Plant Breeding, **135**, 399–404 (2016)

doi:10.1111/pbr.12369

Association mapping of leaf traits in spinach (*Spinacia oleracea* L.)

Jianbing Ma[1], Ainong Shi[1,5], Beiquan Mou[2,5], Michael Evans[1], John R. Clark[1], Dennis Motes[3], Jim C. Correll[4], Haizheng Xiong[1], Jun Qin[1], Jessica Chitwood[1] and Yuejin Weng[1]

[1]Department of Horticulture, University of Arkansas, Fayetteville, AR 72701, USA; [2]Crop Improvement and Protection Research Unit, USDA-ARS, Salinas, CA 93905, USA; [3]Vegetable Research Center, University of Arkansas, Alma, AR 72921, USA; [4]Department of Plant Pathology, University of Arkansas, Fayetteville, AR 72701, USA; [5]Corresponding authors, E-mails: ashi@uark.edu; beiquan.mou@ars.usda.gov

With 3 figures and 3 tables

Received October 10, 2015 / Accepted February 13, 2016
Communicated by M. Havey

Abstract

Spinach (*Spinacia oleracea* L.) is an important leafy vegetable crop grown worldwide. Leaf traits, surface texture (smooth, savoy or semi-savoy), petiole colour (different shades of green vs. purple) and edge shape (serrate vs. entire), are important commercial traits of spinach. Association mapping for the three traits was conducted on 323 USDA spinach germplasm accessions, originally collected from 33 countries and representing the entire USDA spinach germplasm collection. The majority of accessions were from Europe (36.3%), Asia (25.3%) and North America (15.8%). The majority of the spinach accessions (82.0%) were smooth (unwrinkled types), whereas the savoy and semi-savoy types (wrinkled types) accounted for 18.0%. The collection contained 74.9% green petiole types, while the purple petioles consisted of 25.1%. The collection consisted of 27.2% serrated leaf types and 72.8% entire leaf edge types. Genotyping-by-sequencing (GBS) was used for single nucleotide polymorphism (SNP) discovery, and SNPs were used as genotypic data to conduct genetic diversity and association mapping of the three leaf traits. Five genetic subpopulations and principal components (PCs) were postulated by STRUCTURE 2 and JMP Genomics 7 for this association panel. Five, seven and 14 SNPs were identified to be associated with surface texture, edge shape and petiole colour, respectively. This study provides us an approach to identify SNP markers through association analysis in spinach and thus leads to select these three leaf traits through marker-assisted selection in spinach breeding programme.

Key words: spinach — *Spinacia oleracea* — association mapping — single nucleotide polymorphism — genotyping-by-sequencing — leaf surface texture — petiole colour — edge shape

In molecular plant breeding, marker-assisted selection (MAS) utilizes molecular markers that are tightly linked to target loci that control phenotypic traits to more accurately and efficiently select for these traits. Thus, the knowledge of the genetic basis of phenotypic traits is very important for MAS. Mapping of quantitative trait loci (QTL), often referred to as 'family mapping', aims to identify the QTL that are responsible for phenotypic variation and usually involves the development of the segregating populations derived from parents with contrasting phenotypes (Myles et al. 2009). However, the main downside for utilizing linkage mapping approach is the relatively low mapping resolution and the limited recombination within the biparental population (Hall et al. 2010). In association mapping, also known as 'population mapping' (Myles et al. 2009), however, a natural population with unknown relatedness is evaluated to determine the marker–trait associations based on linkage disequilibrium (LD, the non-random association of alleles at different loci) (Zondervan and Cardon 2004). In comparison with linkage mapping, the main advantage of association mapping is that it is based on the phenotypic variation in the collections of natural genetic resources and assesses the entire genome for trait-associated variants rather than analysing specific genes/QTL, thus allowing detection of genes/QTL which would escape from linkage-based studies, resulting in a higher mapping resolution (Neale and Savolainen 2004). The disadvantage of association mapping is that the structure in the large-scale population can produce spurious marker–trait associations without physical linkage information (Pritchard et al. 2000, Buckler and Thornsberry 2002). However, by combining the relatedness among individuals (e.g. kinship) and the population structure, the chances for false correlation between markers and phenotypic traits have been reduced (Aranzana et al. 2005, Yu et al. 2006). Association mapping becomes the alternative strategy for conventional linkage-based mapping to study the genetic basis of the phenotypic variation in different plants (Goldstein and Weale 2001). To date, many plant association studies for different phenotypic traits have been reported, such as flowering time (Zhao et al. 2007), leaf and plant architecture (Tian et al. 2011, Wei et al. 2014) and fruit quality (Xu et al. 2013).

