

甘蓝型矮秆直立株型油菜株高性状 基因的初步定位

李盼^{1,2}, 李超^{1,2}, 张瑞茂², 张显强^{1,3}, 罗京^{1,2}

(¹ 贵州师范大学生命科学学院, 贵阳 550025; ² 贵州省农业科学院油菜研究所, 贵阳 550006; ³ 贵州警察学院, 贵阳 550005)

摘要: 定位甘蓝型矮秆直立株型油菜的株高基因, 解析矮秆性状的遗传规律, 对油菜育种的产量具有重要意义。通过甘蓝型三系油菜(不育系 5824ci × 恢复系 5771r) 杂交, 在 F₁ 中发现变异单株, 命名“DW871”。经多代自交, 从群体中选取纯合矮秆和对应纯合高秆杂交, 从杂交后代分离群体中分别选取 47 个矮秆、47 个高秆和 10 个纯合矮秆材料, 分别构建基因池。利用 BSA-seq 简化基因组测序技术, 以 47 个矮秆与 47 个高秆, 10 个纯合矮秆与 47 个高秆进行关联分析 (BSA), 从 47 个矮秆与 47 个高秆基因混合池间检测到差异 SNP 共 121998 个, 非同义突变 SNP 共 1582 个, 在 ChrA10 染色体上定位 1 个显著关联区域, 区间长度 6.39 Mb, 含 1405 个候选基因; 从 10 个纯合矮秆与 47 个高秆基因混合池间共获得 1752011 个 SNP, 非同义突变 SNP 共 27723 个, InDel 518420 个, 在 ChrA03、ChrA04、ChrA06、ChrA07、ChrA10 和 ChrC03 上定位共 19 个与性状相关的候选区域, 区间长度 5.35 Mb, 含 1143 个候选基因。然而由于未达到理论阈值, 这个区域很可能是假阳性区域, 需要进一步验证。47 个矮秆基因混池和 10 个纯矮秆基因混池在 ChrA10 染色体上定位的关联区域互相重合, 重合区间为 1.83 Mb。本研究在 ChrA10 染色体上定位 1 个与 DW871 株高性状显著关联区间, 这一研究结果为精确定位甘蓝型矮秆直立株型油菜株高性状基因奠定了良好基础。

关键词: 甘蓝型矮秆油菜; 株高基因; BSA-seq 技术

Low-resolution Mapping of a Qualitative Trait Controlling Plant Height in *Brassica napus* L.

LI Pan^{1,2}, LI Chao^{1,2}, ZHANG Rui-mao², ZHANG Xian-qiang^{1,3}, LUO Jing^{1,2}

(¹ College of Life Sciences, Guizhou Normal University, Guiyang 550025; ² Institute of Rape Research, Guizhou Academy of Agricultural Sciences, Guiyang 550006; ³ Guizhou Police College, Guiyang 550005)

Abstract: Genetic identification of the functional genes causing the erect dwarf-type are of great significance in academic study and rapeseed breeding. We previously identified a naturally-occurred dwarf mutant (named DW871) from a single F₁ plant by crosses of sterile line 5824 × recovery line 5771R. By making crosses between the homozygous mutant with a homozygous wild-type line, 47 dwarf rods (DR), 47 high rods (HR) and 10 homozygous dwarf rod lines (HDR) were identified for a genetic mapping study. Three bulks were subjected

收稿日期: 2019-05-14 修回日期: 2019-07-11 网络出版日期: 2019-09-06

URL: <http://doi.org/10.13430/j.cnki.jpgr.20190514001>

第一作者研究方向为植物生态, E-mail: 1547649286@qq.com

通信作者: 张显强, 研究方向为喀斯特植物生理生态, E-mail: zhangxiangqiang@126.com

李超, 研究方向为植物遗传育种, E-mail: gzlichao@126.com

基金项目: 贵州省科学技术基金重点项目 (黔科合基础 [2017] 1415); 贵州省优秀青年科技人才项目 (黔科合平台人才 [2017] 5642); 贵州省支撑计划重点项目 (黔科合支撑 [2017] 2568 号); 贵州省人才基地 (RCJD2018-14); 贵州植物保育技术应用省级工程研究中心

Foundation project: Guizhou Science and Technology Foundation Key Project (Guizhou Science and Technology Foundation [2017] 1415), Guizhou Excellent Youth Science and Technology Talents Project (Guizhou Science and Technology Cooperation Platform Talents [2017] 5642), Guizhou Provincial Support Plan Key Project (Guizhou Science and Technology Cooperation Support [2017] 2568), Guizhou Talent Base (RCJD2018-14), Guizhou Provincial Engineering Research Center for Plant Conservation Technology Application

for simplified genome sequencing technology (BSA-seq), in order to achieve the preliminary mapping of the underlying functional genes. A total of 121998 SNPs were detected between 47 dwarves and 47 high-stem mixed gene pools. 1582 SNPs resulting in non-synonymous mutations were located on ChrA10 chromosome, Interval length is 6.39 Mb, including 1405 candidate genes. Out of 1752011 SNPs that were obtained between HR and HDR pools 27723 SNPs led to the non-synonymous mutations. A total of 19 candidate loci were mapped on ChrA03, ChrA04, ChrA06, ChrA07, ChrA10 and ChrC03, Interval length is 5.35 Mb, including 1143 candidate genes. However, these locus showed un-significant associations and need to be further verified. Interestingly, the association region on the ChrA10 chromosome was revealed by cases and a coincidence interval of 1.83 Mb was suggested. As a summary, the qualitative locus controlling the plant height in DW871 was mapped on ChrA10 chromosome, and this result laid a good foundation for the future fine mapping of the functional gene in *Brassica napus* L.

Key words: *Brassica napus* L.; plant height gene; BSA-seq technology

油菜 (*Brassica napus* L.) 是重要的油料作物, 是植物油脂的第三大来源。株高是产量相关的重要农艺性状, 也是影响油菜倒伏的因素之一。油菜倒伏后对产量和品质有重要影响, 贺亚军等^[1]和郑本川等^[2]分析油菜倒伏后产量与品质变化, 一般减产 10%~30%, 严重时可达到 60%, 倒伏油菜含油量也比正常油菜低 10%~30%。张书芬等^[3]研究表明, 株高为 1.2~1.3 m 的矮秆油菜不仅增强了抗倒伏性而且更适合机械化收获, 一些相对较矮的油菜品种 (如中双 11 和圣光 127) 已经在生产上推广和应用, 这些品种抗倒伏能力强。目前, 发现并鉴定的油菜矮秆材料较少, 可利用的矮秆有效基因不多, 油菜株高调控的分子基础研究有限, 导致矮化新品种选育及大面积推广进展缓慢。因此, 进一步挖掘和鉴定油菜矮秆突变体, 定位控制矮秆油菜株高性状基因, 解析株高形成的分子基础, 对油菜矮化育种及推广应用具有重要的促进作用^[4-5]。

