

葡萄果粒形状简化基因组关联分析

张恒^{1,2,*}, 刘众杰^{1,*}, 樊秀彩², 张川¹, 崔力文¹, 刘崇怀^{2,**}, 房经贵^{1,**}

(¹南京农业大学园艺学院, 南京 210095; ²中国农业科学院郑州果树研究所, 郑州 450009)

摘要: 为了寻找控制葡萄果粒形状的基因位点, 本研究中对 304 个葡萄种质的果形性状进行了简化基因组关联分析。对葡萄种质材料的基因组 DNA 进行测序, 平均测序深度为 8.13×, 共开发 466 618 个 SLAF 标签, 其中多态性 SLAF 标签有 392 374 个。比对测序分析获得了 4 241 729 个 SNP 标记, 进一步根据完整度大于 0.8 以及次要基因型频率大于 0.05 的标准进行过滤, 共得到 481 192 个 SNP。本次检测到的控制葡萄果粒形状的相关位点为首次报道, 除基于一般线性模型的 SNP rs26137624 外, 其余检测表型遗传变异贡献率均大于 11%。基于一般线性模型和混合线性模型分别关联到 7 个和 3 个显著性 SNP 标记, 且均位于 2 号染色体上。通过筛选每个显著性 SNP 标记上下游区域, 对关联区域进行功能基因挖掘, 共得到 10 个有预测功能注释的基因和 7 个未知功能的基因。

关键词: 葡萄; 果粒形状; SLAF; 基因组关联分析

中图分类号: S 663.1

文献标志码: A

文章编号: 0513-353X (2017) 10-1959-10

Genome-wide Association Mapping of Berry Shape Traits via the Reduced Representation Sequencing in Grape

ZHANG Heng^{1,2,*}, LIU Zhongjie^{1,*}, FAN Xiucui², ZHANG Chuan¹, CUI Liwen¹, LIU Chonghuai^{2,**}, and FANG Jinggui^{1,**}

(¹College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China; ²Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, Zhengzhou 450009, China)

Abstract: Three hundred and four grape germplasms were selected to do the reduced representation sequencing for genome-wide association mapping to detect the genetic loci controlling berry shape. The DNA samples isolated from the grape germplasms were sequenced with a 8.13-fold sequencing depth for each sample, and a total of 466 618 high quality SLAF (specific locus amplified fragment sequencing) labels were acquired, among which 392 374 were polymorphicones. Alignment analysis to the polymorphic SLAF labels identified 4 241 729 SNP, from which 481 192 SNP were obtained with the screening based on a proportion of genotyped efficiency over 0.8 and minor allele frequencies higher than 0.05. In this study, the loci controlling grape berry shape were the first reported, and the devotions of the gene variances for berry shape were higher than 11% except of the SNP rs26137624 based on GLM. Further analysis using GLM and MLM model detected 7 and 3 SNPs, respectively, and all these 10 SNPs

收稿日期: 2017-05-20; **修回日期:** 2017-09-12

基金项目: 江苏省科技支撑计划(农业)项目(BE2013431); 国家自然科学基金项目(31672131); 国家现代农业产业技术体系建设专项资金项目(CARS-30); 农业部作物种质资源保护项目(NB2012-2130135-34)

* 并列第一作者

** 通信作者 Author for correspondence (E-mail: fanggg@njau.edu.cn; liuchonghuai@caas.cn)

located on chromosome 2. Through filtering the associated upstream and downstream areas around each significant SNP, and detecting these related regions for functional genes, 10 genes with predicted functional annotations and 7 without named protein product predicted were found.

Keywords: grape; berry shape; SLAF; genome-wide association analysis

果粒形状是葡萄的重要农艺性状之一，它影响着葡萄的商品价值与观赏性，一直是葡萄选育的重要目标性状之一。果粒形状除了受基因调控外，还受环境条件（高温、干旱等）、植株营养水平，内源激素和植物生长调节剂等因素的直接或间接影响（Houel et al., 2013; 房经贵和刘崇怀, 2014a）。

目前关于葡萄果粒形状形成的分子机理研究较少，探究果粒形状的遗传基础以及进行基因定位将为分子育种工作带来重要帮助。随着 RFLP（限制性片段长度多态性，Restriction fragment length polymorphism）、SSR（简单重复序列，Simple sequence repeat）和 SNP（单核苷酸多态性，Single nucleotide polymorphisms）等分子标记技术在植物遗传学研究中的广泛应用，通过分子标记辅助选择育种的手段可以实现目标性状的早期鉴定。用传统遗传图谱方法研究控制性状基因位点需要构建遗传群体，耗费时间较长，同时常由于群体较少，作图精度低，定位到的等位基因较少。近年来，随着葡萄全基因组测序的完成，葡萄基因组学的研究已经呈现出由简单质量性状向复杂的数量性状转移的趋势，特别是大量 SNP 标记的开发以及生物信息学的迅猛发展，基于连锁不平衡的关联分析为研究葡萄复杂数量性状提供了新途径。全基因组关联分析（Genome-wide association study, GWAS）已经发现了拟南芥、玉米、水稻等的一些重要性状控制基因（Huang et al., 2010; Yan et al., 2011; Jiao et al., 2012; Verslues et al., 2014）。虽然全基因组关联分析的最有效、最直接的方法是全基因组测序，但是昂贵的测序成本是该技术被广泛采用的障碍。而成本较低、关联分析效果较好的 GBS（基因型测，Genotyping by Sequencing）（Elshire et al., 2011）、RAD（限制位点相关性 DNA，Restriction-site Associated DNA）（Chen et al., 2013）、SLAF（特异性位点扩增片段测序，Specific Locus Amplified Fragment Sequencing）（Zhang et al., 2013）等简化基因组测序技术成为应用于功能基因挖掘更广泛的措施（Han et al., 2015; Sonah et al., 2015; Zhou et al., 2015）。本试验中对 304 份葡萄种质材料进行了果粒形状的分析，并利用 SLAF 技术开发的葡萄全基因组 SNP 标记对果粒形状进行了关联分析，以期获得控制果粒形状基因的标记、基因或相关候选基因。

