

矮丛蓝莓品种“美登”叶片 转录组测序和生物信息学分析

王云生¹, 文光琴², 聂飞³

(1. 凯里学院 环境与生命科学学院, 贵州 凯里 556011; 2. 贵州省植物园, 贵州 贵阳 550004;

3. 贵州省生物研究所, 贵州 贵阳 550000)

摘要:以矮丛蓝莓品种“美登”为试材,采用第二代测序技术对其嫩叶转录组进行了测序及生物信息学分析,研究了矮丛蓝莓功能基因组信息,以期增加对矮丛蓝莓生长发育分子机理的了解,并为矮丛蓝莓的分子标记辅助育种奠定基础。结果表明:在获得的约1 G 纯净数据,共拼装了41 929条 Unigenes,其中被GO数据库成功注释 Unigene 分别涉及到生物学过程、细胞成分及分子功能相关的共45类生理代谢功能;KOG数据库注释到19 149条 Unigenes,共涉及25条代谢通路;KEGG数据库成功注释13 072条 Unigenes,共涉及到5个大类、18个中类、129条代谢通路;Nr数据库成功注释30 738条 Unigenes,这些 Unigenes 共对应32 949条蛋白质的编码区。同时,生物信息学分析还显示,全部 Unigenes 中,有884条编码转录因子,涉及到55个家族;2 261条编码抗性基因,涉及18个家族;检测到6 591个 SSR 位点,其中2碱基重复的 SSR 数目最多,占总 SSR 位点数的70%。

关键词:生物信息学;矮丛蓝莓;转录组;高通量测序

中图分类号:S 663.203.6 **文献标识码:**A **文章编号:**1001-0009(2017)01-0101-07

蓝莓营养丰富,富含花青素、类黄酮等抗氧化活性物质,具有极高的食疗保健价值,被誉为“小浆果之王”,国际粮农组织将蓝莓列为人类五大健康食品之首^[1-2]。近年来,国内外蓝莓产业发展迅速,种植面积不断扩大,产量不断增加^[3]。栽培蓝莓包括3个种类:高丛蓝莓、兔眼蓝莓和矮丛蓝莓^[4]。其中矮丛蓝莓富含多酚、纤维、 β -谷甾醇、锰、维生素等^[5-6]。同时,矮丛蓝莓在3种蓝莓中还是花青素含量最高的种类,其抗氧化和细胞保护作用尤其突出^[7]。研究表明,矮丛蓝莓能够显著减轻部分代谢综合征及炎症等症状^[8-10],食用矮丛蓝莓可有效增加胃肠道的

有益菌群种类和数量^[11-13]。

“美登”(‘Blomidon’)是矮丛蓝莓的代表性品种,具有抵御-40℃的抗寒能力,适宜在大兴安岭、小兴安岭和长白山等高寒地区种植,是目前我国出口与加工最理想的品种之一^[14]。该研究运用第二代高通量测序仪对矮丛蓝莓“美登”的叶片转录组进行了测序,并对测序数据进行生物信息学分析,旨在更多揭示矮丛蓝莓的功能基因组信息,挖掘有用基因,开发分子标记如EST-SSR,为矮丛蓝莓生长发育的分子机理研究、利用基因工程和分子辅助育种技术创造新的优良矮丛蓝莓品种奠定基础。

1 材料与方法

1.1 试验材料

2015年5月取种植于贵州省麻江县瑞蓝果业种苗公司苗圃的2年生蓝莓品种“美登”扦插苗的嫩叶为供试材料。

1.2 试验方法

1.2.1 总RNA提取,文库构建及测序 使用生工生物工程(上海)股份有限公司生产的UNIQ-10柱式

第一作者简介:王云生(1975-),男,博士,副教授,现主要从事遗传育种与园艺学及基因组学等研究工作。E-mail:wys3269@126.com

责任作者:聂飞(1965-),男,硕士,研究员,现主要从事蓝莓栽培推广等研究工作。E-mail:1134477929@qq.com

基金项目:国家自然科学基金资助项目(31560091);贵州省科技厅重点资助项目(黔教合KY字[2013]186);贵州省科技成果重点推广计划资助项目(黔科合成果字[2014]5007-1号)。

