

杜灿灿,曾生元,景德道,等. 基于 BSA-seq 的一个苗期黄化转绿突变体基因定位[J]. 江苏农业科学,2024,52(12):53-60.
doi:10.15889/j.issn.1002-1302.2024.12.007

基于 BSA-seq 的一个苗期黄化转绿突变体基因定位

杜灿灿,曾生元,景德道,胡庆峰,李 闯,钱华飞,林添资,余 波,孙立亭,

周义文,杨 军,韩华新,龚红兵

(江苏丘陵地区镇江农业科学研究所,江苏句容 212400)

摘要:叶片是植物进行光合作用的重要场所,叶色突变往往制约水稻具有充足的源,研究叶色突变体对选育高光效品种以及进一步解析叶色调控的分子机制均具有重要意义。以育种中间材料苗期叶片黄化转绿突变体 *gry* 为对象,对其进行主要农艺性状分析,并对其突变性状进行遗传分析、基因定位。结果显示,该突变体 3 叶期前植株呈淡黄色至白色,4 叶期变绿,除成熟期株高显著高于野生型以外,其他主要农艺性状均无显著差异。通过遗传分析确定该突变体的黄化转绿表型受隐性单基因控制。利用 *gry* 突变体与野生型的 F₂ 群体构建混池,通过 BSA-seq 的 Δ SNP-index 和 ED 2 种算法进行基因定位,结果显示 *gry* 基因被定位在 22.23 ~ 26.77 Mb 区间内。在该区间内开发 KASP 标记对 62 个 F₂ 黄化转绿单株进行验证与连锁分析,验证 *gry* 基因在 21.18 ~ 25.30 Mb 区间内。基于定位区间,结合基因注释数据库与水稻基因变异数据库推测候选基因为 *LOC_Os03g40020*,测序表明该基因在第 1 768 位发生单碱基 T 缺失,导致编码的氨基酸大量改变而产生功能变化。

关键词:苗期黄化转绿突变体;BSA-seq; Δ SNP-index 算法;ED 算法

中图分类号:S511.01 **文献标志码:**A **文章编号:**1002-1302(2024)12-0053-07

水稻是全球主要的粮食来源。水稻产量形成主要受“源、库、流”及其相互之间的协调关系影响^[1]。水稻叶片是主要的源性状,它是植物进行光合作用的重要场所,水稻产量的 90% 以上来自叶片的光合作用,叶色突变往往制约水稻具有充足的源,并且叶色突变体是研究叶绿体发育、叶绿素生物合成、信号转导、光合作用机制的理想材料,叶色突变的分子机制又是复杂多样的^[2]。因此,水稻叶色突变体研究不仅可以为选育强源的高光效品种提供理论依据,还可以进一步解析叶色调控的分子机制。

有关水稻叶色突变体的研究是广泛而持续的,水稻叶色突变体包括黄化、斑马、斑点、类病斑、滞绿、白化、紫色、白化转绿等不同类型的表型^[3-8]。目前已经克隆了 100 多个相关基因,如参与叶绿体发育的 *SPP*(编码基质加工肽酶,表型为白化)、*ALI*

(编码质体核糖体 L12 蛋白,表型为白化致死)、*ClpP*(编码 Clp 蛋白酶水解亚基,表型为黄化转绿)、*OsALBL*(编码三角状五肽蛋白,表型为白化),参与光捕获的 *OsLIL3*(编码捕光色素蛋白,表型为黄绿叶)和 *DYE1*(编码光系统 I 中捕光复合物 I 的一个亚基 Lhca4,表型为持绿);参与叶绿素生物合成的 *OsCAO1*(编码叶绿素 a 加氧酶,表型为淡绿叶)、*OsCHLD*(编码镁螯合酶的 CHLD 亚基,表型为黄化叶)和 *YGL1*(编码叶绿素合成酶,表型为黄化转绿);参与叶绿素降解的 *OsSGR*(编码镁离子去螯合酶,表型为持绿)、*YPDI*(编码 LRR-Like1 蛋白,表型为黄化)和 *NYCI*(编码叶绿素 b 还原酶,表型为滞绿)^[9-18]。其中 *PPR*(*pentatricopeptide repeat*, *PPR*)基因表达受阻易形成白化表型,该基因编码的五肽蛋白在真核生物中无处不在,在原核生物中也有报道^[19-20]。大多数植物三角状五肽蛋白被定位在线粒体、质体(叶绿体)上,也有研究将 *PPR* 定位于细胞质、细胞核。在水稻中,大约每 1.6 Mb 就存在一个 *PPR* 基因,目前水稻中有 400 多个非冗余的 *PPR* 蛋白,并且不同位置的 *PPR* 蛋白具有不同的功能。例如位于线粒体的 *PPR*(*FLO18*)可以调控线粒体功能和水稻胚乳发育,位于细胞质中的 *PPR*(*OsCPPRI*)有助于花粉发育,位于叶绿体的 *PPR*(*YSA*)缺失对不

收稿日期:2023-06-30

基金项目:句容市农业科技支撑计划(编号:ZA32221);镇江市农业科学院青年基金(编号:QNJJ2018002)。

作者简介:杜灿灿(1989—),女,江苏徐州人,硕士,助理研究员,主要从事水稻遗传育种及推广应用研究。E-mail:20162804@jaas.ac.cn。

通信作者:龚红兵,硕士,研究员,主要从事水稻新品种选育及推广应用研究。E-mail:hongbinggong973@sina.com。

同温度和光照条件下雄性不育性不产生影响,并对杂种 F_1 的产量、农艺性状没有负面影响^[21-24]。

随着高通量测序技术的发展,BSA-seq 已经逐渐成熟,并且准确性有了明显提高、成本也大幅降低^[25-26]。将 BSA-seq 与传统定位方法相结合可以快速、有效地获得/缩小定位区间^[27-28]。本研究利用该方法,将苗期黄化转绿基因快速定位在 22.23 ~ 26.77 Mb,并推测候选基因为 *LOC_Os03g40020*,该基因编码 1 个与叶绿体发育相关的三角状五肽重复蛋白,以期为选育高光效品种及解析叶色调控的分子机制提供参考意义。