1 材料与方法

1.1 试验材料

试验材料全部来自中国农业科学院郑州果树研究所国家果树种质郑州葡萄圃。共选取 304 份种质材料（表 1），包括 199 份欧亚种、93 份欧美杂种和 12 份野生种。常规管理，树势基本相同。于 2015 年 4 月初萌芽期取其嫩叶，用液氮速冻后放 -80 °C 冰箱保存，用 CTAB 法（Murray & Wf, 1980）提取基因组 DNA。

1.2 果粒形状调查

果粒形状调查及测量参照刘崇怀等（2006）编著的《葡萄种质资源描述规范和数据标准》进行，2013—2015 年连续 3 年对 304 份材料的果形性状进行系统观察、测量。

1.3 参考基因组信息

预计每个个体平均测序深度为 8×, 进行群体内的 SNP 标记开发。选取葡萄基因组 (PN40024) 作为参考基因组进行酶切预测。参考基因组具体信息: (1) 葡萄实际基因组大小约为 416 Mb, GC 含量为 39.08%; (2) 葡萄 (PN40024) 基因组 (Jaillon et al., 2007) 组装出的基因组大小为 486 Mb, GC 含量为 33.41% (下载地址: ftp://ftp.ensemblgenomes.org/pub/release-23/plants/fasta/vitis_vinifera/dna/)。

1.4 酶切方案确定、试验建库及高通量测序

利用电子酶切对参考基因组 (PN40024) 进行电子酶切预测, 选择最适酶切方案。选择原则: (1) 酶切片段长度与具体试验体系的吻合程度; (2) 酶切片段在基因组上尽量均匀分布; (3) 位于重复序列的酶切片段比例尽可能低; (4) 最终获得酶切片段 (SLAF 标签) 数满足预期标签数。

根据选定的最适酶切方案, 对检测合格的各样品基因组 DNA 分别进行酶切建库。每个样品的 DNA 总量为 10 μg, 分别加入寡切酶 I, T₄ DNA 连接酶, ATP 在 37 °C 下孵化 20 min, 再加入寡切酶 II 37 °C 下孵化 20 min, 得到酶切片段。对得到的酶切片段 (SLAF 标签) 进行 3'端加 A 处理, 连接 Dual-index (Kozich et al., 2013) 测序接头, PCR 扩增, DNA 纯化, 混样, 切胶选取目的片段, 再进行文库质检, 合格后用 Illumina HiSeq™ 2500 上机测序。

1.5 序列的处理、信息分析与利用

利用 Dual-index 对测序得到的原始数据进行识别, 得到各个样品的 raw reads。过滤掉测序 raw reads 的接头后, 得到 clean reads, 然后进行测序质量和数据量的评估。高质量的双端 clean reads 用 SOAP₂ (Li et al., 2009) 和 BWA 软件 (Li et al., 2009) 比对到参考基因组 (PN40024) 上, 用 GATK (McKenna et al., 2010) 和 SAMtools (Li et al., 2009) 软件开发全基因组范围的 SNP 标记。

全基因组关联分析, 是基于 SNP 利用 TASSEL (Bradbury et al., 2007) 软件的一般线性模型 (General Linear Model, GLM) 和混合线性模型 (Compressed Mixed Linear Model, MLM) 得到关联值, $Y = \alpha X + \beta Q + \mu K + e$, α 、 β 、 μ 和 e 为相关参数。其中, 通过 admixture 软件 (Alexander et al., 2009) 计算样品群体结构 Q , 通过 SPAGeDi (Hardy & Vekemans, 2002) 软件计算样品间亲缘关系 K , 一般线性模型使用 Q 信息, 而混合线性模型使用 $Q + K$ 信息。 Y 为表型, X 为基因型, 最终每个 SNP 位点都能得到一个关联值 (P)。 P 值用 Bonferroni 方法校正: $\alpha \leq 0.01$ 和 $\alpha \leq 0.05$ (分别为 $P \leq 2.70 \times 10^{-7}$ 和 $P \leq 1.35 \times 10^{-6}$)。

2 结果与分析

2.1 葡萄种质资源的果粒形状

对 304 份葡萄种质资源果实形状性状进行调查, 其果粒形状共有 10 种 (图 1, 表 1)。其中, 椭圆形果粒的种质资源最多, 占总数的 44.1%; 钝卵圆形仅有 3 个、3 个长圆形、弯形 2 个和束腰形 1 个。欧美杂种中果粒形状多样性 (6 种) 较欧亚种 (10 种) 低。另外, 欧美杂种以椭圆形和圆形的果粒为主, 分别占欧美杂种总数的 46.2% 和 34.4%; 欧亚种中以椭圆形果粒为主, 占欧亚种总数的 45.7%, 其次是圆形和长椭圆形, 分别占欧亚种总数的 23.6% 和 17.1%; 野生种葡萄的果粒形状基本都是圆形。