Single nucleotide polymorphisms (SNPs) are the most common types of DNA polymorphism in plant genomes (Mammadov et al. 2012, Thomson 2014). Using SNPs derived from the known genomic locations enables one to accelerate the understanding of the nucleotide diversity levels, the background patterns of LD and the relatedness among individuals within specific populations (Mandel et al. 2013). Therefore, SNP markers have the potential to allow the genetic dissection for crop phenotypes and ultimately facilitate marker-assisted breeding (Collard and Mackill 2008). Nowadays, due to the increasing availability of inexpensive DNA sequencing and genotyping methods (e.g. genotyping-by-sequencing), SNPs have been extensively used in genetic studies, including association analysis of candidate genes in phenotypic variation in rice (Cao et al. 2006) and linkage disequilibrium-based association mapping in potato (Achenbach et al. 2009) and sunflower (Mandel et al. 2013). Genotyping-by-sequencing (GBS) is one of the next-generation sequencing platforms for genomewide SNP discovery which can be used in association mapping (Elshire et al. 2011, Sonah et al. 2013).

Spinach (*Spinacia oleracea* L.) is an important leafy vegetable crop that is grown worldwide both in temperate regions and in the cooler parts of the tropic regions (Siemonsma and Piluek 1993). In the United States, 48390 acres of spinach were harvested in 2015, and the gross production value (fresh and processing) was about $272.8 million (NAAS, 2015). Leaf

morphological traits are critical phenotypic traits with considerable commercial value, particularly for precleaned package salad mixes. Currently, there are three major spinach leaf types in the US market: (i) savoy type with wrinkled and curly leaves, (ii) semi-savoy type with slightly crinkled leaves, and (iii) smooth type with flat and unwrinkled leaves. Savoy and semi-savoy leaf types of spinach are popular for fresh-market consumption, but smooth types of spinach are more suitable for processing for bagged salad and freeze products because the uncrinkled leaves are easier to wash. Other leaf traits such as the petiole colour (purple or green) and edge shape (entire or serrated) are also appealing for consumers and should be considered during spinach breeding. A few genetic studies have been conducted to evaluate the genetic variability of spinach germplasm collections using either target region amplification polymorphism (TRAP) markers or simple sequence repeat (SSR) markers (Hu et al. 2007, Khattak et al. 2007). However, association analysis between SNPs markers and important leaf phenotypic traits in spinach has not been performed. Therefore, the objective of this study was to conduct association mapping for spinach leaf morphology using the USDA spinach germplasm collections.

Materials and Methods

Association panel: A panel of 323 spinach accessions, which was obtained from the USDA Germplasm Resources Information Network (USDA-GRIN), was used for association analysis. These accessions were collected from 33 different countries and regions and represented the entire USDA spinach germplasm collection. The majority of accessions were from Europe (36.3%), Asia (25.3%) and North America (15.8%).

Phenotypic evaluation: All spinach accessions were planted in the field at the Vegetable Research Station, University of Arkansas during the 2013–2014 winter season. Each accession was planted in a single row with 5 m length and 1 m width and the plant distance in the row was approximately 8 cm apart. Leaf morphological traits including leaf petiole colour (purple or green), leaf edge shape (entire or serrated) and leaf texture (savoy, semi-savoy or smooth) were recorded. In this study, leaf edge shape was determined as entire if the most extended leaves at the base presented a smooth edge; otherwise, leaf edge shape was considered as serrated. Leaf texture was classified based on the most extended leaves on the top.

DNA extraction and genotyping-by-sequencing (GBS): Leaf sample was taken from a bulk of 10 plants with uniform leaf morphological traits from each entry. Genomic DNA was extracted following the CTAB (cetyltrimethylammonium bromide) method (Hulbert and Bennetzen 1991). DNA concentrations were determined using a NanoDrop 2000c spectrophotometer (Thermo Scientific, Wilmington, DE, USA). DNA qualities were checked on 1% agarose gels with EtBr (ethidium bromide) gel stain. DNA library was prepared using the restriction enzyme *ApeKI* following the GBS protocol described by Elshire et al. (2011). The 90-bp double-end sequencing was performed on each spinach line using GBS protocol by HiSeq 2000 in BGI Genomics Research Institute-Hong Kong.

Genotypic data analysis: GBS data were analysed for SNPs by SOAPsnp pipeline, which was a member of the Short Oligonucleotide Analysis Package (SOAP) (http://soap.genomics.org.cn/index.html) using the spinach genome sequences AYZV01 and AYZV02 as reference (http://www.ncbi.nlm.nih.gov/Traces/wgs/?val=AYZV01 and http://www.ncbi.nlm.nih.gov/Traces/wgs/?val=AYZV02). SNP markers with minimum allele frequencies (MAF) lower than 5%, more than 5% missing data or heterozygous genotype >50% were discarded from the statistical analysis. The remaining high-quality SNP markers were used for population structure and marker–trait association analyses.