在油菜基因组公布及测序技术高速发展之前, 油菜及其两个亚种白菜和甘蓝相关的遗传连锁图谱的构建为油菜功能基因的定位提供了便利。在对矮秆突变体矮秆表型的遗传属性鉴定的基础上, 研究者采用 BSA 法结合遗传图谱上的标记分析开展了矮秆基因的定位与克隆工作。Muangprom 等^[6-7]和唐丁等^[8]采用 BSA 方法将白菜矮秆基因 *DWF₂* 定位在了白菜 A6 (IM) 连锁图的底端。2010 年, Liu 等^[9]利用油菜 92-B10 双单倍体的 EMS 诱变体, 从 A6 染色体上克隆了一个编码 DELLA 蛋白并作为赤霉素信号受体的半显性矮秆突变体基因 *ds-1* (BnRGA)。2011 年, Zeng 等^[10]将突变体 *bnaCdwf* 矮秆表型的矮秆基因 *BnaCDWF* 定位到油菜 C8 (N18) 连锁群上,

其与左右两侧最近的 AFLP 标记 *EAI0MC09* 和 *EAI2MC02* 的图距分别为 0.2 cM 和 0.05 cM。杨朋娜等^[11]和 Sun 等^[12]通过 SSR 和 SRAP 标记结合 BSA 法将矮秆基因 *ndf1* 定位在油菜 C3 (N13) 连锁群上, *ndf1* 与最近的 SSR 连锁标记 Na12-E02 的图距为 6.34 cM。Wang 等^[13]用 EMS 诱变油菜纯系 NJ7982 的显性矮秆突变体 *Bndwf1* 和中双 11 衍生的回交分离群体, 将 *BnDWF* 基因该位点精细定位到 A9 染色体 152 kb 的区段, 并从 14 个注释基因中鉴定了 7 个候选基因。2016 年, Wang 等^[14]利用油菜的 SNP 芯片结合 QTL 作图分析, 将控制油菜矮秆突变体 *Bndwf1* 矮化表型的主效基因 *BnDWF1* 和突变体 *Bndwf/dcl1* 的主效矮化基因 *BnDWF/DCL1* 定位在油菜 A9 染色体 2.4 cM 和 C5 染色体 6.58 cM 的遗传区间。进一步利用目标区间的 scaffold 序列信息开发 SSR 标记将候选区段分别缩小到 A9 染色体 152 kb 和 C5 染色体 175 kb 的物理区间。2017 年, 赵波^[15]和 Zhao 等^[16]以甘蓝型油菜矮秆突变体 *ds-3* 为材料, 对株高基因 *DS-3* 进行遗传定位并克隆了编码赤霉素 (GA) 信号转导抑制因子 DELLA 蛋白的 *BnaC07.rga-ds* 基因; 对油菜矮秆突变体 *ds-4* 的矮秆基因进行了精细定位, 结果将 *DS-4* 锁定在 SNP32 和 SNP40 之间, 对应于甘蓝型油菜 “*Darmor-bzh*” 基因组 C05 染色体上长度约为 475 kb 的区间内。

从上述内容可知目前克隆的油菜矮秆基因很少, 多数处于定位阶段。究其原因主要是甘蓝型油菜基因组数据库不够完善以及甘蓝型油菜基因组的复杂性, 这很大程度上降低了连锁标记开发的成功率, 影响了定位进程。但随着芯片技术和高通量测序的发展, 以 SNP 单核苷酸多态性和 InDel 插入缺

失标记为主的定位标记的应用及其基因型分析效率的提高,将大大加快油菜功能基因定位和克隆速度。甘蓝型矮秆油菜 DW871 具有整个生育期处于直立状态,株高 130~140 cm,株型紧凑,完全抗倒伏,结角密度集中,与现有报道的矮秆油菜完全不属于同一类型,产量潜力大,是在我国具有特色,有利于机械化栽培的矮秆种质资源^[17-18]。拟通过 BSA-seq 方法研究 DW871 株高株型性状的基因遗传机制,为进一步基因克隆及综合利用奠定基础。

1 材料与方法

1.1 植物材料

2008 年,从甘蓝型三系杂交油菜[不育系(5824ci)×恢复系(5771r)]品系试验 F₁ 中发现的变异单株(命名为 DW871),后经过 4 代自交纯合,育成稳定遗传的甘蓝型油菜矮秆新材料 DW871。该株系矮秆单株后代性状表现稳定,农艺性状表现为主序花蕾段比同源正常高秆系平均长 12.5 cm,花序生长特性、开花习性等与甘蓝类似的标记性状,节间短,木质部厚,抗倒性特强,株型紧凑,分枝位低,一次有效分枝数多,果枝硬,着果密度达 1.8 cm 左右,收获指数为 0.37,其品质表现为双低,杂黄籽,平均含油量 47.6%,丰产性与同源高秆系相当,是甘蓝型油菜新型矮秆种质资源。

张瑞茂等^[17]分别以 DW871 为父母本,通过构建不同世代的群体对该品系的株高性状进行遗传分析,结果表明 DW871 株高性状是由 1 对显性主基因控制,不存在细胞质效应,但是存在修饰基因的影响,表现出数量性状的遗传特征。

1.2 株高(矮秆、高秆)基因池构建

2015 年,将杂合矮秆进行套袋自交(第 7 代)。2016 年春季,根据田间植株性状,从分离的后代中,分别提取株高性状为矮秆型和高秆型植株各 47 株的全基因组 DNA,矮秆编号 DW₁(1~47),高秆编号 HR(1~47);提取纯合矮秆植株全基因组 DNA,编号 DW₂(1~10),分别构建矮秆群体 DW₁(1~47)、DW₂(1~10)和高秆群体 HR(1~47)3 个基因混合池。

使用十六烷基三甲基溴化铵(CTAB)方法^[19]对供试材料进行全基因组 DNA 提取。用 Nanodrop 2000 UV-Vis 分光光度计(NanoDrop, Wilmington, DE, USA)分析 DNA 浓度和质量,并将浓度调至 50 ng/μL 后,等体积混合后建成 2 个株高表现极端的子代基因组混合池,其中矮秆性状基因池 1 个,纯

合矮秆基因池 1 个,纯合高秆性状基因池 1 个。

1.3 测序文库构建

选择已测序完成的甘蓝型油菜(*Brassica napus* L.)基因组作为参考基因组(下载地址: <http://www.genoscope.cns.fr/brassicanapus/data/>),根据基因组大小以及 GC 含量等信息作为参考基因组进行电子酶切预测,利用限制性内切酶对基因进行酶切,加上带有 barcode 的接头后,对每个样品进行扩增,然后混合样品。样品检测合格后,用超声波的方法将 DNA 随机打断成 350 bp 的片段, DNA 片段经末端修复、3' 端加 A、加测序接头、纯化、PCR 扩增完成测序文库的构建。文库经质检合格后通过北京的 Biomarker Technologies 公司利用 Illumina HiseqTM 2500 (Illumina, Inc; San Diego, CA, USA) 测序平台进行双末端测序。群体株高基因混合池测序深度 ≥50X。在测序过程中,对每个循环进行实时监控,分析 Q30 和 GC 含量,保证测序质量。

1.4 欧氏距离关联分析方法

欧氏距离(ED, Euclidean Distance)算法,是利用测序数据寻找混池间存在的显著差异标记,并以此评估与性状关联区域的方法^[20]。理论上,群体分离分析法(BSA, bulked segregant analysis)项目构建的 2 个混池间除了目标性状相关位点存在差异,其他位点均趋向于一致,因此非目标位点的欧氏距离值应趋向于 0。欧氏距离值越大表明该标记在两混池间的差异越大。

