Pak. J. Bot., 49(6): 2263-2272, 2017.

CANDIDATE GENE ASSOCIATION RESEARCH ON FLAVONES CONTENT IN HAWTHORN (*CRATAEGUS PINNATIFIDA* BGE.)

ZHAO YUHUI[#], SU KIA[#], ZHANG LIPING, LI JUNPENG AND GUO YINSHAN[*]

College of Horticulture, Shenyang Agricultural University, Shenyang, P.R. China
**Corresponding author's email: guoyinshan77@126.com*
#These authors contributed equally to this study

Abstract

In this study, 87 hawthorn (*Crataegus pinnatifida* Bge.) accessions were obtained from National Hawthorn Germplasm Nursery in Shenyang Agricultural University of China and flavones contents including vitexin-rhamnoside, vitexin, rutin, hyperin and quercetin were detected by high performance liquid chromatography (HPLC) in 2014 and 2015. Chalcone Synthase genes (*Cp*CHS) of 87 hawthorn accessions were cloned and sequenced, and then SNPs (Single Nucleotide Polymorphisms) among the result of sequencing were screened out, based on thees SNPs, we conducted the association analysis between the SNPs and phenotypic traits of hawthorn flavones content, the significant associated SNPs wre then discovered. The result showed that, in total of 71SNPs were discovered and three putative causal SNPs were associated with vitexin content both in 2014 and 2015, some SNPs were associated with vitexin-rhamnoside, rutin, hyperin and quercetin. This result can be used for molecular assisted breeding of hawthorn flavone content and also provide the reference for woody plants in association analysis filed.

Key words: Hawthorn, Flavones content, SNP marker, Association analyses.

Introduction

Hawthorn (*Crataegus pinnatifida* Bge.) belongs to the family Rosaceae, a deciduous thorny shrub or tree, it has three to five lobed broadly ovate or triangular ovate leaves, the flowers are white, and the fruits vary from red, orange, yellow and yellow-green(Rigelsky & Sweet, 2002). Until now, it has been planted in Heilongjiang, Jilin, Liaoning province and Inner Mongolia, located in north and northeast of China. It is also a typical medical and edible plant which can be used for treating many kinds of cardiovascular diseases, besides, it is also considered as one of the ornamental plants (Rigelsky & Sweet, 2002; Kao *et al*., 2005).

The flavones were considered to be the primary bioactive component of hawthorn (Chang *et al*., 2002). It is a kind of secondary metabolites of polyphenols which has low molecular weight and widely exists in plant kingdom and formed from the process of resisting bad ecological condition, animals and microbes invade of plants in long-term (Dixon & Paiva, 1995; Jahan *et al.*, 2013). Its therapeutic use is reported including scavenging free radical, antioxidant, anticancer and so on (Dixon & Steele, 1999; Zhao *et al*., 2014), hence the flavones are widely researched in the areas of agriculture, chemical industry and medicine. As a key enzyme in flavones biosynthetic pathway, chalcone synthase (CHS) catalyses the first reaction of flavonoid biosynthesis(Koes *et al*., 1989) and obligatory for the formation of naringenin chalcone which is considered as the first intermediate in flavones biosynthesis (Heller & Hahlbrock, 1980). Because of the vital function of this enzyme, it was studied intensively and more than 650 CHS and CHS-like gene sequences have been cloned, sequenced and characterized from monocotyledon to dicotyledon plants (Yang & Guo, 2006), including maize (Wienand *et al*., 1986), parsley (Jahnen & Hahlbrock, 1988), antirrhinum (Lipphardt *et al*., 1988), Arabidopsis (Feinbaum & Ausubel, 1988), Petunia (Meer *et al*., 1990), bean (Schmid *et al*., 1990), Matthiola (Epping *et al*., 1990), *Lilium hybrid* (Akira *et al*., 2003), *Phalaenopsis* Orchid (Han *et al*., 2006).

As the development of molecular quantitative genetics, the association analysis has been widely used in the research of human genetics research (Corder *et al*., 1994; Templeton *et al*., 1992). Thornsberry (2001) was the first to report the use of association analysis in the plant genetics. The researches of these years showed that association analysis was a vital method to make the genotype associated with the phenotype (Thornsberry *et al*., 2001; Kraakman *et al*., 2004; SherryA *et al*., 2005; Zhu *et al*., 2008). Genome-wide association (GWA) analysis and candidate gene association analysis were the two association analysis methods, besides the great preeminence of Genome-wide association analysis (Hirschhorn & Daly, 2005; Atwell *et al*., 2010). Candidate gene association analysis was a very important approach to gene mapping for less complex traits (Tabor *et al*., 2002; Ehrenreich *et al*., 2009).

No papers are available on the genetic diversity of CHS gene affecting flavones content in hawthorn. In this study, we assessed the association of nucleotide variation in the candidate gene CHS with flavones content in hawthorn, the objective of our study were conduct association analysis between individual polymorphisms and flavones content in order to identify putative causal SNPs based on determine flavones monomers content accurately, the result can be helpful for high flavones content selection and breeding of hawthorn accessions and it will also prove some reference for woody plant in the filed of association analysis.

Material and Methods

Plants: Tender leaves were collected from 87 hawthorn accessions held by the National hawthorn germplasm nursery of China in Shenyang Agricultural University (Table 1). This population included by typical hawthorn species came from different provinces such as Liaoning, Shangdong, Shanxi, Henan.

Table 1. 87 hawthorn resource message.

No.	Variety name	No.	Variety name	No.	Variety name
1.	fengshuishanzha	30.	mengyinjinxing	59.	linfenbaiyesheng
2.	Yinyeling NO.1	31.	anshanzirou	60.	huabeixiaoshanzha
3.	chengoudahong	32.	mengyindajinxing	61.	beijingduizhao
4.	jiangxianshanzha	33.	beijingdenglonghong	62.	piposhi
5.	Xinbin NO.7	34.	wanqiushanlihong	63.	tuguzi NO.1
6.	kaiyuanruanzi	35.	fushunshanzha	64.	tongtaibaiyesheng
7.	liangshanhong	36.	xifen NO.4	65.	Huangbaoyu NO.1
8.	qiujinxing	37.	xinghong NO.2	66.	donglingqingkou
9.	xinbinruanzi	38.	jingduanNO.1	67.	niejiayu NO.2
10.	jinxianxiaoye	39.	81-2	68.	yinyeling NO.2
11.	benxi NO.7	40.	fenlishanzha	69.	xiezishi NO.3
12.	benxi NO.4	41.	haitangshanzha	70.	jianchangshanzha
13.	shancheng NO.1	42.	baiquan7901	71.	zhuantaishanzha
14.	yu8002	43.	shandonghongmianzha	72.	tielingshanzha
15.	yubeihong	44.	wolonggang NO.2	73.	benxi NO.2
16.	bairangmian	45.	xuzhoudahuo	74.	jiangxian798203
17.	xifenshisheng	46.	Jiangxian798202	75.	mopan
18.	liaoyangzirou	47.	Jiangou NO.2	76.	shanxitiansheng
19.	shen 2-4	48.	xiajinxing	77.	baili
20.	yiduxiaohuang	49.	zifeng	78.	jiangxian798201
21.	tianxiangyu	50.	yu8001	79.	niuxintai NO.1
22.	tianshui	51.	zimuhong	80.	sishanling
23.	donglingshanzha	52.	baiquan7903	81.	shuanghong
24.	shen 78201	53.	xinglongshisheng	82.	kaiyuanNO.1
25.	xihong	54.	huixiandahong	83.	dawuleng
26.	linxianshangkou	55.	magangdajinxing	84.	xifen NO.1
27.	beijingzaosheng	56.	yidutedahuangmianzha	85.	guajiayu NO.1
28.	duanzhishanlihong	57.	majiafenrou	86.	huixiandahongkongqi
29.	fushunshangzhuanbai	58.	majiadadui	87.	yinyeling NO.7

Hawthorn leaves treatment: Hawthorn leaves of 87 accessions were kept at 50ºC in drying baker until constant weight, 1g hawthorn powder of ground leaves was filtered by 20 meshes, weighed and then dissolved with 70% ethyl alcohol up to 30mL in 30 mL volumetric flask. After extraction with 225 W ultrasonic power for 30 minutes, 1 mL solution was absorbed for each hawthorn species solution and filtered with 0.45mL microporous membrane. The filtrate was used for flavonoid monomers content determination directely.

