INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
19–0105/2019/22–4–827–834
DOI: 10.17957/IJAB/15.1137
http://www.fspublishers.org

Full Length Article

Identification of SSR Markers and Putative Genes Associated with Chlorogenic Acid in *Vaccinium uliginosum* through Transcriptome Analysis

Yunsheng Wang[1*†], Fozia Ghouri[2†], Aysha Kiran[3], Muhammad Qasim Shahid[2], Guangqin Wen[4] and Fei Nie[5*]
[1]*College of Life and Health Science, Kaili University, Kaili City, Guizhou Province, 556011, China*
[2]*College of Agriculture, South China Agricultural University, Guangzhou, Guangdong Province, 510642, China*
[3]*Department of Botany, University of Agriculture, Faisalabad, Pakistan*
[4]*Guizhou Botanical Garden, Guiyang City, Guizhou Province, 556000, China*
[5]*Biological institute of Guizhou Province, Guiyang City, Guizhou Province, 556000, China*
[*]For correspondence: wys3269@126.com; 1134477929@qq.com
[†]These authors contributed equally to this work

Abstract

Bog bilberry (*Vaccinium uliginosum*) is a perennial woody deciduous member of the *Vaccinium* genus, subfamily Vacciniaceae, family Ericaceae. *V. uliginosum* has been domesticated as a promising fruit crop for significant healthful functions from the last few decades. In the present study, more than 2G clean data was obtained from the leaf transcriptome of *V. uliginosum*, which was sequenced by using second generation high-throughput technology, and assembled into 43,507 unigenes with an average length of 588 bp. Of these, 30,689 (70.538%) unigenes were annotated by at least one of Nr, Swiss-Port, KOG or KEGG databases, and 30,641, 21,518, 19,273 and 13,334 unigenes had a significant hit ($p<1E-5$) in these databases, respectively. In addition, 2,339 unigenes were identified as candidate resistance genes, which belong to 19 families. A total of 9,680 SSR loci from 7,780 unigenes were identified with AG/CT repeat motifs accounted for 72.19%. But most importantly, eight candidate unigenes that encoded five key enzymes and responsible for the bio-synthesis of chlorogenic acid were identified, which is an important bioactive component with high medicinal function. The molecular markers developed in this study would be helpful for further studies about the molecular genetics, molecular ecology and marker assisted breeding of *V. uliginosum*. The putative resistance genes and chlorogenic acid identified in this study would help to deepen the understanding on molecular mechanism of ecological adaptation and medicinal functions of *V. uliginosum*. © 2019 Friends Science Publishers

Key words: Blueberry; Bioinformatics; Functional annotation; Transcriptome; Unigenes

Introduction

Vaccinium uliginosum, commonly called bog bilberry or blueberry, is a member of family Ericacea. It is a deciduous shrub (Kähkönen *et al.*, 2001). Its typical habitat is open biotopes like wet heaths and bogs in circumboreal regions and it tends to grow on acidic, poorly drained and wet soils (Jacquemart, 1992). Compared to its two relatives, bilberry and blueberries, *V. uliginosum* is less utilized in commercial berry products, but recently great interest developed worldwide for their high anthocyanin content as well as flavonols, especially myricetin and quercetin (Latti *et al.*, 2010; Li *et al.*, 2011). The fruit of *V. uliginosum* has been used as edible berries, and added into juices, jams, pies, jelly, and wine. *V. uliginosum* has been used as folk medicinal materials for centuries in some circumboreal countries (Stefko *et al.,* 2014; Wang *et al.,* 2014). Anthocyanin in *V. uliginosum* have many health benefits, such as anti-memory decline effects (Krikorian *et al.,* 2010), anti-light damage to the retina (Yin *et al.,* 2012) antioxidant activity (Kähkönen *et al.*, 2001) and anticancer, anti-vascular disease effects (Vendrame *et al.*, 2014). Li *et al.* (2011) reported these berries as a good source of food colorants. Leaves of *V. uliginosum* contain essential oils, hyperoside (also known as quercetin-3-O-galactoside) (Wang *et al.,* 2008; Chen *et al.,* 2015), and chlorogenic acid (also known as 3-O-caffeoylquinic acid, Chlorogenate, Caffeoyl quinic acid, trans-5-O-Caffeoyl-D-quinate), and these three bioactive compounds are very beneficial to the human health (Umezu, 2012; Zhang *et al.,* 2014). Among them, the biosynthetic pathways of chlorogenic acid have been revealed, and three metabolic pathways that could generate chlorogenic acid have been found (Campa *et al.,* 2003).

To cite this paper: Wang, Y., F. Ghouri, A. Kiran, M.Q. Shahid, G. Wen and F. Nie, 2019. Identification of SSR markers and putative genes associated with chlorogenic acid in *Vaccinium uliginosum* through transcriptome analysis. *Intl. J. Agric. Biol*., 22: 827−834

V. uliginosum has been cultivated for about two decades, and the studies on *V. uliginosum* have increased during that time, including ecology (Auffret *et al.*,, 2010), population genetics (Albert *et al.,* 2005), botany (Rossum *et al.,* 2012; Stephens *et al.,* 2012), breeding (Lyrene and Olmstead, 2012), molecular biology (Primetta *et al.,* 2015), physiology (Boesgaard *et al.,* 2012), phyto-chemistry (Kellogg *et al.,* 2010; Su *et al.,* 2016), and pharmacology or pharmaceutics (Han *et al.,* 2012; Kim and Choung, 2014; Park *et al.,* 2016; Yoon *et al.,* 2016). However, there is no genomic information on *V. uliginosum*, which could delay the study of molecular biology or genetics of *V. uliginosum*.

Transcriptome sequencing has been widely used for various crop plants to detect molecular markers at large scale. Molecular markers are essential tools of modern agriculture to conduct studies on various aspects of plants (Varshney *et al.,* 2005; Zheng *et al.,* 2017; Nadeem *et al.,* 2018). This is the first report on the leaf transcriptome of *V. uliginosum* by next generation high-throughput sequencing technology. Our analysis result would not only provide a lot of information about functional genes to deepen the understanding about the medicinal functions of *V. uliginosum*, the SSR marker we identified from transcriptome data would also offer more molecular tools for further studies on the molecular genetics, molecular ecology and marker assisted breeding of *V. uliginosum*.