1.5 多态性位点检测

多态性位点检测主要使用 GATK^[21]软件工具包实现。根据过滤后测序序列在参考基因组的定位结果,使用 Picard^[22]进行去重复(Mark Duplicates)、GATK 进行局部重比对(Local Realignment)、碱基质量值校正(Base Recalibration)等预处理,以保证检测得到的 SNP 准确性,再使用 GATK 进行多态性位点的检测,过滤,并得到最终的多态性位点。

1.6 候选区域内基因注释

利用 BLAST^[23]软件对候选区间内编码基因进行 GO^[24]、COG^[25]、KEGG^[26]、NG、NT、trEMBL、Swiss-Prot 等多个数据库深度注释,通过详细的注释,快速筛选候选基因。

2 结果与分析

2.1 测序数据统计与评估

对 3 个样本 DNA 进行测序, DW₁ 混池获得 721.39 Mb 测序序列, DW₂ 混池获得 133.90 Mb 测

序序列;HR 混池获得 648.32 Mb 测序序列,3 个样品混池平均 Q30 比例为 94.66%,平均 GC 含量为 39.05%。依据甘蓝型油菜 (*Brassica napus* L.) 基因组进行基因比对, DW₁ 混池和 HR 混池基因组覆盖深度分别是 79.00X、70.00X,基因组覆盖度分别为 95.55%、95.35%。DW₂ 混池基因组覆盖度约为 94.24%,基因组覆盖深度约为 18 X(表 1),

DW₂ 混池覆盖深度较其他 2 个混池小,可能原因是由于矮秆样本量少。根据染色体各位点的覆盖深度情况进行作图,基因组覆盖深度在染色体上的分布比较均匀,认为测序随机性比较好。3 个基因池的染色体覆盖深度分布分别见图 1、图 2、图 3,图上深度不均一的地方可能是由于重复序列、PCR 偏好性引起的。

表 1 样品测序数据统计

Table 1 Statistics of the sequencing datasets

样品 Sample	测序数据 (Mb) Read number	过滤后数据 (Mb) Clean Reads	过滤后碱基数 (Mb) Clean Base	Q30 (%)	GC (%)	平均覆盖深度 (X) Av-depth	基因组覆盖度 (%) Genome coverage
HR	648.32	324.16	81656.93	95.92	40.12	70.00	95.35
DW ₁	721.39	360.69	90857.89	95.78	38.99	79.00	95.55
DW ₂	133.90	66.95	20056.12	92.27	38.05	18.00	94.24

Q30: 质量值大于等于 30 的碱基占总碱基数的百分比; GC: 样品 GC 含量; 基因组覆盖度: 参考基因组上被测序序列覆盖到的碱基数占基因组的百分比。

Q30: the percentage of bases with a mass value greater than or equal to 30 as a percentage of the total number of bases, GC : Sample GC content, Genome coverage: the percentage of bases covered by the sequence in the reference genome

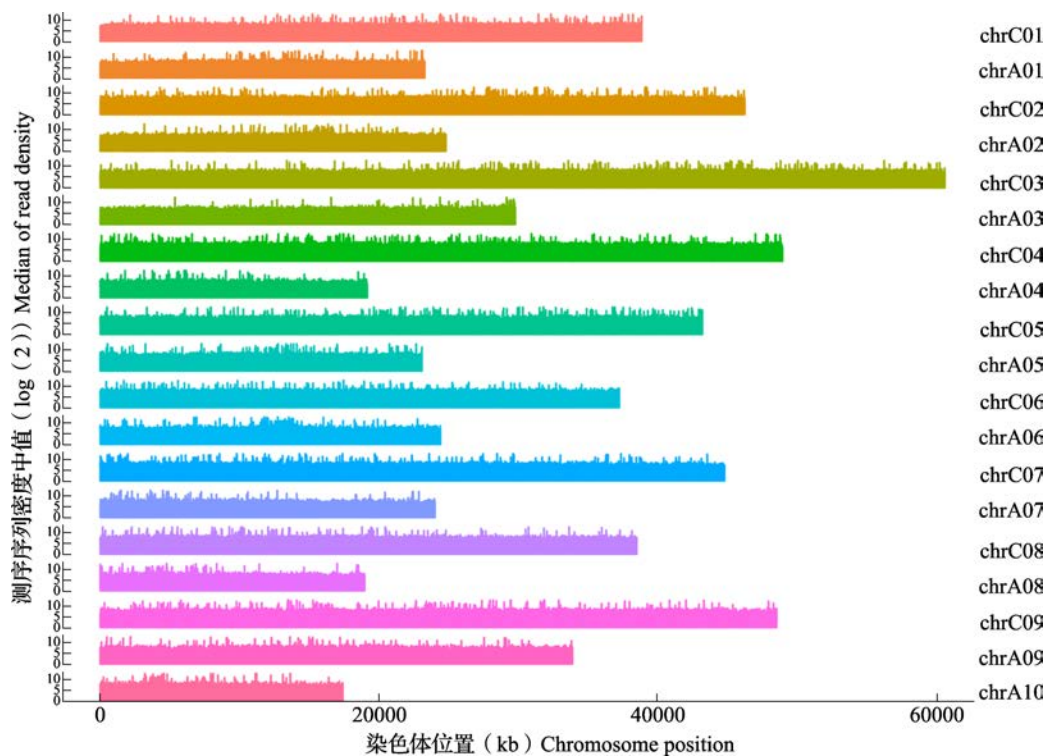


图 1 HR 混池基因组覆盖深度在染色体上分布

Fig.1 Distribution of genome coverage depth on chromosomes in HR mixing pool

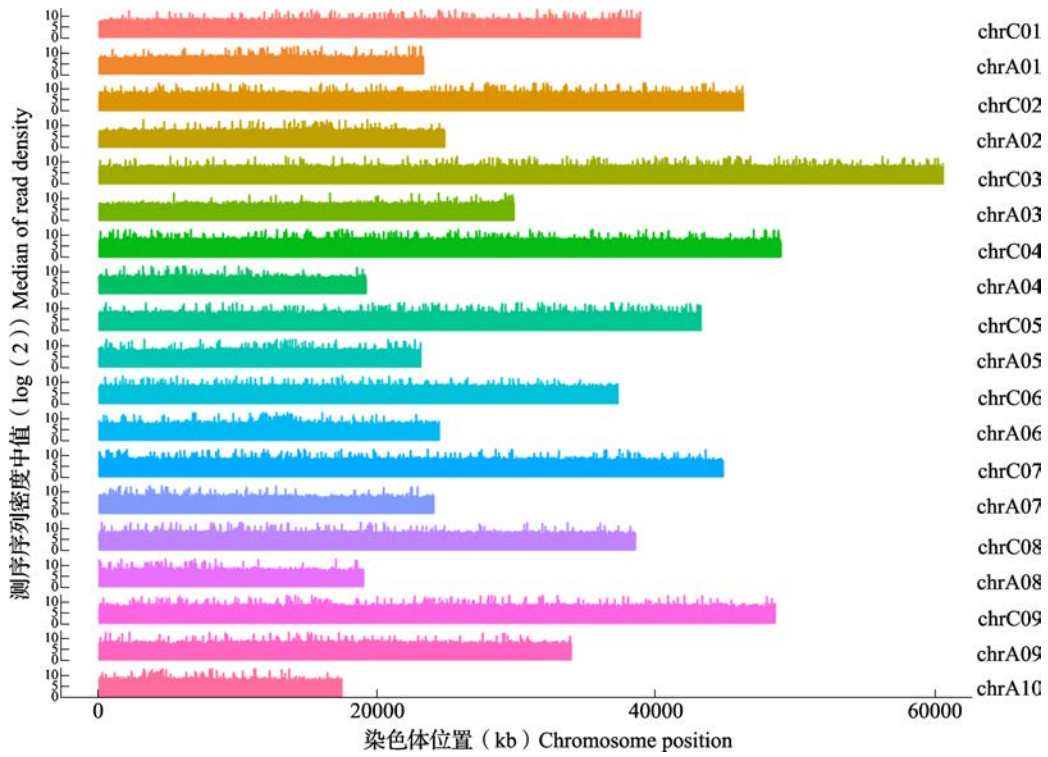


图 2 DW₁ 混池基因组覆盖深度在染色体上分布
Fig.2 Chromosomal distribution of sequencing coverage depth in DW₁ pool

