Euphytica (2016) 212:17–28
DOI 10.1007/s10681-016-1740-0

Association analysis for oxalate concentration in spinach

Ainong Shi · Beiquan Mou · James C. Correll

Received: 16 May 2016 / Accepted: 21 June 2016 / Published online: 1 July 2016

Abstract Reducing oxalate content of spinach is a major breeding objective. The aim of this research was to conduct association analysis and identify SNP markers associated with oxalate concentration in spinach germplasm. A total of 310 spinach genotypes, including 300 USDA germplasm accessions and ten commercial cultivars, were used for the association analysis of oxalate concentration. Genotyping by sequencing was used to identify 841 SNPs among the genotypes examined for the association analysis. The distribution of oxalate concentration showed a near normal distribution with a wide range in concentrations from 647.2 to 1286.9 mg/100 g on a fresh weight basis and 53.4 to 108.8 mg/g on a dry weight basis. The range in oxalate concentration in spinach suggests that it is a complex quantitative trait which may be controlled by multiple genes, each with a minor effect among the tested spinach panel. Association analysis indicated that six SNP markers (AYZV02031464_116, AYZV02031464_117, AYZV02031464_95, AYZV02283363_2707, AYZV02287123_2830, and AYZV02296293_852) were associated with the oxalate concentration. The SNP markers may be useful for breeders to select germplasm for reduced oxalate concentrations in spinach breeding programs through marker-assisted selection.

Keywords Association mapping · Germplasm · Breeding · Genotyping by sequencing (GBS) · Oxalate concentration · Single nucleotide polymorphism (SNP) · *Spinacia oleracea*

Electronic supplementary material The online version of this article (doi:10.1007/s10681-016-1740-0) contains supplementary material, which is available to authorized users.

A. Shi (✉)
Department of Horticulture, University of Arkansas, Fayetteville, AR 72701, USA
e-mail: ashi@uark.edu

B. Mou (✉)
Crop Improvement and Protection Research Unit, USDA-ARS, 1636 E. Alisal Street, Salinas, CA 93905, USA
e-mail: beiquan.mou@ars.usda.gov

J. C. Correll (✉)
Department of Plant Pathology, University of Arkansas, Fayetteville, AR 72701, USA
e-mail: jcorrell@uark.edu

Introduction

Spinach (*Spinacia oleracea* L.) is an economically important vegetable crop worldwide and considered one of the healthiest vegetables in the human diet due to its high concentration of nutrients and health-promoting compounds such as beta carotene (pro vitamin A), lutein, folate, vitamin C, calcium, iron, phosphorous, and potassium (Dicoteau 2000; Lester et al. 2013; Morelock and Correll 2008; Correll et al. 2011). However, spinach also contains a greater amount of oxalic acid than most crops (Holmes and Kennedy 2000; Kitchen et al. 1964; Mou 2008;

Noonan and Savage 1999). Oxalic acid (also referred to as oxalates) is a naturally occurring chemical in plants including spinach. The major issue surrounding oxalic acid in food is whether or not it contributes to human health concerns such as the formation of kidney stones (Noonan and Savage 1999; Solberg et al. 2015). Oxalate can react with calcium, iron and other minerals, forming crystals which then inhibit mineral absorption (Franceschi and Horner 1980; Noonan and Savage 1999; Bohn et al. 2004) or create calcium oxalate (Holmes and Kennedy 2000; Oke 1969) deposited in the kidneys of certain people as a common form of kidney stones (Massey et al. 1993). Food such as beets, rhubarb, strawberries, nuts, chocolate, tea, wheat bran, and all dry beans (fresh, canned, or cooked), excluding lima and green beans, are known to increase oxalate in the urine and may contribute to kidney stone formation. High oxalate concentrations are most commonly found in vegetables from the Chenopodiaceae, but also the Polygonaceae (Noonan and Savage 1999; Solberg et al. 2015).