Population structure: Selected SNP markers from GBS were used to run the following programs to determine the appropriate population structure and relative kinship. Principal component analysis (PCA) and multidimensional scaling (MDS) were performed to estimate the genetic relatedness between 323 spinach accessions and *Q*-matrix using JMP® Genomics 7 (SAS Institute, Cary NC). The first 10 eigenvectors of PCA were calculated from the correlation matrix derived from SNP genotypes. STRUCTURE v2.3.4 program (Pritchard et al. 2000) was used to assess the population structure. The number of subgroups (K) was evaluated from 1 to 10 with the initial burning period set to 100 000 with 100 000 MCMC (Markov chain Monte Carlo) repeats. The number of population was determined based on the highest likelihood value of $L(K)$, which is represented as $LnP(D)$ in STRUCTURE (Evanno et al. 2005). The minimum subgroup number was confirmed by the MDS analysis carried out by JMP® Genomics 7 (SAS Institute, 2010).

Association analysis: All marker–trait association analyses were performed by JMP® Genomics 7 (SAS Institute 2010), and Genome Association and Prediction Integrated Tool (GAPIT), which is a statistical package that is run in the R software environment (Lipka et al. 2012). Two different association mapping models were tested for each trait in which the general linear model (GLM) with population structure (*Q*-matrix) and the mixed linear model (MLM) combining kinship (*K*-matrix) with population structure (*Q*-matrix) (Yu et al. 2006) were run in JMP® Genomics 7, and the compressed MLM (CMLM) (Zhang et al. 2010) was run in GAPIT. The P values to determine the significant marker–trait associations were set as lower than 5×10^{-3} for GLM, MLM and CMLM. The quantile–quantile (QQ) plots which plotted the observed P values from the association analysis against an expected (cumulative) probability distribution were illustrated for linear model testing. The models were considered to have fewer false positives and produce more significant results than expected by chance if followed the expected line more closely.

Results

Phenotypic variation

The majority of the spinach accessions (82.0%) were smooth and unwrinkled type, and the savoy and semi-savoy types together accounted for 18.0% (Table 1). For petiole colour, 242 lines (74.9%) from the entire collection showed green petioles, while the remaining 81 lines (25.1%) had purple petioles. The numbers of accessions with serrated and entire leaf edges were 88 and 235, which accounted for 27.2% and 72.8%, respectively.

GBS output

After the pooled GBS library was sequenced, a total of 204 024 SNP markers were discovered when MAF ≥ 5% for the 323 lines. After excluding the SNPs with more than 5% missing data and heterogeneous loci > 50%, a total of 4077 SNP markers were used to perform population structure analysis

Table 1: Three leaf phenotypic traits and their percentages among 323 spinach (*Spinacia oleracea* L.) accessions

Surface Texture			Petiole colour			Edge shape		
Smooth	265	82.0%	Green	242	74.9%	Serrated	88	27.2%
Savoy or semi-savoy	58	18.0%	Purple	81	25.1%	Entire	235	72.8%

and association mapping. Of the remaining high-quality SNPs, the majority of markers revealed a MAF between 0.05 and 0.5 (Fig. 1.).

Population structure analysis

Using 4077 SNP markers, the population structure for the panel of 323 spinach accessions was analysed by STRUCTURE and determined by the highest likelihood value of $L(K)$. After plotting the model values $L(K)$ with each likelihood subpopulation K (1–10), the model value was shown to be $K = 5$, which corresponded to the population number. The number for the minimum subgroup was confirmed by the MDS analysis as five dimensions (Fig. 2). Therefore, we used $K = 5$ as the minimum number of groups for our association analysis. Due to the imperfect Q-matrix results from STRUCTURE, an accession was assigned to the subpopulation 1 (Q1) to the subpopulation 5 (Q2) when at least 50% of the genome information (Q value ≥ 0.5) was evaluated to belong to one group. The numbers of accessions for Q1, Q2, Q3, Q4, and Q5 were 90, 8, 56, 36 and 30, respectively. The ungrouped 103 spinach accessions were assigned to group 6 (G6). PCA was used to access the internal patterns of population structure for the USDA germplasm collections (Fig. 2.). The top two principal components (PCs), which were determined by the covariance matrix, explained 30.14% and 19.9% of the total variance among the spinach collections, respectively. Except the matrix group $Q = 2$, the remaining matrix groups, $Q = 3$, 4, 5 and group 6, belonged to principal component 1 (PC1), and $Q = 1$ was assigned to principal component 2 (PC2), respectively.