Phenotypic data of flavonoid monomers content: In 2014 and 2015, the content of five flavonoid monomers, vitexin-rhamnoside, vitexin, rutin, hyperoside, and quercetin, was detected with an Agilent 1100 HPLC with a DAD detector and a C18 column (250 × 4.6 mm, 5 μm). Acetonitrile (A) and 0.5% phosphoric acid solution (B) were used as the mobile phase, with the following gradient elution protocol: 0–9 min, 18–20%(A); 9–25 min, 20–50%(A); 25–33 min, 50–18%(A); 33–38 min, 18%(A). The maximum absorption spectrum was 345 nm, the column temperature was 25 °C, the flow rate was 1.0 $mL \cdot min^{-1}$, and the injection volume was 20 μL.

RNA extraction and cDNA synthesis: The tender leaves were sampled from the filed in May 2015; they were frozen by liquid nitrogen separately and then stored at -80°C for RNA extraction. Using modified CTAB method to extract the total RNA. The 260/280 nm ratio of RNA concentration were determined by spectrophotometer before and after DNase I digestion. Integrity (?) of RNA was checked by electrophoresis on 1.0% agarose gels in 0.5X TBE (90 mM Tris-acetate, 2 mM EDTA). The synthesis of first strand cDNA from the total RNA were made by Prime Script TM RT reagent Kit with gDNA Eraser (Takara) according to the manufacturer's instructions.

***Cp*CHS gene clone and sequencing:** A cDNA pool of each material was used as the template for full-ORF *Cp*CHS cDNA amplification. The full-ORF *Cp*CHS cDNA was amplified through PCR with the forward primer: 5'-ATGGTTACCGTCGAGGAAGTTC-3' and the reverse primer: 5'-TCAAGCACCCACAC TGTGAAGCA-3', PCR program condition were: 94ºC as first denaturation step for 5 min followed by 35 cycles of 94ºC for 30s, 56ºC for 30s, 72ºC for 2min. The final elongation step was 72ºC for 10min. Using 1% agarose gel to resolve the target gene

of *Cp*CHS through gel electrophoresis and then the PCR production were purified from agarose gel by using TaKaRa MiniBEST Agarose Gel DNA Extraction Kit Ver.4.0 (TaKaRa, Dalian) according to manufacturer's instructions. Target gene fragments were then subcloned in to pMD18-T vector(pMDTM18-T Vectot Cloning Kit, TaKaRa, Dalian) and *E.coli* One Shot TOP10 was employed as the host strain for target gene manipulation. Colonies containing *Cp*CHS alleles A and B of 87 hawthorn resource were used for sequencing after colony PCR screening.

Nucleotide polymorphisms and diversity of the candidate gene *Cp*CHS: Based on the CpCHS genes of 87 hawthorn resource sequenced, the SNPs of *Cp*CHS were defined by using the software of DnaSP (Rozas *et al*., 2003). Parameter of π was used for evaluating the nucleotide diversity which represent the average number of nucleotide differences per site between two sequences (Nei *et al*., 1979).Parameter of θ was used for calculated the neutral mutation from the total number of mutations θ (Watterson, 1979). Tajima's D test and Fu and Li's D* test were used to estimate the neutrality of SNP polymorphisms (Francesco *et al*., 2010).

Population structure and association analysis: In this study, 249 SRAP molecular markers were used to infer the population structure by using STRUCTURE 2.1 (Pritchard *et al*., 2000; Falush *et al*., 2003) software aimed to avoid the false positive associations and get the Q-matrix (Barnaud *et al*., 2006). The SNPs, Q-matrix and flavonoid content phenotypic data were then used to conduct the association analysis undergoing General Liner Model(GLM) in TASSEL2.1 software (Bradbury *et al*., 2007).

Results

Phenotypic data analyses: 5 flavonoid monomers detected in this study were vitexin-rhamnoside, vitexin, rutin, hyperin and quercetin, according to the flavonoid content data achieved in the year of 2014 and 2015 (Table 2), vitexin-rhamnoside was the major monomer among these 5 with 78.15 and 84.54 percentage of total flavonoid content in 2014 and 2015 respectively. Vitexin was 1.4% and 2.10%, rutin was 3.52% and 3.40%, hyperin was 16.64% and 9.62%, quercetin was 0.29% and 0.30%. Vitexin-rhamnoside content ranged from 0.04% to 0.52% in 2014 and 0.12% to 0.64% in 2015 (Fig. 1); vitexin content ranged from 0.002% to 0.090% in 2014 and 0.001% to 0.068% in 2015 (Fig. 2); Rutin content ranged from 0.004% to 0.108% in 2014 and 0.0007% to 0.05% in 2015 (Fig. 3); hyperin content ranged from 0.0% to 0.52% and 0.003% to 0.232% in 2015 (Fig. 4); quercetin content ranged from 0.00025% to 0.00298% in 2014 and 0.0006% to 0.0134% in 2015 (Fig. 5).

***Cp*CHS gene clone and nucleotide diversity analyses:** We first time finished amplified full-ORF of CHS gene from the cDNA retrotranscribed from the RNA of leaves in 87 hawthorn accessions, two alleles A and B of *Cp*CHS were then cloned and sequenced. The full length of *Cp*CHS was 1170bp, with its protein sequence containing 389 amino acids. In total of 71 SNPs were identified and they were named according to the position on *Cp*CHS ORF (Table 3). SNP variation among the 87 hawthorn accessions was at an average of 1 SNP every 16.5bp. The ratio of synonymous to non-synonymous was 1.45:1. Nucleotide diversity (π=0.0123 θ=0.0460) were achieved as the SNPs in the entire region of *Cp*CHS. Tajima's D test revealed the significant deviation from the neutral expectation and resulted in a slightly negative value. Fu and Li's D* was also statistically significant with the positive value (Table 4).

Population structure analyses: Population structure analysis was used to avoid the possibility of false association result, in this study, 87 hawthorn accessions were grouped into 4 sub-populations (k=4) and the Q matrix was then used as covariate to conduct the association analysis.

Candidate gene association analyses: In total of 8 SNPs were associated with 5 flavonoid monomers content in 2014 and 2015. In 2014, 4 SNPs were associated with vitexin and quercetin content, thereinto, S18 was associated with vitexin content which at the site of 195($p<0.05$), S12 and S14 were significantly associated with vitexin content ($p<0.01$). S55 was associated with quercetin content. In 2015, 12 SNPs were associated with the content of vitexin-rhamnoside, vitexin, rutin, and hyperin. S13 was associated with vitexin-rhamnoside and vitexin content at site 108, S12, S13,S14,S18, S56 were associated with vitexin content at site 102, 108, 138, 195 and 768 respectively, among these SNPs, S56 was significantly associated vitexin content. S12 and S14 were also associated with rutin content, there into S12 reached at significant level. S18, S30, S32 and S56 were associated with hyperin content at site 195, 441, 492 and 768. S18 has also reached at the significant level (Table 5).