1 材料与方法

1.1 试验材料

gry 是镇稻 18 号中间材料的一个自然突变,经过多代种植表明黄化转绿表型可以稳定遗传。*gry* 与 9311 构建的 F_2 群体作为分析群体,*gry* 与野生型构建的 F_2 群体作为定位群体。试验材料均种植于江苏丘陵地区镇江农业科学研究所行香创新基地,5 月 10 日播种,6 月 15 日移栽,单本种植,株行距 13.3 cm × 25.0 cm。在拔节孕穗期取叶片提取 DNA 用于 BSA-seq。在成熟期,分别选取 *gry* 与野生型各 20 株进行农艺性状测定,包括株高、有效穗数、穗长、结实率、千粒重。

1.2 试验方法

1.2.1 BSA-seq 混池构建及测序 播种后 10 d,用挂牌区分 *gry* 与野生型构建的 F_2 群体中黄化与正常表型个体,移栽时分开种植。根据 Takagi 等的方法,在拔节孕穗期取 *gry* 与野生型亲本各 20 株, F_2 群体突变与正常个体各 40 株,提取 DNA 并检测其浓度,分别等量混合构建 P1 池(*gry*)、P2 池(镇稻 18 号中间材料)、R 池(黄化转绿型 F_2)、D 池(正常

表型 F_2)^[29]。然后对应构建文库,并采用 Illumina NovaSeq 6000 测序平台进行全基因组重测序,测序深度 30 ×。

1.2.2 测序数据分析 首先利用 FastQC 0.11.5 对测序结果(Raw reads)进行质量控制得到 Cleaned reads。再利用 BWA 0.7.9 将有效数据(Cleaned reads)比对到参考基因组(http://plants.ensembl.org/Oryza_sativa/Info/Index)上^[30]。然后利用 SAMtools 1.0 比对对结果进行排序,利用 Picard 1.120 标记重复 reads^[31]。利用 Picard V4.0 局部重新比对与碱基质量校正。最后利用重复标记后的比对结果进行覆盖度、深度等统计。利用 SAMtools 1.0 检测、过滤 SNP 及 InDel 信息;利用 VEP 4.3 注释单核苷核多态性(SNP)及 InDel 结果。基于以上获得的信息,利用 MutMap pipeline 进行数据分析,计算 SNP 在 R 池和 D 混池出现的频率,对 Δ SNP-index (2 个子代池的 SNP-index 的差值)进行滑窗作图后,出现 1 个峰,该处就是连锁区域,即可得到和突变表型连锁的染色体区段和可能的突变位点^[32]。同时,本研究还采用 ED(euclidean distance)算法,通过计算不同混池间各突变型的频率距离,采用距离差异来反映标记与目标区域的连锁强度^[33]。

1.2.3 定位区间分析验证 基于 BSA-seq 区间,结合基因注释与 RiceVarMap2 (ncpgr.cn)公布的 4 726 个水稻品种的变异位点进行分析,推测候选基因。然后选取候选基因内的变异位点与高峰两侧的变异位点各 1 个,设计引物开发 KASP 标记,选取 62 个 F_2 单株进行验证。引物利用 SNPWay (<http://www.snpway.com/>)进行设计,引物信息见表 1,KASP 检测委托武汉市景肽生物科技有限公司进行。

表 1 用于区间验证的引物序列

位置	引物名称	引物序列(5'→3')
Chr03-21187102	3-2118Re	GAAGGTGACCAAGTTCATGCTTTTTCGATCTACTTAGTGAAACAATTC
	3-2118Ra	GAAGGTCGGAGTCAACGGATTATTTTCGATCTACTTAGTGAAACAATTA
	3-2118F	GTGCCTACAAATCAGGCCTC
Chr03-22236828	3-2223Ra	GAAGGTGACCAAGTTCATGCTACAGGTTGCTTTGAAGGTACTCA
	3-2223Rg	GAAGGTCGGACTCAACGGATTACAGTTGCTTTGAAGGTACTCG
	3-2223F	TCCGTCTGAAGAGAGACTGGAG
Chr03-25304267	3-2530Rt	GAAGGTGACCAAGTTCATGCTGCCTGCTTGCTTTACCCTACTT
	3-2530Re	GAAGGTCGGAGTCAACGGATTGCCTGCTTGCTCTACCCTACTC
	3-2530F	GCTCCGACGATGATGATGCT

1.3 表型数据分析

利用 IBM SPSS Statistics 21 对突变体与野生型、籼稻 93 - 11 相互杂交的 F₂ 群体进行 χ^2 检验,对突变体与野生型的株高、有效穗数、千粒重、结实率等表型数据进行 *t* 检验。