Association mapping

Association between SNP marker and leaf traits was tested using three different models: a general linear model (GLM) and two different mixed linear models (MLM, CMLM) analysed using JMP® Genomics 7 and GAPIT package, respectively. The results indicated that under different significance levels, three different models detected different numbers of SNP markers that were strongly associated with leaf traits (Table 2). GLM and CMLM inclined to detect more markers under all the significant levels in comparison with MLM (Table 2). Using GLM, 133 markers were found to be associated with leaf surface attributes, petiole colour and edge shape when $P < 0.01$. However, only 69 markers were detected by MLM at $P < 0.01$. Comparing with GLM, when accounting for population structure to correct for spurious associations, the numbers of markers associated with all leaf traits tested by MLM were obviously reduced under all significant levels. Without considering population structure in GLM, the discovered markers may be spurious and related to the population structure. GLM and CMLM (Tables S1) were able to discover leaf-trait association markers under more stringent $P < 0.001$ level in comparison with MLM. Generally, fewer markers associated with traits were detected with the increasing stringent P significant level.

For each leaf trait, the three association models were graphed using quantile–quantile (QQ) plots as structure (Q), population structure as measured by STRUCTURE plus kinship (Q+K), and the compressed MLM (CMLM) (Fig. 3.). As presented in QQ plots of the observed P values vs. the expected P values for each of the three models as well as a naïve model that does not account for Q or K, the distribution of P values for naïve model generally was distant from the expected distribution. However, the distribution of P values for other models was close to the expected distribution except for the CMLM in leaf surface analysis. As a result, the mixed model accounting for kinship (K) as well as population structure (Q) that was analysed by JMP® Genomics 7 indicated the optimum in reducing confusing population structure and relatedness bias, and the distribution of P values suggested a much-similar uniform distribution (Fig. 3.). Therefore, the significant SNPs discovered by MLM performed

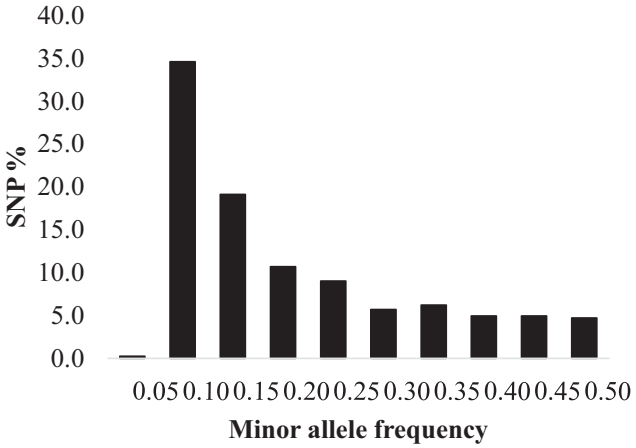

Fig. 1: The distribution of SNPs based on their minor allele frequency (MAF). SNPs with MAF < 5% were excluded from the analysis.

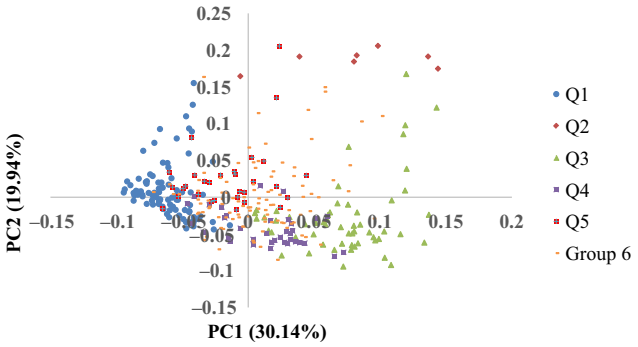

Fig. 2: Principal component analysis (PCA) plot for all 323 spinach accessions. Subgroups are indicated by Q-matrix (Q1–Q5) from STRUCTURE. Group 6 indicates the ungrouped spinach accessions.

Table 2: The numbers of SNPs associated with three traits using three statistical models (Q+K, Q and CMLM) at different significance levels

	Q+K			Q			CMLM		
Trait	P < 0.01	P < 0.005	P < 0.001	P < 0.01	P < 0.005	P < 0.001	P < 0.01	P < 0.005	P < 0.001
Surface Texture	17	5	0	36	18	5	11	7	3
Petiole Colour	34	14	2	47	29	11	51	26	4
Edge Shape	18	7	1	50	21	7	54	29	5