Materials and Methods

Plant Material and Total RNA Extraction

A fully grown plant of *V. uliginosum* (two years old with about 0.8 m high), collected from Greater Khingan Range and planted at the research center of blueberry engineering technology, Majiang county, Guizhou province, China. Fresh leaves of *V. uliginosum* were used for total RNA isolation from by classic TRIzol method. RNA was quantified by using Bio-analyzer 2100 (Agilent Technologies, Santa Clara C.A.), and quality was checked on agarose gel electrophoresis.

Construction of cDNA Library and Sequencing

To construct the cDNA sequencing library, the following protocol was used. Sera-mag Magnetic Oligo (dT) Beads (Illumina) were used to purify the poly (A) RNA from 20 mg total RNA. Purified mRNA was treated with fragmentation buffer. Random hexamer primers were used for cDNA synthesis. cDNA was synthesized by using QIA rapid PCR kit (QIAGEN). The cDNA was amplified and sent for sequencing to Illumina sequencing platform (Illumina HiSeq 2000, Guangzhou, China). Average length of sequencing read was 2*125 bp. Whole set of raw sequencing reads was deposited into the Short Read Archive (SRA) database of the National Center for Biotechnology Information (NCBI) [accession number SRP157131].

Data Filtering and *De Novo* Assembly

Low-quality reads (reads with Q-value ≤ 20 containing more than 50% bases) were discarded and adaptors were also trimmed. De-novo assembly of remaining high-quality sequences into unigenes was done by using the short reads assembling program Trinity (https://github.com/trinityrnaseq) with the default parameters.

Functional Annotation

Assembled unigenes were searched for functional annotation against the non-redundant protein database (Nr, http://www.ncbi.nlm.nih.gov/) by using BLASTx and the Swiss-Port database (http://www.expasy.ch/sprot), with a cut off E-value of 10^{-5}. Gene ontology (GO) enrichment analysis by Blast2GO (Conesa *et al.,* 2005) used to classify the unigenes into functional categories and WEGO program (Ye *et al.,* 2006) was used to generate GO trees. Then the unigenes were queried against the STRING database v. 9.05 (http://string-db.org) for forecasting about clusters of orthologous groups (COGs) and possible functions using BLASTx with a cut off E-value of 10^{-10}. Finally, complex biological functions and the metabolic pathways associated with the unigenes were searched against Kyoto Encyclopedia of Genes and Genomes database (KEGG, http://www.genome.jp/kegg) (Kanehisa and Goto, 2000)

Detection of SSR Markers and Primer Design

SSR markers were identify by using MISA program (http://pgrc.ipk-gatersleben.de /misa/) among all assembled genes with the following parameters: motifs ranging from the minimum number of repeat units (6 for di-, 5 for tri-, and 4 for tetra-, penta- and hexa-nucleotides), and the maximum interruption length between two SSRs of 100 bases. Primer 3 (http:// primer3.ut.ee/) were used to design SSR primers. GC contents ranged from 40% to 60%, and the expected PCR product sizes ranging from 100 to 280 bp.

Detection of Candidate Genes Encoding Enzymes Involved in Biosynthesis of Chlorogenic Acid

Physiological pathways for chlorogenic acid was accessed at KEGG database which was specifically annotated to *V. uliginosum* transcriptome. Biosynthetic route map concluded by Campa *et al.* (2003) and Lepelley *et al.* (2007) were followed for the detection of homologous unigenes. Corresponding unigenes were confirmed by blast search at NCBI database.

Results

Sequencing and *De Novo* Assembly

Illumina HiSeq produced 20,356,360 raw reads, from

Table 1: Summary of RNA-sequencing of *V. uliginosum*

	Total reads	Total reads nt	Max/min length	Mean length	Q20 (percentage)	Q30 (percentage)	GC (percentage)
Raw data	20356908	2544613500	125/125	125	2440185074 (95.90%)	2349684086 (92.34%)	1188117625 (46.69%)
Clean data	20214388	2508184666	125/50	123	2428839064 (96.84%)	2341476440 (93.35%)	1170793118 (46.68%)

Table 2: Assembled results of the leaf transcriptome of *V. uliginosum*

Total number (>200 bp)	Total number (>500 bp)	Total number (>1kp)	Total number (>2kp)	Total number (3kp)	GC ratio	N50 number	Total size	Mean length
43507	16406	6657	1153	206	44.2273	9198	25615894	588

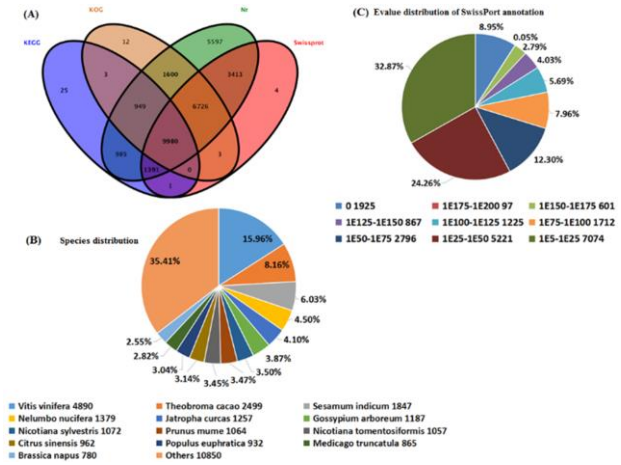

Fig. 1 (A): Venn diagram of BLAST hits for unigenes against KOG, Nr, Swiss-Port and KEGG database (E-value≤1.0e-05). Numbers in the circles indicate the number of unigenes annotated by single or multiple databases, (**B**) Species distribution of the top BLAST hits for the assembled unigenes (E-value of 1.0e-05). (**C**) E-value distribution of BLAST hits Swiss-Port dataset for each unique sequence (E-value ≤1.0e-05)