收稿日期:2016-09-27

Trizol 总 RNA 抽提试剂盒并严格遵照说明书提供的相应步骤提取叶片总 RNA。在提取的总 RNA 加入 DNase I, 37 °C 水浴 30 min 去除残余 DNA。使用 2100 Bio-analyzer(Agilent Technologies, Santa Clara, CA)及 RNA 琼脂糖凝胶电泳检测总 RNA 的质量和浓度。用 Oligo-dT 磁珠(QIAGEN)将 mRNA 从总 RNA 中分离出来,之后再将 mRNA 片段加入碎片缓冲液(fragmentation buffer),使之分解成短片段。然后将 mRNA 反转录成 cDNA;具体步骤利用随机 6 bp 引物反转录生成第一链,使用 RNase H 和 DNA 聚合酶合成互补链。生成的 cDNA 片段使用 QIA 快速 PCR 试剂盒提取纯化。纯化 cDNA 片段经 EB 缓冲液洗涤后修复末端 poly-A,然后连接上测序接头,进行琼脂糖凝胶电泳纯化。最后利用聚合酶链反应 PCR 技术对连接接头的 cDNA 进行扩增获得测序文库。测序文库由 Illumina 测序平台(Illumina HiSeq™ 2000)进行测序(广州,中国),测序读长为 125 bp。测序数据上传 GenBank(SUB1709548)。

1.2.2 转录组生物信息学分析 1)统计测序得到的原始序列(Raw reads)数据量,包括序列数量和总碱基获得量;2)利用自写的 Perl 程序鉴定并去除低质量序列(质量低于 20 的碱基超过 50%的、未读取碱基信息超过 5%及仅含接头的序列),保留达到质量要求的有效序列(Clean reads)并统计其数量;3)使用 Trinity 软件(<http://trinityrnaseq.sourceforge.net/>)进行非冗余特异基因序列(Unigenes)组装,并统计其数量,长度分布;4)Unigenes 功能注释。利用 Blast 软件(<http://blast.ncbi.nlm.nih.gov/Blast.cgi>)与 GO、COG、KEGG 和 Nr 相应数据库进行比对获得 Unigenes 的功能信息;5)与植物转录因子数据库(<http://planttfdb.cbi.pku.edu.cn/>)进行比对,挖掘转录因子候选序列并进行统计;6)与抗性基因数据库(http://prgdb.crg.eu/wiki/Main_Page)比对,识别抗性基因的候选 Unigenes 并进行统计;7)利用网络软件 WebSat(<http://purl.oclc.org/NET/web-sat/>)识别 SSR 位点并进行统计。

2 结果与分析

2.1 原始序列及拼装

从表 1 可以看出,测序共得到 8 989 210 条原始序列,总碱基量达到 1.1 G,碱基中 GC 含量为 47.10%。其中只含测序接头的序列为 3 328 条,占 0.037%,质量不达标的为 134 646 条,占 1.5%,其余为质量合格的序列,约占全部原始序列数的 98.5%,

表明测序质量较好。从表 2 可知,对测序得到的 8 851 236 条高质量序列进行拼装,共得到了 41 929 条非冗余的独立基因(Unigenes),全部 Unigenes 的总碱基数 24.733 4 M,GC 含量为 44.9%,Unigenes 的长度最低为 201 bp,最长达到 11 980 bp,平均长度达到 589 bp,这些 Unigenes 的长度如果以 100 bp 为间距进行分组,长度 200~299 bp 范围内的数目最多,达到 15 031 条,占 35.85%,以后随着长度的增加,区间内的 Unigenes 数目逐渐递减,超过 1 000 bp 的 Unigenes 条数为 6 231,占 14.86%,超过 2 000 bp 的 Unigenes 条数为 1 377 条,仅占 3.28%(图 1)。从组装 Unigenes 的原始序列数目来看,在 11~100 个的 Unigenes 数目最多,高达 19 330 条,占 46.10%,其次是 1~10 个的 Unigenes,数目为 7 805 条,占 18.61%;此后随着原始序列数目的提高,相应的 Unigenes 数目呈递减趋势,其中有 48 条 Unigenes 组装的原始序列超过 10 000 个(图 2)。

表 1 原始序列数据统计

Table 1 Summary of raw reads

原始序列总数 Number of total raw reads/条	接头序列数 Number of adaptor reads/条	低质量序列数 Number of low quality reads/条	高质量序列数 Number of High quality reads/条	总碱基数 Total number of bases /nt	GC 含量 GC content /%
8 989 210	3 328	134 646	8 851 236	1 106 404 500	47.10

表 2 Unigenes 统计

Table 2 Summary of Unigenes

Unigenes 数量 Number of Unigenes/条	总碱基数 Total bases /nt	GC 含量 GC content /%	Unigenes 长度范围 Length range of Unigenes/bp	Unigenes 平均长度 Average length of Unigenes/bp
41 929	24 733 400	44.9	201~11 980	589