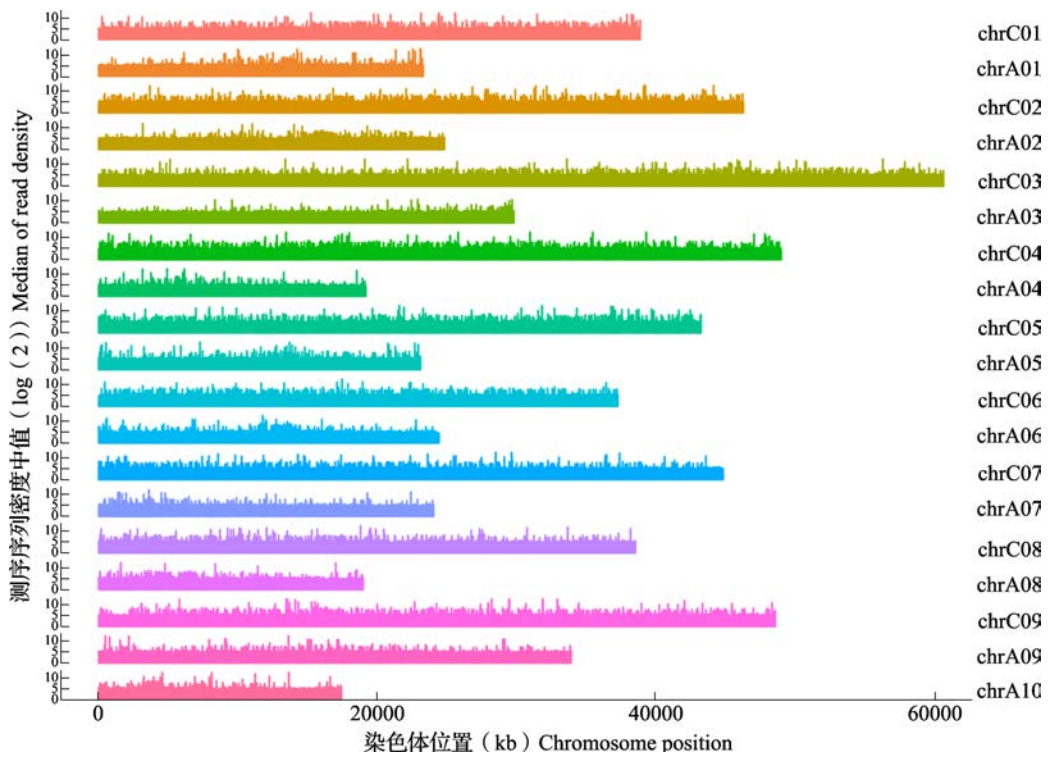


图 3 DW₂ 混池基因组覆盖深度在染色体上分布
Fig.3 Chromosomal distribution of sequencing coverage depth in DW₂ pool

2.2 多态性位点检测与注释

根据样品过滤后测序序列与参考基因组的比对及 SnpEff^[27] 注释结果, 汇总 3 个基因池之间差异的变异位点, DW₁ 混池与 HR 混池间共检测差异 SNP 121998 个, 其中基因内 1529 个, 引起非同义突变的 SNP 共 1582 个; DW₁ 混池与 HR 混池间差异 InDel

共有 43319 个, 其中基因内 1232 个, 发生移码突变 211 个。DW₂ 混池与 HR 混池间共获得 1752011 个 SNP, 其中基因内 23848 个, 引起非同义突变的 SNP 共 27723 个; DW₂ 混池与 HR 混池间共获得 518420 个 InDel, 其中基因内 15749 个, 发生移码突变 3015 个(表 2)。

表 2 SNP、Indel 统计及注释结果

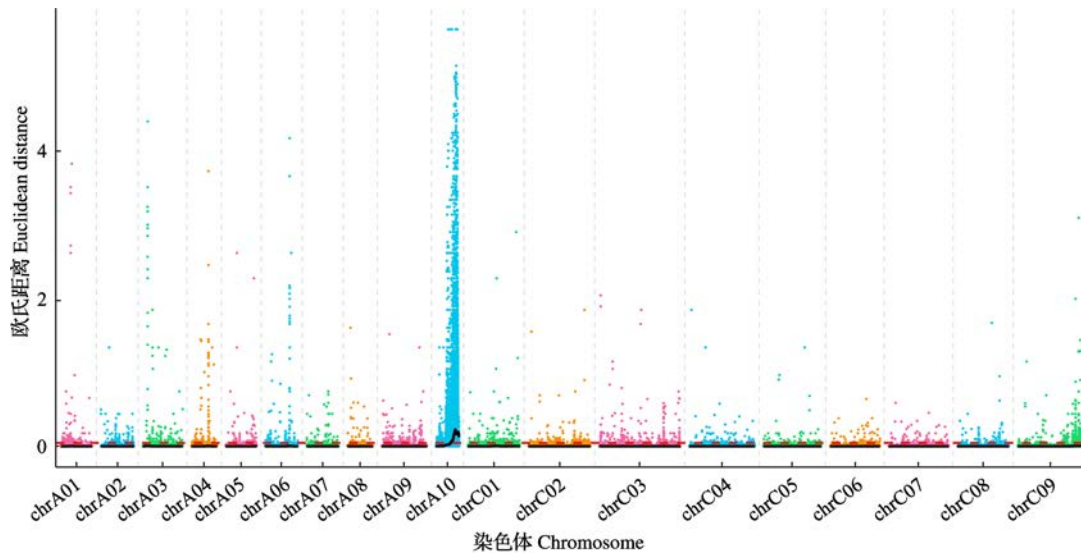
Table 2 Statistics of SNP and Indel polymorphisms

类型 Type	单核苷酸多态性 SNP		插入缺失标记 Indel	
	DW ₁ 与 HR 基因混池 DW ₁ vs HR gene pool	DW ₂ 与 HR 基因混池 DW ₂ vs HR gene pool	DW ₁ 与 HR 基因混池 DW ₁ vs HR gene pool	DW ₂ 与 HR 基因混池 DW ₂ vs HR gene pool
基因间区 Intergenic	57069	799849	17968	206153
基因内 Intragenic	1529	23848	1232	15749
内含子 Intron	1396	22778	974	12035
基因上游区域 Upstream	21407	303205	9141	106155
基因下游区域 Downstream	20128	289770	8187	100210
剪切受体突变 Splice_site_acceptor	6	85	3	59
剪切供体突变 Splice_site_donor	7	72	5	89
剪切区域突变 Splice_site_region	54	960	38	572
起始密码子丢失 Start_lost	4	70	4	57
同义编码突变 Synonymous_coding	2965			
非同义编码突变 Non_synonymous_coding	1582	27723		
同义终止密码突变 Synonymous_stop	5	83		
非同义起始密码子突变 Non_synonymous_start		12		
移码突变 Frame_shift			211	3015
密码子插入 Codon_insertion			69	1018
密码子删除 Codon_deletion			57	863
外显子删除 Exon_deleted				1
非密码子插入			23	396
Codon_change_plus_codon_insertion				
非密码子删除			37	564
Codon_change_plus_codon_deletion				
终止密码子获得 Stop_gained	21	454	8	101
终止密码子丢失 Stop_lost	9	166	6	59
其他 Other	15816	242538	5356	71324
总数 Total	121998	1752011	43319	518420