图 1 葡萄果粒形状
Fig. 1 Berry shape for grape

表 1 304 份葡萄种质果粒形状调查表
Table 1 The berry shape of 304 grape germplasms

种类 Species	果粒形状 Berry shape	数量 Number	种质资源 Germplasm resource
欧美杂种 Hybrids	扁圆形 Flat	4	红萨福尔克 Suffolk Red, 蜜汁 Mizhi, 昆诺无核 Queenora, 戈尔比 Geerbi
<i>V. vinifera</i> × <i>V. labrusca</i>	倒卵形 Obovate	12	峰寿 Fengshou, 红富士 Beni Fuji, 红瑞宝 Beni Zuiho, 红伊豆 Beniizu, 吉峰 Jifeng, 龙宝 Ryuho, 蜜红 Honey Red, 摩尔多瓦 Moldova, 茉莉香 Molixiang, 尼亚加拉 Niagara, 田野红 Tano Red, 沈农香丰 Shennong Xiangfeng
	圆形 Round	32	金星无核 Venus, 巴士 Bath, 哈佛德 Hartford, 黑奥林 Black Olympia, 红星 Hongxing, 琥珀 Hupo, 京亚 Jingya, 京优 Jingyou, 康拜尔 Campbell, 状元红 Zhuangyuanhong, 纽约玫瑰香 New York Mascat, 罗曼尔 Rommel, 美洲白 Meizhoubai, 蜜尔紫 Mills, 申秀 Shenxiu, 维金娜斯 Vergennes, 香悦 Xiangyue, 紫早 Zizao, 沪培 2 号 Hupei 2, 俄罗斯康可 Russian Concord, 加拿大玫瑰 Canada muscat, 玛尔斯 Mars, 无核蜜 Honey Seedless, 天康玫瑰 Tiankang Meigui, 巨峰玫瑰 Jufeng Meigui, 华夫葡萄 Huafu, 黑丰 Heifeng, 黑佳酿 Heijianiang, Reliance, Rosebell, 11-5, 16-8
	鸡心形 Ovate	1	贵妇人 Guifuren
	长椭圆形 Long elliptic	1	翠峰 Cuifeng
	椭圆形 Elliptic	43	艾尔威因 Aierweiyin, 白香蕉 Baixiangjiao, 蓓蕾玫瑰 Muscat Bailey, 碧绿珠 Lady Washington, 脆红 Cuihong, 大阪 48202 Daban 48202, 峰后 Fenghou, 高墨 Takasumi, 高砂 Takasago, 贵妃玫瑰 Guifei Rose, 黑贝蒂 Herbert, 黑蜜 Honey Black, 红奥林 Red Olympia, 红双味 Hongshaungwei, 红香蕉 Hongxiangjiao, 户太 8 号 Hutai 8, 金玫瑰 Golden Muscat, 巨峰 Kyoho, 白奥林 Baiaolin, 巨玫瑰 Jumeigui, 卡托巴 Catawba, 莎加蜜 Sagami, 斯蒂本 Steuben, 藤稔 Fujiminori, 天使玫瑰香 Muscat Angel, 田野黑 Tano Black, 夕阳红 Xiyanghong, 先峰 Xianfeng, 信浓红 Xinnonghong, 伊豆锦 Izu Nishiki, 着色香 Zhuosexiang, 紫玉 Shigyoku, 紫珍香 Zizhenxiang, 二伯娜 Urbana, 音田 Otda, 布朗无核 Bronx Seedless, 范纳萨无核 Vanerssa Seedless, 康能无核 Canadice, 累克芒特 Lakemont, 希姆劳德 Ximulaode, 醉人香 Zuirenxiang, 玫瑰魁 Governor Rose, 沈农硕丰 Shennong Shuofeng
欧亚种 <i>V. vinifera</i> L.	椭圆形 Elliptic	91	阿古西 Angour siokh, 安娜玛丽亚 Annamaria, 白拉查基 White Rozaki, 阿弗阿丽 Afuali, 比昂扣 Rosario Bianco, 比赛尔 Julski Beaser, 登瓦斯玫瑰 Misket Dounvaski, 芳香拉查基 Mathias Aromatic, 绯红 Cardinal, 粉红阿里蔓登 Fenhong Alimandeng, 粉红拉查基 Fenhong Lachaji, 佛斯玫瑰 Fusi Meigui, 和田红 Hetianhong, 黑彭斯 Pensal Blanco, 红古沙 Gousale kara, 红罗莎里奥 Rosario Rosso, 红亚历山大 Flame Muscat, 基拉尔 Jilaer, 济南早红 Jinan Zaohong, 库特赛塔 Kutesaita, 李子香 Lizixiang, 龙眼 Longyan, 洛迪 Rodi, 马琳格尔 Precoce de Malingre, 玫瑰香 Mascat Hamburg, 那多尔 Naduoer, 派克斯 Aromatic of pees, 葡萄园皇后 Queen of Vineyard, 普列文玫瑰 Mascat Plevenski, 巧吾什 Qiaowushi, 琼尼 Joanna Charnice, 秋白 Qiubai, 沈阳玫瑰 Shenyang Rose, 胜利 Shengli, 胜利花药 2 号 Shenglihuayao 2, 特别黑大粒 Tebieheidali, 吾家克阿依 Wujiakeayi, 西莱特 Selecta, 夏白 Xiabai, 洋葡萄 Yang Grape, 伊丽莎白 Elisabeth Grape, 依斯比沙里 Ispissar, 意大利 Italia, 意大利玫瑰 Rose D Italia, 早黑宝 Zaoheibao, 泽香 Zexiang, 紫峰 Zifeng, 紫鸡心 Zijixin, 紫桃 Zitao, 紫珍珠 Zizhenzhu, 埃木布腊德 Emberad, 爱神玫瑰 Aishen Rose, 波发尔 Madh pa fare, 红宝石无核 Ruby Seedless, 红脸无核 Blush Seedless, 红无籽露 Hongwuzilu, 京紫晶