2 结果与分析

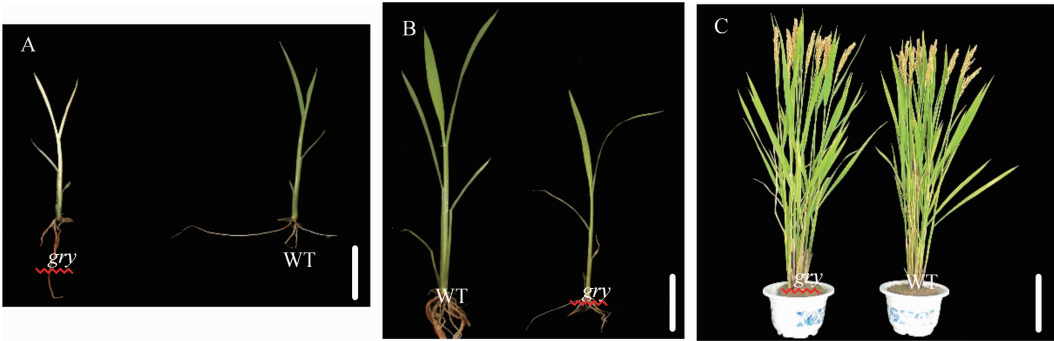
2.1 表型分析

突变体叶子在三叶期前植株呈淡黄色至白色(图 1 - A),四叶期变绿(图 1 - B),成熟期与野生型一样呈绿色(图 1 - C)。成熟期对突变体与野生

型进行农艺性状调查,结果显示突变体株高显著高于野生型,穗长、每穗总粒数、结实率、千粒重、单株有效穗数无显著差异(表 2)。

2.2 gry 突变性状的遗传分析

gry 突变体分别与绿色野生型(粳稻)、93 - 11(籼稻)杂交,F₁ 代均表现为绿色,F₂ 代均出现黄化转绿表型与绿色表型分离现象。经卡方(χ^2)检验,黄化转绿植株与绿色植株数量符合 3 : 1 的分离比(表 3)。遗传分析结果表明,*gry* 突变体的黄化转绿表型是隐性的,受单基因控制。



A: 播种后 20 d 突变体与野生型表型, 比例尺 2 cm; B: 播种后 28 d 突变体与野生型表型, 比例尺 2 cm; C: 播种后 141 d 突变体与野生型表型, 比例尺 15 cm

图1 突变体(*gry*)与野生型(WT)表型

表 2 突变体和野生型农艺性状比较

类型	穗长 (cm)	每穗总粒数 (粒)	结实率 (%)	千粒重 (g)	株高 (cm)	单株有效穗数
<i>gry</i>	16.88 ± 2.65	146.10 ± 9.84	84.35 ± 5.89	23.26 ± 0.17	95.60 ± 2.41 *	9.20 ± 1.69
WT	16.39 ± 2.78	140.68 ± 11.14	84.30 ± 6.70	23.08 ± 0.26	93.10 ± 2.69	10.80 ± 2.04

注: * 表示在 0.05 水平上差异显著。

表 3 4 个不同 F₂ 群体中野生型与突变型分离比

杂交	野生型	突变体	总计	χ^2 (3 : 1)	<i>P</i> 值
突变体 × 野生型	684	210	894	1.087	0.297
野生型 × 突变体	1 763	564	2 327	0.722	0.395
突变体 × 9311	1 158	344	1 502	3.523	0.061
9311 × 突变体	926	297	1 223	0.334	0.563

2.3 测序质量分析

对来自 *gry* × 野生型的 F₂ 群体极端个体及双亲分别构建的 P1 池(*gry*)、P2 池(镇稻 18 号中间材料)、R 池(黄化转绿型 F₂)、D 池(正常表型 F₂) 进行全基因组重测序,共获得 23.05 Gbp 数据量,过滤后数据 22.85 Gbp,并且过滤后数据质量主要分布在 Q30 ≥ 88% 以上,GC 含量在 40% ~ 41%。另外,P1、P2、R 池、D 池中分别有 98.00%、99.00%、

98.73%、98.63% 的 Clean Reads 数比对到参考基因组(表 4),这说明测序数量与质量均合格,可用于后续分析。

表 4 测序数据质量情况统计

样本	原始数据	原始数 (Gbp)	质控后 数据	质控后 数据 (Gbp)	Q20 (%)	Q30 (%)	GC (%)
P1	38 553 898	5.78	38 290 020	5.74	95.79	89.14	41.00
P2	38 022 660	5.70	37 762 171	5.66	95.84	89.24	40.00
D 池	38 208 454	5.73	37 909 046	5.68	95.60	88.85	41.00
R 池	38 880 802	5.83	38 620 163	5.78	95.89	89.36	40.50

注: Q20、Q30 表示 Phred 数值大于 20、30 的碱基占总体碱基的百分比。

2.4 SNP 和 InDel 多态性分析

经过检验、过滤,如表 5 所示,共鉴定出 4 266 个 SNPs、435 个 InDels 在双亲间有多态性。其中,

SNP 和 InDel 标记在 12 条染色体上的平均密度分别为 11.43 SNP/Mb 和 1.17 InDel/Mb, 每条染色体都有充分多、高质量的标记可以用于 BSA - seq 分析。

表 5 鉴定出的 SNPs 和 InDels 在染色体上的分布

染色体	长度 (bp)	SNP		InDel	
		数量 (个)	密度 (个/Mb)	数量 (个)	密度 (个/Mb)
Chr1	43 270 923	400	9.24	49	1.13
Chr2	35 937 250	107	2.98	17	0.47
Chr3	36 413 819	833	22.88	88	2.42
Chr4	35 502 694	133	3.75	14	0.39
Chr5	29 958 434	94	3.14	9	0.3
Chr6	31 248 787	117	3.74	11	0.35
Chr7	29 697 621	99	3.33	11	0.37
Chr8	28 443 022	167	5.87	12	0.42
Chr9	23 012 720	110	4.78	8	0.35
Chr10	23 207 287	107	4.61	8	0.34
Chr11	29 021 106	1 966	67.74	203	6.99
Chr12	27 531 856	133	4.83	5	0.18
Total	373 245 519	4 266	11.43	435	1.17