2.3 关联分析

通过计算 2 个混池间等位基因的基因型频率确定与目标性状关联的区域。进行关联分析前, 首先对 SNP 和 Indel 进行过滤, 过滤标准如下: 首先过滤掉有多个基因型的 SNP 和 Indel 位点, 其次过滤掉序列支持度小于 4 的 SNP 和 Indel 位点, 再次过滤掉混池之间基因型一致的 SNP 和 Indel 位点。在进行分析时, 利用两混池间基因型存在差异的 SNP 位点, 统计各个碱基在不同混池中的深度, 并计算每个位点 ED 值, ED 值越高, 代表该点关联效果越好, DW₁ 混池与 HR 混池间共得到高质量的可靠 SNP

位点 121998 个, 根据 Hill 等^[20] 的 ED 值计算方法, 由 ED 算出的关联值分布如图 4 所示, 取所有位点拟合值的中间值与 3 倍标准差之和作为分析的关联阈值, 计算得 0.0458, chrA10 染色体拟合后的 ED 值超过了显著性关联阈值(图 4), 说明 chrA10 染色体上存在着与 DW871 株高性状有关联的基因。根据关联阈值判定, 在 ChrA10 染色体上 ED 值表现最高, 获得 1 个位于 ChrA10 染色体的关联区域, 总长度为 6.39 Mb, 共包含 1405 个基因(候选区间 11.00~17.39 Mb), 其中非同义突变 SNP 位点的基因共 201 个。



彩色的点代表每个 SNP 位点的 ED 值,黑色的线为拟合后的 ED 值,红色的虚线代表显著性关联阈值。下同
 The color point represents the ED value of each SNP locus, the black line represents the fitted ED value, and the red dotted line represents the significant correlation threshold. The same as below

图 4 DW₁ 与 HR 基因混池间 ED 值在染色体上的分布
 Fig.4 Distribution of ED value of DW₁and HR mixed pool on chromosome

DW₂ 与 HR 混池间筛选到高质量的可信 SNP 位点 2871863 个, InDel 位点 835978 个。取所有 SNP 位点拟合值的中间值与 3 倍标准差之和作为分析的关联阈值, 计算得 4.07。取所有 InDel 位点拟合值的中间值与 3 倍标准差之和作为分析的关联阈值, 计算得 3.92, 理论上, 目标位点及其附近的连锁位点应趋近于该阈值, 因此显著关联的区域附近应该出现一个较高的峰值。但从结果上看, 没有超过理论阈值的区域, 说明本试验中没有发现显著的定位结果。为了充分利用数据, 将阈值降低以寻找

比较可能的定位区域, DW₂ 混池与 HR 混池间 SNP 关联阈值即为 3.25 (图 5), 最终得到 13 个区域, 分别是在 ChrA03、ChrA04、ChrA06、ChrA07、ChrA10 和 ChrC03 上, 总长度为 6.31 Mb, 共包含 1337 个基因, 其中非同义突变 SNP 位点的基因 231 个 (表 3); DW₂ 与 HR 混池间 InDel 关联阈值降低后为 3.24 (图 6), 最终得到 16 个区域, 总长度为 6.29 Mb, 分别是在 ChrA03、ChrA04、ChrA06、ChrA07、ChrA10 和 ChrC03 上, 共包含 1240 个基因, 其中移码突变 InDel 位点的基因 48 个 (表 4)。

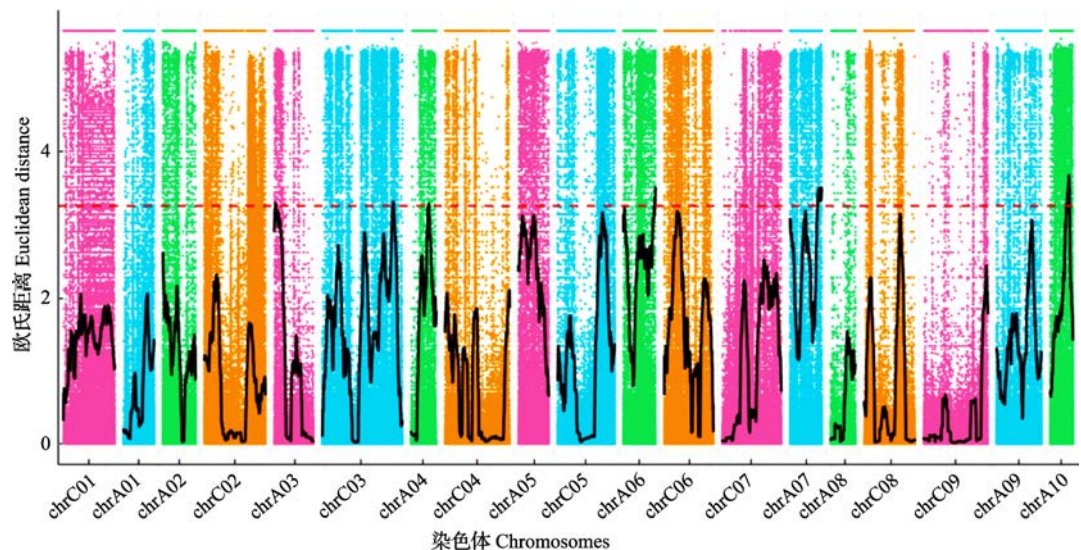
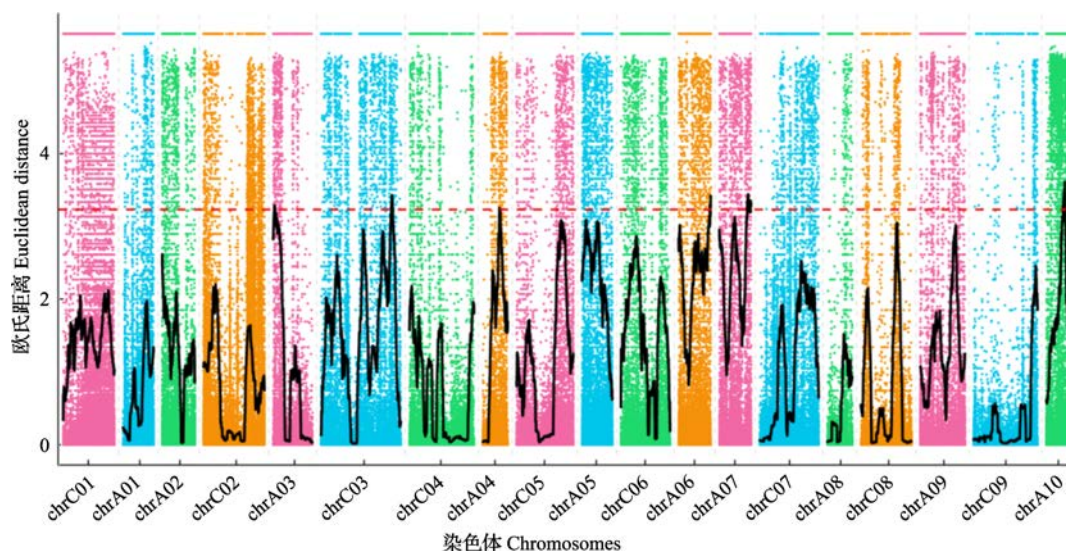


