

引用格式:

陈聪, 刘忠松. 甘蓝型油菜品质相关性状全基因组关联分析[J]. 湖南农业大学学报(自然科学版), 2020, 46(2): 138–143.

CHEN C, LIU Z S. Genome-wide association analysis of quality-related traits in *Brassica napus*[J]. Journal of Hunan Agricultural University(Natural Sciences), 2020, 46(2): 138–143.

投稿网址: <http://xb.hunau.edu.cn>



甘蓝型油菜品质相关性状全基因组关联分析

陈聪, 刘忠松*

(湖南农业大学农学院, 湖南 长沙 410128)

摘要:采用近红外方法, 分析 104 份甘蓝型油菜种子中的芥酸、油酸与硫代葡萄糖苷含量; 运用简化基因组技术(ddRADseq)对全部材料进行基因分型; 结合性状与基因型的数据, 利用 Tassel 混合线性模型进行全基因组关联分析。结果表明: 芥酸含量与硫代葡萄糖苷呈极显著正相关, 油酸含量与芥酸、硫代葡萄糖苷呈极显著负相关; 群体结构分析将 104 份材料划分为 4 个亚群; 通过全基因组关联分析发现, 与芥酸、油酸含量显著关联的 SNP 标记主要分布在 A08 与 C03 染色体上, 与硫代葡萄糖苷含量显著关联的标记分布在 A05、A02 与 C09 染色体上。

关键词:甘蓝型油菜; 全基因组关联分析; 简化基因组测序; 群体结构; 品质性状

中图分类号: S565.403.53 文献标志码: A 文章编号: 1007-1032(2020)02-0138-06

Genome-wide association analysis of quality-related traits in *Brassica napus*

CHEN Cong, LIU Zhongsong*

(College of Agronomy, Hunan Agricultural University, Changsha, Hunan 410128, China)

Abstract: *Brassica napus* is the main oil crop and the main source of edible vegetable oil in China. High-quality "double low" rapeseed is an important goal for the genetic breeding of *Brassica napus*. In this study, the content of erucic acid, oleic acid, and glucosinolate in 104 *Brassica napus* L. seeds has been analyzed by near-infrared method, and all materials have been genotyped by simplified genome technology(ddRADseq). Combining traits with genotype data, a Tassel mixed linear model has been used for genome-wide association analysis. The results showed that there was a significant positive correlation between erucic acid content and glucosinolates, and a significant negative correlation between oleic acid content and erucic acid and glucosinolates. The population structure analysis divided 104 materials into 4 subgroups. It was found that SNP markers significantly associated with erucic acid and the oleic acid content were mainly distributed on the A08 and C03 chromosomes, and markers associated with glucosinolate content were distributed on chromosomes A05, A02 and C09.

Keywords: *Brassica napus*; genome-wide association analysis; reduced-representation genome sequencing; population structure; quality-related traits

低芥酸、零芥酸是油菜育种的长期目标, 也是评价油菜品质的重要指标。菜籽饼粕中硫代葡萄糖苷含量也是“双低”油菜育种的重要目标^[1]。菜籽饼粕因富含蛋白质等营养物质, 具有较高的饲用价值, 因而受到许多研究人员的重视。但硫代葡萄糖

苷经动物食用后, 在胃肠条件下会水解生成硫氰酸盐等有毒物质^[2], 从而引起食欲不振等健康问题, 影响动物的正常发育。

王继变等^[3]通过构建甘蓝型油菜 DH 群体(doubled haploid, 双单倍群体), QTL(quantitative

收稿日期: 2019-03-06

修回日期: 2019-04-03

基金项目: 国家自然科学基金项目(3157101920)

作者简介: 陈聪(1992—), 男, 汉族, 湖南祁阳人, 硕士研究生, 主要从事油菜分子育种研究, 1291153987@qq.com; *通信作者, 刘忠松, 博士, 教授, 主要从事油菜分子育种研究, zslu48@sohu.com

trait locus, 数量性状基因座)分析发现, 控制芥酸含量的 QTL 位点位于 A2、A7、A9 和 C1 上, 控制油酸合成的 QTL 位点位于 A9 和 C7 上。QIU 等^[4]与 ECKE 等^[5]通过构建甘蓝型油菜 DH 群体与 QTL 分析, 均发现在 A8 与 C3 的相同位置上存在控制芥酸合成的 QTL。荐红举等^[6]通过 QTL 定位甘蓝型油菜种子硫代葡萄糖苷含量 QTL, 结合高密度 SNP 遗传图谱, 检测到 5 个 QTL, 分布在 A3、A9、C2 染色体上。王金芳^[7]通过构建 DH 作图群体, 定位到 2 个控制油酸合成的 QTL, 分布在 A5 与 A9 染色体上, 分别能解释 75.9%、3%~5% 的表型变异。