续表 1

种类	果粒形状	数量	种质资源
Species	Berry shape	Number	Germplasm resource
			Jingzijing, 康耐诺 Kangnainuo, 黎明无核 Dawn Seedless, 那赛里 Nasaili, 秋无核 Autumn Seedless, 桑多哈尼 Sangduohani, 无核白 Thompsons Seedless, 无核玫瑰 Rose Seedless, 无核紫 Black Monukka, 西营 Xiyang, 瓦尔瑟 Waerse, 利伐尔 Lival, 红斯地 Hongside, 斯帕汉堡 Sipahanbao, 早康宝 Zaokangbao, 秋红宝 Qihongbao, 京秀 Jingxiu, Waltham Cross, 秋黑宝 Qiuheibao, Sivan, 金田翡翠 Jintian Feicui, 皇家无核 Huangjia Seedless, 金田红 Jintianhong, 布杰苏里 Bujisuli, 沈农金皇后 Shennong Jinhuanghou, Gold, 白比诺 Pinot Blanc, 佳利酿 Carignan, 琼瑶浆 Traminer, Crimson Seedless, 87-1, Gold star, Black Diamond, Hod, Mutant of Thompson Seedless
	扁圆形 Flat	4	米哈尔 Mihaer, 香妃 Xaingfei, 红光无核 Flame Seedless, 90-1
	倒卵形 Obovate	8	芳香 Fangxiang, 芳香葡萄 Fangxiang Grape, 高蓓蕾 High Bailey, 京可晶 Jingkejing, 京丰无核 Jingfeng Seedless, 美丽无核 Beauty Seedless, 杨格尔 Yanggeer, 西拉 Syrah
	钝卵圆形 Obtuse-ovate	2	泽玉 Zeyu, Star Light
	弯形 Arched	2	驴奶 Lunai, 金手指 Golden Finger
	圆形 Round	47	秀特玫瑰 Rose Cioutat, 11-43, 安吉文 Madeleine Angevine, 奥利文 Irsay Oliver, 白达拉依 Baidalayi, 布拉金涅 Bulajinnie, 布列其罗然西 Bulieqiluoranxi, 德苏拉乌苏姆 Djoura ousioum, 凤凰 51 Fenghuang 51, 瑰宝 Guibao, 哈特巴尔 Hatebaer, 黑汉 Black Hamburg, 黑墨 Heimo, 黑夏尼 Heixiani, 红加利亚 Hungaria, 黄蜜斯 Huangmisi, 罗马尼亚 Gros Colman, 牡丹红 Mudanhong, 派卡尔 Paikaer, 巧保 2 号 Qiaobao 2, 苏丹尼 Sultanick Rose, 索罗门 Madeleine solomon, 小白葡萄 Xiaobai, 谢花红 Muscat Mathiasz Janosne, 伊尔玛 Kocsias Irma, 伊犁香葡萄 Yilixiang, 早金香 Zaojinxiang, 早玫瑰 Zaomeigui, 郑州早红 Zhengzhou Zaohong, 波尔莱特 Perlette, 碧香无核 Bixiang Seedless, 青紫 Qingzi, SG, 安娜玛利亚 Anmamaliya, 红地球 Red Globe, 白玉霓 Ugni Blanc, 赤霞珠 Cabernet Sauvignon, 法国兰 Blue French, 黑比诺 Pinot Noir, 贵人香 Italian Riesling, 灰比诺 Pinot Gris, 雷司令 Riesling, 梅鹿辄 Merlot, 品丽珠 Cabernet franc, 蛇龙珠 Cabernet Gernischet, 霞多丽 Chardonnay, 小白玫瑰 Muscat Blanc
	长椭圆形 Long elliptic	34	红莲子 Honglianzi, Jade seedless, 奥古斯特 Augusta, 奥林匹亚 Olympia, 白哈丽丽 Khalili Blanc, 白胡沙涅 Khoussaine Blanc, 保尔加尔 Boulgal, 斯堪地拜格 Skendberg, 超宝 Chaobao, 格拉卡 Greaca, 红意大利 Ruby Okuyama, 克里木考尼松 Kelimukaonisong, 里扎马特 Rizamat, 马奶 Manai, 京玉 Jingyu, 牛心 Niuxin, 潘诺尼亚 Panuoniya, 齐查卡普列 Qichakapulie, 恰齐瓦赫 Qiaqiwahe, 日鲁捷维 Riluwiejie, 森田尼 Centenial, 圣诞玫瑰 Christmas Rose, 苏 46 号 Su 46, 微红白 Weihongbai, 维多利亚 Victoria, 早玛瑙 Zaomanao, 蒙丽莎无核 Melissa Seedless, 新葡 2 号 Xinpu 2, 西瓦兹 Xiwazi, 黑玫瑰 Black Muscat, 14-6, 15-7, 紫地球 Zidiqiu, Mystery
	长圆形 Cylindric	3	克林巴马克 Xycahke Kejihm Bapmak, 玛瑙 Manao, 美人指 Manicure Finger
	束腰形 Bottle	1	瓶儿 Pinger
	鸡心形 Ovate	7	白鸡心 Baijixin, 粉红亚依苏娜 Fenhong Yayisuna, 黑鸡心 Heijixin, 奥迪亚 Otilia, 京早晶 Jingzaojing, Kamea, 粒丽特 Lilit
野生种 Chinese wild Vitis	圆形 Round	10	葛藟葡萄 <i>V. flexuosa</i> , 变叶葡萄 <i>V. piasezkii</i> , 桦叶葡萄 <i>V. betulifolia</i> , 山葡萄 <i>V. amurensis</i> , 桑叶葡萄 <i>V. ficifolia</i> , 毛葡萄 <i>V. heyneana</i> , 华东葡萄 <i>V. pseudoreticulata</i> , 蓂莫 <i>V. bryoniifolia</i> , 秋葡萄 <i>V. romanetii</i> , 刺葡萄 <i>V. davidii</i>
	钝卵圆形 Obtuse-ovate	1	腺枝葡萄 <i>V. adenoclada</i>
	扁圆形 Flat	1	菱叶葡萄 <i>V. hancockii</i>
总计 Total		304	

2.2 酶切均匀性评估与 SLAF 标签开发

对葡萄参考基因组 (PN40024) 进行电子酶切预测, 根据酶切选择原则, 发现 *Rsa* I 与 *Hae* III 两个寡切酶的酶切组合可以使葡萄基因组重复序列的酶切片段比例最低, 酶切片段长度与 SLAF 标签吻合程度最高, 预测可得到的 SLAF 标签超过 160 000 个, 满足试验标签数 (Sun et al., 2013), 所以最终确定 *Rsa* I + *Hae* III 为最佳酶切组合。