图 5 DW₂ 与 HR 混池间 SNP ED 值在染色体上的分布
 Fig.5 Distribution of ED values of SNPs in DW₂ and HR mixed poles on chromosomes

表 3 DW₂ 与 HR 混池间 SNP 关联区域信息统计Table 3 statistics of SNP associated region information between DW₂ and HR mixed pool

染色体编号 Chromosome ID	开始位置 Start	结束位置 End	大小 (Mb) Size	基因数 Gene number
chrA03	1210000	1330000	0.12	19
chrA03	870000	1150000	0.28	62
chrA04	13460000	13500000	0.04	5
chrA06	23650000	23970000	0.32	60
chrA06	24100000	24390000	0.29	50
chrA07	20960000	24000000	3.04	674
chrA10	12390000	12390000	0	1
chrA10	12440000	12470000	0.03	7
chrA10	12820000	14730000	1.91	419
chrC03	53480000	53490000	0.01	2
chrC03	53520000	53520000	0	1
chrC03	53620000	53810000	0.19	30
chrC03	53920000	53990000	0.07	7
总数 Total	—	—	6.31	1337



彩色的点代表每个 InDel 位点的 ED 值,黑色的线为拟合后的 ED 值,红色的虚线代表显著性关联阈值

The color point represents the ED value of each InDel locus, the black line represents the fitted ED value, and the red dotted line represents the significant correlation threshold

图 6 DW₂ 与 HR 混池间 InDel ED 值在染色体上的分布Fig.6 Distribution of ED association value of InDel in DW₂ and HR mixed pool on chromosomes表 4 DW₂ 与 HR 混池间 InDel 关联区域信息统计Table 4 InDel correlation area information statistics between DW₂ and HR mixed pool

染色体编号 Chromosome ID	开始位置 Start	结束位置 End	大小 (Mb) Size	基因数 Gene number
chrA03	850000	1340000	0.49	96
chrA04	12730000	12770000	0.04	6
chrA04	13460000	13470000	0.01	2
chrA06	23660000	23840000	0.18	32
chrA06	23890000	23940000	0.05	10
chrA06	24210000	24380000	0.17	26
chrA07	21040000	22920000	1.88	405
chrA07	23080000	23110000	0.03	7
chrA07	23290000	23320000	0.03	8
chrA07	23340000	23340000	0	1
chrA07	23360000	23360000	0	1
chrA07	23380000	23380000	0	1
chrA07	23400000	23510000	0.11	24
chrA07	23610000	24000000	0.39	109
chrA10	12870000	14700000	1.83	396
chrC03	53230000	54310000	1.08	116
总数 Total	—	—	6.29	1240

对 DW₂ 与 HR 混池间 SNP 和 InDel 关联区域进行筛选, DW₂ 与 HR 混池间 SNP 和 InDel 关联区域结果取交集, 在 ChrA03、ChrA04、ChrA06、ChrA07、ChrA10 和 ChrC03 上共得到 19 个与性状相关的候选区域, 总长度为 5.35 Mb (表 5)。关联区域内共包含基因 1143 个, 然而由于未达到理论阈值, 这个区域很可能是假阳性区域, 需要进一步验证。

根据 DW₁ 基因混池和 DW₂ 基因混池获得的候选区域结果, DW₁ 基因混池和 DW₂ 基因混池间获得的关联区域重合, 重合区间为 ChrA10 上 12.87~14.70 Mb 处, 长度为 1.83 Mb。因此, 认为控制 DW871 矮秆基因极可能位于 ChrA10 关联区域 12.87~14.70 Mb 处。

表 5 DW₂ 与 HR 基因混池间关联区域信息统计
Table 5 Statistics of associated area information in DW₂ and HR mixed pool

染色体编号 Chromosome ID	开始位置 Start	结束位置 End	大小 (Mb) Size	基因数 Gene number
chrA03	1210000	1330000	0.12	19
chrA03	870000	1150000	0.28	62
chrA04	13460000	13470000	0.01	2
chrA06	23660000	23840000	0.18	32
chrA06	23890000	23940000	0.05	10
chrA06	24210000	24380000	0.17	26
chrA07	21040000	22920000	1.88	405
chrA07	23080000	23110000	0.03	7
chrA07	23290000	23320000	0.03	8
chrA07	23340000	23340000	0	1
chrA07	23360000	23360000	0	1
chrA07	23380000	23380000	0	1
chrA07	23400000	23510000	0.11	24
chrA07	23610000	24000000	0.39	109
chrA10	12870000	14700000	1.83	396
chrC03	53480000	53490000	0.01	2
chrC03	53520000	53520000	0	1
chrC03	53620000	53810000	0.19	30
chrC03	53920000	53990000	0.07	7
总数 Total	—	—	5.35	1143

2.4 候选区域内基因注释

根据 BLAST^[23] 软件对候选区间内编码基因进行注释筛选, DW₁ 基因混池候选区域内共注释到 1405 个基因, 其中在基因混池间存在非同义突变基因共注释到 201 个; DW₂ 基因混池候选区域内共注释到 1142 个基因, 其中在基因混池间存在非同义突变基因共注释到 202 个, 移码突变基因共注释到 46 个(表 6)。

表 6 DW₁ 基因混池和 DW₂ 基因混池候选区域内基因注释结果统计
Table 6 Statistics of gene annotations in candidate areas of DW₁ mixing pool and DW₂ mixing pool

功能注释数据库 Annotated_databases	基因数 Gene number		非同义突变基因数 Non-Syn gene number		移码突变基因数 Frame-shift gene number
	DW ₁	DW ₂	DW ₁	DW ₂	DW ₂
GO	1233	990	166	169	34
KEGG	351	285	39	26	4
COG	519	431	52	46	10
NR	1403	1141	201	202	46
NT	1391	1128	199	199	45
trEMBL	1403	1141	201	202	46
SwissProt	1064	852	147	144	32
总数 Total	1405	1142	201	202	46

3 讨论

本研究将 BSA-seq 测序技术, 成功用于甘蓝型矮秆油菜株高性状候选基因的检测。初步定位了甘蓝型矮秆直立株型油菜株高性状基因区域, 在 DW₁ 与 HR 基因混池间检测到 1 个候选区域, 候选区间位于染色体 ChrA10 上, 候选区间范围为 11.00~17.39 Mb, 长度是 6.39 Mb; 在 DW₂ 与 HR 基因混池间共获得 19 个与性状相关的候选区域, 分别位于 ChrA03、ChrA04、ChrA06、ChrA07、ChrA10 和 ChrC03 上, 总长度为 5.35 Mb。关联区域内共包含基因 1143 个, 然而由于未达到理论阈值, 这个区域很可能是假阳性区域, 需要进一步验证。根据 DW₁ 基因混池和 DW₂ 基因混池获得的候选区域结果, DW₁ 基因混池和 DW₂ 基因混池间获得的关联区域重合, 重合区间为 ChrA10 上 12.87~14.70 Mb, 长度为 1.83 Mb。本研究在 ChrA10 染色体上定位 1 个与 DW871 株高性状显著关联区间, 这一研究结果为精细定位甘蓝型矮秆直立株型油菜株高性状基因奠定了良好基础。