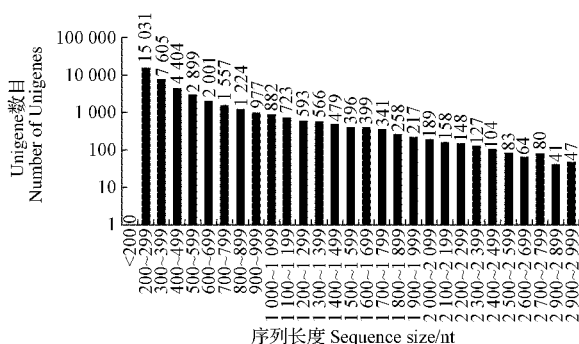
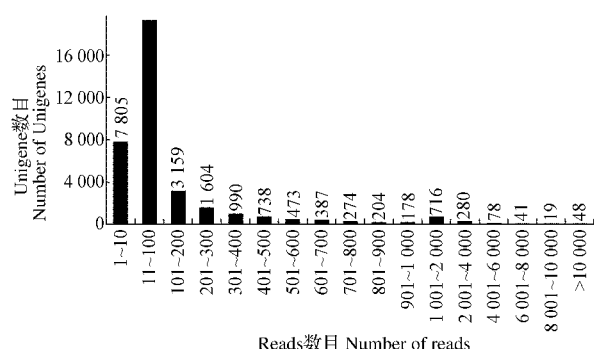


图 1 Unigenes 的长度分布式样

Fig. 1 Length distributing pattern of Unigenes

2.2 Unigenes 注释

2.2.1 GO 注释 GO 注释结果表明,Unigenes 共涉及到生物过程、细胞成分和分子功能三大类生物学功能,其中涉及生物学过程的又可以细分为代谢过



注: X轴表示组装 Unigene 的测序序列的数量, Y轴表示相应数量区间的 Unigene 数目。

图2 Unigene 基因的测序覆盖度

Fig. 2 Sequencing coverage of Unigenes

表3 基于 GO 注释的 Unigenes 功能类别及数量

Table 3 Functional categories and corresponding number of Unigenes according to GO annotation

生物过程	Unigene 数量	细胞成分	Unigene 数量	分子功能	Unigene 数量
Biological process	Number of Unigenes	Cell component	Number of Unigenes	Molecular function	Number of Unigenes
代谢过程	12 601	细胞	9 029	催化活性	11 993
细胞过程	12 073	细胞成分	9 026	结合	11 178
单一的生物过程	9 373	细胞器	7 189	转运活性	1 116
生物调节	4 185	膜	4 580	结构分子活性	453
应激反应	3 774	膜成分	3 166	核酸结合转录因子活性	392
定位	3 517	细胞器成分	2 969	分子转导活性	224
细胞组成的组织或合成	2 590	大分子复合物	1 973	分子功能调节器	177
发育过程	2 005	细胞连接	516	酶调节活性	147
多细胞生物过程	1 370	胞外区	236	抗氧化活性	95
信号	1 161	膜蛋白	182	蛋白质结合转录因子活性	74
生殖过程	1 017	病毒成分	94	鸟嘌呤核苷酸交换因子活性	30
多生物过程	616	病毒	94	电子载体活性	22
免疫系统的过程	327	胞外区成分	12		
生长	286	类核	11		
生物粘附	115	细胞外基质	10		
繁殖	111				
节律过程	57				
运动	28				
细胞的杀伤	4				

表4 基于 KOG 注释的 Unigenes 功能类别及相应数量

Table 4 Functional categories and corresponding number of Unigenes according to KOG annotation

序号	功能	Unigene 数量	序号	功能	Unigene 数量
Code	Function	Number of Unigenes/条	Code	Function	Number of Unigenes/条
A	RNA 加工和修饰	2 116	N	细胞运动	21
B	染色质结构和动态	495	O	蛋白质翻译后修饰、折叠、分子伴侣	3 649
C	能量生产和转化	1 199	P	无机离子转运和代谢	616
D	细胞周期调控, 细胞分裂, 染色体分区	747	Q	次生代谢物的生物合成, 运输及分解代谢	962
E	氨基酸的转运和代谢	852	R	常规功能	5 628
F	核苷酸的转运和代谢	277	S	未知功能	1 505
G	碳水化合物转运和代谢	1 312	T	信号转导机制	3 479
H	辅酶的转运和代谢	228	U	细胞内运输, 分泌和囊泡运输	1 806
I	脂质的转运和代谢	1 053	V	防御机制	210
J	翻译, 核糖体结构和生物合成	1 482	W	细胞外结构	80
K	转录	1 805	Y	核结构	152
L	复制, 重组和修复	790	Z	细胞骨架	647
M	细胞壁/膜/胞外被膜	386			