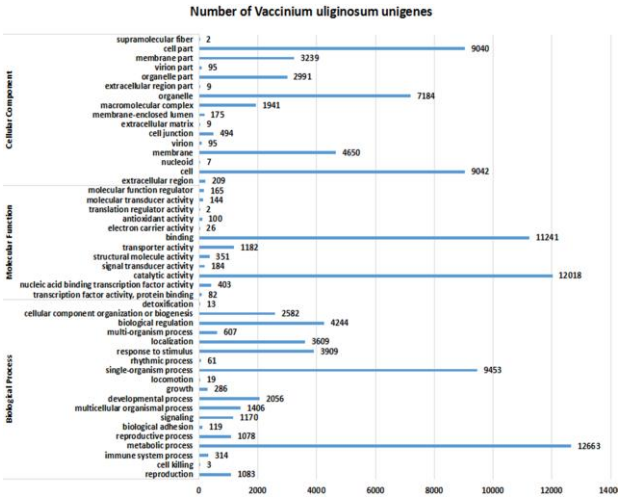

Fig. 2: GO functional classification of *V. uliginosum* Unigenes

which 20,214,388 high quality sequences were obtained after filtration of low quality 141,972 reads with adaptors, accounted for 0.70% of total raw data (Table 1). Average length of 123 bp was found in clean reads (total nucleotides: 2,508,184,666 bp), 96.84% bases reached to the Q20 quality standard and 93.35% bases reached to the Q30 quality standard. The GC content of total clean data was 46.68% (Table 1).

High quality sequences were assembled into 43507 unigenes with a length more than 200bp. Among them, 27,101 (62.29%) were 200–500 bp in length, 16,406 (37.71%) unigenes were more than 500 bp, 6,657 (15.30%) unigenes were more than 1000 bp while, 1153 (2.65%) unigenes were more than 2000 bp and only 206 (0.47%) unigenes were more than 3000 bp long. The size of N50 was 807 bp and its number was 9189. The total nucleotides of whole assembled unigenes were 25,615,894, and the average length of unigenes was 588 bp (Table 2).

Functional Annotation

Among 43,507 assembled unigenes, there were 30,641 (70.43%) unigenes which were showed significant similarity to the known protein in Nr data base, while and 21,518 (49.46%) unigenes showed significant similarity in Swiss-Port datasets. Furthermore, 19,273 (44.30%) and 13,334 (30.65%) unigenes could be annotated by the KOG and KEGG databases for *V. uliginosum*, respectively (Fig. 1A). A total of 30,689 unigenes accounted for 70.54% of the total unigenes that were annotated by at least one of the above datasets and the remaining 12,818 accounted for 29.46% were not annotated to above datasets (Fig. 1A). The number of known proteins showed significant similarities with *V. uliginosum* unigenes, and the 10 top-hit species with unigenes hit were *Vitis vinifera* (4890, 15.96%), *Theobroma cacao* (2499, 8.16%), *Sesamum indicum* (1847, 6.03%), *Nelumbo nucifera* (1379, 4.50%), *Jatropha curcas* (1257, 4.10%), *Gossypium arboretum* (1187, 3.87%), *Nicotiana sylvestris* (1072, 3.50%), *Prunus mume* (1064, 3.47%), *N. tomentosiformis* (1057, 3.45%) and *Citrus sinensis* (962, 3.14%) (Fig. 1B). 42.87% annotated sequences with high identity with their best hits (smaller than 1e-50) were found in the Swiss-Port database when compared in the E-value distribution, whereas 24.26% ranged from 1e-25 to 1e-50 and another 32.87% ranged from 1e-25 to 1e-5 (Fig. 1C).

GO classification assigned 109,755 GO terms to 21,782 unigenes with BLAST hits to annotated proteins in the Nr database (Fig. 2). These GO terms were classified into three main GO categories and 47 sub-categories by WEGO. 19 sub-categories belong to biological process, 12 sub-categories belong to molecular function and 16 sub-categories belong to cellular component, indicating a diverse range of functional genes in the *V. uliginosum*. Among sub-categories belonging to biological process, the unigenes were more frequent in metabolic process and single-organism process and the numbers of corresponding GO terms were 12,663 and 9,453. Among the sub-categories belonging to molecular function, the unigenes were most frequent in the catalytic activity and binding categories, and the number of corresponding GO terms reached to 12,018 and 11,241, respectively. Cell part and cell were more frequent sub-categories in the cellular component category, and the numbers of GO terms were 9,042 and 9,040, respectively (Fig. 2).

Subsequently, 19,273 unigenes were aligned to the appropriate COG clusters and 31,461 functional annotations were obtained and grouped into 25 functional categories, while some unigenes exhibited multiple COG functional annotations. Among these categories, the number of unigenes involving general functional category were the most frequent, followed by posttranslational modification, protein turnover, chaperones and signal transduction mechanisms (Fig. 3).

Finally, to represent active biological pathways in *V. uliginosum*, 30,641 annotated unigenes were aligned to the reference plant pathways in the KEGG database. Among them, 13,334 unigenes were assigned to 129 KEGG pathways, illustrating the broad overview of the groups of genes located in transcriptome. The top five pathways with most unigenes were "Carbon metabolism" (427 members), "RNA transport" (364 members), "Spliceosome" (361 member), "Plant-pathogen interaction" (339 members) and "Biosynthesis of amino acids" (572 members). All KEGG annotation information of unigenes is presented in Table 3.

Resistance gene (R-gene) is the key molecular mechanism that plants adapt to cope with the environment. In the current study, we identified 2,339 unigenes as candidate R-genes belonging to 19 gene families. Among them, 731 unigenes belong to the RLP family, which accounted for 21.56% of total candidate R-gene, followed by N, NL and TNL families, which accounted for 15.86, 13.25 and 9.53% respectively, and the rest 15 families accounted for 30.10% of total candidate R-gene unigenes (Table 4).