2.5 基因定位及候选基因预测

根据 ΔSNP - index 方法分析, 以 95% 置信区间作为阈值, 在第 3 号染色体上鉴定出 1 个候选区间; 根据 ED 方法分析, 在第 3 号、第 9 号染色体上共鉴定到 2 个区间。由于黄化转绿表型由单基因控制, 结合 2 种算法, 本研究将 2 种方法重叠的 1 个区间 (22.23 ~ 26.77 Mb) 认为是最终的候选区间 (图 2)。

根据 BSA - seq 测序结果分析, 候选区间内共有 684 个 SNP 变异位点和 156 个 InDel 变异, 这些变异位于 58 个基因。结合 MSU - RGAP 和 RAP - DB 2 个数据库的基因注释筛选出 LOC_Os03g40020、LOC_Os03g44850、LOC_Os03g44880、LOC_Os03g44890、LOC_Os03g44900 5 个基因。然后通过 RiceVarMap2 (ncpgr.cn) 公布的 4 726 个水稻品种的变异位点与 5 个初筛基因中的变异位点的对比分析, 排除了 4 个基因中的变异位点, 推测候选基因为 LOC_Os03g40020^[33]。该基因编码 1 个与叶绿体发育相关的三角状五肽重复蛋白, 该基因仅由 1 个外显子组成, 通过突变体与野生型的测序比对表明: 该基因在第 1 768 位发生单碱基 T 缺失, 突变位点位于外显子上, 属于移码突变, 导致该基因编码的 742 个氨基酸中 153 个发生改变 (图 3), 从而影响了其功能。因此将该基因确定为本研究的候选基因。

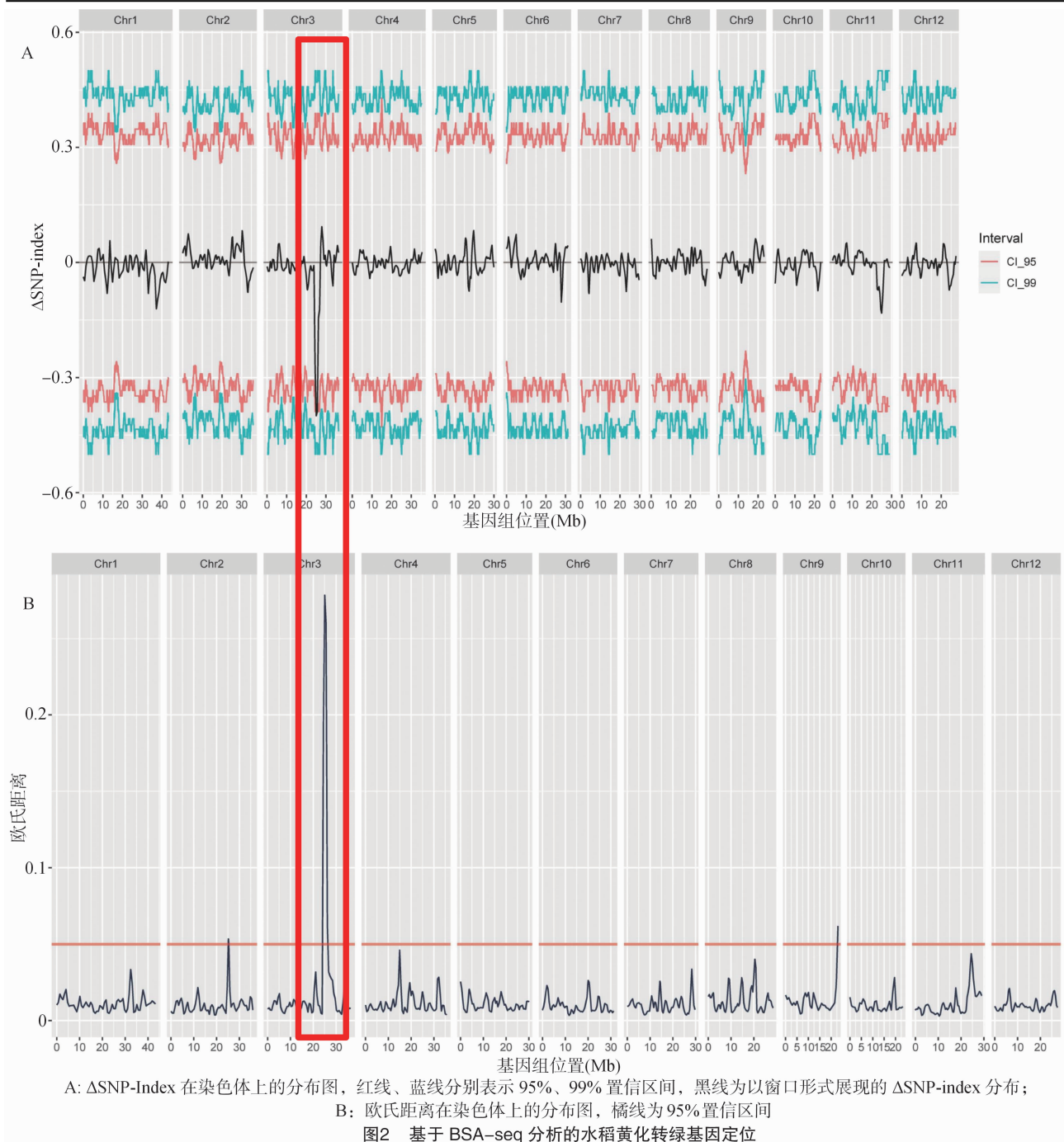
2.6 目标区间及变异位点验证

本研究选取候选基因 LOC_Os03g40020 内的变异位点与高峰两侧的变异位点各 1 个, 设计 KASP 标记 3 - 2223、3 - 2118、3 - 2530, 基因分型结果显示, 3 个标记可以用于区分不同基因型。然后利用 F₂ 群体中的突变表型单株进行验证, 结果发现标记 3 - 2118 处有 1 个交换株, 标记 3 - 2223 处有 0 个交换株, 标记 3 - 2530 处有 27 个交换株, 并且左右两边交换株不同, 这表明 gry 基因位于 3 - 2118 (21.18 Mb) 和 3 - 2530 (25.30 Mb) 之间, 且 3 - 2118 为共分离标记 (图 4)。结合标记 3 - 2118 所在变异位点为基因 LOC_Os03g40020 内引起的功能变异, 本研究认为, 黄化转绿表型是由 LOC_Os03g40020 基因突变引起的。

