5	144	114.72	0.80	14.73
6	125	114.46	0.92	7.75
7	151	128.06	0.85	11.96
8	132	113.11	0.86	19.60
9	116	94.48	0.81	7.68
10	95	68.13	0.72	5.23
11	68	60.18	0.89	5.24
12	112	98.79	0.88	6.07
合计 Total	1 612	1 327.82	0.82	19.60

共线性分析结果 (图 5)显示,各个连锁群上大 i 部分标记与在基因组上保持一致,共线性较好,图

谱质量高。



左侧刻度为相对遗传距离; Chr 1~12: 染色体编号; Lg 1~12: 连锁群编号; 红色为连锁群, 蓝色为参考基因组 R498 染色体

The left scale is relative genetic distance; Chr 1–12: Chromosome number; Lg 1–12: Linkage group number; Red area is linkage group and blue area is chromosome of reference genome R498

图 5 遗传图谱和物理图谱的共线性分析

Fig. 5 Collinearity analysis between genetic map and physical map

3 讨论与结论

高通量测序技术的迅速发展加速了水稻重要基因的挖掘和应用,在育种上也可以应用相关分子标记提高材料选育的准确性,加快品种选育。运用高通量测序获得高质量 SNPs 并构建遗传图谱进行QTL 鉴定是目前的研究热点。

亲本的选择是获得高质量图谱的第一步,对后续的 QTL 定位也至关重要。亲本间的 DNA 具有多态性是选材的首要条件,在某一表型或者多个表型具有极端差异的 2 个材料是理想的选择。在研究产量[10-12]、粒形[13-14] 等性状时,亲本的目的性状要存在极端差异。前期观察分析发现籼稻'MDS'和'R315'在多个农艺性状上均存在明显差异,包括株高、叶色、叶长、叶宽、粒形、穗粒数等。进一步的分析发现双亲在每穗粒数、叶宽、每穗实粒数、粒长、二次枝梗数、株高等性状上均存在极显著差异,预示着双亲中可能存在多个控制水稻产量的基因或者 QTLs。接下来我们将对群体的多个表型数据进行分析,结合构建的高密度遗传图谱开展深入的研究。

基于重测序获得高质量的 SNPs, 再划分 bin 标记能大幅度提高作图的精度与效率, Huang 等^[9] 对 150 个子代的 RIL 群体在 F₈ 代使用 287 个分子标记进行基因分型, 构建的遗传图谱标记间平均遗传距离为 5 cM, 平均物理距离为 1.4 Mb, 需要 3 个人通过 1 年的工作完成; 而在 F₁₁ 代使用重测序, 每 40 kb 就获得 1 个 SNP, 将重组断点的分辨率提高了 35 倍, 仅花费 2 周时间, 将每 15 个 SNPs 划分为 1 个 bin 标记, 标记间平均遗传距离为 2.3 cM, 平均物理距离为 600 kb。本研究在前人基础上构建的图谱标记间平均遗传距离为 0.82 cM, 平均物理距离为 242 kb, 平均每 16 kb 就能产生 1 个 SNP, 进一步提高了重组断点的分辨率。

标记的数量也是遗传图谱的一个重要指标,和 梗杂交能获得丰富的多态性位点^[1],但会定位到大 量的 QTLs,使得后续的研究难以进行。所以进一步 精细定位高级的作图群体,主要指近等基因系类 群体,其通过连续重复回交获得,特征是群体中个 体间遗传背景相似,仅带有少数供体片断,从而消 除背景的干扰和主效基因对微效 QTLs 的掩盖作 用,如导入系 (Introgression lines, ILs) 和替换系 (Substitution lines, SLs)^[15-16],其目的为在保证遗传背景更加相似的情况下得到准确的结果。本研究选择了在表型上具有较大差异的 2 个籼稻品种作为亲本,以期在去除相同的遗传背景后得到一定的多态性遗传标记,准确关联到差异表型的 QTLs。