Discussion

The aim of our study was to investigate the connection between the SNP positions of candidate gene and flavones content based on the association genetics approach.

Flavones content variance analyses and high flavones content samples selection: Because of the influence of environment, 5 flavonoid monomers content were difference between 2014 and 2015. Average content of vitexin-rhamnoside, vitexin, rutin and quercetin in 2015 were 0.4000%, 0.0099%, 0.0161% and 0.0014% and higher than 2014 which were 0.2627%, 0.0084%, 0.0151% and 0.0010%. Hyperin average content in 2014 was 0.0745% higher than 0.0455% in 2014. This indicated that hyperin content responded to the environment factor which was different to the other 4 flavonoid monomers. Combined the phenotypic traits of 2014 and 2015, we selected some hawthorn accessions with high flavonoid monomer content, "fengshuishanzha" had the highest vitexin-rhamnoside content, 0.5370%; "duanzhishanlihong" had the highest vitexin and hyperin content, 0.07565% and 0.3760%; "huangbaoyuNO.1" had the highest rutin content and it was 0.0613%; "beijingdenglonghong" had the highest quercetin content ,0.0077%.

Table 2. Flavones content of hawthorn leaves determined in 2014 and 2015.

No.	Vitexin-rhamnoside (%)		Vitexin (%)		Rutin (%)		Hyperin (%)		Quercetin (%)	
					2014	2015	2014	2015	2014	2015
1.	0.4515	0.6224	0.0017	0.0052	0.0038	0.0177	0.0115	0.0191	0.0015	0.0015
2.	0.2788	0.5027	0.0017	0.0069	0.0121	0.0185	0.0263	0.0408	0.0004	0.0008
3.	0.1521	0.2931	0.0008	0.0064	0.0050	0.0235	0.0538	0.0579	0.0012	0.0008
4.	0.2180	0.4881	0.0017	0.0074	0.0082	0.0129	0.0262	0.0155	0.0003	0.0017
5.	0.1736	0.2428	0.0093	0.0152	0.0084	0.0209	0.0143	0.0404	0.0000	0.0008
6.	0.2231	0.4513	0.0267	0.0394	0.0000	0.0225	0.1672	0.0973	0.0000	0.0000
7.	0.3837	0.5092	0.0112	0.0145	0.0236	0.0234	0.0600	0.0849	0.0020	0.0013
8.	0.2811	0.4119	0.0131	0.0199	0.0138	0.0112	0.0970	0.0259	0.0010	0.0000
9.	0.3663	0.5768	0.0662	0.0540	0.0102	0.0474	0.0038	0.1080	0.0000	0.0000
10.	0.1991	0.5505	0.0000	0.0099	0.0123	0.0204	0.0452	0.0237	0.0016	0.0027
11.	0.3251	0.4085	0.0079	0.0123	0.0058	0.0106	0.0249	0.0271	0.0000	0.0000
12.	0.2204	0.3118	0.0035	0.0054	0.0203	0.0234	0.0732	0.0135	0.0020	0.0007
13.	0.2763	0.3022	0.0122	0.0035	0.0117	0.0007	0.0324	0.0206	0.0023	0.0027
14.	0.3149	0.3361	0.0155	0.0105	0.0257	0.0463	0.0068	0.0264	0.0014	0.0023
15.	0.1828	0.3457	0.0000	0.0478	0.0075	0.0152	0.0274	0.0503	0.0020	0.0013
16.	0.2586	0.5527	0.0004	0.0046	0.0047	0.0059	0.0189	0.0062	0.0012	0.0000
17.	0.1889	0.3120	0.0033	0.0048	0.0080	0.0049	0.0602	0.0071	0.0017	0.0041
18.	0.3063	0.4911	0.0034	0.0060	0.0138	0.0148	0.0327	0.0055	0.0014	0.0000
19.	0.0000	0.3067	0.0000	0.0039	0.0000	0.0160	0.0000	0.0485	0.0000	0.0010
20.	0.2634	0.2792	0.0047	0.0038	0.0295	0.0154	0.0843	0.0136	0.0017	0.0000
21.	0.2161	0.3723	0.0067	0.0091	0.0155	0.0440	0.0357	0.0433	0.0016	0.0000
22.	0.2451	0.4846	0.0000	0.0151	0.0121	0.0166	0.0341	0.0098	0.0012	0.0020
23.	0.2320	0.3287	0.0023	0.0048	0.0200	0.0129	0.0758	0.0255	0.0022	0.0000
24.	0.2370	0.2822	0.0085	0.0089	0.0299	0.0510	0.1715	0.2061	0.0018	0.0011
25.	0.2681	0.4674	0.0028	0.0089	0.0203	0.0061	0.1066	0.0352	0.0019	0.0009
26.	0.2042	0.4850	0.0000	0.0072	0.0108	0.0176	0.0353	0.0421	0.0011	0.0020
27.	0.1721	0.2554	0.0051	0.0082	0.0156	0.0345	0.0808	0.0557	0.0014	0.0075
28.	0.3996	0.4574	0.0826	0.0687	0.0000	0.0000	0.5200	0.2320	0.0016	0.0011
29.	0.2370	0.3683	0.0021	0.0080	0.0112	0.0105	0.0550	0.0136	0.0011	0.0008
30.	0.1518	0.4070	0.0000	0.0061	0.0123	0.0220	0.0567	0.0430	0.0022	0.0016
31.	0.3300	0.3078	0.0063	0.0052	0.0263	0.0226	0.1498	0.0693	0.0017	0.0025
32.	0.2773	0.4447	0.0000	0.0035	0.0088	0.0101	0.0331	0.0106	0.0024	0.0006
33.	0.2483	0.2872	0.0067	0.0078	0.0224	0.0277	0.0917	0.0970	0.0020	0.0134
34.	0.3829	0.3975	0.0181	0.0106	0.0305	0.0000	0.0606	0.0714	0.0016	0.0010
35.	0.3098	0.5459	0.0000	0.0052	0.0108	0.0171	0.0692	0.0227	0.0012	0.0013
36.	0.1596	0.2599	0.0037	0.0049	0.0120	0.0138	0.0429	0.0442	0.0000	0.0062
37.	0.2551	0.2613	0.0267	0.0278	0.0000	0.0013	0.0375	0.0560	0.0000	0.0021
38.	0.3772	0.5568	0.0003	0.0058	0.0105	0.0093	0.0278	0.0370	0.0015	0.0000
39.	0.0979	0.1583	0.0030	0.0050	0.0040	0.0026	0.0363	0.0226	0.0000	0.0012
40.	0.5220	0.3278	0.0175	0.0217	0.0245	0.0127	0.1095	0.1663	0.0000	0.0012
41.	0.3996	0.3314	0.0039	0.0032	0.0167	0.0079	0.0850	0.0194	0.0000	0.0014
42.	0.2756	0.4513	0.0094	0.0091	0.0119	0.0084	0.0308	0.0136	0.0030	0.0018
43.	0.3145	0.2118	0.0169	0.0150	0.0348	0.0374	0.2152	0.0701	0.0011	0.0008

Table 2. (Cont'd.).