3 讨论

水稻叶片是水稻最主要的源器官, 提高源器官的光合效率能有效提高水稻单产。水稻中已经发现许多叶色相关的基因, 但是多数叶色相关基因会伴有叶片早衰、生育期延迟、产量降低等不良性状^[8,34]。因此, 挖掘不影响水稻综合农艺性状的叶色相关基因具有重要意义。本研究的突变体除幼苗期表现明显白化外, 后期产量性状与野生型相比无显著差异, 具有潜在利用价值。因此, 本研究的目的是研究苗期黄化转绿突变体的遗传基础, 并通过 BSA - seq 分析确定候选基因, 助力解析高光效品种强源机制及叶色调控的分子机制, 同时为设计育种选育强源、高光效的高单产品种奠定基础。

与传统基因定位相比, BSA - seq 更高效, 但是分析过程中使用不同的计算方式、置信区间阈值等会使定位区间产生波动, 所以还需要结合传统定位方式确定候选基因。例如, qFLT9 通过 BSA - seq 方法被定位在 9 号染色体 19.10 ~ 20.03 Mb 区域, 并通过连锁图谱分析将候选区域缩小到 0.928 Mb 的区间^[35]。郭微利用 BSA - seq 中的 ED 算法将耐碱 QTL 定位在 5 条染色体上, InDel - index 算法将该 QTL 定位在第 2 号染色体, 综合这 2 种方法与传统 QTL 定位结果进一步确定了候选区域^[36]。本研究中, 利用 ΔSNP - index 方法分析, 以 95% 置信区间作为阈值, 在第 3 号染色体上鉴定出 1 个候选区间; 利用 ED 方法分析, 以 95% 置信区间作为阈值, 在第 3 号染色体、第 9 号染色体上鉴定到 2 个区间。由于黄化转绿表型由单基因控制, 本研究将 2 种方法



重叠的一个区间认为是最终的候选区间。这也说明,若利用 BSA-seq 进行 QTL 鉴定需要通过传统方式验证,以排除算法等分析造成的影响。

另外,随着测序技术的发展,大量数据库被发布,一些包含大量水稻基因组序列或者表型序列的数据库为水稻相关研究提供了便利。本研究利用 RiceVarMap2 (ncpgr.cn) 数据库,通过查询数据库中碱基在指定位置的出现频率排除变异位点。例如 *LOC_Os03g44850* 基因在 25 304 267 bp 处有 A

(突变型)、G(野生型)2 种类型,通过数据查询,4 726 个水稻品种中 56.10% 为 A 型,43.30% 为 G 型,说明这个位点是一个常见变异位点。而本研究的黄化转绿自然突变表型是一个非常规表型,因此可以将该位点排除。该数据库的利用促进了本研究的进行,为普通水稻功能基因组研究者和水稻育种者提供了便捷有效的途径。