程、细胞过程、单一的生物学过程等 19 类更为具体的功能。涉及细胞成分又可分为细胞、细胞成分、细胞器等 15 类更为具体的功能; 分子功能又分为催化活性、结合、转运活性等 12 类更为具体的功能。涉及各种功能的 Unigene 数目不等, 最多的是代谢过程, 达 12 601 条, 其次是细胞过程, 达 12 073 条, 分别约占总 Unigene 数目的 30% (表 3)。

2.2.2 KOG 注释 从表 4 可知, KOG 注释到 19 149 条 Unigenes, 占总 Unigenes 数量的 45.67%。这些注释到的 Unigenes 分别涉及 25 种主要代谢功能。除了 5 628 条涉及不具体的常规功能, 涉及信号转导机制蛋白质翻译后修饰、折叠, 分子伴侣的 Unigenes 数量最多, 为 3 649 条, 占注释 Unigenes 的

19.06%,以后依次是信号转导机制的3 479条、RNA加工和修饰的2 116条。涉及细胞外结构的Unigenes数量最少,仅80条。

2.2.3 KEGG注释 从表5可知,KEGG数据库共注释到13 072条Unigenes,占总Unigenes数量的31.18%。涉及5个大的功能类别(A类),18个中等功能类别(B类),129条具体的代谢通路。5个A类分别是:1)细胞过程。只包含运输与代谢这个B类,共4条具体的代谢通路;2)环境信息处理。包括膜转运和信号转导2个B类,共3条具体的代谢途径;3)遗传信息处理。包含转录、折叠、分拣和降解、复制和修复3个中类,共21条具体的代谢途径;4)代谢类。包括碳水化合物代谢、全球概况、氨基酸代谢、脂质代谢、能量代谢、核苷酸代谢、其它次生代谢产物的生物合成、辅助因子和维生素代谢、其它氨基酸的代谢、糖链的生物合成与代谢、萜类和酮类化合物的代谢共11个B类共99条具体代谢通路;5)生物系统。包括环境适应这一B类,只涉及到2条代谢途径。涉及各个代谢途径的Unigene数量从几十到上千不等。

表5 KEGG注释的Unigenes功能类别及相应数量

Table 5 Functional categories and corresponding number of Unigenes according to KEGG annotation

KEGG A类 KEGG_A_class	KEGG B类 KEGG_B_class	通路 Pathway	Unigene 数量 Number of Unigenes/条
细胞过程	运输和代谢	4	658
环境信息处理	信号转导	2	358
	膜转运	1	71
遗传信息处理	转录	8	1 811
	折叠、分拣和降解	7	1 043
	复制和修复	6	404
代谢	碳水化合物代谢	15	1 430
	全球概况	5	982
	氨基酸代谢	14	880
	脂质代谢	14	720
	能量代谢	6	485
	核苷酸代谢	2	386
	其它次生代谢产物的生物合成	12	355
	辅助因子和维生素代谢	11	351
	其它氨基酸的代谢	5	294
	糖链的生物合成与代谢	7	257
	萜类和酮类化合物的代谢	8	238
生物系统	环境适应	2	404

2.2.4 Nr注释 从表6可知,在共组装的41 929条Unigenes中,30 738条能够被Nr数据库成功注释,它们共对应到1 960条蛋白质编码序列(coding sequence,CDS)。对应蛋白质编码序列长度以100 bp

梯度为一个区间,分布在200~299 bp的数目最多,达到1 163条,长度越长,相应区间的蛋白质编码序列数量呈下降趋势,超过1 000 bp的仅11条,占5.05%,没有检测到超过1 400 bp的蛋白质编码序列。

表6 Nr数据库注释的Unigenes对应长度区间的CDS数量

Table 6 Unigenes annotated by Nr dataset and the number of corresponding CDS in certain length range

长度区间 Length range /bp	数量 Count /条	长度区间 Length range /bp	数量 Count /条	长度区间 Length range /bp	数量 Count /条
<200	81	600~699	41	1 100~1 199	3
200~299	1 163	700~799	12	1 200~1 299	4
300~399	404	800~899	18	1 300~1 399	2
400~499	149	900~999	6	>1 400	0
500~599	75	1 000~1 099	2		