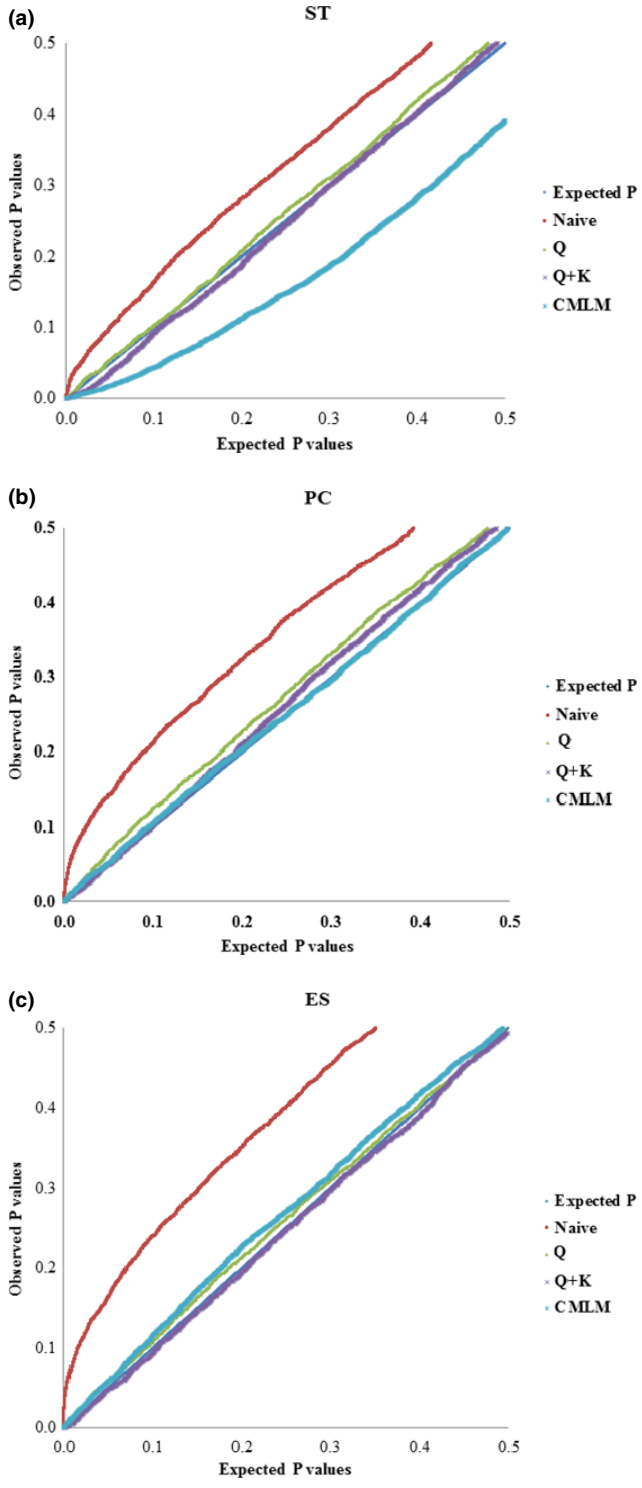

Fig. 3: Comparison of QQ plots for different models of association for leaf surface texture (ST), petiole colour (PC) and edge shape (ES). The naïve association (red line) is a one-way ANOVA without accounting for Q or K. The simple linear model (Q, green line) and K+Q (purple line) were analysed by JMP® Genomics 7. The CMLM (blue line) was analysed by GAPIT. The axes are restricted to a maximum of 0.5 to better exhibit the comparison of different models.

by JMP® Genomics 7 were presented for leaf-trait association (Table 3).

The significant SNP markers (including marker types, the reference contigs and position) that associated with different leaf traits analysed by MLM are listed in Table 3. When $P < 0.005$, the numbers of the markers associated with leaf surface texture, petiole colour and edge shape were 5, 14 and 7, respectively. The majority of markers were [A/C] type. The R^2 values for markers detected by MLM ranged from 2.6% to 4.4%.

Discussion

In this study, two association mapping models (GLM and MLM), run by JMP® Genomics 7, and CMLM, run by GAPIT, were used to analyse SNP marker–trait association in the panel of 323 spinach accessions. Generally, TASSEL software is used to perform marker–trait association analysis (Bradbury et al., 2007). However, as the three spinach leaf traits in this study were binary and not quantitative, JMP® Genomics 7 was more suitable for the analysis. We also found that the SNP markers analysed from JPM Genomics 7 were more reasonable than those from GAPIT based on the optimum in reducing confusing population structure and relatedness bias, and the distribution of P values suggested a much-similar uniform distribution (Fig. 3).

The GLM only accounts for population structure (Q-matrix), but the MLM accounts for both the kinship (*K*-matrix) and the population structure (*Q*-matrix) and allows one to overcome the issue of false positives in marker–trait association (Pritchard et al. 2000, Zhu et al. 2000, Price et al. 2006, Yu et al. 2006). The comparisons among different marker–trait association models for association mapping studies were reported by previous researchers, and MLM discovered less significant markers associated with traits of interest in comparison with GLM (Pritchard et al. 2000, Evanno et al. 2005, Price et al. 2006, Yu et al. 2006, Mandel et al. 2013, Xu et al. 2013, Wei et al. 2014), which was supported by this study. Using the quantile–quantile (QQ) plots, the model which produces more significant results than expected by chance could be found for marker–trait association analysis (Bastien et al. 2014). In this study, MLM indicated a better match for expected P values, compared with the CMLM conducted by GAPIT. GAPIT is an R-based program (R Development Core Team 2011), and the CMLM performed by GAPIT uses a group kinship matrix calculated from clustered individuals, which is more computationally efficient (Zhang et al. 2010). Although GAPIT is generally implemented in genomewide association studies (GWAS) and genomic prediction and selection (GS), it can be performed for marker–trait association.