In spinach, oxalate concentration has been observed to be higher in older leaves (Okutani and Sugiyama 1994). However, these results were not consistent with the results from purslane (Palaniswamy et al. 2004). Noonan and Savage (1999) also reported that soluble oxalate was the predominant form of oxalate in leaves and petioles and the oxalate concentration in leaves was much higher than that in petioles in a range of plants including spinach. Kawazu et al. (2003) observed that spinach cultivars with round leaf blades generally contained less oxalate than those with lobate blades, while other studies showed that oxalate concentration had little correlation with leaf type (Mou 2008; Solberg et al. 2015). The oxalate content in spinach is also affected by growth rate and the type of production system. Kaminishi and Kita (2006) reported that fast-growing cultivars contained higher nitrate and lower oxalate, whereas slow-growing cultivars had lower nitrate and higher oxalate concentrations. Kawazu et al. (2003) also reported the cropping season under summer, autumn, and winter affected oxalate concentrations in spinach. Oguchi et al. (1996) reported that selection and breeding of cultivars with a high leaf/petiole ratio is recommended for the production of better quality spinach with regard to oxalates. Koh et al. (2012) reported no significant effects on oxalate content of spinach produced under certified-organic versus conventional cropping systems.

Genetic variation in spinach oxalate content has been reported (Kitchen et al.1964). They found significant differences among 39 breeding lines, hybrids, and F2 populations in the amount of anhydrous oxalic acid present. In a study of 182 open-pollinated and F1 hybrid cultivars and breeding lines available in Japan, Solberg et al. (2015) reported that the oxalate content varied from 59 to 531 mg/g based on fresh weight. It was also reported that no differences were detected between older versus newer spinach cultivars or between open-pollinated cultivars and F1-hybrids, but did identify some accessions to potentially breed low-nitrate/low-oxalate spinach. Although Kaminishi and Kita (2006) observed that fast-growing cultivars contained lower oxalate, Kohman (1939) observed no significant differences in the oxalate content in 53 commercial and experimental genotypes. However, Kaminishi and Kita (2006) and Mou (2008) identified a 1.9-fold variation in oxalate content among spinach accessions. Murakami et al. (2009) also have produced mutation spinach lines with 17–33 % lower oxalate levels compared to control material. These data indicate that there is a potential to use the genetic diversity of spinach to breed spinach for reduced oxalate levels.

Based on the variation in reports of oxalate concentrations in spinach tissue and germplasm, oxalate apparently is a complex trait affected by genetics and environmental conditions. Due to the complexity of oxalate production in spinach, and the difficulty to select for this trait, it would be difficult and time-consuming to select for low oxalate spinach through a classical plant breeding approach. However, molecular plant breeding may be an efficient way to select quantitative traits through marker assisted selection (MAS). Single nucleotide polymorphisms (SNPs), with the abundance, cost efficiency, and high-throughput screening, has become a powerful tool in genome mapping, association studies, diversity analysis, and tagging of important genes in plant genomics (Collard et al. 2005; Collard and Mackill 2008; Xu and Crouch 2008). Therefore, identification of SNP markers associated with oxalate concentration will provide breeders with a useful tool to assist in selecting for low-oxalate in spinach breeding programs. Genotyping by sequencing (GBS) is one of the next-generation sequencing (NGS) platforms to discover SNPs without prior knowledge of

the genome in spinach (Elshire et al. 2011; He et al. 2014; Sonah et al. 2013). In recent years, GBS has been widely used in quantitative trait loci (QTL) and association mapping (He et al. 2014; Iquira et al. 2015; Liu et al. 2014; Nimmakayala et al. 2014; Sonah et al. 2015). Therefore, using an NGS platform will be a good approach for QTL and association mapping and MAS in spinach. The spinach genome sequences AYZV01 and AYZV02 are available to the public (http://www.ncbi.nlm.nih.gov/Traces/wgs/?val=AYZV01 and http://www.ncbi.nlm.nih.gov/Traces/wgs/?val=AYZV02), and represent approximately half of the spinach genome (Dohm et al. 2014; Minoche et al. 2015). These resources provide a reference for SNP discovery and association analysis in spinach.