全基因组关联分析(genome whole association study, GWAS), 也称为连锁不平衡作图(linkage disequilibrium mapping), 是近 10 年来鉴定控制农作物复杂数量性状基因座的重要方法, 已广泛应用于谷子^[8]、小麦^[9]、玉米^[10]、油菜^[11]等农作物。目前, 在油菜领域内, GWAS 已成功应用于含油量^[12]、开花时间^[13]、株高^[14]、角果长度^[15]、种子萌发与活力^[16]等重要性状的研究。

本研究以 104 份甘蓝型自交系群体为研究材料, 对芥酸、油酸、硫代葡萄糖苷含量等品质性状进行分析考察, 采用 ddRADseq 测序技术, 运用 GWAS 方法将性状与 SNP 标记关联, 以寻找与性状显著关联的标记, 为高品质“双低”甘蓝型油菜品种的培育提供依据。

1 材料与方法

1.1 材料

以甘蓝型油菜高代自交系群体为关联分析群体, 总共 104 份材料, 其中黄籽材料 44 份, 黑籽材料 60 份。简化基因组测序中用到的参考基因组为法国冬油菜 Dammar v4.1。

1.2 方法

1.2.1 田间试验设计与性状考察

2017 年 9 月下旬, 将 104 份甘蓝型油菜自交系播种于湖南农业大学耘园基地。每份材料种植 1 个小区, 每个小区 5 行, 行距 40 cm。10 月中下旬间苗, 每行定苗 10~12 株, 株距 20~30 cm, 其他管理同大田常规管理。

油菜成熟后每个小区收获 3 份单株自交种子,

采用近红外光谱法测定芥酸、油酸、硫代葡萄糖苷含量, 每份材料用近红外仪 Foss NIR systems 5000 扫描 3 次, 取其平均值。利用 SPSS 22.0 对所得数据进行统计分析相关性分析。

1.2.2 简化基因组测序分型与群体结构分析

由武汉基诺赛克科技有限公司采用基于新一代测序平台 Illumina HiSeq 的 ddRADseq (double digested restriction-site associated DNA Sequencing) 技术进行简化基因组测序。基因组 DNA 经检测合格后, 用限制性内切酶 Sac I 和 Mse I 进行双酶切, 经 PCR 扩增、纯化, 筛选插入长度在 300~400 bp 的片段进行测序^[17], 原始测序读长为 Paired-end 150 bp。对于下机的原始数据(raw reads), 用 Cutadapt 与 Trimmomatic 进行数据质控, 然后将得到的 clean data 比对到参考基因组(Darmor v4.1)^[18]上, 进行变异检测(包含 SNP 与 InDel), 得到原始的包含变异信息的 vcf 格式文件。利用 Plink1.9^[19]对标记进行筛选, 过滤掉缺失率(missing rate)大于 20% 的标记, 同时剔除次要等位基因频率(MAF)小于 5% 的标记。

采用 Admixture1.3.0^[20]计算甘蓝型油菜群体结构 Q, 通过 CV error 值(cross-validation error rate, 交叉验证错误率)判断最佳分群数。使用 Plink1.9 (-indep-pairwise 50 5 0.5)过滤掉邻近连锁较强的 SNP, 用于群体结构分析。获得的 θ 矩阵采用 Excel 作 CV error 值, 获得趋势图, 用 R 语言作群体结构图。利用 Tassel 5.2.40 对自交系群体 104 份油菜自交系进行亲缘关系评估, 得到亲缘关系矩阵, 依据 YU 等^[21]的定义, 当计算结果出现亲缘关系值为负数时, 则定义为 0。