No.	Vitexin-rhamnoside (%)		Vitexin (%)		Rutin (%)		Hyperin (%)		Quercetin (%)	
					2014	2015	2014	2015	2014	2015
44.	0.2338	0.2553	0.0000	0.0033	0.0250	0.0155	0.1439	0.0634	0.0021	0.0009
45.	0.1889	0.4759	0.0004	0.0273	0.0120	0.0065	0.0583	0.0179	0.0025	0.0021
46.	0.2793	0.3453	0.0041	0.0042	0.0110	0.0049	0.0209	0.0219	0.0022	0.0009
47.	0.1766	0.1555	0.0094	0.0048	0.0073	0.0190	0.0772	0.0780	0.0015	0.0020
48.	0.2682	0.4356	0.0039	0.0058	0.0188	0.0149	0.0824	0.0358	0.0027	0.0012
49.	0.2567	0.3982	0.0033	0.0045	0.0158	0.0051	0.0388	0.0109	0.0024	0.0000
50.	0.2624	0.2726	0.0074	0.0026	0.0046	0.0181	0.0211	0.0747	0.0017	0.0011
51.	0.3029	0.2245	0.0043	0.0071	0.0057	0.0167	0.0250	0.0389	0.0000	0.0012
52.	0.1430	0.3583	0.0034	0.0053	0.0050	0.0113	0.0980	0.0193	0.0026	0.0000
53.	0.2560	0.4060	0.0030	0.0052	0.0105	0.0062	0.0307	0.0139	0.0014	0.0000
54.	0.1759	0.4965	0.0000	0.0031	0.0071	0.0072	0.0685	0.0241	0.0018	0.0000
55.	0.3445	0.5484	0.0052	0.0080	0.0384	0.0338	0.1671	0.0771	0.0017	0.0010
56.	0.0875	0.1216	0.0034	0.0057	0.0000	0.0355	0.0253	0.0359	0.0000	0.0027
57.	0.2407	0.3602	0.0009	0.0040	0.0120	0.0089	0.0419	0.0346	0.0000	0.0007
58.	0.2264	0.5112	0.0010	0.0063	0.0126	0.0143	0.0473	0.0317	0.0000	0.0023
59.	0.1662	0.2495	0.0024	0.0032	0.0051	0.0021	0.0118	0.0065	0.0000	0.0000
60.	0.3612	0.6105	0.0083	0.0051	0.0134	0.0188	0.0855	0.0539	0.0007	0.0012
61.	0.3625	0.3921	0.0073	0.0060	0.0050	0.0127	0.0253	0.0677	0.0000	0.0000
62.	0.2088	0.4282	0.0044	0.0092	0.0175	0.0257	0.0397	0.0863	0.0000	0.0009
63.	0.4044	0.4547	0.0059	0.0045	0.0077	0.0123	0.0561	0.0549	0.0011	0.0009
64.	0.2280	0.6431	0.0011	0.0105	0.0121	0.0237	0.0520	0.0483	0.0000	0.0023
65.	0.2765	0.3470	0.0075	0.0053	0.1085	0.0142	0.1946	0.0359	0.0000	0.0014
66.	0.3128	0.4505	0.0063	0.0072	0.0069	0.0257	0.0219	0.0702	0.0004	0.0011
67.	0.2953	0.4236	0.0072	0.0065	0.0355	0.0172	0.1887	0.0537	0.0015	0.0000
68.	0.1780	0.4473	0.0022	0.0045	0.0175	0.0072	0.1171	0.0182	0.0000	0.0000
69.	0.2409	0.4356	0.0045	0.0063	0.0085	0.0073	0.0146	0.0144	0.0005	0.0020
70.	0.3468	0.5611	0.0055	0.0087	0.0123	0.0151	0.0355	0.0212	0.0008	0.0011
71.	0.1807	0.2661	0.0061	0.0054	0.0321	0.0255	0.2169	0.0768	0.0021	0.0000
72.	0.2152	0.3744	0.0089	0.0116	0.0055	0.0115	0.0617	0.0666	0.0000	0.0000
73.	0.0475	0.5424	0.0916	0.0262	0.0000	0.0047	0.1256	0.0118	0.0000	0.0007
74.	0.1773	0.4824	0.0000	0.0070	0.0093	0.0139	0.0634	0.0575	0.0000	0.0014
75.	0.2652	0.4424	0.0005	0.0030	0.0075	0.0175	0.0295	0.1185	0.0000	0.0008
76.	0.2431	0.4821	0.0048	0.0110	0.0128	0.0073	0.0627	0.0244	0.0012	0.0000
77.	0.2297	0.5433	0.0043	0.0101	0.0085	0.0116	0.0196	0.0217	0.0000	0.0007
78.	0.2334	0.4347	0.0009	0.0052	0.0056	0.0104	0.0227	0.0126	0.0004	0.0012
79.	0.3456	0.3429	0.0083	0.0065	0.0385	0.0214	0.2531	0.0858	0.0011	0.0009
80.	0.3235	0.4667	0.0054	0.0089	0.0137	0.0242	0.0251	0.0399	0.0006	0.0009
81.	0.2304	0.4175	0.0077	0.0068	0.0151	0.0099	0.0780	0.0225	0.0000	0.0000
82.	0.4427	0.4369	0.0071	0.0085	0.0184	0.0165	0.0608	0.0445	0.0014	0.0011
83.	0.2830	0.1987	0.0096	0.0013	0.0520	0.0138	0.2223	0.0235	0.0009	0.0000
84.	0.3076	0.4733	0.0015	0.0081	0.0111	0.0121	0.0595	0.0241	0.0000	0.0015
85.	0.3043	0.4273	0.0041	0.0044	0.0168	0.0103	0.0856	0.0302	0.0000	0.0009
86.	0.3315	0.4442	0.0085	0.0085	0.0185	0.0122	0.1581	0.0031	0.0007	0.0012
87.	0.2530	0.4676	0.0038	0.0067	0.0133	0.0168	0.0474	0.0486	0.0000	0.0118

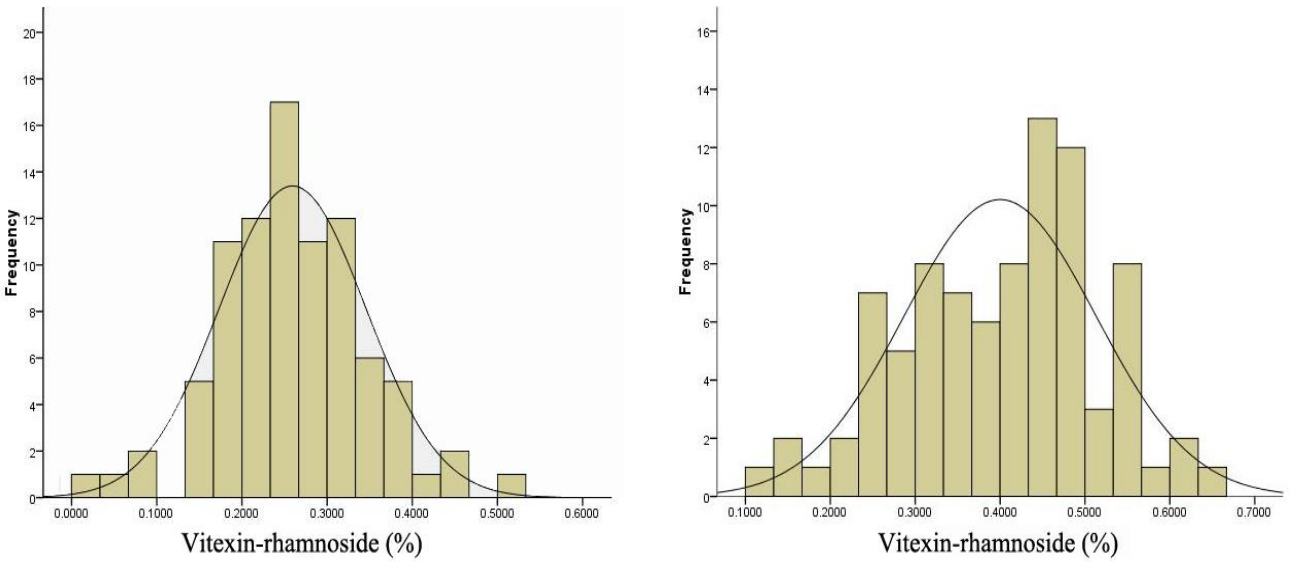

Fig. 1. Vitexin-rhamnoside content frequency in 2014 and 2015.