在本研究中, 在 DW₁ 基因混合池与 HR 基因混合池间 ChrA10 上获得 1 个候选区域, 含 1405 个候选基因, 候选区域较大, 候选基因数多, 暂不能确定控制株高基因在染色上的精确位置; 在 DW₂ 基因混合池中, 发现 ChrA03、ChrA04、ChrA06、ChrA10 和 ChrC03 上共有 19 个与性状相关的候选区域, 关联区域较多。2 个基因混池分析结果存在很大差异, 原因可能是 DW₂ 纯合矮秆材料较少; 其次是 DW₁ 基因池中既含有高秆基因也含有矮秆基因, 而纯合矮秆没有高秆基因, 造成 2 种分析结果有差异。在 2 个基因混合池的 ChrA10 上同时定位到候选区

域,且根据本研究结果关联区域为 12.87~14.70 Mb 核苷酸处,说明在 ChrA10 上含有控制甘蓝型矮秆直立株型油菜矮秆基因。另外,DW871 这个矮秆性状是株高性状还是花序性状的连锁值得进一步研究,因为只要有特殊的花序性状就是矮秆材料,故可以理解为一因多效;那么反过来,是矮秆性状导致花序性状?也是一因多效?都值得商榷。

株高受环境条件的影响较大而表型难以准确鉴定。在甘蓝型油菜 19 个连锁群上都报道有 QTL 分布。定位到最多的 QTL 为 ChrA03 (25 个次)、ChrA01 (24 个次)、ChrC06 (24 个次),最少的为 ChrC05 (1 个次)、ChrC04 (2 个次)、ChrC01 (2 个次)^[28-29]。在油菜 ChrA10 上也曾报道有油菜株高 QTL,在 ChrA10 连锁群上的主效 QTL 区域内预测到 2 个株高候选基因 *BnaA10g07740D* 和 *BnaA10g12020D*^[30],但这 2 个株高候选基因不在本研究热点区域内。Basunanda 等^[31]利用 DH 系群体在 Chr10 连锁群上定位了株高 QTL,但其贡献率均较小。张凤启等^[32]在 ChrA10 连锁群上定位株高 *QTLqPH10-1* (贡献率 9.23%)。Liu 等^[9]和 Muangprom 等^[33]克隆了 *Brrgal-d* 和 *BnRGA* 基因。赵波^[15]和 Zhao 等^[16]克隆了控制矮秆性状的矮秆基因 *BnaC07.rga-ds*。本研究在 ChrA03、ChrA04、ChrA06、ChrA10 和 ChrC03 上检测到共 19 个候选区域。目前,油菜株高遗传研究整体上还处于 QTL 定位阶段,只有少数油菜矮秆基因被成功克隆,不过随着芯片技术和高通量测序的发展,以 SNP 和 InDel 为主的定位标记的应用及其基因型分析效率的提高,将大大加快油菜功能基因定位和克隆速度。

参考文献

- [1] 贺亚军,吴道明,傅鹰,钱伟. 利用 DH 和 IF₂ 群体检测甘蓝型油菜株高相关性状 QTL. 作物学报, 2018, 44 (4): 533-541
He Y J, Wu D M, Fu Y, Qian W. Detection of QTLs for plant height related traits in *Brassica napus* L. using DH and IF₂ populations. Acta Agronomica Sinica, 2018, 44 (4): 533-541
- [2] 郑本川,崔成,张锦芳,李浩杰,柴靓,蒋俊,蒋梁材. 甘蓝型油菜育种亲本单株产量与农艺性状相关性分析. 植物遗传资源学报, 2019, 20 (1): 113-121
Zheng B C, Cui C, Zhang J F, Li H J, Chai L, Jiang J, Jiang L C. Correlation analysis between yield and agronomic traits of *Brassica napus* L. breeding parents. Journal of Plant Genetic Resources, 2019, 20 (1): 113-121
- [3] 张书芬,马朝芝,傅廷栋,钱伟. 甘蓝型油菜主要农艺和产量性状的杂种优势及其分离世代分析. 中国油料作物学报, 2007 (2): 121-125
Zhang S F, Ma C Z, Fu T D, Qian W. The heterosis of main agronomic and yield traits of *Brassica napus* L. and analysis of its separation generation. Chinese Journal of Oil Crop Sciences, 2007 (2): 121-125
- [4] 范成明,田建华,胡赞民,王珏,吕慧颖,葛毅强,魏珣,邓向东,张蕾颖,杨维才. 油菜育种行业创新动态与发展趋势. 植物遗传资源学报, 2018, 19 (3): 447-454
Fan C M, Tian J H, Hu Z M, Wang J, Lv H Y, Ge Y Q, Wei X, Deng X D, Zhang L Y, Yang W C. Innovation trends and development trends of rape breeding industry. Journal of Plant Genetic Resources, 2018, 19 (3): 447-454
- [5] 梁翰文,吕慧颖,葛毅强,魏珣,邓向东,杨维才,田志喜. 作物育种关键技术发展态势. 植物遗传资源学报, 2018, 19 (3): 390-398
Liang H W, Lv H Y, Ge Y Q, Wei X, Deng X D, Yang W C, Tian Z X. Development trends of key technologies in crop breeding. Journal of Plant Genetic Resources, 2018, 19 (3): 390-398
- [6] Muangprom A, Osborn T C. Characterization of a dwarf gene in *Brassica rapa*, including the identification of a candidate gene. Theoretical and Applied Genetics, 2004, 108 (7): 1378-1384
- [7] Muangprom A, Mauriera I, Osborn T C. Transfer of a dwarf gene from *Brassica rapa* to oilseed *B. napus*, effects on agronomic traits, and development of a perfect marker for selection. Molecular Breeding, 2006, 17: 101-110
- [8] 唐丁,吕慧颖,王珏,葛毅强,魏珣,杨维才,程祝宽. 作物基因组学研究进展. 植物遗传资源学报, 2018, 19 (3): 383-389
Tang D, Lv H Y, Wang J, Ge Y Q, Wei X, Yang W C, Cheng Z K. Progress in crop genomics research. Journal of Plant Genetic Resources, 2018, 19 (3): 383-389
- [9] Liu C, Wang J, Huang T, Wang F, Yuan F, Cheng X M, Zhang Y, Shi S W, Wu J S, Liu K. A missense mutation in the VHYNP motif of a DELLA protein causes a semi-dwarf mutant phenotype in *Brassica napus*. Theoretical and Applied Genetics, 2010, 121 (2): 249-258
- [10] Zeng X, Zhu L, Chen Y, Qi L, Pu Y, Wen J, Yi B, Shen J, Ma C, Fu T. Identification, fine mapping and characterisation of a dwarf mutant (*bnac.dwf*) in *Brassica napus*. Theoretical and Applied Genetics, 2011, 122 (2): 421-428
- [11] 杨朋娜,张璐,史慧娟,李清,王茂林. 甘蓝型油菜矮秆突变基因 *ndf1* 的分子标记连锁遗传作图. 四川大学学报: 自然科学版, 2015, 52 (4): 889-894
Yang P N, Zhang L, Shi H J, Li Q, Wang M L. Molecular marker linked genetic mapping of dwarf mutant gene *ndf1* in *Brassica napus* L. Journal of Sichuan University: Natural Science Edition, 2015, 52 (4): 889-894
- [12] Sun C, Wang B, Yan L, Hu K, Liu S, Zhou Y, Guan C, Zhang Z, Li J, Zhang J, Chen S, Wen J, Ma C, Tu J, Shen J, Fu T, Yi B. Genome-wide association study provides insight into the genetic control of plant height in Rapeseed (*Brassica napus* L.). Frontiers in Plant Science, 2016, 7 (7): 1102
- [13] Wang X, Wang H, Long Y, Liu L, Zhao Y, Tian J, Zhao W, Li B, Chen L, Chao H, Li M. Dynamic and comparative QTL analysis for plant height in different developmental stages of *Brassica napus* L. Theoretical and Applied Genetics, 2015, 128 (6): 1175-1192
- [14] Wang Y, Chen W, Chu P, Wan S, Yang M, Wang M, Guan