EST-SSR Kinds and Number

SSRs are one of the most popular molecular markers, which widely used in the life science studies such as molecular genetics, molecular ecology, association study and marker assisted breeding. From the *V. uliginosum* leaf transcriptome, we identified 9680 SSR sequence segments from 7780 unigenes. Of these SSR, two, three, four, five and six base repeat types of SSRs were 7423, 1907, 147, 76 and 127 that accounted for 76.68%, 19.70%, 1.52%, 0.08% and 1.31%, respectively. The highest repeat number of two base type was 6, three base type was 5 and four, five and six base types were four (Table 5). Among all detected SSRs, the AG/CT motif kind exhibited the highest frequency, which accounted for 72.19%, followed by AAG/CTT, AC/GT, and ACC/GGT motif kind that accounted for 6.42, 3.36 and 3.33%, respectively (Table 6).

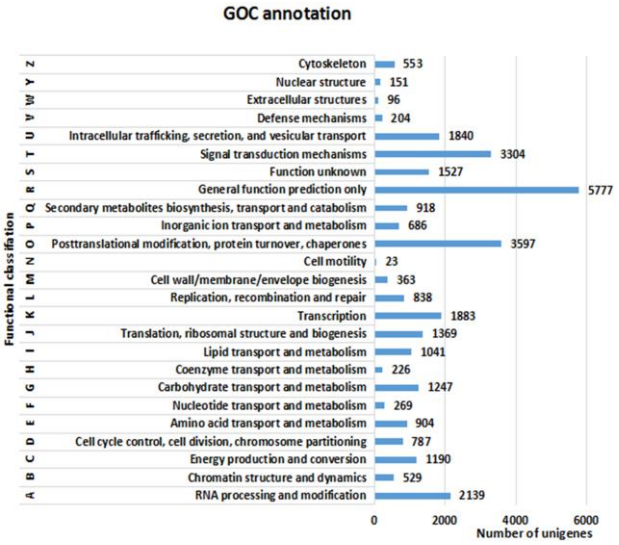

Fig. 3: KOG functional classification of *V. uliginosum* Unigenes

The Candidate Gene Involving in Chlorogenic Acid Biosynthesis

KEGG database (http://www.kegg.jp/kegg/kegg1.html) was explored by query 'chlorogenic acid' under the retrieve item 'DBGET', and three 'KEGG compound' encoded as C00852, C10468 and C17417 related to the chlorogenic acid were found. C00852 was found to be involved in three pathways including 'Phenylpropanoid biosynthesis' (map00940), 'Flavonoid biosynthesis' (map00941), 'Stilbenoid, diarylheptanoid and gingerol biosynthesis' (map00945) that associated with the biosynthesis of chlorogenic acid. Homologous unigenes were detected after annotation which encode five key enzymes for bio-synthesis of chlorogenic acid, including Unigene0000447, Unigene0000448 and Unigene0000449 encoded phenylalanine ammonia-lyase (PAL), Unigene0008683 encoded 4-coumarate-CoA ligase (4CL), Unigene0014113 and Unigene0014114 encoded trans-cinnamate 4-monooxygenase (CYP73A), Unigene0015838 encoded shikimate O-hydroxycinnamoyltransferase (HCT) and Unigene0035395 encoded coumaroylquinate

Table 3: KEGG functional classification of *Vaccinium uliginosum* Unigenes

KEGG_A_class	KEGG_B_class	Pathway Count (7036)
Cellular Processes	Transport and catabolism	Endocytosis (270), Peroxisome (132), Phagosome (126), Regulation of autophagy (74)
Environmental Information Processing	Membrane transport	ABC transporters (84)
Environmental Information Processing	Signal transduction	Plant hormone signal transduction (265), Phosphatidylinositol signaling system (102)
Genetic Information Processing	Folding, sorting and degradation	Protein processing in endoplasmic reticulum 346 Proteasome (81), Protein export (78), RNA degradation (232), SNARE interactions in vesicular transport (45), Sulfur relay system (21), Ubiquitin mediated proteolysis (259)
	Replication and repair	Base excision repair (59), DNA replication (64), Homologous recombination (71), Mismatch repair (70), Non-homologous end-joining (15), Nucleotide excision repair (123).
	Transcription	Basal transcription factors (84), RNA polymerase (89), Spliceosome (361), Aminoacyl-tRNA biosynthesis (108), mRNA surveillance pathway (266), Ribosome (293), Ribosome biogenesis in eukaryotes (221), RNA transport (364)
Metabolism	Amino acid metabolism	Alanine, aspartate and glutamate metabolism (82), Arginine and proline metabolism (75), Arginine biosynthesis (41), Cysteine and methionine metabolism(125), Glycine, serine and threonine metabolism (124), Histidine metabolism (25), Lysine biosynthesis (22), Lysine degradation (62), Phenylalanine metabolism (68), Phenylalanine, tyrosine and tryptophan biosynthesis (63), Tryptophan metabolism (36), Tyrosine metabolism (58), Valine, leucine and isoleucine biosynthesis (28), Valine, leucine and isoleucine degradation (79)
	Biosynthesis of other secondary metabolites	Anthocyanin biosynthesis (26), Betalain biosynthesis (1), Caffeine metabolism (6), Flavone and flavonol biosynthesis (9), Glucosinolate biosynthesis (2), Flavonoid biosynthesis (39), Isoquinoline alkaloid biosynthesis (40), Isoflavonoid biosynthesis (6), Monobactam biosynthesis (18), Phenylpropanoid biosynthesis (135), Stilbenoid, diarylheptanoid and gingerol biosynthesis (24), Tropane, piperidine and pyridine alkaloid biosynthesis (35)
	Carbohydrate metabolism	Amino sugar and nucleotide sugar metabolism (188), Ascorbate and aldarate metabolism (66), Butanoate metabolism (27), Citrate cycle (TCA cycle) (92), C5-Branched dibasic acid metabolism (13), Starch and sucrose metabolism (257), Galactose metabolism (110), Fructose and mannose metabolism (89), Glycolysis / Gluconeogenesis (193), Glyoxylate and dicarboxylate metabolism (102), Inositol phosphate metabolism (112), Pentose and glucuronate interconversions (57), Pentose phosphate pathway (74), Propanoate metabolism (66), Pyruvate metabolism (144)
	Energy metabolism	Carbon fixation in photosynthetic organisms (125), Nitrogen metabolism (36), Oxidative phosphorylation (207), Photosynthesis (76), Photosynthesis - antenna proteins (22), Sulfur metabolism (52)
	Global and Overview	Biosynthesis of amino acids (334), Carbon metabolism (427), Degradation of aromatic compounds (10), Fatty acid metabolism (122), 2-Oxocarboxylic acid metabolism (68)
	Glycan biosynthesis and metabolism	Glycosylphosphatidylinositol(GPI)-anchor biosynthesis (38), Glycosaminoglycan degradation (35), Glycosphingolipid biosynthesis - ganglio series (16), Glycosphingolipid biosynthesis - globo series (17), N-Glycan biosynthesis (86), Other types of O-glycan biosynthesis (9), Other glycan degradation (51)
	Lipid metabolism	alpha-Linolenic acid metabolism (63), Arachidonic acid metabolism (26), Biosynthesis of unsaturated fatty acids (48), Cutin, suberine and wax biosynthesis (13), Ether lipid metabolism (45), Fatty acid biosynthesis (70), Fatty acid elongation (41), Fatty acid degradation (71), Glycerolipid metabolism (100), Glycerophospholipid metabolism (135), Linoleic acid metabolism (28), Sphingolipid metabolism (59), Steroid biosynthesis (39), Synthesis and degradation of ketone bodies (8)
	Metabolism of cofactors and vitamins	Biotin metabolism (26), Folate biosynthesis (26), Lipoic acid metabolism (6), Nicotinate and nicotinamide metabolism (42), One carbon pool by folate (35), Pantothenate and CoA biosynthesis (33), Porphyrin and chlorophyll metabolism (72), Thiamine metabolism (28), Riboflavin metabolism (16), Ubiquinone and other terpenoid-quinone biosynthesis (66), Vitamin B6 metabolism (26)
	Metabolism of other amino acids	beta-Alanine metabolism (75), Glutathione metabolism (102), Cyanoamino acid metabolism (54), Selenocompound metabolism (25), Taurine and hypotaurine metabolism (16),
	Metabolism of terpenoids and polyketides	Brassinosteroid biosynthesis (25), Carotenoid biosynthesis (39), Diterpenoid biosynthesis (26), Limonene and pinene degradation (13), Monoterpenoid biosynthesis (18), Sesquiterpenoid and triterpenoid biosynthesis (21), Terpenoid backbone biosynthesis (84), Zeatin biosynthesis (21)
	Nucleotide metabolism	Purine metabolism (259), Pyrimidine metabolism (192)
Organismal Systems	Environmental adaptation	Circadian rhythm - plant (77), Plant-pathogen interaction (339)