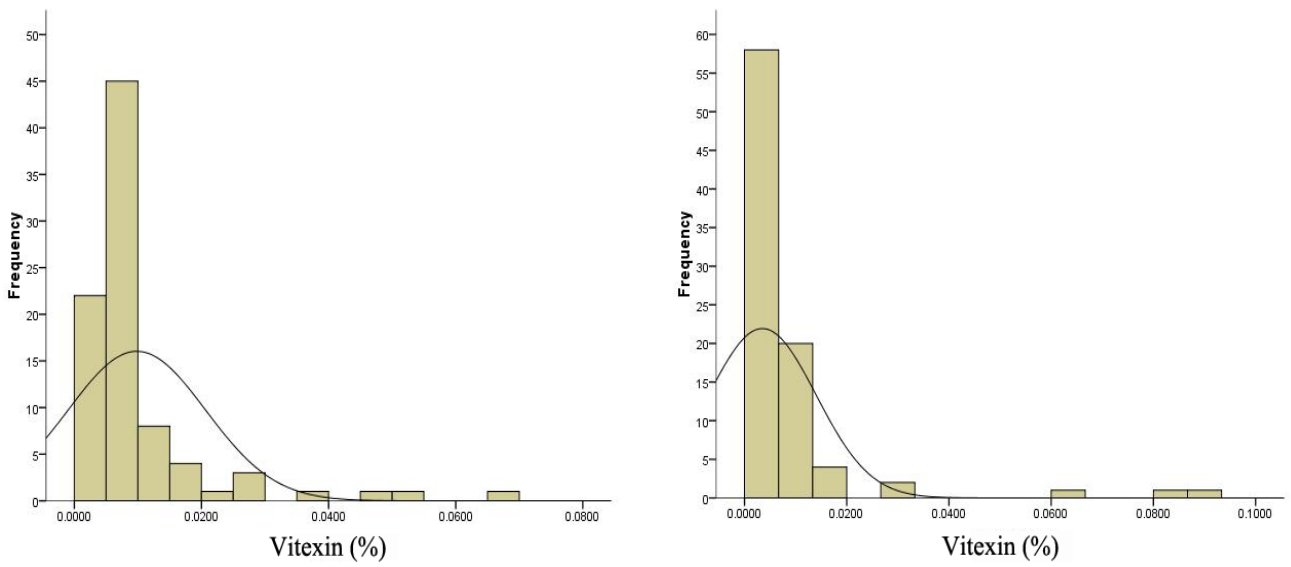

Fig. 2. Vitexin content frequency in 2014 and 2015.

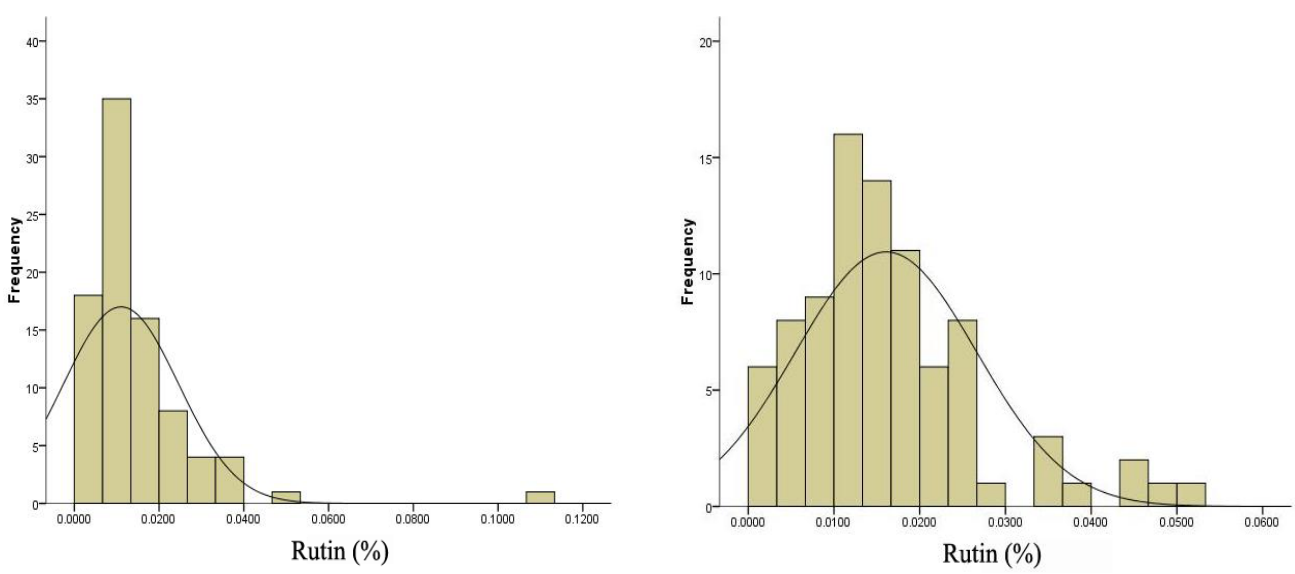

Fig. 3. Rutin content frequency in 2014 and 2015.

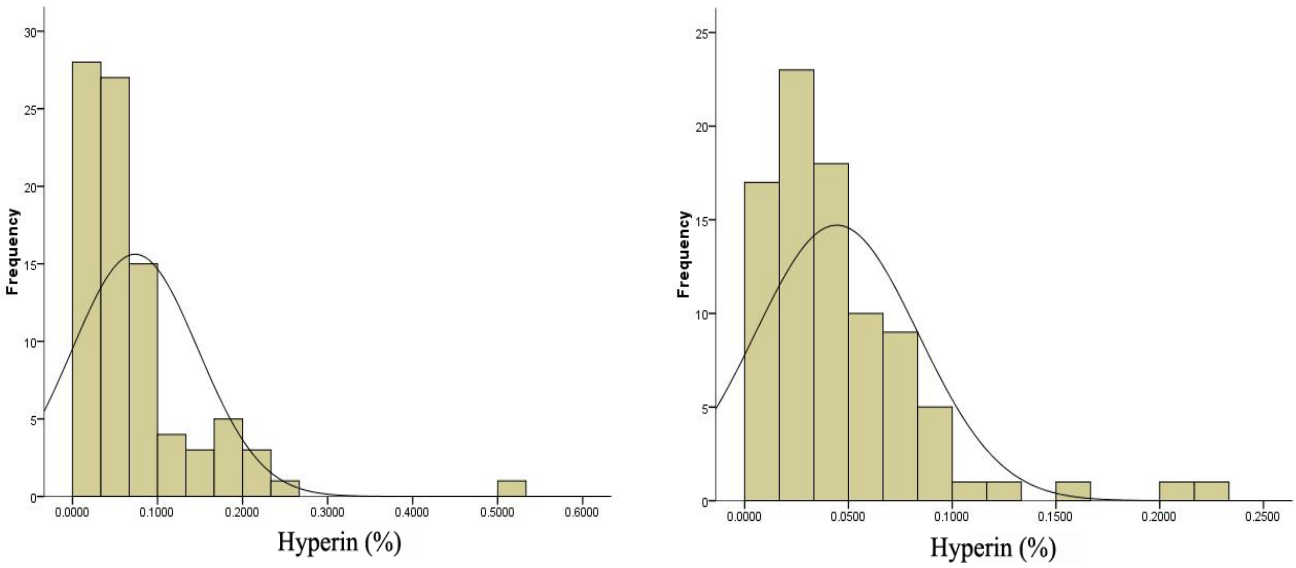

Fig. 4. Hyperin content frequency in 2014 and 2015.