2.3 特异Unigene挖掘

2.3.1 转录因子 从表7可知,该研究获得的在41 929条Unigenes中,884条具有转录因子特征序列,共涉及55个家族,其中包含Unigenes数量较多的10类转录因子家族分别为GRAS、MYB_related、ERF、bHLH、WRKY、C3H、NAC、Nin-like、FAR1和C2H2,它们占全部可能的转录因子Unigene的4.19%~7.01%。

2.3.2 抗性基因 从表8可知,在该研究获得的41 929条Unigenes中,2 261条具有抗性基因序列特征,共涉及18个家族,其中RLP、N、NL、TNL、CNL、RLK及PLK-GNK2家族最多,分别占转录因子Unigene总数的34.94%、15.75%、14.64%、11.10%、9.86%、9.24%及5.13%。

2.3.3 EST-SSR 从表9可知,该研究获得的41 929条Unigenes中,包含6 591条简单重复序列(single simple sequence repeat,SSR)位点,其中基本序列重复6次的Unigenes数量最多,为1 538条,占总数的23.33%,其次分别是基本序列重复7次和5次的,分别为992条和962条,各占总数的15.11%和14.60%。含有SSR的Unigenes中,基本序列为2碱基的最多,其余依次是3碱基、4碱基、6碱基和5碱基,分别为4 614、1 692、128、93和64条,各占70.00%、25.67%、1.94%、1.41%和0.97%。在2碱基重复的SSR序列中,又以AG/CT为最高,占2碱基重复的95%,占总数的66.5%,其余基本序列类型均少于10%,AAG/CTT重复的序列也较多,达到8.2%(图3)。

表 7 转录因子家族及相应的 Unigene 数量

Table 7 Families of transcript factor and corresponding number of Unigene 条									
家族名称 Family name	Unigene 数量 Number of Unigenes	家族名称 Family name	Unigene 数量 Number of Unigenes	家族名称 Family name	Unigene 数量 Number of Unigenes	家族名称 Family name	Unigene 数量 Number of Unigenes	家族名称 Family name	Unigene 数量 Number of Unigenes
GRAS	62	bZIP	30	TALE	12	M-type	5	EIL	3
MYB_related	59	Tribelix	26	NF-YC	11	DBB	5	RAV	2
ERF	55	G2-like	26	HB-other	11	CAMTA	5	NF-X1	2
bHLH	54	Dof	23	LSD	9	AP2	4	E2F/DP	2
WRKY	46	HD-ZIP	22	CO-like	8	Whirly	4	WOX	2
C3H	46	GATA	21	ARF	7	VOZ	3	HRT-like	1
NAC	43	TCP	16	NF-YA	6	GeBP	3	SlFa-like	1
Nin-like	43	HSF	15	BBR-BPC	6	YABBY	3	HB-PHD	1
FAR1	39	B3	15	BES1	6	ARR-B	3	SRS	1
C2H2	37	MIKC	12	NF-YB	6	CPP	3	EIL	3
MYB	33	SBP	12	LBD	6	ZF-HD	3	RAV	2

表 8 抗性基因类别及相应的 Unigene 数量

Table 8 Categories of R-gene and corresponding number of Unigene 条					
家族名称 Family name	Unigene 数量 Number of Unigenes	家族名称 Family name	Unigene 数量 Number of Unigenes	家族名称 Family name	Unigene 数量 Number of Unigenes
RLP	790	RLK-GN2	116	L	13
N	356	CN	61	RPW8-NL	6
NL	331	Other	48	RLK-Pto-like	4
TNL	251	T	42	RLP-Malectin	3
CNL	223	Pto-like	30	TN	2
RLK	209	Mlo-like	25	RLP-Malectina	2

表 9 EST-SSR 类型及相应的 Unigenes 数量

Table 9 Categories of EST-SSR and corresponding number of Unigenes 条						
基序重复次数 Number of repeat unit	重复基序碱基数量 Base number of repeat unit					总数 Total
	2 碱基 Di-	3 碱基 Tri-	4 碱基 Tetra-	5 碱基 Penta-	6 碱基 Hexa-	
4	0	0	103	41	77	221
5	0	921	19	18	4	962
6	1 116	413	3	3	3	1 538
7	779	208	1	1	3	992
8	714	74	0	0	4	792
9	638	28	0	1	0	667
10	510	11	1	0	1	523
11	256	6	0	0	1	263
12	110	7	0	0	0	117
13	39	5	0	0	0	44
14	5	5	1	0	0	11
≥15	447	14	0	0	0	461
总计 Total	4 614	1 692	128	64	93	6 591