本研究利用 BSA-seq 定位了 1 个隐性苗期黄化转绿突变体基因 *gry*,并利用 KASP 标记进行了验

野生型	MPRVCAAPRAPPPPCPCHVGVGLRPRWRASRHGPLRAAGQEQLLTALREQPDPAALRMLNAALARDDF	70
<i>ysa</i>	MPRVCAAPRAPPPPCPCHVGVGLRPRWRASRHGPLRAAGQEQLLTALREQPDPAALRMLNAALARDDF	70
<i>gry</i>	MPRVCAAPRAPPPPCPCHVGVGLRPRWRASRHGPLRAAGQEQLLTALREQPDPAALRMLNAALARDDF	70
野生型	APGPEVYEEIIRKLGAVGALDMKVLVAEMRREGHQVKLGVVHSFLDSYEGQQLFDDAVDLILNQLQPLF	140
<i>ysa</i>	APGPEVYEEIIRKLGAVGALDMKVLVAEMRREGHQVKLGVVHSFLDSYEGQQLFDDAVDLILNQLQPLF	140
<i>gry</i>	APGPEVYEEIIRKLGAVGALDMKVLVAEMRREGHQVKLGVVHSFLDSYEGQQLFDDAVDLILNQLQPLF	140
野生型	GIQADTVVYNHLLNVLVEGSKMKLLESVYSEMARGIKPDVVTNTLMKALCRAHQVRTAVLMLEEMSSR	210
<i>ysa</i>	GIQADTVVYNHLLNVLVEGSKMKLLESVYSEMARGIKPDVVTNTLMKALCRAHQVRTAVLMLEEMSSR	210
<i>gry</i>	GIQADTVVYNHLLNVLVEGSKMKLLESVYSEMARGIKPDVVTNTLMKALCRAHQVRTAVLMLEEMSSR	210
野生型	GVAPDETTFTTLMQGFVEEGSIEAALRVKARMLEMGCSATKVTNVNLINGCYCKLGRVEDALGYIQQEIAD	280
<i>ysa</i>	GVAPDETTFTTLMQGFVEEGSIEAALRVKARMLEMGCSATKVTNVNLINGCYCKLGRVEDALGYIQQEIAD	280
<i>gry</i>	GVAPDETTFTTLMQGFVEEGSIEAALRVKARMLEMGCSATKVTNVNLINGCYCKLGRVEDALGYIQQEIAD	280
野生型	GFEPDQITYNTFVNGLCQNDHVGHALKVMDVMVQEGHDPDVFTYNIIVNCLCKNGQLEAAGILNQMVDR	350
<i>ysa</i>	GFEPDQITYNTFVNGLCQNDHVGHALKVMDVMVQEGHDPDVFTYNIIVNCLCKNGQLEAAGILNQMVDR	350
<i>gry</i>	GFEPDQITYNTFVNGLCQNDHVGHALKVMDVMVQEGHDPDVFTYNIIVNCLCKNGQLEAAGILNQMVDR	350
野生型	GCLPDITTFNTLIAALCTGNRLEEALDLARQVTVKGVSPDVYTFNIIINALCKVGDPHALARLFEEMKNS	420
<i>ysa</i>	GCLPDITTFNTLIAALCTGNRLEEALDLARQVTVKGVSPDVYTFNIIINALCKVGDPHALARLFEEMKNS	420
<i>gry</i>	GCLPDITTFNTLIAALCTGNRLEEALDLARQVTVKGVSPDVYTFNIIINALCKVGDPHALARLFEEMKNS	420
野生型	GCTPDEVTYNTLIDNLCSLGKLGKALDLLKDMESTGCPRSTIITYNTIIDGLCKKMRIEAEVEFDQMDLQ	490
<i>ysa</i>	GCTPDEVTYNTLIDNLCSLGKLGKALDLLKDIHWLSSKYNIIHYNRVMQENENRRSRFSNNGSARHFEB	490
<i>gry</i>	GCTPDEVTYNTLIDNLCSLGKLGKALDLLKDMESTGCPRSTIITYNTIIDGLCKKMRIEAEVEFDQMDLQ	490
野生型	GISRNAITFNTLIDGLCKDKKIDDAFELINQMISEGLQPNNTYNSILTHYCKQGDIKKAADILETMTAN	560
<i>ysa</i>	CNHIQYSHRWFGQGKDCFAYSNDKRVATQYHQQFYSNLLQARHKKCCGYFRNYDCKWISGCCYVRYSD	560
<i>gry</i>	GISRNAITFNTLIDGLCKDKKIDDAFELINQMISEGLQPNNTYNSILTHYCKQGDIKKAADILETMTAN	560
野生型	GFEVDVVTYGTGLINGLCKAGRTQVALKVLRGMRIKMRPTPKAYNPVLQSLFRNRNIRDAISLREMAEV	630
<i>ysa</i>	RSMQGWDTGCFEQTQRYADKRDEAYSLSLSCAPVSLQTEYQRCPESEFGDGRGWASCFDIDCFESWALSW	630
<i>gry</i>	GFEVDVVTYGTGLINGLCKAGRTQVALKVLEVCGKGGLLQKLTILCSSLSSDGTISMPVFSGRVQRLVSL	630
野生型	GEPPDALTYKIVFRGLCRGGPIKEADDEMLEMVDKGFIPFSSFRMLAEGLLNLGMDIYFIRAEIIME	700
<i>ysa</i>	WRMYRSFFHVGDDGVHTRVLILPYASRSIKPGYGLLHSHNHGKGRPQRVCFCKNGKISQDPQILCISNLW	700
<i>gry</i>	LMHHRILFFVGSVVVEGLLKKLISQWRMLIRGSYQSSHPVCLKVYTWWMITSLEPLKSWKRSTESL	700
野生型	KVDIRSDVSAIRGYLKIRKFYDALAIFGRFLINNPQWSYR	742
<i>ysa</i>	PFPGDAQPSMEIPMX-----	715
<i>gry</i>	MFLQGDISRSANFMMHQPLAVSWRSTILNGVTIX-----	734

完全、半保守的氨基酸分别用黑色、灰色阴影表示

图3 野生型 LOC_Os03g40020、*gry*、*ysa* 基因的氨基酸序列

样本	3-2118	3-2223	3-2530	表型	样本	3-2118	3-2223	3-2530	表型
G310	2	2	2	绿	A01-03	1	1	3	黄化转绿
G309	1	1	1	黄化转绿	A01-11	1	1	3	黄化转绿
A01-17	1	1	3	黄化转绿	A01-19	1	1	3	黄化转绿
A01-25	1	1	3	黄化转绿	A01-27	1	1	3	黄化转绿
A01-33	1	1	3	黄化转绿	A01-35	1	1	3	黄化转绿
A01-41	1	1	3	黄化转绿	A01-43	1	1	3	黄化转绿
A01-49	1	1	3	黄化转绿	A01-59	1	1	3	黄化转绿
A01-57	1	1	3	黄化转绿	A01-04	1	1	3	黄化转绿
A01-02	1	1	3	黄化转绿	A01-12	1	1	3	黄化转绿
A01-10	1	1	3	黄化转绿	A01-20	1	1	3	黄化转绿
A01-18	3	1	1	黄化转绿	A01-28	1	1	3	黄化转绿
A01-26	1	1	3	黄化转绿	A01-36	1	1	3	黄化转绿
A01-34	1	1	3	黄化转绿	A01-44	1	1	3	黄化转绿
A01-42	1	1	3	黄化转绿	A01-50	1	1	2	黄化转绿
A01-58	1	1	3	黄化转绿	A01-51	1	1	2	黄化转绿

1—*grygry* 基因型；2—*GRYGRY* 基因型；3—*GRYgry* 基因型

图4 *gry* 基因精细定位

证,将区间确定在 21.18 ~ 25.30 Mb 区间内。通过基因注释数据库与水稻基因变异数据库推测候选基因为 *LOC_Os03g40020*,通过查阅文献发现 *gry* 基因与前人克隆的 *ysa* 基因^[24]等位。但是通过对 DNA 序列和氨基酸序列进行比对发现,*gry* 基因为 1 个碱基缺失,*ysa* 基因为 5 bp 缺失,并且缺失位置不同,*ysa* 基因突变导致更多氨基酸的改变(图 3)。因此,本研究定位的 *gry* 基因是一个具有新的变异位点的 *LOC_Os03g40020* 等位基因。