对 304 份葡萄样品进行测序, 共得到 663.59 M clean reads。用 SOAP₂ 和 BWA 两种软件将每个样品的 clean reads 比对到葡萄 (PN40024) 参考基因组上, 可靠的 clean reads 为双端都比对到参考基因组上的 clean reads。SLAF 标签为比对到基因组唯一位置的双端序列, 根据比对结果, 选取平均深度在 2 以上的标签定义为 SLAF 标记。统计每 100 kb 范围内 SLAF 标记个数, 得到 SLAF 标记在每条染色体上的分布情况。结果表明 SLAF 标记在葡萄基因组各染色体上分布基本均匀。通过统计

共获得 466 618 个 SLAF 标签, 多态性 SLAF 标签有 392 374 个, 样品平均测序深度为 8.13 \times 。

2.3 SNP 标记的挖掘与分析

以葡萄(PN40024)基因组作为参考基因组,用 SOAP₂ 软件和 BWA 软件将测序得到的 clean reads 比对到参考基因组上,并使用 GATK 和 SAMtools 两种软件开发 SNP,以两种软件开发获得的 SNP 交集作为最终可靠的 SNP 数据集。经过比对与分析共获得 4 241 729 个 SNP 标记,进一步根据完整度大于 0.8,次要基因型频率大于 0.05 过滤,共得到高一致性的群体 SNP 481 192 个。按照葡萄约有 4.16 亿碱基计算,本试验中大约每 1 kb DNA 序列长度上有一个 SNP。SLAF 标签在基因组各染色体上分布基本均匀且 SNP 密度达到全基因组关联分析的要求 (Sun et al., 2013),简化基因组测序正常,可以用于全基因组关联分析。

2.4 GWAS 关联分析

关联分析应选取目标表型有明显差异的个体,且表型的性状分布基本符合正态分布,保证有足够的多态性。从调查的 304 份葡萄种质资源果形的情况看,符合关联分析对目标性状多样性的要求。进一步利用从 304 份种质资源中开发得到的 481 192 个 SNP,采用一般线性模型 (Generalized Linear Models, GLM) 和混合线性模型 (Mixed Linear Models, MLM) 进行果粒形状的关联分析 (图 2, 表 2),分别关联到 7 个 SNP 和 3 个 SNP,且一般线性模型得到的 7 个显著关联 SNP 包含混合线性

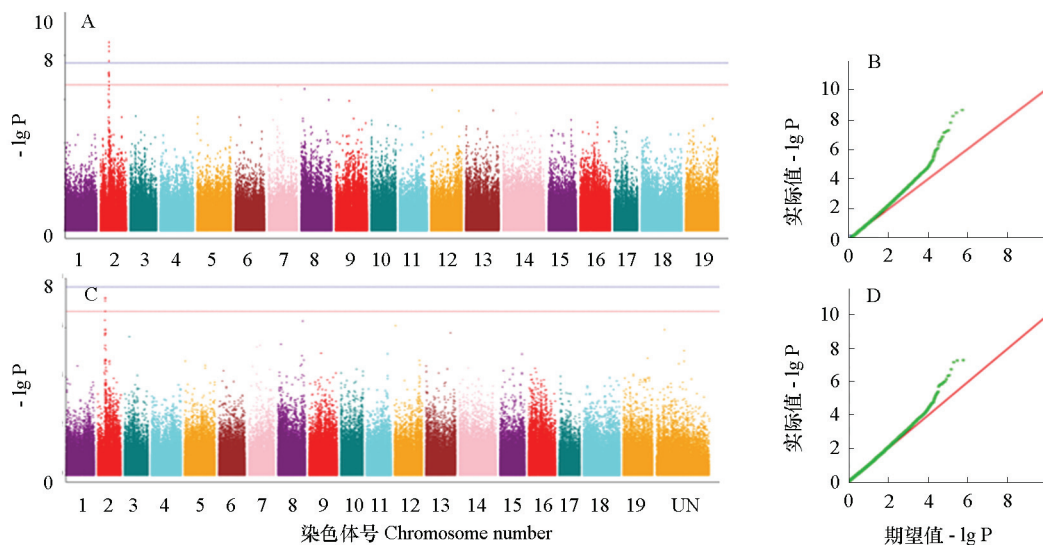


图 2 果粒形状关联分析结果

A, C 分别为一般线性模型和混合线性模型的曼哈顿图,纵坐标为关联值的 $-\lg P$ 值,横坐标为染色体。一个点代表一个 SNP 位点;蓝色虚线位置为多重假设检验 (FDR 检验) 的显著性阈值,高于阈值的点为关联显著位置;红色虚线为 0.05/SNP 总数的负对数。

B, D 为 QQ-plot 图,实线是预测的阈值,SNP 与预测的阈值在前端贴合一致,仅在末端翘起,代表 SNP 假阳性低,关联结果可靠。

Fig. 2 The result from genome-wide association scanning for berry shapes

A, C show the Manhattan plots for the generalized linear model and compressed mixed linear model; The X-axis shows the single nucleotide polymorphism (SNP) along each chromosome, the Y-axis is the $-\lg P$ for the association, a dot represents a SNP loci, the blue dashed line is significance threshold for FDR inspection, higher than the threshold value of points is the associated prominently; the red dotted line is 0.05 SNP objective negative logarithm. B, D show quantile-quantile plot for GLM and MLM, solid lines are predicted threshold, SNP and predicted threshold on the front joint is consistent, only in the end is perk, the associated results reliable.