3 讨论

近年来,国内外对矮丛蓝莓的研究日渐增加,在国内,主要集中在组培快繁^[15-17]、栽培技术^[18]和采收等方面^[19]。在国际上,研究内容包括组培^[20]、采后生理^[21]、药用价值^[22]、生物学^[23]、栽培学^[24]、遗传多样性^[25-27]等。随着基因组测序技术和生物信息学的

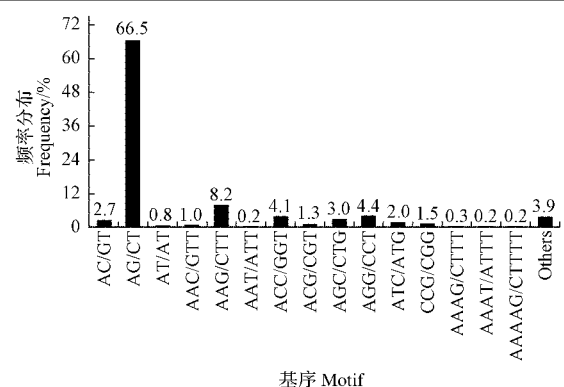


图 3 EST-SSR 的基序类型及相应的比例

Fig. 3 Motif kinds of EST-SSR and corresponding percentage 发展进步,果树研究逐渐步入基因组时代,利用基因组测序来解释果树生长发育分子机理等方面已经成为研究热点^[28-31]。蓝莓研究也已经进入基因组时代,相关的研究甚多,特别是揭示蓝莓成熟耐冷机理方面,及果实发育成熟分子机理方面,但研究对象主要是高丛蓝莓^[32]。对矮丛蓝莓的相关基因学研究不多。该研究对矮丛代表性品种的叶片转录组进行测序和生物信息学分析,将为矮丛蓝莓生长发育、代谢生理等的分子机理研究及分子育种等提供必要的基础信息。

参考文献

- [1] PRIOR R L, CAO G, MATIN A, et al. Antioxidant capacity as influenced by total phenolic and anthocyanin content, maturity, and variety of *Vaccinium* species [J]. Journal of Agricultural and Food Chemistry, 1998, 46: 2686-2693.
- [2] KALT W, JOSEPH J A, SHUKITT-HALE B. Blueberries and human health; a review of current research [J]. Journal of the American Chemical Society, 2007, 61: 151-160.
- [3] 李丽敏, 赵春雷, 郝庆升. 中外蓝莓产业比较研究 [J]. 中国农学通报, 2010, 26(23): 354-359.
- [4] 陈宏毅. 蓝莓的种群划分及品种特性研究 [J]. 现代农业科技,

2011(22):154-158.