参考文献:

- [1] 邵奎添. 水稻库源基因 *NALI* 的生理功能研究[D]. 沈阳:沈阳农业大学,2022.
- [2] 贺治洲,尹明,谢振宇,等. 水稻转绿型叶色突变体研究进展[J]. 热带农业科学,2014,34(8):30-36,42.
- [3] 杨颜格,黄纤纤,赵亚男,等. 水稻叶色基因克隆与分子机制研究进展[J]. 植物遗传资源学报,2020,21(4):794-803.
- [4] 王丰. 水稻类病斑突变体 *lrd25* 的基因定位及特征特性分析[D]. 沈阳:沈阳农业大学,2022.
- [5] 廉院训,韦子芸,张强,等. 水稻斑马叶突变体 *z17* 的鉴定与基因的精细定位[J]. 中国水稻科学,2023,37(2):113-124.
- [6] 刘亚萍. 水稻深绿穗基因 *DGPI* 的克隆与功能研究[D]. 金华:浙江师范大学,2017.
- [7] 李洋洋. 微型紫叶观赏水稻新种质“紫薇”的创制[D]. 重庆:西南大学,2020.
- [8] 胡婷婷,何弯弯,王友霜,等. 水稻白化转绿突变体 *al14* 的表型分析及基因定位[J]. 江苏农业学报,2021,37(6):1361-1369.
- [9] Yue R Q, Wang X F, Chen J Y, et al. A rice stromal processing peptidase regulates chloroplast and root development[J]. Plant and Cell Physiology,2010,51(3):475-485.
- [10] 赵冬生. 水稻幼苗白化致死基因 *ALI* 和粒形基因 *GS9* 的克隆与功能分析[D]. 扬州:扬州大学,2016.
- [11] Li W, Wu C, Hu G C, et al. Characterization and fine mapping of a novel rice narrow leaf mutant *nal9*[J]. Journal of Integrative Plant Biology,2013,55(11):1016-1025.
- [12] 李家飞. 水稻白化突变体 *albl* 和 *alblhc2* 的光合特性和基因克隆[D]. 上海:上海师范大学,2011.
- [13] 李春梅. 两个水稻黄绿叶突变基因的图位克隆及功能分析[D]. 雅安:四川农业大学,2019.
- [14] Yamatani H, Kohzuma K, Nakano M, et al. Impairment of Lhca4, a subunit of LHCI, causes high accumulation of chlorophyll and the stay-green phenotype in rice[J]. Journal of Experimental Botany, 2018,69(5):1027-1035.
- [15] Yang Y L, Xu J, Huang L C, et al. PGL, encoding chlorophyllide a oxygenase 1, impacts leaf senescence and indirectly affects grain yield and quality in rice[J]. Journal of Experimental Botany,2016, 67(5):1297-1310.
- [16] Deng X J, Zhang H Q, Wang Y, et al. Mapped clone and functional analysis of leaf-color gene *Ygl7* in a rice hybrid (*Oryza sativa* L. ssp. *indica*) [J]. PLoS One,2014,9(6):e99564.
- [17] Wu Z M, Zhang X, He B, et al. A chlorophyll-deficient rice mutant with impaired chlorophyllide esterification in chlorophyll biosynthesis[J]. Plant Physiology,2007,145(1):29-40.
- [18] Shin D, Lee S, Kim T H, et al. Natural variations at the Stay-Green gene promoter control lifespan and yield in rice cultivars[J]. Nature Communications,2020,11:2819.
- [19] Pusnik M, Small I, Read L K, et al. Pentatricopeptide repeat proteins in *Trypanosoma brucei* function in mitochondrial ribosomes[J]. Molecular and Cellular Biology,2007,27(19):6876-6888.
- [20] 李景芳,王宝祥,刘艳,等. PPR 蛋白在水稻生长发育中的功能研究进展[J]. 植物遗传资源学报,2022,23(2):358-367.
- [21] Li S B, Sun Q P, Hu M H, et al. Phylogenetic genome-wide comparisons of the pentatricopeptide repeat gene family in indica and japonica rice[J]. Biochemical Genetics,2012,50(11):978-989.
- [22] Yu M Z, Wu M M, Ren Y L, et al. Rice FLOURY ENDOSPERM 18 encodes a pentatricopeptide repeat protein required for 5' processing of mitochondrial nad5 messenger RNA and endosperm development[J]. Journal of Integrative Plant Biology,2021,63(5):834-847.
- [23] Zheng S Y, Dong J F, Lu J Q, et al. A cytosolic pentatricopeptide repeat protein is essential for tapetal plastid development by regulating *OsGLK1* transcript levels in rice[J]. New Phytologist, 2022,234(5):1678-1695.
- [24] Su N, Hu M L, Wu D X, et al. Disruption of a rice pentatricopeptide repeat protein causes a seedling-specific albino phenotype and its utilization to enhance seed purity in hybrid rice production[J]. Plant Physiology,2012,159(1):227-238.
- [25] 李辉平,骆昕,侯子强,等. 基于 BSA-seq 技术挖掘糙皮侧耳抗螨候选基因[J]. 江苏农业学报,2022,38(6):1648-1656.
- [26] Liang T M, Chi W C, Huang L K, et al. Bulk segregant analysis coupled with whole-genome sequencing (BSA-seq) mapping identifies a novel pi21 haplotype conferring basal resistance to rice blast disease[J]. International Journal of Molecular Sciences, 2020,21(6):2162.
- [27] Xin W, Liu H L, Yang L M, et al. BSA-seq and fine linkage mapping for the identification of a novel locus (qPH9) for mature plant height in rice (*Oryza sativa*) [J]. Rice,2022,15(1):26.
- [28] Kaur G, Yadav I S, Bhatia D, et al. BSA-seq identifies a major locus on chromosome 6 for root-knot nematode (*Meloidogyne graminicola*) resistance from *Oryza glaberrima* [J]. Frontiers in Genetics,2022,13:871833.
- [29] Takagi H, Abe A, Yoshida K, et al. QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations[J]. The Plant Journal,2013,74(1):174-183.
- [30] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform[J]. Bioinformatics,2009,25(14):1754-1760.
- [31] Abe A, Kosugi S, Yoshida K, et al. Genome sequencing reveals agronomically important loci in rice using MutMap[J]. Nature Biotechnology,2012,30(2):174-178.