In association analysis, the LD measures the non-random associations between alleles at different loci. The resolution of the association mapping is influenced by LD in such a way that the resolution will be low, but fewer markers will be needed if LD is high (Rafalski 2002). Because of the potential higher homozygosity at certain loci, the autogamous species incline to have lower efficient recombination rate than allogamous species. Therefore, LD is generally expected to be higher in autogamous species than in allogamous species (Flint-Garcia et al. 2003). Thus, more markers are needed for association analysis in allogamous species. However, the challenge of conducting association mapping using allogamous species, such as spinach, was the high heterozygosity at different loci. In this study, the heterozygous rate for different loci ranged from 0 to 93.81%, and the average heterozygosity was 21.23%. This may explain the imperfect results for population structure analysis for the 323 spinach accessions in which only 220 (68.11%) accessions were assigned to five different subpopulation groups. In addition, the PCA for population structure also indicated an uncertain grouping, which may be due to the high heterozygosity rate in these 323 spinach accessions.

Table 3: The overview of the significant SNP markers (P < 0.005) associated with three different leaf traits using Q+K model

Trait	Contig	Position	SNP type	MAF	P value	Marker R^2 (%)
Surface texture	AYZV01123718	240	[A/T]	0.0739	0.001	3.5
	AYZV01033090	7799	[A/C]	0.187	0.0019	4.0
	AYZV01083680	1654	[A/T]	0.4799	0.0024	4.0
	AYZV01176546	280	[C/T]	0.2385	0.0039	2.7
	AYZV01110266	939	[C/T]	0.3013	0.0043	2.7
Petiole colour	AYZV01046525	2300	[C/T]	0.3468	0.0004	4.2
	AYZV01064222	13663	[A/C]	0.1912	0.0008	3.7
	AYZV01064222	13486	[A/G]	0.2159	0.001	3.5
	AYZV01167386	459	[C/T]	0.1309	0.001	3.4
	AYZV01065783	1382	[C/T]	0.3764	0.0012	4.4
	AYZV01053164	12114	[A/C]	0.4056	0.0017	3.1
	AYZV01123866	10080	[A/C]	0.1209	0.0019	3.1
	AYZV01034140	6661	[A/G]	0.346	0.0021	4.1
	AYZV01008945	5709	[C/T]	0.1293	0.0023	3.9
	AYZV01115526	333	[C/G]	0.3356	0.0023	3.1
	AYZV01082115	3553	[C/T]	0.0751	0.0028	2.9
	AYZV01218231	95	[A/T]	0.207	0.003	2.9
	AYZV01148500	3204	[A/T]	0.2606	0.0036	3.7
	AYZV01155706	1703	[A/G]	0.3956	0.0044	2.7
Edge shape	AYZV01009983	15542	[A/T]	0.154	0.0007	3.8
	AYZV01082596	10562	[C/G]	0.13	0.0017	3.2
	AYZV01010088	1875	[C/T]	0.2269	0.0029	2.9
	AYZV01006392	336	[A/G]	0.1721	0.0031	2.8
	AYZV01098173	195	[C/T]	0.0645	0.0034	2.7
	AYZV01090260	36157	[C/T]	0.3834	0.0039	3.5
	AYZV01174425	808	[A/C]	0.1277	0.0041	2.6