模型得到的 3 个显著 SNP, 这几个 SNP 位于 2 号染色体 6 125 174 ~ 6 137 940 和 6 125 174 ~ 6 140 580 区间。这些显著相关的 SNP 解释果粒形状的遗传变异的贡献率 (R^2) 为 6% ~ 13%, 基于一般线性模型, 有 3 个 SNP (rs26137940、rs26125174、rs215463958) 属于极显著水平 ($P \leq 2.70 \times 10^{-7}$)。

表 2 关联分析中果粒形状关联结果

Table 2 Details of loci associated with berry shape traits identified via GWAS

重要 SNP Significant SNP	染色体 Chromosome	位置 Position	-lg P		R ² (贡献率) /%		次要基因型 频率 MAF
			混合线性模型	一般线性模型	混合线性模型	一般线性模型	
			MLM	GLM	MLM	GLM	
rs26137940	2	6 137 940	7.22	8.6	12	13	0.18
rs26137624	2	6 137 624	7.19	5.18	12	6	0.27
rs26125174	2	6 125 174	7.09	8.41	12	13	0.17
rs215463958	2	15 463 958	—	8.19	—	13	0.06
rs26137630	2	6 137 630	—	7.74	—	12	0.12
rs26140580	2	6 140 580	—	7.22	—	11	0.08
rs25847252	2	5 847 252	—	7.19	—	11	0.43

根据不同染色体位置衰减水平不同, 在显著相关 SNP 的上下游区域, 对关联区域进行功能基因挖掘, 在这些显著关联 SNP 区域共获得 16 个与葡萄果粒形状关联的基因 (表 3)。包括 7 个未知功

表 3 果粒形状显著关联区域的候选基因

Table 3 The candidate genes of significantly associated regions for berry shape

基因名称 Gene name	染色体号 Chromosome	基因起始 位置 Gene-start	基因结束 位置 Gene-end	关联 SNP 位置 SNP_location	预测的基因功能注释 Annotation
VIT_02s0012g00150.t01	2	5 785 719	5 790 810	5 847 252	含有 GDPDL 2 亚型 X1 的 Gpd 蛋白激酶结构域 Glycerophosphodiester phosphodiesterase protein kinase domain-containing GDPDL2-like isoform X1
VIT_02s0012g00160.t01	2	5 798 038	5 801 982	5 847 252	预测蛋白 VITISV_038964 Hypothetical protein VITISV_038964
VIT_02s0012g00170.t01	2	5 804 094	5 806 987	5 847 252	酰基载体蛋白 3, 线粒体亚型 X2 Acyl carrier protein 3, mitochondrial isoform X2
VIT_02s0012g00180.t01	2	5 811 458	5 820 423	5 847 252	S-noroclaurine synthase 1 isoform X1
VIT_02s0012g00190.t01	2	5 826 774	5 866 553	5 847 252	未知蛋白 LOC100263809 Uncharacterized protein LOC100263809
VIT_02s0012g00240.t01	2	5 887 055	5 890 114	5 847 252	未知 Unnamed protein product
VIT_02s0012g00300.t01	2	6 039 133	6 077 860	6 125 174	未知 Unnamed protein product
VIT_02s0012g00310.t01	2	6 092 736	6 111 900	6 125 174	未知 Unnamed protein product
VIT_02s0012g00320.t01	2	6 113 176	6 114 475	6 125 174	未知蛋白 LOC100242112 Uncharacterized protein LOC100242112
VIT_02s0012g00350.t01	2	6 121 047	6 124 367	6 125 174	SNF1 相关蛋白激酶调节亚基 β -2 SNF1-related protein kinase regulatory subunit beta-2-like
VIT_02s0012g00360.t01	2	6 126 853	6 130 300	6 125 174	可能为受体蛋白激酶 At5g39020 Probable receptor-like protein kinase At5g39020
VIT_02s0012g00390.t01	2	6 174 307	6 176 342	6 125 174	未知 Unnamed protein product
VIT_02s0012g00400.t01	2	6 184 190	6 187 378	6 125 174	未知 Unnamed protein product
VIT_02s0012g00410.t01	2	6 193 683	6 198 742	6 125 174	硫酸化表面糖蛋白 185 亚型 X1 Sulfated surface glycoprotein 185 isoform X1
VIT_02s0012g00430.t01	2	6 208 485	6 208 852	6 125 174	未知 Unnamed protein product
VIT_02s0012g00440.t01	2	6 209 000	6 209 468	6 125 174	未知 Unnamed protein product
VIT_02s0012g00450.t01	2	6 211 851	6 213 556	6 125 174	含有 GDPDL 2 亚型 X1 的 Gpd 蛋白激酶结构域 Glycerophosphodiester phosphodiesterase protein kinase domain-containing GDPDL2-like isoform X1

能的基因以及 9 个预测基因功能注释的基因。这 9 个有功能预测的基因分别编码载体蛋白、蛋白激酶、受体蛋白等，可能是果粒形状调控网络中的关键酶或者功能蛋白。

3 讨论

葡萄果粒形状是数量性状，性状受多基因位点控制，遗传基础复杂，对其研究比较困难。育种实践表明优良的葡萄品种常出自少数亲本及其组合，许多新品种多来自有限数量的骨干亲本，其狭窄的遗传基础是进一步提高葡萄产量、品质的障碍，而有效地利用葡萄种质资源遗传多样性是克服其品种遗传基础狭窄问题的关键（房经贵和刘崇怀，2014b）。关联分析以连锁不平衡为基础，分析标记候选基因的遗传变异与某一群体内目标性状关系的方法。选用不同来源的自然群体作为分析对象，相对于连锁分析所研究的等位基因更丰富，因此，能够对数量性状的多影响因素进行细分，更容易解释性状关联的遗传基础。Sonah 等（2015）采用简化基因组技术对大豆的 8 个重要农艺性状进行关联分析，结果表明大多数的关联位点与前人报道的相同，也发现了一些新的位点，同时也说明了全基因组关联分析对于发现由多基因控制的复杂性状基因位点是一个有效的工具。