(coumaroylshikimate) 3'-monooxygenase (CYP98A, C3'H), and codes of these enzymes in KEGG dataset were 4.3.1.24, 6.2.1.12, 1.14.13.11, 2.3.1.133 and 1.14.13.36, respectively (Table 7)

Discussion

Vaccinium L. (Ericaceae) comprising of approximately 450 species is a morphologically diverse genus of terrestrial or epiphytic shrubs and lianas. Primarily these species occur in the cooler areas of northern hemisphere. *Vaccinium* species have been highly concerned in the recent years because the genus contains three kinds of new fruit with excellent health care values *viz*., blueberry, cranberry and lingonberry. Among three kinds of domesticated *Vaccinium* fruits, the blueberry genomics research has been widely conducted, and these studies have greatly enhanced the blueberry

Table 4: Candidate R-gene families and corresponding number of unigenes in *V. uliginosum*

Class	Count	Class	Count	Class	Count	Class	Count	Class	Count
RLP	731	N	371	NL	310	TNL	223	CNL	189
RLK	188	RLK-GNK2	104	CN	63	Other	43	T	41
Mlo-like	26	Pto-like	24	L	10	RPW8-NL	6	RLP-Malectin	5
RLK-Malectina	2	RLP-Malectina	1	RLK-Pto-like	1	RLK-Malectin	1	Total	2339

Table 5: SSR types identified from the *V. uliginosum* leaf transcriptome

Number of repeat unit	Di-	Tri-	Tetra-	Penta-	Hexa-	Total
4	0	0	100	60	79	239
5	0	1044	34	13	29	1120
6	1238	467	12	2	4	1723
7	1033	215	1	0	8	1257
8	1030	43	0	1	7	1081
9	993	28	0	0	0	1021
10	799	34	0	0	0	833
11	361	9	0	0	0	370
12	69	10	0	0	0	79
13	17	20	0	0	0	37
14	104	11	0	0	0	115
>=15	1779	26	0	0	0	1805
Total	7423	1907	147	76	127	9680

Table 6: Motif kinds of SSR and its count in *V. uliginosum* leaf transcriptome

Motif kinds	Count	Frequency	Motif kinds	Count	Frequency
AC/GT	325	0.0335743801652893	ACT/AGT	18	0.0018595041322314
AG/CT	6988	0.721900826446281	AGC/CTG	203	0.0209710743801653
AT/AT	98	0.0101239669421488	AGG/CCT	339	0.0350206611570248
AAC/GTT	56	0.00578512396694215	ATC/ATG	133	0.0137396694214876
AAG/CTT	621	0.0641528925619835	CCG/CGG	104	0.0107438016528926
AAT/ATT	40	0.00413223140495868	AAAG/CTTT	28	0.00289256198347107
ACC/GGT	322	0.0332644628099174	AAAT/ATTT	20	0.00206611570247934
ACG/CGT	71	0.00733471074380165	others	314	0.0324380165289256

Table 7: The key enzymes responsible for the bio-synthesis of chlorogenic acid and corresponding unigenes identified in *V. uliginosum*

Code in KEGG dataset	Name	Description	ORTHOLOG	Coresponding unigenes code
4.3.1.24	PAL	phenylalanine ammonia-lyase	K10775	Unigene0000447 Unigene0000448 Unigene0000449
6.2.1.12	4CL	4-coumarate--CoA ligase	K01904	Unigene0008683
1.14.13.11	CYP73A	trans-cinnamate 4-monooxygenase	K00487	Unigene0014113 Unigene0014114
2.3.1.133	HCT	shikimate O-hydroxycinnamoyltransferase	K13065	Unigene0015838
1.14.13.36	CYP98A, C3'H	Coumaroylquinate (coumaroylshikimate) 3'-monooxygenase	K09754	Unigene0035395

growth, physiological active components metabolism, molecular breeding of blueberry, biological engineering technology and understanding the molecular mechanism of cold tolerance (Rowland *et al.,* 2012). Genome draft of high bush blueberry has also been published as integrated genome browser (Gupta *et al.,* 2015), and genome database of American cranberry (*V. macrocarpon*) also available (Polashock *et al.,* 2014). However, there is no report about the genome and transcriptome analysis of *V. uliginosum*, and this is the first corresponding report until now.