Three leaf traits, surface texture (smooth, savoy, or semi-savoy), petiole colour (different shades of green vs. purple) and edge shape (serrate vs. entire), were analysed for association analysis in this study. Five SNP markers, AYZV01123718_240 (where 'AYZV01123718' is the contig based on AYZV01 and AYZV02 spinach genome sequences and '240' is the SNP position located at AYZV01123718, which is also used for other SNP markers), AYZV01033090_7799, AYZV01083680_1654, AYZV01176546_280 and AYZV01110266_939, were identified to be associated with leaf surface texture; 14 SNP markers, AYZV01046525_2300, AYZV01064222_13663, AYZV010642 22_13486, AYZV01167386_459, AYZV01065783_1382, AYZ V01053164_12114, AYZV01123866_10080, AYZV01034140_6 661, AYZV01008945_5709, AYZV01115526_333, AYZV0108 2115_3553, AYZV01218231_95, AYZV01148500_3204 and AYZV01155706_1703, associated with petiole colour; and seven SNP markers, AYZV01009983_15542, AYZV01082596_10562, AYZV01010088_1875, AYZV01006392_336, AYZV0109817 3_195, AYZV01090260_36157 and AYZV01174425_808, associated with edge shape (Table 3). So far, there is few report about the heritance of the three leaf traits in spinach. From our previous knowledge, the leaf traits should be controlled by single gene or several major genes, but the low R^2 values and not very low P value of the associated SNP markers identified from this research do not support that there existed major genes in this spinach panel. It will be ideal if we can map the major genes or QTL of the spinach leaf traits by SNP markers into the spinach chromosomes or linkage groups. However, as there are no full-scale genomic sequencing data available for public to use and without SNP genetic maps available in spinach, it is very difficult for us to determine the specific genomic location for discovered SNPs and determine whether there exist major genes in the spinach for controlling the three leaf traits. However, in future, if the spinach genome sequence information is available, then we can do a follow-up study to locate our discovered SNPs to certain chromosome, which will be very informative for spinach breeding. Meanwhile, we are developing segregating populations to build the consensus genetic maps in spinach with SNP markers. After we have populations, we will study the genetics of the leaf traits and map the genes into the specific location of the chromosomes.

To our knowledge, this is the first study performed using association mapping analysis for spinach leaf traits using a panel of 323 USDA germplasm collections. So far, we have not seen any article that reported association mapping using SNP markers in spinach. Although the 26 SNP markers identified from this study are not very strong markers with low R^2 values and not very low P values, this study provides us a feasible approach to identify SNP markers through association analysis in spinach and thus leads to select these three leaf traits, leaf surface texture, petiole colour and edge shape, through marker-assisted selection in spinach breeding programme.

Acknowledgements

This work was supported, in part, by the USDA National Institute of Food and Agriculture Hatch Project accession number 1002423.

References

Achenbach, U., J. Paulo, E. Ilarionova, J. Lübeck, J. Strahwald, E. Tacke, H. R. Hofferbert, and C. Gebhardt, 2009: Using SNP markers to dissect linkage disequilibrium at a major quantitative trait locus for resistance to the potato cyst nematode *Globodera pallida* on potato chromosome V. Theor. Appl. Genet. **118**, 619—629.

Aranzana, M. J., S. Kim, K. Zhao, E. Bakker, M. Horton, K. Jakob, C. Lister, J. Molitor, C. Shindo, C. Tang, C. Toomajian, B. Traw, H. Zheng, J. Bergelson, C. Dean, P. Marjoram, and M. Nordborg, 2005: Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. PLoS Genet. **1**, e60. Epub.

Bastien, M., H. Sonah, and F. Belzile, 2014: Genome wide association mapping of *Sclerotinia sclerotiorum* resistance in soybean with a Genotyping-by-Sequencing approach. Plant Genome **7**, 1—13.

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**, 2633—2635.

Buckler, E. S., and J. M. Thornsberry, 2002: Plant molecular diversity and applications to genomics. Curr. Opin. Plant Biol. **5**, 107—111.

Cao, Q., B. R. Lu, H. Xia, J. R. Francesco, A. Spada, and F. Grassi, 2006: Genetic diversity and origin of weedy rice (*Oryza sativa* f. *spontanea*) populations found in North-eastern China revealed by simple sequence repeat (SSR) markers. Ann. Bot. **98**, 1241—1252.

Collard, B. C. Y., and D. J. Mackill, 2008: Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos. Trans. R. Soc. Lond. B Biol. Sci. **12**, 557—572.

Elshire, R. J., J. C. Glaubitz, and Q. Sun, 2011: A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One **6**, e19379.

Evanno, G., S. Regnaut, and J. Goudet, 2005: Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. **14**, 2611—2620.

Flint-Garcia, S. A., J. M. Thornsberry, and E. S. Buckler, 2003: Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. **54**, 357—374.

Goldstein, D. B., and M. E. Weale, 2001: Population genomics: linkage disequilibrium holds the key. Curr. Biol. **11**, R576—R579.

Hall, D., C. Tegström, and P. K. Ingvarsson, 2010: Using association mapping to dissect the genetic basis of complex traits in plants. Brief. Funct. Genomics **9**, 157—165.

Hu, J., B. Mou, and B. A. Vick, 2007: Genetic diversity of 38 spinach (*Spinacia oleracea* L.) germplasm accessions and 10 commercial hybrids assessed by TRAP markers. Genet. Resour. Crop Evol. **54**, 1667—1674.

Hulbert, S. H., and J. L. Bennetzen, 1991: Recombination at the *Rp1* locus of maize. Mol. Gen. Genet. **226**, 377—382.

Khattak, J. Z. K., J. L. Christiansen, A. M. Torp, and S. B. Andersen, 2007: Genic microsatellite markers for discrimination of spinach cultivars. Plant Breeding **126**, 454—456.