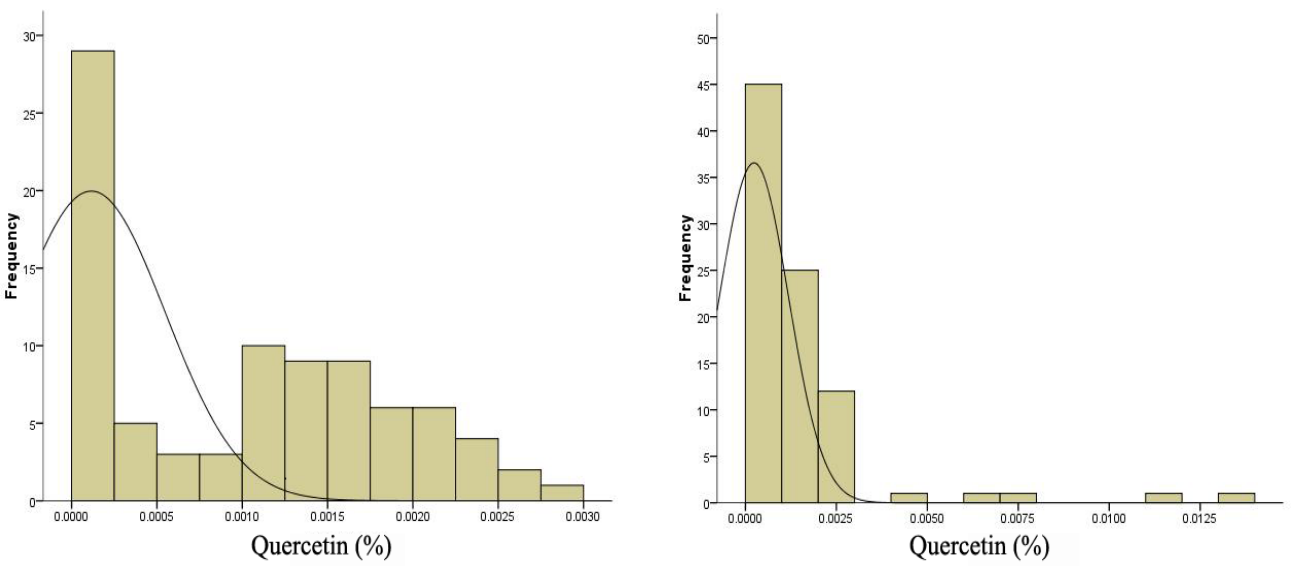

Fig. 5. Quercetin content frequency in 2014 and 2015.

The SNP frequency of *Cp*CHS: In total of 71 SNPs were detected in the ORF region of CHS cDNA, the frequency was 1SNP every 16.5bp which was higher than tartary buckwheat (Xia *et al*., 2012), sunflower (Fusari *et al*., 2008) and the same with tea plant (Zhang *et al*., 2014). And it was also slightly higher than other candidate genes reported in association analyses, such as the VvDXS in grapevine described by Lijavetsky *et al*., in 2007(1 SNP every 49bp) and Cunff *et al*., in 2008(1 SNP every 64bp) and Francesco *et al*., in 2010 (1 SNP every 51bp).

Candidate gene association analysis and excellent allele development: Molecular assisted selection (MAS) was an important approach to improve the breeding efficiency and so, developing molecular markers was a key procedure for molecular assisted selection. The aim of candidate gene association analysis was to discover the polymorphic nucleotide site leading to the phenotypic variation. In this study, we found three SNP positions S12, S14 and S18 which were associated with vitexin content both in 2014 and 2015, combined with phenotypic variance analysis, these SNPs may be recognized as stable positions which can generated different vitexin content. For S12, T/C was higher average vitexin content genotype than T/T, 23 hawthorn samples contained the genotype of T/C, the other 64 samples were T/T; For S14, T/C was higher average vitexin content genotype than T/T, 22 hawthorn samples which contained the genotype of T/C, the other 65 samples were T/T; for S18, A/A was higher average vitexin content genotype than C/C and A/C, 3 hawthorn samples contained the genotype of A/A, 6 samples were A/C and 78 samples were C/C. According to these three SNPs, S14 has the highest contribution rate, especially in 2015, the contribution rate can reach to 41.90%, so it can recognized as the major mutation site to help selecting the high vitexin content hawthorn samples. In our study, "duanzhishanlihong" was the selected hawthorn sample which had high vitexin content, but its genotype at S12 and S14 were T/T not T/C, this might be because as quantitative trait, the mutations in other candidate genes in flavones synthesis pathway could also contribute to the flavones content such as CHI, F3H, FLS, DFR, ANS and UFGT(Honda *et al*., 2002) and it was also influenced by the interaction between genes and environment.

Table 3. Full-ORFcDNA SNP message of CpCHS.

No.	Position (bp)	Genotype (number)	No.	Position (bp)	Genotype (number)
S01	33	A/A(70) G/G(17)	S37	522	C/C(73)T/C(14)
S02	34	C/C(68)A/C(19)	S38	537	G/G(71)A/G(16)
S03	51	A/A(64) A/C (23)	S39	555	T/T(71)T/C(16)
S04	54	T/T(64) A/T (21) G/T (2)	S40	610	G/G(69)A/G(18)
S05	57	T/T(64)T/C(21)A/T(1)	S41	621	C/C(72)T/C(15)
S06	60	G/G(68)A/G(19)	S42	633	T/T(72)T/C(15)
S07	63	C/C(65)T/C(22）	S43	639	C/C(72)T/C(15)
S08	72	A/A(69)A/T(18)	S44	648	C/C(72)T/C(15)
S09	87	C/C(66)T/C(21)	S45	676	T/T(70)G/T(17)
S10	93	G/G(66) A/G(21)	S46	685	G/G(71)G/T(16)
S11	99	A/A(73)A/G(14)	S47	696	C/C(71)T/C(16)
S12	102	T/T(64)T/C(23)	S48	708	G/G(74)A/G(13)
S13	108	C/C(55)T/T(27)T/C(5)	S49	714	A/A(72)A/G(15)
S14	138	T/T(65)T/C(22)	S50	715	T/T(72)T/C(15)
S15	148	G/G(65)A/G(22)	S51	723	G/G(71)G/T(15)A/G(1)
S16	149	C/C(65)T/C(22)	S52	726	A/A(70)A/G(2)A/T(15)
S17	168	C/C(66)T/C(21)	S53	729	A/A(72)T/T(1)A/C(14)
S18	195	C/C(78)A/A(3)A/C(6)	S54	738	T/T(71)T/C(16)
S19	198	A/A(65)A/G(22)	S55	741	T/C(11)G/G(6)T/T(26)G/T(44)
S20	229	T/T(68)T/C(19)	S56	768	C/C(64)A/C(18)T/C(3) T/T(2)
S21	237	G/G(70)A/G(17)	S57	774	C/C(70)T/C(16)A/T(1)
S22	243	A/A(70)A/G(17)	S58	783	A/A(71)C/C(2) A/T(14)
S23	270	A/A(11)G/G(30)A/G(46)	S59	804	C/C(64)G/G(11)S(11)A/C(1)
S24	324	G/G(71)A/G(16)	S60	820	A/A(69)A/C(17)G/C(1)
S25	327	T/T(71)G/T(16)	S61	828	C/C(70)G/C(15)A/C(2)
S26	339	C/C(71)A/C(16)	S62	834	T/T(69)T/C(18)
S27	417	C/C(72)T/C(15)	S63	837	T/T(69)T/C(18)
S28	423	T/T(71) T/C(14) C/C(2)	S64	876	T/T(71)T/C(16)
S29	429	A/A(71) A/G(14) G/G(1)A/C(1)	S65	957	T/T(68) G/T(17)C/C(1)G/C(1)
S30	441	C/C(72) A/C(15)	S66	1042	C/C(26)G/G(8)G/C(53)
S31	489	T/T(69)C/C(1)T/C(17)	S67	1044	A/A(70)A/G(17)
S32	492	C/C(72) T/C(15)	S68	1062	T/T(60)C/C(1)T/C(26)
S33	495	C/C(73)T/C(14)	S69	1076	A/A(9)A/G(18)
S34	498	C/C(72)T/C(15)	S70	1107	A/A(70)A/G(17)
S35	501	C/C(71)A/C(16)	S71	1116	A/A(70)A/G(17)
S36	504	G/G(72)G/C(15)			