1.2.3 全基因组关联分析

运用 Tassel 5.2.40 进行关联分析, 采用混合线性模型, 对群体结构与亲缘关系进行校正。关联分析完成后, 用 R 语言 CMplot 包绘制 Q-Q 图与 Manhattan 图, 以 $P=0.01/N$ (N 为关联分析时用到的标记的总数) 为阈值, 筛选与性状显著关联的位点。

2 结果与分析

2.1 品质性状的统计分析

运用近红外仪测定了 104 份材料的芥酸、油酸、硫代葡萄糖苷含量, 相关性分析结果(表 1)表明:

芥酸含量与硫代葡萄糖苷含量之间呈极显著正相关, 相关系数为 0.349($P=0.008\ 4$); 油酸与芥酸、硫代葡萄糖苷含量之间呈极显著负相关, 相关系数分别为 -0.894 ($P=0.006\ 3$)、 -0.428 ($P=0.005\ 2$)。在 104 份材料中, 芥酸含量变化范围为 0.00%~46.99%, 平均值为 6.04%, 变异系数高达 214.94%; 油酸含量平均值为 63.32%, 变化范围为 8.27%~93.16%, 变异系数为 30.89%; 硫代葡萄糖苷含量平均值为 27.45 $\mu\text{mol/g}$, 变化范围为 3.56~60.16 $\mu\text{mol/g}$, 变异系数为 36.46%, 说明供试材料变异较丰富。

表 1 供试材料 3 个品质性状的平均值及变异系数

Table 1 The mean value and coefficient of variation of three quality traits

项目	最小值	最大值	平均值	变异系数
芥酸含量/%	0.00	46.99	6.04 $\pm$ 12.99	214.94
油酸含量/%	8.27	93.16	63.32 $\pm$ 19.56	30.89
硫代葡萄糖苷含量/($\mu\text{mol}\cdot\text{g}^{-1}$)	3.56	60.16	27.45 $\pm$ 10.01	36.46

2.2 基因分型与群体结构分析

2.2.1 简化基因组测序结果

基于 ddRADseq 技术, 对 104 份材料进行了简化基因组测序, 大部分测序位点碱基质量分值都在 Q30 以上, 将获得的 clean reads 序列与参考基因组(法国冬油菜 Dammar v4.1)进行比对, 每份材料的平均 reads 数为 1 917 862, 平均覆盖度为 2.48%,

平均覆盖深度为 11.14 倍。简化基因组测序获得了 2 106 617 个 SNP 标记, 经过滤获得 28 127 个高质量的 SNP 标记。

2.2.2 群体结构与亲缘关系

运用 Admixture1.3.0 计算群体的遗传结构, 根据连锁不平衡强度, 筛选 4 778 个均匀分布在染色体上的 SNP 标记, 对 K 值从 1 到 10 进行模拟运算, 得到 CV error 值(图 1)。当 $K=4$ 时, CV error 值最小。将 104 份甘蓝型自交系群体划分为 4 个亚群, 分群情况见图 2。P1 亚群包含 15 个甘蓝型油菜品系; P2 亚群包含 28 个甘蓝型油菜品系; P3 亚群包含 38 个甘蓝型油菜品系; P4 亚群包含 23 个甘蓝型油菜品系。