Because its unique geography distribution, *V. uliginosum* is often used as a representative plant to study the ecology, physiology and evolution of polar/Arctic plants (Alsos *et al.,* 2003; Albert *et al.,* 2005; Boesgaard *et al.,* 2012; Faubert *et al.,* 2012; Preece and Phoenix, 2014). During the last few decades, molecular markers, revealing polymorphism at the DNA level, have played an important role in studies of plant biotechnology and their genetics and ecology (Kumar *et al.,* 2009; Nadeem *et al.,* 2018). In the recent years, some SSR markers have been developed using pyrosequencing method for *V. uliginosum* (Mayer *et al.,* 2014); however, the number was limited. In this study, we exploited huge numbers of EST-SSRs markers from whole transcriptome data. The molecular markers play indispensable roles in many research areas such as molecular breeding, conservation genetics, molecular ecology and association studies (Li *et al*., 2016; Nadeem *et al*., 2018; Wang *et al*., 2019).

Therefore, the results of this study would greatly help for above mentioned studies about *V. uliginosum*.

R-genes evolved from a phylogenetically ancient form of immunity, which is common in animals and plants (Chisholm *et al.,* 2006). It is a major protection way to defend themselves against the attack of pathogens based on the mechanism of gene-for-gene interactions between plants and their infectious agents (Flor, 1971). The R-gene identification had attracted the attention of scientists for its potential value for improving crop adaptability (Harris *et al.,* 2013; Dong *et al*., 2019). In this study, we detected 2339 unigenes related to 19 families as the candidate of resistance gene, and this would greatly deepen our understanding about the molecular mechanism of resistance in *V. uliginosum*. Moreover, corresponding unigenes are potential molecular tool for the resistance breeding and genetic engineering of *V. uliginosum* and its relative species.

Not only the berries of *V. uliginosum* are rich in antioxidants, especially anthocyanins and other polyphenolic compounds. They are medicinally important for cardiovascular diseases, obesity, aging diseases, urinary tract infections and periodontal disease, diabetes and in reduction of the risk of cancers or assist in treatment of cancers (Su, 2012). Its leaves are also rich in bioactive ingredients such as chlorogenic acid (3-O-caffeoylquinic acid) (Chen *et al.,* 2015). Chlorogenic acid (CGA) is the ester of caffeic acid and quinic acid, functioning as an intermediate in lignin biosynthesis (Boerjan *et al.,* 2003). Three pathways of CGA bio-synthesis have been exploited, from the primary substrate 'Phenylalanine' to primary production 'Caffeoyl quinic acid' CGA, and there are few key enzymes including PAL, C4H, 4CL, C3'H, CQT, COMT and CCoACMT (Campa *et al.,* 2003). From leave transcriptome data of *V. uliginosum*, we have identified homologous genes which encoded key enzymes such as PAL, C4H, 4CL, C3'H and CYP73A that responsible for the biosynthesis of CA. Moreover, we also have detected some unigenes that were involved into other biosynthesis pathways of antioxidation components of other secondary metabolites, such as Anthocyanins and flavonoids (Table 3). These results are important to deepen our knowledge about the molecular mechanism of healthful functions hidden in *V. uliginosum* leaves.

Conclusion

In this study, about 2G transcriptome data of *V. uliginosum* was obtained and analyzed for unigenes assemblage and functional annotation. To date this is the first functional information reported on *V. uliginosum*. The transcriptome data assembled into 43,507 unigenes with an average length of 588 bp of these, 30,689 (70.538%) unigenes were annotated by Nr, Swiss-Port, KOG or KEGG databases, and 30,641, 21,518, 19,273 and 13,334 unigenes had a significant hit ($p<1E-5$) in above four databases, respectively. 2339 unigenes were identified as candidate resistance genes, which belong to 19 families. A total of 7780 unigenes was identified as 9680 SSR loci, and AG/CT repeat motifs was the most common SSR. Overall, eight candidate unigenes encoded five key enzymes that were responsible for the bio-synthesis metabolic pathway of chlorogenic acid, which is one of the important bioactive components with high medicinal function. This transcriptome analysis is a significant contribution for molecular marker development, functional gene discovery and studies of differential gene expressions in *V. uliginosum.*

Acknowledgement

This work was funded by the key project of the Education Department of Guizhou Province [KY(2013)186], Science Department of Guizhou Province [(2015)6013] and [(2016)5202] and the National Natural Science Foundation of China [31560091].

References

Albert, T., O. Raspé and A. Jacquemart, 2005. Diversity and spatial structure of clones in *Vaccinium uliginosum* populations. *Can. J. Bot*., 83: 211–218

Alsos, I.G., S. Spjelkavik and T. Engelskjøn, 2003. Seed bank size and composition of *Betula nana*, *Vaccinium uliginosum*, and *Campanula rotundifolia* habitats in Svalbard and northern Norway. *Can. J. Bot*., 81: 220–231

Auffret, A.G., E. Meineri, H.H. Bruun, R. Ejrnæs and B.J. Graae, 2010. Ontogenetic niche shifts in three *Vaccinium* species on a sub-alpine mountain side. *Plant Ecol. Divers*., 3: 131–139

Boerjan, W., J. Ralph and M. Baucher, 2003. Lignin biosynthesis. *Ann. Rev. Plant Biol*., 54: 519–546