Note: 1. "NO." means number of polymorphic SNPs; 2. "Position" means polymorphic SNP site in CHS gene; 3. "Genotype number" means different SNP genotype number of 87 hawthorn accessions

Table 4. Gene diversity analyses of *Cp*CHS.

Paremeter	cDNA ORF	SNP position	SNP frequency	Synonymous VS Nonsynonymous	π/θ	Tajima D	Fu and Li's D *
Value	1170 bp	71	1/16.5	1.45/1	0.0123/0.0460	-2.31 (p<0.001)	10.11 (p<0.02)

Some SNP positions were just associated with phenotypic flavones content in 2014 or 2015. S12 and S14 were associated with rutin content in 2015, and the association analyses of S12 reached to significant level (P=0.0088). The contribution rate was higher than S14, so the genotype of T/T at S12 could be developed as an important SNP site to identify high rutin content hawthorn samples. S13 was the only SNP site, associated with vitexin-rhamnoside content in 2015, C/C was the highest vitexin-rhamnoside content genotype, the highest vitexin-rhamnoside content sample "fengshuishanzha" was also present in this group, besides, S13 was also associated with vitexin content in 2015, but its contribution rate was lower than S12 ,S14 and S18. S18, S30, S32 and S56 were associated with hyperin content in 2015, among them, S56 shared the highest contribution rate, C/C was the highest hyperin genotype, the highest hyperin content sample "duanzhishanlihong" was also present in this group, so the genotype of C/C at S56 could be developed as an important SNP site to identify high hyperrin content hawthorn samples. S55 was just detected to be associated with quercetin content in 2014, the genotype of T/T shared the highest quercetin content, but the highest quercetin content sample "beijingdenglonghong" was not contained in this group.

Table 5 Association analyses of SNP markers with flavones content in 2014 and 2015.

Site	Trait	P value		Phenotypic difference (%)								Contribution (%)	
		2014	2015	2014	2015	2014	2015	2014	2015	2014	2015	2014	2015
S12	Vitexin	0.0022**	0.0487*	T/T(64) 0.0072	T/T(64) 0.0098	T/C(23) 0.0166	T/C(23) 0.0103					8.59	4.00
	Rutin		0.0088**		T/T(65) 0.0165		T/C(22) 0.0148						10.45
S13	Vitexin-rhamnoside		0.0482*		C/C(55) 0.4058		T/C(5) 0.3706		T/T(27) 0.3934				3.89
	Vitexin		0.0482*		C/C(55) 0.0203		T/C(5) 0.0107		T/T(27) 0.0065				3.89
S14	Vitexin	0.0019**	0.0431*	T/T(65) 0.0070	T/T(65) 0.0098	T/C(22) 0.0112	T/C(22) 0.0103					8.67	41.90
	Rutin		0.0114*		T/T(65) 0.0170		T/C(22) 0.0134						9.69
S18	Vitexin	0.024*	0.01*	A/A(3) 0.0296	A/A(3) 0.0266	A/C(6) 0.0064	A/C(6) 0.0076	C/C(78) 0.0073	C/C(78) 0.0095			5.34	5.05
	Hyperin		0.0084*		A/A(3) 0.1148		A/C(6) 0.0373		C/C(78) 0.0435				6.96
S30	Hyperin		0.0484*		A/C(15) 0.0409		C/C(72) 0.0465						4.77
S32	Hyperin		0.0484*		T/C(15) 0.0409		C/C(72) 0.0465						4.77
S55	Quercetin	0.0191*		T/T(26) 0.0080		T/C(11) 0.0008		T/G(44) 0.0011		G/G(6) 0.0017		20.08	
S56	Vitexin		0.0022**		A/C(18) 0.0076		T/C(3) 0.0091		T/T(2) 0.0467		C/C(64) 0.0095		12.99
	Hyperin		0.0434*		A/C(18) 0.0279		T/C(3) 0.0414		T/T(2) 0.1027		C/C(64) 0.0489		8.57

Note: 1: "Phenotypic difference" showed the alleles respond to average flavones content; 2:"*"means association level at $p<0.05$; "**" means association level at $p<0.01$

Acknowledgement

Research supported by the National Natural Science Funds of China (grant #31101515) and Natural Science Founds of Liaoning Province of China (grant #20170540804).

References

Akira, N., I. Yoko and Y. Masumi. 2003. Spatial and temporal expression of chalcone synthase and dihydroflavonol 4-reductase genes in the Asiatic hybrid lily. *Plant Sci*., 165(4): 759-767.

Atwell, S., Y.S. Huang, B.J. Vilhjálmsson, G, Willems, A.M. Tarone, T.T. Hu, R. Jiang, N.W. Muliyati, X. Zhang, M.A. Amer, I. Baxter, B. Brachi, J. Chory, C. Dean, M. Debieu, J. Meaux, J.R. Ecker, N. Faure, J.M. Kniskern, J.D. Jones, T. Michael, A. Nemri, F. Roux, D.E. Salt, C. Tang, M. Todesco, M.B. Traw, D. Weigel, P. Marjoram, J.O. Borevitz, J. Bergelson and M. Nordborg. 2010. Genome-wide association study of 107 phenotypes in a common set of Arabidopsis thaliana inbred lines. *Nature*, 465: 627-631.

Barnaud, A., T. Lacombe and A. Doligez. 2006. Linkage disequilibrium in cultivated grapevine *Vitis vinifera* L. *Theor. Appl. Genet*., 112(4): 708-716.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss and E.S. Buckler. 2007. Tassel: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19): 2633-2635.

Chang, Q., Z. Zuo, F. Harrison and M.S.S. Chow. 2002. Hawthorn. *Eur. J. Clin. Pharmacology*, 42(6): 605-612.

Corder, E.H., A.M. Saunders, N.J. Risch, W.J. Strittmatter, D.E. Schmechel and G.P. Jr. 1994. Protective effect of Apo lipoprotein E type 2 allele for late ons*et al*zheimer disease. *Nat. Genet*., 7(2): 180-184.

Cunff, L.L., A. Fournier-Level, V. Laucou, S. Vezzulli, T. Lacombe, A.F.A. Blondon, J.M. Boursiquot and P. This. 2008. Construction of nested genetic core collections to optimize the exploitation of natural diversity in *Vitis vinifera* L. subsp. *sativa*. *BMC Plant Biol*., 8(1): 31.

Dixon, R.A. and C.L. Steele. 1999. Flavonoids and isoflavonoids - a gold mine for metabolic engineering. *Trends. Plant Sci*., 4(10): 394-400.