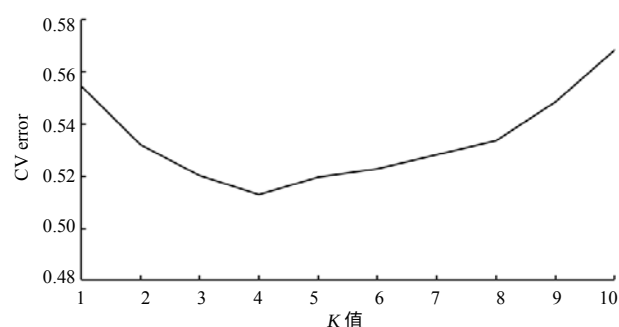


图1 不同 K 值的CV error

Fig.1 CV error with different K value

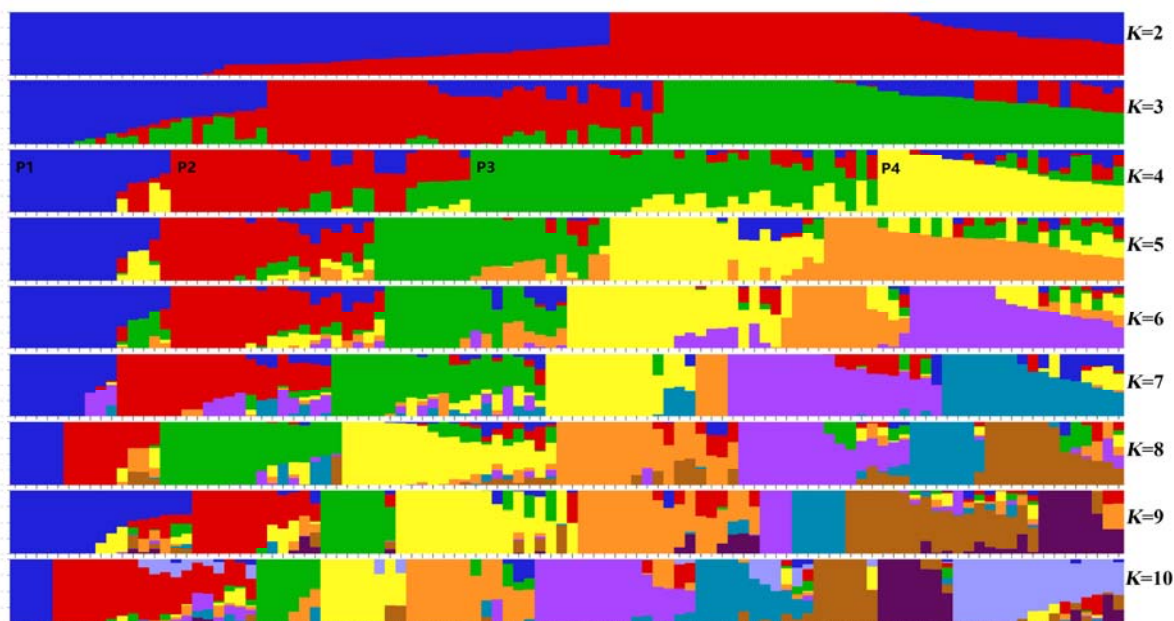


图2 104份甘蓝型油菜自交系群体结构

Fig.2 Population structure of 104 *Brassica napus*

2.3 全基因组关联分析

采用混合线性模型(mixed linear model, MLM), 运用 Tassel 5.2.40 对 104 份自交系群体材料种子中的芥酸、油酸和硫代葡萄糖苷含量进行 GWAS 分析(图 3), 检测到与芥酸、油酸、硫代葡萄糖苷含量显著关联的 SNP 位点数分别为 42、61、27。与芥酸含量显著关联的 SNP 位点在 A08、C02、C03、C05 与 C08 染色体上均有分布, 其中 C03 染色体上分布的 SNP 位点最多, 为 19 个, 单个 SNP 对表型

的解释率为 4.85%~24.45%; 与油酸含量显著关联的 SNP 位点分布在 A05、A08、C03、C05、C07 与 C08 染色体上, 其中 C03 染色体分布最多, 存在 29 个位点, 单个 SNP 对表型的解释率为 1.31%~30.62%; 关联分析检测到与硫代葡萄糖苷含量显著关联的 SNP 标记最少, 为 27 个, 分布在 A02、A09、C02、C05、C06 与 C09 染色体上, 单个标记对表型的解释率为 1.62%~25.92%。