- [5] WU X L, BEECHER G R, HOLDEN J M, et al. Concentrations of anthocyanins in common foods in the United States and estimation of normal consumption[J]. Journal of Agricultural and Food Chemistry, 2006, 54(11):4069-4075.
- [6] POSSEMIERS S, BOLCA S, VERSTRAETE W, et al. The intestinal microbiome: A separate organ inside the body with the metabolic potential to influence the bioactivity of botanicals[J]. Fitoterapia, 2011, 82:53-66.
- [7] WOLFE K L, LIU R H. Structure-activity relationships of flavonoids in the cellular antioxidant activity assay[J]. Journal of Agricultural and Food Chemistry, 2008, 56(18):8404-8411.
- [8] CANI P D, DELZENNE N M. The gut microbiome as therapeutic target[J]. Pharmacology and Therapeutics, 2011, 130:202-212.
- [9] CONERNO L, FAVA F, VIOLA R, et al. Obesity and the gut microbiota: does up-regulating colonic fermentation protect against obesity and metabolic disease?[J]. Genes and Nutrition, 2011, 6:241-260.
- [10] VENDRAME S, DAUGHERTY A, KRISTO AS, et al. Wild blueberry (*Vaccinium angustifolium*) consumption improves inflammatory status in the obese Zucker rat model of the metabolic syndrome[J]. Journal of Nutritional Biochemistry, 2013, 24(8):1508-1512.
- [11] VENDRAME S, GUGLIELMETTI S, RISO P, et al. Six week consumption of a wild blueberry powder drink increases Bifidobacteria in the human gut[J]. Journal of Agricultural and Food Chemistry, 2011, 59:12815-12820.
- [12] DUYNHOVEN J, VAUGHAN E E, JACOBS D M, et al. Metabolic fate of polyphenols in the human superorganism[J]. Proceedings of the National Academy of Sciences USA, 2011, 108:4531-4538.
- [13] SERAFINI M, PELUSO I, RAGUZZINI A. Antioxidants and the immune system flavonoids as anti-inflammatory agents[J]. Proceedings of the Nutrition Society, 2010, 69:273-278.
- [14] 张志东, 李亚东, 吴林, 等. 适宜高寒山区栽培的越橘优良品种-美登[J]. 园艺学报, 1999, 26(2):135.
- [15] 孙晓梅, 王新苗, 杨宏光, 等. 矮丛蓝莓北极星启动培养研究[J]. 北方园艺, 2010(1):23-26.
- [16] 韩婷婷, 孙周平, 吴媛媛. 矮丛蓝莓茎段再生植株研究[J]. 中国农学通报, 2010, 26(5):164-168.
- [17] 韩婷婷, 孙周平. 矮丛蓝莓叶片的愈伤组织诱导及植株再生[J]. 西北植物学报, 2010, 30(3):615-620.
- [18] 张含生, 张健全, 杨春梅, 等. 伊春地区矮丛蓝莓越冬防寒方法比较试验[J]. 黑龙江农业科学, 2013(12):76-78.
- [19] 郭艳玲, 鲍玉冬, 何培庄, 等. 手推式矮丛蓝莓采摘机设计与试验[J]. 农业工程学报, 2012, 28(7):40-45.
- [20] DEBNATH S C. Adventitious shoot regeneration in a bioreactor system and EST-PCR based clonal fidelity in lowbush blueberry (*Vaccinium angustifolium* Ait.)[J]. Scientia Horticulturae, 2011, 128:124-130.
- [21] FRACASSETTI D, DELBO C, SIMONETTI P, et al. Effect of time and storage temperature on anthocyanin decay and antioxidant activity in wild blueberry (*Vaccinium angustifolium*) powder[J]. Journal of Agricultural and Food Chemistry, 2013, 61:2999-3005.
- [22] MCGOVERN K B, ANNIS S L, YARBOROUGH D E. Efficacy of organically acceptable materials for control of mummy berry disease on lowbush blueberries in maine[J]. International Journal of Fruit Science, 2012, 12:188-204.
- [23] BELL D J, DRUMMOND F A, ROWLAND L J. Evidence of functional gender polymorphisms in a population of the hermaphroditic lowbush blueberry (*Vaccinium angustifolium*) [J]. Botany, 2012, 90:393-399.
- [24] BOYD N S, WHITE S, RAO K, et al. Inputs for winter bentgrass- (*Agrostis hyemalis*) infested lowbush blueberry fields[J]. Weed Technology, 2014, 28:527-534.
- [25] BELL D J, ROWLAND L J, ZHANG D, et al. Spatial genetic structure of lowbush blueberry, *Vaccinium angustifolium*, in four fields in Maine[J]. Botany, 2009, 87:932-946.
- [26] BELL D J, ROWLAND L J, STOMMEL J, et al. Yield variation among clones of lowbush blueberry as a function of genetic similarity and self-compatibility[J]. Journal of the American Society for Horticultural Science, 2010, 135:259-270.
- [27] BELL D J, DRUMMOND F A, ROWLAND L J. Fine-scale spatial genetic structure associated with *Vaccinium angustifolium* aiton (Ericaceae)[J]. International Journal of Modern Botany, 2012, 2(4):72-82.
- [28] 乔鑫, 李梦, 殷豪, 等. 果树全基因组测序研究进展[J]. 园艺学报, 2014, 41(1):165-177.
- [29] XU Q, LIU C Y, BISWAS M K, et al. Recent advances in fruit crop genomics[J]. Frontiers of Agricultural Science and Engineering, 2014(1):21-27.
- [30] 冯超, 朱长青, 徐昌杰, 等. RNA-Seq 在果树学研究中的应用[J]. 果树学报, 2014, 31(1):115-124.
- [31] 王云生, 聂飞, 林顺权. 果树高通量测序的最新研究进展[J]. 基因组学与应用生物学, 2015, 34(9):2034-2043.
- [32] ROWLAND L J, BELL D, ALKHAROUF N, et al. Generating genomic tools for blueberry improvement[J]. International Journal of Fruit Science, 2012, 12:276-287.