Lipka, A. E., F. Tian, Q. Wang, J. Peiffer, M. Li, P. J. Bradbur, M. A. Gore, E. S. Buckler, and Z. Zhang, 2012: GAPIT: genome association and prediction integrated tool. Bioinformatics **28**, 2397—2399.

Mammadov, J., R. Aggarwal, R. Buyyarapu, and S. Kumpatla, 2012: SNP markers and their impact on plant breeding. Int. J. Plant Genomics Article ID 728398, 11 pages.

Mandel, J. R., S. Nambeesan, J. E. Bowers, L. F. Marek, D. Ebert, L. H. Rieseberg, S. J. Knapp, and J. M. Burke, 2013: Association mapping and the genomic consequences of selection in sunflower. PLoS Genet. **9**, e1003378.

Myles, S., J. Peiffer, P. J. Brown, S. E. Ersoz, Z. Zhang, E. D. Costich, and S. E. Buckler, 2009: Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell **21**, 2194—2202.

Neale, D. B., and O. Savolainen, 2004: Association genetics of complex traits in conifers. Trends Plant Sci. **9**, 325—330.

NASS (National Agricultural Statistics Service), 2015: http://www.nass.usda.gov/Statistics_by_Subject/result.php?1AD8F2E7-A87C-3823-AE11-D60E193D42A1&sector=CROPS&group=VEGETABLE-S&comm=SPINACH (last accessed on Febrary 27, 2016).

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich, 2006: Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. **38**, 904—909.

Pritchard, J. K., M. Stephens, and P. Donnelly, 2000: Inference of population structure using multilocus genotype data. Genetics **155**, 945—959.

R Development Core Team, 2011: R: A Language and Environment for Statistical Computing. The R Foundation for Statistical Computing, Vienna, Austria. ISBN: 3-900051-07-0. Available at: http://www.R-project.org/.

Rafalski, A., 2002: Applications of single nucleotide polymorphisms in crop genetics. Curr. Opin. Plant Biol. **5**, 94—100.

Siemonsma, J. S., and K. Piluek, 1993: Plant Resources of South-East Asia No.8. Vegetables. Pudoc Scientific Publishers, Wageningen, Netherlands. 412.

Sonah, H., M. Bastien, E. Iquira, A. Tardivel, and G. Legare, 2013: An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PLoS One **8**, e54603.

Thomson, M. J., 2014: High-throughput SNP genotyping to accelerate crop improvement. Plant Breeding Biotech. **2**, 195—212.

Tian, F., P. J. Bradbury, P. J. Brown, H. Hung, Q. Sun, S. Flint-Garcia, T. R. Rocheford, M. D. McMullen, J. B. Holland, and E. S. Buckler, 2011: Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat. Genet. **43**, 159—162.

Wei, Z., G. Zhang, Q. Du, J. Zhang, B. Li, and D. Zhang, 2014: Association mapping for morphological and physiological traits in *Populus simonii*. BMC Genet. **15**(Suppl 1), S3.

Xu, J., N. Ranc, S. Muños, S. Rolland, J. P. Bouchet, N. Desplat, M. C. Le Paslier, Y. Liang, D. Brunel, and M. Causse, 2013: Phenotypic diversity and association mapping for fruit quality traits in cultivated tomato and related species. Theor. Appl. Genet. **126**, 567—581.

Yu, J., G. Pressoir, W. H. Briggs, I. Vroh Bi, M. Yamasaki, J. F. Doebley, M. D. McMullen, B. S. Gaut, D. M. Nielsen, J. B. Holland, S. Kresovich, and E. S. Buckler, 2006: A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. **38**, 203—208.

Zhang, Z., E. Ersoz, C. Q. Lai, R. J. Todhunter, H. K. Tiwari, M. A. Gore, P. J. Bradbury, J. Yu, D. K. Arnett, J. M. Ordovas, and E. S. Buckler, 2010: Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. **42**, 355—360.

Zhao, J., M. J. Paulo, D. Jamar, P. Lou, F. van Eeuwijk, G. Bonnema, D. Vreugdenhil, and M. Koornneef, 2007: Association mapping of leaf traits, flowering time, and phytate content in *Brassica rapa*. Genome **50**,47 963—973.

Zhu, C., M. Gore, E. S. Buckler, and J. Yu, 2000: Status and prospects of association mapping in plants. Plant Genome **1**, 5—20.

Zhu, C., M. Gore, E. Buckler, and J. Yu, 2008: Status and prospects of association mapping in plants. Plant Genome **1**, 5—20.

Zondervan, K. T., and L. R. Cardon, 2004: The complex interplay among factors that influence allelic association. Nat. Rev. Genet. **5**, 89—100.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Information of the significant SNP markers ($P < 0.001$) associated with three different leaf traits using CMLM.