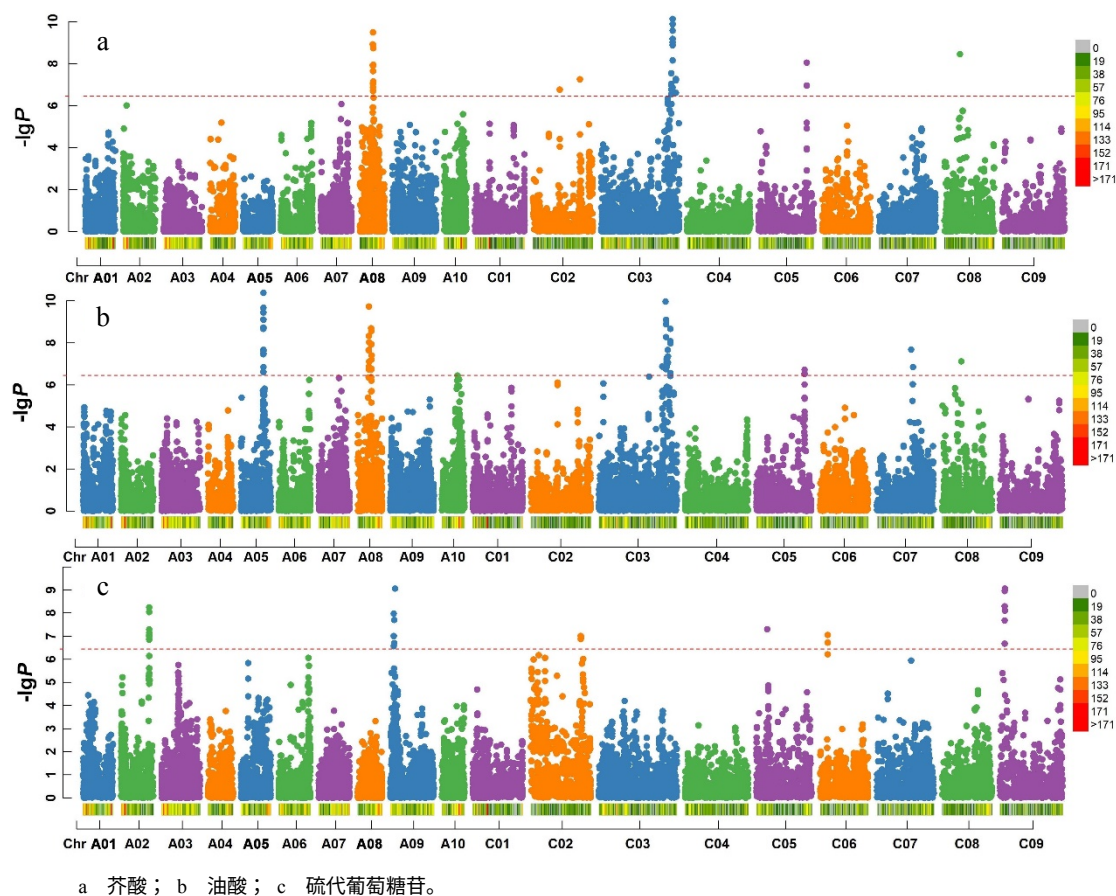


图3 3个品质性状全基因组关联分析

Fig.3 The genome-wide association analysis of three quality traits

3 结论与讨论

通过分析 104 份甘蓝型油菜自交系的芥酸、油酸与硫代葡萄糖苷含量 3 个品质性状, 发现性状的变异系数均在 30%以上; 相关性分析表明, 芥酸含量与硫代葡萄糖苷含量呈极显著正相关, 油酸含量与芥酸、硫代葡萄糖苷含量呈极显著负相关, 这与王健胜等^[22]的研究结果一致; 因此, 在甘蓝型油菜品种选育中, 可考虑以高芥酸与高硫代葡萄糖苷含量, 高油酸与低芥酸、低硫代葡萄糖苷含量的材料作为育种亲本。

前人^[23]在进行全基因组关联分析时发现, 群体内部的亲缘关系矩阵与群体结构会导致较高假阳性结果的出现; 因此, 需采用群体结构矩阵 Q 与亲缘关系矩阵作为协方差矩阵, 对关联分析进行校正。本研究通过群体结构分析, 可将 104 份材料分为 4 个亚群。采用混合线性模型分析性状与标记之间的关联程度, 发现与芥酸、油酸、硫代葡萄糖苷含量显著关联的 SNP 位点数分别为 42、61、27。

本研究通过 GWAS 检测到与芥酸、油酸含量显著关联的位点主要集中在 A08、C03 染色体上, 与

硫代葡萄糖苷含量显著关联的位点主要分布在 A02、A09、C09 染色体上。叶桑等^[24]通过对重组自交系群体(RIL)构建高密度的遗传连锁图谱,发现在 A05、A08 与 C03 染色体上存在多种脂肪酸含量的 QTL“富集区域”。LI 等^[25]通过关联分析,检测到 A08 与 C03 染色体上存在 2 个控制芥酸含量的基因。李施蒙^[26]利用 SNP 芯片对 520 份甘蓝型油菜的油酸含量进行关联分析,结果发现主要在 A08、C03 染色体上检测到与油酸含量显著关联的位点。本研究中检测到与油酸显著关联的位点主要集中在 A08、C03 上,这表明在 A08、C03 染色体上确实存在控制油酸含量的基因位点,可作为改良油菜籽中脂肪酸组成的重要区段。