王乙淋,李宏宇,韦共宇,等. 基于转录组和代谢组分析八角黄酮类化合物合成途径关键基因[J]. 江苏农业科学,2024,52(12):60-68.
doi:10.15889/j.issn.1002-1302.2024.12.008

基于转录组和代谢组分析八角黄酮类化合物合成途径关键基因

王乙淋¹, 李宏宇¹, 韦共宇¹, 蒙泳成¹, 黄 勇^{1,2,3}

(1. 广西中医药大学,广西南宁 530200; 2. 广西壮瑶药重点实验室,广西南宁 530200;

3. 广西高校中药民族药资源保护与利用重点实验室,广西南宁 530200)

摘要:为解析八角(*Illicium verum* Hook. f)黄酮类成分合成途径的关键基因,以八角果实和叶片为试验材料,进行转录组测序分析和 UPLC-ESI-MS/MS 代谢组分析,并将两组学进行联合分析。八角代谢组分析获得 12 类共 1 292 种化合物,不同组织次生代谢产物的累积有明显的差异,共有 571 个差异代谢物(DAMs),包括 331 个上调基因、240 个下调基因,其中黄酮类化合物占比最大;测序共获得 41.04 Gb 的 clean data,各样本 Q30 碱基占 91.37% 及以上。八角果和叶中检测并筛选得到 4 506 个差异基因(DEGs),包括 2 035 个上调基因、2 471 个下调基因,其中有 132 个与黄酮合成相关的差异基因;两组学联合分析筛选得到 25 个代谢物和 33 个基因,将其进行相关性分析发现,CAH、CHS、CHI、F3H、F3'H、F3'5'H 等基因的表达水平与黄酮代谢物的积累显著相关,表明这些基因参与调控黄酮类化合物的生物合成。本研究首次阐释了八角黄酮类成分的合成途径和相关基因,为利用生物工程技术生产其黄酮类化合物提供了依据,对于扩大用药资源具有重要意义。

关键词:八角;转录组;代谢组;黄酮类;关键基因

中图分类号:R282.71;R284.3

文献标志码:A

文章编号:1002-1302(2024)12-0060-09

药用植物是许多化学药物的重要原料,目前临床用药来源于植物代谢物的占 1/3 以上^[1]。植物在发育和生长过程中,不同器官和组织的发育阶段以及不同生长环境影响着其代谢产物和转录本的积

累和表达模式,导致基因和代谢物存在一定的差异,基于这些差异进行植物功能基因、活性成分的生物合成途径的分子机制、药用植物新资源开发、资源可持续利用及分子标记辅助育种等研究具有深远意义^[2-3]。

药用植物中含有大量多样的次生代谢产物,其积累与时间、空间、环境条件等密切相关^[4]。次生代谢物的合成是一个复杂的调控网络,其中包括相关基因的表达、转录后修饰、蛋白质翻译和翻译后再修饰等过程,而单一的组学技术难以充分揭示如此复杂的植物生物学过程^[5-6]。通过多组学整合,建立起基因调控网络,能够在分子水平上对各分子间的调控以及因果关系进行更深层的研究,进而解析不同条件下植物的基因功能以及相互作用网

收稿日期:2023-08-02

基金项目:广西自然科学基金(编号:2020GXNSFAA238017、2023GXNSFDA026065、2019GXNSFAA245084);广西壮瑶药重点实验室项目(编号:GXZYKF2022-20);广西中医药大学研究生教育创新计划(编号:YCSY2023006);广西研究生联合培养基地项目(编号:桂学位[2021]6号);广西高校大学生创新创业训练计划(编号:S202210600060、S202210600063)。

作者简介:王乙淋(1999—),女,广西河池人,硕士研究生,主要从事中药(壮瑶药)鉴定研究。E-mail:2460228324@qq.com。

通信作者:黄 勇,博士,教授,主要从事中药资源和分子生药研究。E-mail:huangykiz@163.com。

[32] Hill J T, Demarest B L, Bisgrove B W, et al. MMAPP: mutation mapping analysis pipeline for pooled RNA-seq[J]. Genome Research, 2013, 23(4): 687-697.

[33] Zhao H, Li J C, Yang L, et al. An inferred functional impact map of genetic variants in rice[J]. Molecular Plant, 2021, 14(9): 1584-1599.

[34] 姚晓云,蓝海军,邓 伟,等. 水稻淡泊叶突变体的叶绿素含量

测定及农艺性状比较分析[J]. 江西农业学报, 2020, 32(12): 12-15.

[35] Chen D G, Zhou X Q, Chen K, et al. Fine-mapping and candidate gene analysis of a major locus controlling leaf thickness in rice (*Oryza sativa* L.)[J]. Molecular Breeding, 2022, 42(2): 6.

[36] 郭 微. 利用 QTL 定位和 BSA-seq 分析鉴定碱胁迫下水稻产量相关性状的候选基因[D]. 哈尔滨:东北农业大学, 2019.