SLAF 技术由于其测序深度高、成本低、标记数量等优点，已经在大田作物（大豆，Zhao et al., 2015）、蔬菜（黄瓜，Wei et al., 2014）等上成功应用。例如，Chen 等（2013）基于 SLAF 技术对长穗偃麦草（ $2n = 2x$ ）、普通小麦进行多态性标记开发，获得 7e 染色体特异性片段 518 个，从中任意选择 135 条片段设计引物，开发了 89 个长穗偃麦草特异分子标记，成功率达到 65.9%。这些特异标记在长穗偃麦草以及其他小麦品种等材料中进行验证，结果表明具有很高的特异性、重复性和稳定性。本研究中参考葡萄基因组序列，利用 SLAF 简化基因组测序技术，获得测序品种在全基因组水平的 SNP，采用一般线性模型和混合线性模型进行关联分析，鉴定出与果粒形状关联的 SNP。

果树上进行关联分析研究的报道较少。曹珂等（2012）用来源于中国 6 个生态群的 104 份材料，采用 53 对 SSR（Simple Sequence Repeats）标记进行了单果质量和 6 个物候期性状的全基因组的关联分析，结果表明单果质量物候期存在遗传相关性。麻国升等（2014）利用不同黏离核性状的 45 份桃种质资源为试材，采用重测序技术开发全基因组内的 SNP，结果暗示 ppa006839m 可能与黏离核性状存在紧密的联系。本试验基于一般线性模型和混合线性模型分别获得 7 个和 3 个显著性关联 SNP，且利用一般线性模型得到的 7 个包含混合线性模型得到的 3 个。由于混合线性模型比一般线性模型要严格，一般线性模型只考虑了群体结构分析得出的 Q 值，而混合线性模型，不仅考虑了 Q 值，还考虑群体间亲属关系，所以得到的显著 SNP 较一般线性模型少。在关联到 3 个极显著 SNP 上下游共检测到其余 9 个有预测功能注释的基因，可能参与果粒形状的调控，具体如何参与调控还需要进一步试验验证。

植物果实形状是多基因控制的数量性状，它不仅受环境条件影响，而且遗传机理非常复杂，受基因间相互作用及微效基因影响（宋伟 等，2010；高慧君 等，2016）。前人的研究发现葡萄果粒形状同样是数量遗传性状（王勇 等，2015），这决定了葡萄果粒形状丰富多样性。但是，有关葡萄果粒形状遗传机理的研究还很不明确。多少个基因参与了葡萄果形性状的调控、相关基因的功能、调控生物网络以及各种环境条件的影响等多方面都有待深入研究。随着遗传分析及生物技术手段的不断完善，葡萄果粒形状的遗传基础与形成机理会得到越来越清晰认识。

References

- Alexander D H, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19 (9): 1655 - 1664.
- Bradbury P J, Zhang Z, Kroon D E, Casstevens T M, Ramdoss Y, Buckler E S. 2007. Tassel: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23 (19): 2633 - 2635.
- Cao Ke, Wang Li-rong, Zhu Geng-rui, Fang Wei-chao, Chen Chang-wen, Zhao Juan. 2012. Genetic association of fruit weight and six phenological traits in peach (*Prunus persica* L.). *Scientia Agricultura Sinica*, 45 (2): 311 - 319. (in Chinese)
曹 珂, 王力荣, 朱更瑞, 方伟超, 陈昌文, 赵 娟. 2012. 桃单果重 6 个物候期性状的遗传关联分析. *中国农业科学*, 45 (2): 311 - 319.
- Chen S, Huang Z, Dai Y, Qin S, Gao Y, Zhang L. 2013. The development of 7e chromosome-specific molecular markers for *Thinopyrum elongatum* based on SLAF-seq technology. *PLoS ONE*, 8 (6): e65122 - e65122.
- Elshire R J, Glaubitz J C, Sun Q, Poland J A, Kawamoto K, Buckler E S. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, 6 (5): e19379.
- Fang Jing-gui, Liu Chong-huai. 2014a. *Grape molecular biology*. Beijing: Science Press. (in Chinese)
房经贵, 刘崇怀. 2014a. 葡萄分子生物学. 北京: 科学出版社.
- Fang Jing-gui, Liu Chong-huai. 2014b. *Grape genetics, breeding and genomics*. Nanjing: Jiangsu Phoenix Science Press. (in Chinese)
房经贵, 刘崇怀. 2014b. 葡萄遗传育种与基因组学. 南京: 江苏凤凰科学技术出版社.
- Gao Hui-jun, Li Peng, Dong Tao. 2016. Advances in the study of actors influencing fruit shape. *Journal of Tropical Biology*, 7 (2): 279 - 284. (in Chinese)
高慧君, 李 鹏, 董 涛. 2016. 影响果实形状发育相关因素的概述. *热带生物学报*, 7 (2): 279 - 284.
- Han Y, Zhao X, Cao G, Wang Y, Li Y, Liu D. 2015. Genetic characteristics of soybean resistance to HG type 0 and HG type 1.2.3.5.7 of the cyst nematode analyzed by genome-wide association mapping. *BMC Genomics*, 16 (1): 1 - 11.
- Hardy O J, Vekemans X. 2002. SPAGeDi: a versatile computer program to analyze spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, 2 (4): 618 - 620.
- Houel C, Martin-Magniette M L, Nicolas S D, Lacombe T, Cunff L L, Franck D. 2013. Genetic variability of berry size in the grapevine (*Vitis vinifera* L.). *Australian Journal of Grape and Wine Research*, 19 (2): 208 - 220.
- Huang X H, Wei X H, Sang T, Qiang Z, Qi F, Yan Z. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics*, 42 (11): 961 - 967.
- Jaillon O, Aury J M, Noel B, Policriti A, Clepet C, Casagrande A. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, 449 (7161): 463 - 467.
- Jiao Y, Zhao H, Ren L, Song W, Zeng B, Guo J. 2012. Genome-wide genetic changes during modern breeding of maize. *Nature Genetics*, 46 (9): 1039 - 1040.
- Kozich J J, Westcott S L, Baxter N T, Highlander S K, Schloss P D. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq Illumina sequencing platform. *Applied & Environmental Microbiology*, 79 (17): 5112 - 5120.
- Li H, Durbin R. 2010. Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics*, 25 (5): 1754 - 1760.
- Li R, Yu C, Li Y, Lam T W, Yiu S M, Kristiansen K. 2009. SOAP₂: an improved ultrafast tool for short read alignment. *Bioinformatics*, 25 (15): 1966 - 1967.
- Li Chong-huai, Shen Yu-jie, Chen Jun. 2006. *Descriptors and data standard for grape (Vitis L.)*. Beijing: China Agriculture Press.
刘崇怀, 沈育杰, 陈 俊. 2006. 葡萄种质资源描述规范和数据标准. 北京: 中国农业出版社.
- Ma Guo-sheng, Wang Li-rong, Cao Ke, Zhu Geng-rui, Fang Wei-chao, Chen Chang-wen, Wang Xin-wei. 2014. Polymorphism of SNPs in two endo-PG genes and its association analysis for flesh adhesion trait in peach. *Journal of Fruit Science*, 31 (3): 345 - 352. (in Chinese)
麻国升, 王力荣, 曹 珂, 朱更瑞, 方伟超, 陈昌文, 王新卫. 2014. 桃 2 个 endo-PG 基因序列 SNPs 的遗传多样性及其与果实黏离核