参考文献:

- [1] ALI I, FARHATULLAH. Environmental and induced mutations influences on oil content, oleic acid and erucic acid in rapeseed[J]. Sarhad Journal of Agriculture, 2013, 29(2), 187–193.
- [2] WALLIG M A, BELYEA R L, TUMBLESOM M E. Effect of pelleting on glucosinolate content of Crambe meal[J]. Animal Feed Science and Technology, 2002, 99(1/2/3/4): 205–214.
- [3] 王继变, 陈洪成, 陈玉波, 等. 通过条件 QTL 定位分析油菜不同脂肪酸组分对含油量性状的影响[J]. 浙江大学学报(农业与生命科学版), 2013, 39(5): 504–512. WANG J B, CHEN H C, CHEN Y B, et al. Molecular dissection of oil content with respect to fatty acid compositions by conditional QTL analysis in oilseed rape[J]. Journal of Zhejiang University(Agriculture and Life Sciences), 2013, 39(5): 504–512.
- [4] QIU D, MORGAN C, SHI J, et al. A comparative linkage map of oilseed rape and its use for QTL analysis of seed oil and erucic acid content[J]. Theoretical and Applied Genetics, 2006, 114(1): 67–80.
- [5] ECKE W, UZUNOVA M, WEIBLER K. Mapping the genome of rapeseed (*Brassica napus* L.) II. Localization of genes controlling erucic acid synthesis and seed oil content[J]. Theoretical and Applied Genetics, 1995, 91(6/7): 972–977.
- [6] 荐红举, 魏丽娟, 李加纳, 等. 利用 SNP 高密度遗传连锁图谱定位甘蓝型油菜种子硫代葡萄糖苷含量的 QTL[J]. 作物学报, 2014, 40(8): 1386–1391. JIAN H J, WEI L J, LI J N, et al. Mapping quantitative traits loci for seed glucosinolate content in *Brassica napus* using high-density SNP map[J]. Acta Agronomica Sinica, 2014, 40(8): 1386–1391.
- [7] 王金芳. 甘蓝型油菜中一个新高油酸位点的遗传分析[D]. 武汉: 华中农业大学, 2014. WANG J F. Identification of a new high oleic acid locus in *Brassica napus*[D]. Wuhan: Huazhong Agricultural University, 2014.
- [8] 贾小平, 张博, 全建章, 等. 98 份谷子材料穗部性状的全基因组 SNP 关联分析[J]. 湖南农业大学学报(自然科学版), 2019, 45(1): 25–29. JIA X P, ZHANG B, QUAN J Z, et al. Genome-wide association analysis between SNP and panicle traits for 98 foxtail millet materials[J]. Journal of Hunan Agricultural University(Natural Sciences), 2019, 45(1): 25–29.
- [9] 翟俊鹏, 李海霞, 毕惠惠, 等. 普通小麦主要农艺性状的全基因组关联分析[J]. 作物学报, 2019, 45(10): 1488–1502. ZHAI J P, LI H X, BI H H, et al. Genome-wide association study for main agronomic traits in common wheat[J]. Acta Agronomica Sinica, 2019, 45(10): 1488–1502.
- [10] 张小琼, 郭剑, 代书桃, 等. 玉米花期根系结构的表型变异与全基因组关联分析[J]. 中国农业科学, 2019, 52(14): 2391–2405. ZHANG X Q, GUO J, DAI S T, et al. Phenotypic variation and genome-wide association analysis of root architecture at maize flowering stage[J]. Scientia Agricultura Sinica, 2019, 52(14): 2391–2405.
- [11] 李阳阳, 荆蓉蓉, 吕蓉蓉, 等. 甘蓝型油菜湿害胁迫响应性状的全基因组关联分析及候选基因预测[J]. 作物学报, 2019, 45(12): 1806–1821. LI Y Y, JING R R, LYU R R, et al. Genome-wide association analysis and candidate genes prediction of waterlogging-responding traits in *Brassica napus* L.[J]. Acta Agronomica Sinica, 2019, 45(12): 1806–1821.
- [12] 魏大勇, 崔艺馨, 梅家琴, 等. 油菜种子含油量 GWAS 分析及位点整合系统构建[J]. 作物学报, 2018, 44(9): 1311–1319. WEI D Y, CUI Y X, MEI J Q, et al. Genome-wide association study on seed oil content in rapeseed and construction of integration system for oil content loci[J]. Acta Agronomica Sinica, 2018, 44(9): 1311–1319.
- [13] XU L P, HU K N, ZHANG Z Q, et al. Genome-wide association study reveals the genetic architecture of flowering time in rapeseed (*Brassica napus* L.)[J]. DNA Research, 2015, 23(1): 43–52.
- [14] 唐敏强, 程晓晖, 童超波, 等. 甘蓝型油菜株高性状的全基因组关联分析[J]. 作物学报, 2015, 41(7): 1121–1126. TANG M Q, CHENG X H, TONG C B, et al. Genome-wide association analysis of plant height in rapeseed (*Brassica napus*)[J]. Acta Agronomica Sinica, 2015,