Leaf Transcriptome Sequencing and Bioinformatics Analysis of Lowbush Blueberry Variety ‘Blomidon’

WANG Yunsheng¹, WEN Guangqin², NIE Fei³

(1. School of Environment and Life Science, Kaili University, Kaili, Guizhou 556011; 2. Botanic Institute of Guizhou Province, Guiyang, Guizhou 550004; 3. Biological Institute of Guizhou Province, Guiyang, Guizhou 550000)

Abstract: In the hope of increasing the understanding of molecular mechanism of the growth and development of *Vaccinium angustifolium*, and establishing the foundation for the molecular marker assisted breeding of dwarf

黑刺菝葜中 6-酮型甾体皂苷含量测定方法

祁雅雯¹, 王 荣¹, 赵秋琳², 张存莉¹

(1. 西北农林科技大学 生命科技学院, 陕西 杨凌 712100; 2. 沂南县检验检测中心, 山东 临沂 276300)

摘 要:以黑刺菝葜为试材,应用紫外可见分光光度法,研究测定其 6-酮型甾体皂苷类化合物含量的方法;以 LG-C 为标准品,依据显色剂对 LG-C 反应的专属性和灵敏度,确定紫外可见分光光度法的显色体系和最大吸收波长;在单因素试验的基础上,通过星点设计-响应面法,确定最适显色反应条件。结果表明:紫外可见分光光度法的显色体系为茴香醛-硫酸法Ⅱ,最大吸收波长为 429 nm;最佳显色温度 82.4 ℃,反应时间 24 min,显色剂用量 1.2 mL。优化反应条件后,在 7.51~50.90 μg 范围内,LG-C 与吸光度值呈良好线性关系, $r=0.996\ 9$;平均加样回收率为 103.37%,RSD 2.2%。该方法灵敏度高、专属性强,重复性好,可以作为黑刺菝葜中 6-酮型甾体皂苷含量的测定方法。

关键词:黑刺菝葜;6-酮型甾体皂苷;紫外可见分光光度法;星点设计-响应面法

中图分类号:R 282.71 **文献标识码:**A **文章编号:**1001-0009(2017)01-0107-07

黑刺菝葜(*Smilax scobinicaulis* C. H. Wright)属百合科菝葜属药用植物,通常以其根茎入药,具有除湿、活血、解毒、镇惊息风、抗癌等功效^[1],其主要

第一作者简介:祁雅雯(1991-),女,硕士研究生,研究方向为天然产物化学及其资源利用。E-mail: yawen_qi@nwsuaf.edu.cn.

责任作者:张存莉(1967-),女,博士,教授,博士生导师,现主要从事天然产物化学和植物资源利用的教学与科研工作。E-mail: cunli_zhang@nwsuaf.edu.cn.

基金项目:国家 863 项目子课题资助项目(2013AA102902-8)。

收稿日期:2016-08-05

的化学成分是拉肖皂苷元((25R)-3 β -羟基-5 α -螺甾-6-酮,La)及其糖苷 A(拉肖皂苷元-3-O- β -D-吡喃葡萄糖基, LG-A)、B(拉肖皂苷元-3-O- $[\beta$ -D-吡喃葡萄糖基-(1 \rightarrow 6)]- α -L-吡喃阿拉伯糖基, LG-B)和 C(拉肖皂苷元-3-O- $[\beta$ -D-吡喃葡萄糖基-(1 \rightarrow 4)-O- $[\alpha$ -L-吡喃阿拉伯糖基-(1 \rightarrow 6)]- β -D-吡喃葡萄糖基, LG-C)等甾体皂苷类化合物,其中 LG-C 含量最高^[2-3]。这些皂苷具有强烈的抑制癌细胞^[4]、磷酸二酯酶^[5]和血小板聚集的活性^[6]。La 和 LG-C 同时还表现出甾体酮型油菜甾醇类植物激素的生物活性,具有显著地促进种子萌发和幼苗生长等作用^[7-9]。为了进一步研究

blueberry, cultivar 'Blomidon' of dwarf blueberry was selected as test material, and its young leave transcriptome was sequenced by the second sequencing technology and then was made bioinformatics analysis, the aim was to study the functional genomic information of lowbush blueberry. The results showed that, total about 1 G clean reads could be assembled into 41 929 Unigenes. Among them, Unigenes were involving 45 kinds metabolism and physiological function of biological process, cellular component and molecular function according to the annotation results of dataset GO; 19 149 Unigenes could be annotated by KOG dataset, which were involving in 25 metabolism pathways; 13 072 Unigenes could be annotated by KEGG dataset, which were involving in five major categories, 18 middle categories, and 129 metabolism pathways; 30 738 Unigenes could be annotated by dataset Nr, which were mapping into 32 949 code regions of protein. In the meanwhile, bioinformatics analysis showed that 884 Unigenes were the transcript factor candidate of 55 families, and 2 261 Unigenes were R gene candidate of 18 families, total of 6 591 SSR loci were detected, of which 2 base repeat SSR accounted for about 70%.

Keywords: bioinformatics; lowbush blueberry; transcriptome; high-throughout sequencing