QTL及基因鉴定的终极目的是应用于育种。 针对水稻的重要性状,当前已经定位和克隆了一些 重要基因,但大部分的 QTLs 或基因都只停留在试 验阶段,没能与育种实践结合起来。其中主要的 1个原因就是部分QTLs受遗传背景影响很大,很 难在不同群体中检测到相同的 QTL 位点, 这很可 能是因为 QTL 的上位性占据了主导地位[17]。主效 QTL 或基因的定位,通常会选择遗传背景有较大差 异的亲本构建群体,例如常用的籼粳杂交群体,对 于群体中发现的有价值的 QTL 想用于选育出好的 品种进行推广,还需要进行多次回交、自交,严重影 响育种进程。与此同时,得到的结果会因为材料或 者鉴定的环境改变而无法在育种中展开,最终导致 定位到的 QTL 或基因与育种实践相脱节[18-19]。本研 究选择的2个籼稻品种,在遗传背景相对相似的情 况下,获得的QTL或基因能更加稳定地表达,而且 选择双亲时要注重在性状上互补,在高世代的 RILs 群体中才可以直接筛选到能稳定遗传的目标 株系用于水稻品种选育,从而达到使 QTL 鉴定与 育种同步进行、互相验证的目的。

参考文献:

- [1] MCCOUCH S R, KOCHERT G, YU Z H, et al. Molecular mapping of rice chromosomes[J]. Theoretical and Applied Genetics, 1988, 76(6): 815-829.
- [2] 徐建龙, 薛庆中, 罗利军, 等. 水稻单株有效穗数和每穗 粒数的 QTL 剖析[J]. 遗传学报, 2001, 28(8): 752-759.
- [3] FOOLAD M R. Genome mapping and molecular breeding of tomato[J]. International Journal of Plant Genomics, 2007, 2007: 64358. doi: 10.1155/2007/64358.
- [4] LIU Z, ZHU H, LIU Y, et al. Construction of a high-density, high-quality genetic map of cultivated lotus (*Nelumbo nucifera*) using next-generation sequencing[J]. BMC Genomics, 2016, 17. doi: 10.1186/S12864-016-2781-4.
- [5] CHEN T X, ZHU Y J, CHEN K, et al. Identification of new QTL for salt tolerance from rice variety Pokkali[J]. Journal of Agronomy and Crop Science, 2020, 206(2): 202-213.

- [6] RAFALSKI A. Applications of single nucleotide polymorphisms in crop genetics[J]. Current Opinion in Plant Biology, 2002, 5(2): 94-100.
- [7] ALLEN G C, FLORES-VERGARA M A, KRASNYANSKI S, et al. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide[J]. Nature Protocols, 2006, 1(5): 2320-2325.
- [8] MCKENNA A, HANNA M, BANKS E, et al. The genome analysis toolkit: A mapreduce framework for analyzing next-generation DNA sequencing data[J]. Genome Research, 2010, 20(9): 1297-1303.
- [9] HUANG X, FENG Q, QIAN Q, et al. High-throughput genotyping by whole-genome resequencing[J]. Genome Research, 2009, 19(6): 1068-1076.
- [10] XUE W, XING Y, WENG X, et al. Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice[J]. Nature Genetics, 2008, 40(6): 761-767.
- [11] ASHIKARI M. Cytokinin oxidase regulates rice grain production[J]. Science, 2005, 309(5735): 741-745.
- [12] HUANG X, QIAN Q, LIU Z, et al. Natural variation at the *DEP1* locus enhances grain yield in rice[J]. Nature Genetics, 2009, 41(4): 494-497.
- [13] WENG J, GU S, WAN X, et al. Isolation and initial characterization of *GW5*, a major QTL associated with rice grain width and weight[J]. Cell Research, 2008, 18(12): 1199-1209.
- [14] FAN C, XING Y, MAO H, et al. *GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein[J]. Theoretical and Applied Genetics, 2006, 112(6): 1164-1171.
- [15] WANG S, WU K, YUAN Q, et al. Control of grain size, shape and quality by *OsSPL16* in rice[J]. Genome Research, 2012, 44(8): 950-955.
- [16] WU Y, WANG Y, MI X, et al. The QTL GNP1 encodes GA20ox1, which increases grain number and yield by increasing cytokinin activity in rice panicle meristems[J]. PLoS Genetics, 2016, 12(10): e1006386.
- [17] JANSEN R C. Studying complex biological systems using multifactorial perturbation[J]. Nature Reviews Genetics, 2003, 4(2): 145-151.
- [18] 王英. 利用回交导入系筛选水稻高产、抗旱和耐盐株系及选择导入系相关性状的 QTL 定位[D]. 北京: 中国农业科学院, 2013.
- [19] 冯博. 水稻抗旱和耐低氮 QTL 定位及优异等位基因的聚合效应评价[D]. 沈阳: 沈阳农业大学, 2018.

【责任编辑 李庆玲】