- 41(7): 1121–1126.
- [15] 周庆红, 周灿, 郑伟, 等. 甘蓝型油菜角果长度全基因组关联分析[J]. 中国农业科学, 2017, 50(2): 228–239. ZHOU Q H, ZHOU C, ZHENG W, et al. Genome-wide association analysis of silique length in *Brassica napus* L. [J]. Scientia Agricultura Sinica, 2017, 50(2): 228–239.
- [16] HATZIG S V, FRISCH M, BREUER F, et al. Genome-wide association mapping unravels the genetic control of seed germination and vigor in *Brassica napus* [J]. Frontiers in Plant Science, 2015, 6: 221.
- [17] WU Z K, WANG B, CHEN X, et al. Evaluation of linkage disequilibrium pattern and association study on seed oil content in *Brassica napus* using ddRAD sequencing[J]. PLoS One, 2016, 11(1): e0146383.
- [18] CHALHOUB B, DENOEUDE F, LIU S Y, et al. Early allopolyploid evolution in the post- Neolithic *Brassica napus* oilseed genome[J]. Science (New York, N. Y.), 2014, 345: 950–953.
- [19] PURCELL S, NEALE B, TODD-BROWN K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses[J]. The American Journal of Human Genetics, 2007, 81(3): 559–575.
- [20] ALEXANDER D H, NOVEMBRE J, LANGE K. Fast model-based estimation of ancestry in unrelated individuals[J]. Genome Research, 2009, 19(9): 1655–1664.
- [21] YU J M, BUCKLER E S. Genetic association mapping and genome organization of maize[J]. Current Opinion in Biotechnology, 2006, 17(2): 155–160.
- [22] 王健胜, 侯桂玲, 李少钦, 等. 甘蓝型油菜育种亲本产量及品质性状遗传效应分析[J]. 江苏农业学报, 2015, 31(3): 489–493. WANG J S, HOU G L, LI S Q, et al. Genetic effects of yield traits and quality traits for breeding parents of *Brassica napus* L [J]. Jiangsu Journal of Agricultural Sciences, 2015, 31(3): 489–493.
- [23] YU J M, PRESSOIR G, BRIGGS W H, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness[J]. Nature Genetics, 2006, 38(2): 203–208.
- [24] 叶桑, 崔翠, 郜欢欢, 等. 基于 SNP 遗传图谱对甘蓝型油菜部分脂肪酸组成性状的 QTL 定位[J]. 中国农业科学, 2019, 52(21): 3733–3747. YE S, CUI C, GAO H H, et al. QTL identification for fatty acid content in *Brassica napus* using the high density SNP genetic map[J]. Scientia Agricultura Sinica, 2019, 52(21): 3733–3747.
- [25] LI F, CHEN B, XU K, et al. Genome-wide association study dissects the genetic architecture of seed weight and seed quality in rapeseed(*Brassica napus* L.) [J]. DNA Research, 2014, 21(4): 355–367.
- [26] 李施蒙. 甘蓝型油菜种子硫代葡萄糖苷含量和脂肪酸组分全基因组关联分析[D]. 重庆: 西南大学, 2016. LI S M. Genome-wide association mapping of seed glucosinolate and fatty acid composition in *Brassica napus* [D]. Chongqing: Southwest University, 2016.
- 责任编辑: 毛友纯
英文编辑: 柳 正