Boesgaard, K.S., K.R. Albert, H. Ro-Poulsen, A. Michelsen, T.N. Mikkelsen and N.M. Schmidt, 2012. Long-term structural canopy changes sustain net photosynthesis per ground area in high arctic *Vaccinium uliginosum* exposed to changes in near-ambient UV-B levels. *Plantarum*, 145: 540–550

Campa, C., M. Noirot, M. Bourgeois, M. Pervent, C.L. Ky, H. Chrestin, S. Hamon and A.D. Kochko, 2003. Genetic mapping of a caffeoyl-coenzyme A 3-O-methyltransferase gene in coffee trees. Impact on chlorogenic acid content. *Theor. Appl. Genet.*, 107: 751–756

Chen, F., X. Du, Y. Zu and L. Yang, 2015. A new approach for preparation of essential oil, followed by chlorogenic acid and hyperoside with microwave-assisted simultaneous distillation and dual extraction (MSDDE) from *Vaccinium uliginosum* leaves. *Ind. Crops Prod.,* 77: 809–826

Chisholm, S., T.G. Coaker, B. Day and B.J. Staskawicz, 2006. Host-microbe interactions: shaping the evolution of the plant immune response. *Cell*, 124: 803–814

Conesa, A., S. Gotz, J.M. Garciagomez, J. Terol, M. Talon and M. Robles, 2005. Blast2 GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21: 3674–3676

Dong, C., X. Zheng, Y. Shi, S.H. Chen, D. Liu, P. Lin and Z. Hu, 2019. Genome-Wide identification, classification and expression of flavonol synthase from *Nelumbo nucifera* in Defense against Various Stresses. *Intl. J. Agric. Biol*., 21: 653–658

Faubert, P., P. Tiiva, A. Michelsen, A. Rinnan, H. Ro-Poulsen and R. Rinnan, 2012. The shift in plant species composition in a subarctic mountain birch forest floor due to climate change would modify the biogenic volatile organic compound emission profile. *Plant Soil*, 352: 199–215

Flor, H.H., 1971. Current status of the gene-for-gene concept. *Annu. Rev. Phytopathol.*, 9: 275–296

Gupta, V., A.D. Estrada, I. Blakley, R. Reid, K. Patel, M.D. Meyer, S.U. Andersen, A.F. Brown, M.A. Lila and A.E. Loraine, 2015. RNASeq analysis and annotation of a draft blueberry genome assembly identifies candidate genes involved in fruit ripening, biosynthesis of bioactive compounds, and stage-specific alternative splicing. *GigaScience*, 4: 1–22

Han, E.K., H.S. Kwon, S.G. Shin, Y.H. Choi, I.J. Kang and C.K. Chung, 2012. Biological effect of *Vaccinium uliginosum* L. on STZ-induced diabetes and lipid metabolism in rats. *J. Kor. Soc. Food Sci. Nutr.*, 41: 1727–1733

Harris, C.J., E.J. Slootweg, A. Goverse and D.C. Baulcombe. 2013. Stepwise artificial evolution of a plant disease resistance gene. *Proc. Natl. Acad. Sci. USA*, 110: 21189–21194

Jacquemart, A.L., 1992. Préliminaires sur la production de nectar chez trois espèces de *Vaccinium*. *Apidologie*, 23: 453–464

Kähkönen, M.P., A.I. Hopia and M. Heinonen, 2001. Berry phenolics and their antioxidant activity. *J. Agric. Food Chem.,* 49: 4076–4082

Kanehisa, M. and S. Goto, 2000. KEGG: Kyoto Encyclopedia of genes and genomes. *Nucl. Acids Res*., 28: 27–30

Kellogg, J., J. Wang, C. Flint, D. Ribnicky, P. Kuhn, E.G.D. Mejia, I. Raskin and M.A. Lila, 2010. Alaskan wild berry resources and human health under the cloud of climate Change. *J. Agric. Food Chem*., 58: 3884–3900

Kim, K.H. and S.Y. Choung, 2014. Oral administration of *Vaccinium uliginosum* L. Extract alleviates DNCB-induced atopic dermatitis in NC/Nga Mice. *J. Med. Food*, 17: 1350–1360

Krikorian, R., M.D. Shidler, T.A. Nash, W. Kalt, M.R. Vinqvist-Tymchuk, B. Shukitt-Hale and J.A. Joseph, 2010. Blueberry supplementation improves memory in older adults. *J. Agric. Food Chem*., 58: 3996–4000

Kumar, P., V.K. Gupta, A.K. Misra, D.R. Modi and B.Y. Pandey, 2009. Potential of molecular markers in plant biotechnology. *Plant Omics J*., 2: 141–162

Latti, A.K., L. Jaakola, K.R. Riihinen and P.S. Kainulainen, 2010. Anthocyanin and flavonol variation in bog bilberries (*Vaccinium uliginosum* L.) in Finland. *J. Agric. Food Chem*., 58: 427–433

Lepelley, M., G. Cheminade, N. Tremillon, A. Simkin, V. Caillet and J. McCarthy, 2007. Chlorogenic acid synthesis in coffee: An analysis of CGA content and real-time RT-PCR expression of HCT, HQT, C3H1 and CCoAOMT1 genes during grain development in *C. canephora*. *Plant Sci*., 172: 978–996

Lyrene P.M. and J.W. Olmstead. 2012. The Use of Inter-Sectional Hybrids in Blueberry Breeding. *Int J Fruit Sci*.,12: 269–275

Li, M., Y. Yang, F. Feng, X. Li, L. Gu, Y. Tian, J. Wang, J. Wang, Z. Li, A. Chen, X. Chen and Z. Zhang, 2016, SSR development and utilization with *Rehmannia glutinosa* transcriptome. *Intl. J. Agric. Biol*., 18: 589–600

Li, R., P. Wang, Q. Guo and Z.Y. Wang, 2011. Anthocyanin composition and contentof the *Vaccinium uliginosum* berry. *Food Chem*., 125: 116–120

Mayer, C., A.L. Jacquemart and O. Raspé, 2014. Development and multiplexing of microsatellite markers using pyrosequencing in a tetraploid plant, *Vaccinium uliginosum* (*Ericaceae*). *Plant Ecol. Evol.,* 147: 285–289