- R. Mapping a major QTL responsible for dwarf architecture in *Brassica napus* using a single-nucleotide polymorphism marker approach. *BMC Plant Biology*, 2016, 16(1): 178
- [15] 赵波. 甘蓝型油菜矮秆基因定位、克隆及功能分析. 武汉: 华中农业大学, 2017: 235-301
- Zhao B. Location, cloning and functional analysis of dwarf gene in *Brassica napus* L.. Wuhan: Huazhong Agricultural University, 2017: 235-301
- [16] Zhao B, Li H, Li J, Wang B, Dai C, Wang J, Liu K D. *Brassica napus* DS-3, encoding a DELLA protein, negatively regulates stem elongation through gibberellin signaling pathway. *Theoretical and Applied Genetics*, 2017, 130(4): 727-741
- [17] 张瑞茂, 李超, 陈大伦, 向阳. 甘蓝型油菜特异矮秆直立株型新品系 DW871 株高的遗传分析. *中国油料作物学报*, 2019, 41(1): 1-9
- Zhang R M, Li C, Chen D L, Xiang Y. Genetic analysis of plant height of *Brassica napus* L. specific dwarf plant type DW871. *Chinese Journal of Oil Crop Sciences*, 2019, 41(1): 1-9
- [18] 张瑞茂, 李超, 陈大伦, 向阳. 甘蓝型油菜矮秆直立株型材料 DW871 的选育. *种子*, 2019, 38(2): 116-120, 123
- Zhang R M, Li C, Chen D L, Xiang Y. Selection of DW871 of *Brassica napus* L. straight plant type material. *Seed*, 2019, 38(2): 116-120, 123
- [19] 孙海波, 王智慧, 刘胜毅, 黄军艳, 董彩华. 一种适用于油菜、大豆、花生、芝麻四种油料作物种子的 RNA 提取方法. *中国油料作物学报*, 2012, 34(4): 353-358
- Sun H B, Wang Z H, Liu S Y, Huang J Y, Dong C H. An RNA extraction method for four oilseed crops of rape, soybean, peanut and sesame seeds. *Chinese Journal of Oil Crop Sciences*, 2012, 34(4): 353-358
- [20] Hill J T, Demarest B L, Bisgrove B W, Gorski B, Su Y C, Yost H J. MMAPP: Mutation mapping analysis pipeline for pooled RNA-seq. *Genome Research*, 2013, 23(4): 687-697
- [21] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo M A. The genome analysis toolkit: a MapReduce framework for analyzing next generation DNA sequencing data. *Genome Research*. 2010, 20(9): 1297-1303
- [22] 刘桢. 基于云计算的基因数据预处理流程优化设计与实现. 广州: 华南理工大学, 2018
- Liu W. Optimization design and implementation of genetic data preprocessing based on cloud computing. Guangzhou: South China University of Technology, 2018
- [23] Altschul S F, Madden T L, Schaffer A A, Zhang J, Zhang Z, Miller W, Lipman D J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 1997, 25(17): 3389-3402
- [24] Ashburner M, Ball C A, Blake J A, Botstein D, Butler H, Cherry J M, Davis A P, Dolinski K, Dwight S S, Eppig J T, Harris M A, Hill D P, Issel-Tarver L, Kasarskis A, Lewis S, Matese J C, Richardson J E, Ringwald M, Rubin G M, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, 2000(25): 25-29
- [25] Tatusov R L, Galperin M Y, Natale D A, Koonin E V. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, 2000, 28(1): 33-36
- [26] Minoru K, Susumu G, Shuichi K, Yasushi O, Masahiro H. The KEGG resource for deciphering the genome. *Nucleic Acids Research*, 2004, 32(D): 277-280
- [27] Cingolani P, Platts A, Wang L L, Coon M, Nguyen T, Wang L. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Orosophila melanogaster* Strain W1118; iso-2; iso-3.. *Fly*, 2012, 6(2): 80-92
- [28] Shi T, Li R, Zhao Z, Ding G, Long Y, Meng J. QTL for yield traits and their association with functional genes in response to phosphorus deficiency in *Brassica napus*. *PLoS One*, 2013, 8(1): e54559
- [29] 唐敏强, 程晓晖, 童超波, 刘越英, 赵传纪, 董彩华, 于景印, 马小良, 黄军艳, 刘胜毅. 甘蓝型油菜株高性状的全基因组关联分析. *作物学报*, 2015, 41(7): 1121-1126
- Tang M Q, Cheng X H, Tong C B, Liu Y Y, Zhao C J, Dong C H, Yu J Y, Ma X G, Huang J Y, Liu S Y. Genome-wide association analysis of plant height traits in *Brassica napus* L.. *Acta Agronomica Sinica*, 2015, 41(7): 1121-1126
- [30] 姜成红, 耿鑫鑫, 魏文辉, 姜慧芳. 甘蓝型油菜株高 QTL 定位及主效 QTL 区间候选基因预测. *河南农业科学*, 2017, 46(8): 27-31
- Jiang C H, Geng X X, Wei W H, Jiang H F. QTL mapping of plant height and candidate gene prediction of major QTL intervals in *Brassica napus* L.. *Journal of Henan Agricultural Sciences*, 2017, 46(8): 27-31
- [31] Basunanda P, Radoev M, Ecke W, Friedt W, Becker H C, Snowden R J. Comparative mapping of quantitative trait loci involved in heterosis for seedling and yield traits in oilseed rape (*Brassica napus* L.). *Theoretical and Applied Genetics*, 2010, 120(2): 271-281
- [32] 张凤启, 刘越英, 程晓晖, 童超波, 董彩华, 唐敏强, 黄军艳, 刘胜毅. 利用高密度 SNP 标记定位甘蓝型油菜株高 QTL. *中国油料作物学报*, 2014, 36(6): 695-700
- Zhang F Q, Liu Y Y, Cheng X H, Tong C B, Dong C H, Tang M Q, Huang J Y, Liu S Y. Study on the plant height QTL of *Brassica napus* L. Using high density SNP markers. *Chinese Journal of Oil Crop Sciences*, 2014, 36(6): 695-700
- [33] Muangprom A, Thomas S G, Sun T P, Osborn T C. A novel dwarfing mutation in a green revolution gene from *Brassica rapa*. *Plant Physiology*, 2005, 137(3): 931-938