Dixon, R.A. and N.L. Paiva. 1995. Stress-induced phenylpropanoid metabolism. *Plant Cell*., 7(7): 1085-1097.

Ehrenreich, I.M., Y. Hanzawa, L. Chou, J.L. Roe, P.X. Kover and M.D. Purugganan. 2009. Candidate gene association mapping of Arabidopsis flowering time. *Genetics*., 183(1): 325-335.

Epping, B., M. Kittel, B. Ruhnau and V. Hemleben. 1990. Isolation and sequence analysis of a chalcone synthase cDNA of matthiola incana, R. Br. (brassicaceae). *Plant Mol. Biol*., 14(6): 1061-1063.

Falush, D., M. Stephens and J.K. Pritchard. 2003. Inference of population structure using multilocus genotype date: Linked loci and correlated allele frequencies. *Genetics*, 164(4): 1567-1587.

Feinbaum, R.L. and F.M. Ausubel. 1988. Transcriptional regulation of the *Arabidopsis thaliana* chalcone synthase gene. *Mol. Cell. Biol*., 8(8): 1985-1992.

Francesco, E., B. Juri, C. Laura, L.L. Cunff, J.M. Boursiquot, P. This and M.S. Grando. 2010. A candidate gene association study on muscat flavor in grapevine (*Vitis vinifera* L.). *BMC Plant Biol*., 10(1): 241.

Fusari, C.M., V.V. Lia, H.E. Hopp, R.A. Heinz and N.B. Paniego. 2008. Identification of single nucleotide polymorphisms and analysis of linkage disequilibrium in sunflower elite inbred lines using the candidate gene approach. *BMC Plant Biology*, 8(1): 1-14.

Han, Y.Y., F. Ming, W. Wang, J.W. Wang and M.M. Ye. 2006. Molecular evolution and functional specialization of chalcone synthase superfamily from Phalaenopsis Orchid. *Genetica*, 128(2): 429-438.

Heller, W. and K. Hahlbrock. 1980. Highly purified "flavanone synthase" from parsley catalyzes the formation of naringenin chalcone. *Arch. Biochem. Biophysics*., 200(2): 617-619.

Hirschhorn, J.N. and M.J. Daly. 2005. Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet*., 6(2): 95-108.

Honda, C., N. Kotoda, M. Wada, S. Kondo, S. Kobayashi, J. Soejima, Z. Zhang, T. Tsuda and T. Moriguchi. 2002. Anthocyanin biosynthetic genes are coordinately expressed during red coloration in apple skin. *Plant Physiol. Bioch.*, 40(11): 955-962.

Jahan, N., U.R. Khalil, S. Ali and M.R. Asi. 2013. Phenolic acid and flavonol contents of gemmo-modified and native extracts of some indigenous medicinal plants. *Pak. J. Bot*., 45(5): 1515-1519.

Jahnen, W. and K. Hahlbrock. 1988. Differential regulation and tissue-specific distribution of enzymes of phenylpropanoid pathways in developing parsley seedlings. *Planta*., 173(4): 453-458.

Kao, E.S., C.J. Wang, W.L. Lin, Y.F. Yin and C.P. Wang. 2005. Anti-inflammatory potential of flavonoid contents from dried fruit of crataegus pinnatifida in vitro and *İn vivo*. *J. Agric. Food Chem*., 53(2): 430-436.

Koes, R.E., C.E. Spelt and J.N.M. Mol. 1989. The chalcone synthase multigene family of petunia hybrida (v30): differential, light-regulated expression during flower development and UV light induction. *Plant. Mol. Biol. Rep*., 12: 213-225.

Kraakman, A.T.W., R.E. Niks, V.D.B. Petra, P. Stam and A.V.E. Fred. 2004. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics*, 168(1): 435-446.

Lijavetzky, D., J.A. Cabezas, A. Ibáñez, V. Rodríguez and J.M. Martinezzapater. 2007. High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and snplex technology. *BMC Genomics*, 8(1): 424.

Lipphardt, S., R. Brettschneider, F. Kreuzaler, J. Schell and J.L. Dandl. 1988. Uv-inducible transient expression in parsley protoplasts identifies regulatory cis-elements of a chimeric antirrhinum majus chalcone synthase gene. *Embo. J*., 7(7): 4027-4033.

Meer, I.M.V.D., C.E. Spelt, J.N. Mol and A.R. Stuitje. 1990. Promoter analysis of the chalcone synthase (chs A) gene of petunia hybrid: a 67 bp promoter region directs flower-specific expression. *Plant Mol. Biol*., 15(1): 95-109.

Nei, M. and W.H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Science*, 76(10): 5269-5273.

Pritchard, J.K., M. Stephens and P. Donnely. 2000. Inference of population structure using multilocus genotype date. *Genetics*, 155: 945-959.

Rigelsky, J.M. and B.V. Sweet. 2002. Hawthorn: Pharmacology and therapeutic uses. *Am J Health-Syst Pharm*., 59: 417-422.

Rozas, J., J.C. Sánchez-DelBarrio, X. Messeguer and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, 19(18): 2496-2497.

Schmid, J., P.W. Doerner, S.D. Clouse, R.A. Dixon and C.J. Lamb. 1990. Developmental and environmental regulation of a bean chalcone synthase promoter in transgenic tobacco. *Plant Cell*., 2(7): 619-631.

SherryA, F.G., T, Anne-Céline, Y. Jianming, P. Gael, S.M. Romero, E.M. Sharon, D. John, K. Stephen, M.G. Major and S.B. Edward. 2005. Maize association population: A high-resolution platform for quantitative trait locus dissection. *Plant J*., 44(6): 1054-1064.

Tabor, H.K., N.J. Risch and R.M. Myers. 2002. Candidate-gene approaches for studying complex genetic traits: Practical considerations. *Nat. Rev.Genet.*, 3(5): 391-397.

Templeton, A.R., K.A. Crandall and C.F. Sing. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. İİİ. cladogram estimation. *Genetics*, 132(2): 619-633.

Thornsberry, J.M., M.M. Goodman, J. Doebley, S. Kresovich, D. Nielsen and E.S. Buckler. 2001. Dwarf8 polymorphisms associate with variation in flowering time. *Nat. Genet*., 28(3): 286-289.

Watterson, G.A. 1975. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol*., 7(2): 256-276.

Wienand, U., U. Weydemann, U. Niesbach-Klösgen, P.A. Peterson and H. Saedier.1986. Molecular cloning of the C_2 locus of zea mays, the gene coding for chalcone synthase. *Mol. Gen. Genet*., 203(3): 202-207.

Xia, N. 2012. Analysis of phenotypic diversity of tartary buckwheat CHS gene diversity and flavonoid content. Shanxi University. (In China).

Yang, J and H.Y. Guo. 2006. Duplication and divergent evolution of the CHS and CHS-like genes in the chalcone synthase (CHS) superfamily. *Chinese. Sci. Bull*., 51(5): 505-509.

Zhang, L.Q., K. Wei, L.Y. Wang, H. Cheng, B.Y. Liu and W.Y. Gong. 2014. The structure and single nucleotide polymorphism analysis of chalcone synthase genes in tea plant (*Camellia sinenesis*). *Scientia Agric. sin*., 47(1): 133-144.

Zhao, Y., K. Su, G. Wang, Z.D. Liu, W.X. Dong and Y.S. Guo. 2014 Genetic diversity of Flavonoid content in leaf of Hawthorn resources. *Pak. J. Bot.*, 46(5): 1543-1548.

Zhu, C.S., G. Michael, S.B. Edward and J.M. Yu. 2008. Status and prospects of association mapping in plants. *Plant Genome*., 1(1): 5-20.

(Received for publiation 7 November 2016)