Nadeem, M.A., M.A. Nawaz, M.Q. Shahid, Y. Doğan, G. Comertpay, M. Yıldız, G. Comertpay, M. Yıldız, R. Hatipoğlu, F. Ahmad, A. Alsaleh, N. Labhane, H. Özkan, G. Chung and F.S. Baloch, 2018. DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnol. Biotechnol. Equipm*., 32: 261–285

Park, C.Y., N. Gu, C.Y. Lim, J.H. Oh, M. Chang, M. Kim and M.Y. Rhee, 2016. The effect of *Vaccinium uliginosum* extract on tablet computer induced asthenopia: randomized placebo-controlled study. *BMC Complem. Altern. Med*., 16: 296

Polashock, J., E. Zelzion, D. Fajardo, J. Zalapa, L. Georgi, D. Bhattacharya and N. Vorsa, 2014. The American cranberry: first insights into the whole genome of a species adapted to bog habitat. *BMC Plant Biol*., 14: 165

Preece, C. and G.K. Phoenix, 2014. Impact of early and late winter icing events on sub-arctic dwarf shrubs. *Plant Biol*., 16: 125–132

Primetta, A.K., K. Karppinen, K.R. Riihinen and L. Jaakola, 2015. Metabolic and molecular analyses of white mutant *Vaccinium* berries show down-regulation of MYBPA1-type R2R3 MYB regulatory factor. *Planta*, 242: 631–643

Rossum, F.V., N.J. Vereecken, E. Bredat and D. Michez, 2012. Pollen dispersal and fruit production in *Vaccinium oxycoccos* and comparison with its sympatric congener *V. uliginosum*. *Plant Biol*., 15: 344–352

Rowland, L.J., D.J. Bell, N. Alkharouf, N.V. Bassil, F.A. Drummond, L. Beers, E.J. Buck, C.E. Finn, J. Graham, S. McCallum, J.F. Hancock, J.J. Polashock and J.W. Olmstead, 2012. Generating genomic tools for blueberry improvement. *Intl. J. Fruit Sci*., 12: 276–287

Stefko, G., S. Hritovski, J.P. Stanoeva, M. Stefova, L. Melovski and S. Kulevanova, 2014. Resource assessment and economic potential of bilberries (*Vaccinium myrtillus* and *Vaccinium uliginosum*) on Osogova Mtn., R. Macedonia. *Ind. Crops Prod.,* 61: 145–150

Stephens, D.T., D.E. Levesque and A.R. Davis, 2012. Pollen-ovule ratios in seven species of *Vaccinium* (Ericaceae) and stamen structure in *Vaccinium myrtilloides* and *Vaccinium vitis-idaea*. *Botany*, 90: 599–614

Su, S., L.J. Wang, C.Y. Feng, Y. Liu, C.H. Li, H. Du, Z.Q. Tang, Y.J. Xu and L.S. Wang, 2016. Fingerprints of anthocyanins and flavonols of *Vaccinium uliginosum* berries from different geographical origins in the Greater Khingan Mountains and their antioxidant capacities. *Food Cont.*, 64: 218–225

Umezu, T., 2012. Evaluation of the effects of plant-derived essential oils on central nervous system function using discrete shuttle-type conditioned avoidance response in mice. *Phytother. Res*., 26; 884–891

Varshney, R.K., A. Graner and M.E. Sorrells, 2005. Genomics-assisted breeding for crop improvement. *Trends Plant Sci.,* 10: 621–630

Vendrame, S., A. Kristo, D.A. Schuschke and D. Klimis-Zacas, 2014. Wild blueberry consumption affects aortic vascular function in the obese Zucker rat. *Appl. Physiol. Nutr. Metab.,* 39: 255–261

Wang, L.J., S. Su, J. Wu, H. Du, S.S. Li, J.W. Huo, Y. Zhang and L. Wang, 2014. Variation of anthocyanins and flavonols in *Vaccinium uliginosum* berry in Lesser Khingan Mountains and its antioxidant activity. *Food Chem*., 160: 357–364

Wang, S., C. Chen, W. Sciarappa, C. Wang and M. Camp, 2008. Fruit quality, antioxidant capacity, and flavonoid content of organically and conventionally grown blueberries. *J. Agric. Food Chem*., 56: 5788–5794

Wang, Y., M.Q. Shahid, F. Ghouri, S. Ercişli, F.S. Baloch and F. Nie, 2019 Transcriptome analysis and annotation: SNPs identified from single copy annotated unigenes of three polyploid blueberry crops. *PLoS One*, 14: e0216299

Ye, J., L. Fang, H. Zheng, Y. Zhang, J. Chen, Z. Zhang, J. Wang, S. Li, R. Li, L. Bolung and J. Wang, 2006. WEGO: a web tool for plotting GO annotations. *Nucl. Acids Res*., 34: 293–297

Yin, L., Y.L. Pi and M.N. Zhang, 2012. The effect of *Vaccinium uliginosum* on rabbit retinal structure and light-induced function damage. *Chin. J. Integr. Med*., 18: 299–303

Yoon, S.M., B.L. Lee, Y.R. Guo and S.Y. Choung, 2016. Preventive effect of *Vaccinium uliginosum* L. extract and its fractions on age-related macular degeneration and its action mechanisms. *Arch. Pharm. Res*., 39: 21–32

Zhang, N., M. Ying, Y. Wu, Z. Zhou, Z. Ye, H. Li and D. Lin, 2014. Hyperoside, a flavonoid compound, inhibits proliferation and stimulates osteogenic differentiation of human osteosarcoma cells. *PLoS One*, 9: e0098973

Zheng, T.T., Z.K. Zhang, M.Q. Shahid, W.L. Wei, F.S. Baloch, J.C. Wu, S.Q. Lin and X.H. Yang, 2017. RNA-Seq reveals differential expression patterns of genes associated with carotenoid accumulation in loquat. *Acta Physiol. Plantarum*, 39: 168

[Received 18 Jan 2019; Accept 18 May 2019; Published (online) 20 Aug 2019]