- 性状的关联分析. 果树学报, 31 (3): 345 - 352.
- Mckenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A. 2010. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20 (9): 1297 - 1303.
- Murray M G, Wf T. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research*, 8 (19): 4321 - 4325.
- Sonah H, O'Donoghue L, Cober E, Rajcan I, Belzile F. 2015. Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnology Journal*, 13 (2): 211 - 221.
- Song Wei, Wang Cai-hong, Tian Yi-ke, Tian Wei, Yin Hao. 2010. Development of SSR molecular markers linked to the fruit shape in pea. *Journal of Qingdao Agricultural University (Natural Science)*, 27 (3): 213 - 215. (in Chinese)
- 宋 伟, 王彩虹, 田义轲, 田 伟, 殷 豪. 2010. 梨果实形状的 SSR 分子标记. 青岛农业大学学报 (自然科学版), 27 (3): 213 - 215.
- Sun X, Liu D, Zhang X, Li W, Liu H, Hong W, Jiang C, Guan N, Ma C, Zeng H, Xu C, Song J, Huang L, Wang C, Shi J, Wang R, Zheng X, Lu C, Wang X, Zheng H. 2013. SLAF-seq: an efficient method of large-scale, de novo, SNP discovery and genotyping using high-throughput sequencing. *PLoS ONE*, 8 (3): e58700.
- Verslues P E, Lasky J R, Juenger T E, Liu T W, Kumar M N. 2014. Genome-wide association mapping combined with reverse genetics identifies new effectors of low water potential-induced proline accumulation in *Arabidopsis*. *Plant Physiology*, 164 (1): 144 - 159.
- Wang Yong, Sunlaiman Aizezi, Li Yu-ling, Sun Feng, Wu Guo-hong, Luo Qiang-wei. 2015. Inheritance trend of the fruit traits of 'Huozhouheiyu' grape. *Acta Botanica Boreali-Occidentalia Sinica*, 35 (2): 275 - 281. (in Chinese)
- 王 勇, 苏来曼·艾则孜, 李玉玲, 孙 锋, 伍国红, 骆强伟. 2015. '火州黑玉' 葡萄杂交后代果实性状遗传倾向分析. 西北植物学报, 35 (2): 275 - 281.
- Wei Q, Wang Y, Qin X, Zhang Y, Zhang Z, Wang J. 2014. An SNP-based saturated genetic map and QTL analysis of fruit-related traits in cucumber using specific-length amplified fragment (SLAF) sequencing. *BMC Genomics*, 15 (1): 1 - 10.
- Yan J, Warburton M, Crouch J. 2011. Association mapping for enhancing maize (*Zea mays* L.) genetic improvement. *Crop Science*, 51 (2): 433 - 449.
- Zhang J, Zhang Q, Cheng T, Yang W, Pan H, Zhong J. 2015. High-density genetic map construction and identification of a locus controlling weeping trait in an ornamental woody plant (*Prunus mume* sieb. et zucc). *DNA Research*, 22 (3): 183.
- Zhang Y, Wang L, Xin H, Li D, Ma C, Ding X. 2013. Construction of a high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing. *BMC Plant Biology*, 13 (1): 1 - 12.
- Zhao X, Han Y, Li Y, Liu D, Sun M, Zhao Y. 2015. Loci and candidate gene identification for resistance to *Sclerotinia sclerotiorum* in soybean (*Glycine max* L. Merr.) via association and linkage maps. *Plant Journal*, 82 (2): 245 - 255.
- Zhou L, Wang S B, Jian J, Geng Q C, Wen J, Song, Q. 2015. Identification of domestication-related loci associated with flowering time and seed size in soybean with the RAD-seq genotyping method. *Scientific Reports*, 5: 9350.

征 订

欢迎订阅 2018 年《烟台果树》

《烟台果树》是烟台市农业科学研究院主办的果树专业性季刊。立足北方水果的主产区山东省烟台市，面向全国。经过 30 多年的发展，《烟台果树》已成为我国广大果树业者新观点、新品种、新技术的交流平台，也是苗木、农药、肥料及各种生产机具等信息发布的平台。期刊发行量大、覆盖面广。每期定价 4 元，全年仅需 16 元。每季度首月 15 日发行。若挂号，全年 28 元。全国各地邮局均可订阅，邮发代号为 24-107；也可随时直接汇款至编辑部订阅。

汇款地址：山东省烟台市芝罘区环山路 145 号 邮编：264008

单位名称：《烟台果树》编辑部 电话：0535-6236524 6615052（传真）

E-mail: ytgsgbjb@163.com ytgsgg@163.com

本刊网址: www.fruitworld.com.cn 或水果世界网