To date, knowledge of the spinach genome is limited and few reports have been published on the use of molecular markers in spinach. Khattak et al. (2006) published a genetic linkage map with six linkage groups, constructing the map with 101 amplified fragment length polymorphisms (AFLPs) and nine simple sequence repeats (SSRs). This genetic map has a total length of 585 cM, with an average distance of 5.18 cM between markers (Khattak et al. 2006), but does not offer a great amount of detail about the linkage groups. The AFLP and SSR markers, while useful, are less robust than SNP markers. Recently, Chan-Navarrete et al. (2016) reported on the first genetic map using SNPs in spinach. Six linkage groups (P01-P06), consisting of 283 SNP markers, ranging in distance from 46 to 116 cM were identified and 39 QTLs were identified that related to nitrogen use efficiency in spinach. The identification of SNP markers for spinach traits of interest will provide breeders with a powerful tool to develop improved spinach cultivars more efficiently. Therefore, the development of robust SNP markers and SNP genetic maps would be a valuable resource for spinach breeding efforts. The objective of this research was to identify SNP markers associated with oxalate concentration and conduct association analysis in spinach germplasm.

Materials and methods

Plant materials

Several experiments on oxalate concentrations in spinach have been conducted and some of these rresults have previously been published (Mou 2008). A total of 310 genotypes were evaluated for the association analysis of oxalate concentration in this study (Supplementary Table S1). The 310 spinach genotypes included 300 USDA spinach germplasm accessions plus 10 commercial hybrids, originally collected from 31 countries, with the majority (60 %) of accessions originating from Turkey (n = 98), United States (n = 37), Afghanistan (n = 22), and Macedonia (n = 21) (Supplementary Table S1). All seed of the 300 USDA germplasm accessions were kindly provided by the North Central Regional Plant Introduction Station, USDA-ARS, Iowa State University, Ames, Iowa. The seed of the commercial cultivars was provided by various companies. The seed of 'Alrite F1' was obtained from American Takii, Salinas, CA; 'Bolero F1', "Bordeaux F1', 'Hellcat F1', 'Melody F1', and Unipack 151 F1' from Seminis Vegetable Seeds, Woodland, CA; 'Indian Summer F1' from Johnny's Selected Seeds, Winslow, ME; 'Lion F1' and 'Whale F1' from Rijk.Zwaan, De Lier, Holland; and 'Nordic IV F1' from Gowan Seed, Salinas, CA.

Oxalate concentration evaluation

To evaluate oxalate concentration in spinach, eight seeds from each genotype were planted in a plastic pot (10 × 10 × 10 cm) filled with field soil in the greenhouse. The field soil was a pasteurized sandy loam soil collected in Salinas, CA. The experiment was conducted with a randomized complete block design (RCBD) with two replications. After plant emergence, each pot was fertilized weekly with 50 mL of a soluble fertilizer as a combination of ammonium phosphate, potassium nitrate, and urea (20 N–8.8P–16.6 K; Nortrace, Ltd., Greeley, CO) at a concentration of 0.8 g/L. The nitrogen component consisted of 2.9 % ammoniac N, 5.0 % nitrate N, and 12.1 % urea N. The air temperature varied between 8 and 27 °C night/day and the day length changed from 10 h 10 min to 11 h 22 min during the experiment. Five weeks after planting, all leaves in a pot were harvested without petioles in the morning and fresh weight of the leaves was determined. Harvested leaves were dried at 60 °C for 24 h before being weighed for dry weight.

The leaves were re-dried at 60 °C overnight, broken into small pieces by hand, and mixed. A 0.01 g leaf sample was homogenized in 5 mL deionized water for

6 min with a homogenizer (Ultra-turrax T25; Janke & Kunkel, IKA-Labortechnik, Staufen, Germany) at 24,000 rpm. The sample was diluted with 5 mL EDTA (10 mM, pH 7.6) and centrifuged at 1,500 rpm for 5 min. The oxalate concentration in the supernatant was determined using an oxalate kit (Procedure No. 591; Trinity Biotech, St. Louis) as described by Palaniswamy et al. (2004). It has been demonstrated that there is a strong linear correlation with oxalic acid concentrations and color using the kit. In addition, a negative control, a urine control with known oxalate concentration, and a known concentration of oxalic acid standard were included in the kit and used under the current testing conditions to authenticate the accuracy of the tests. The oxalate concentrations were calculated on both a fresh and dry weight basis. For fresh weight, the oxalate concentration was recorded in milligrams (mg) per 100 g of fresh spinach leaves and for dry weight, as milligrams (mg) per gram of dried spinach leaves.

The oxalate concentration data were subjected to analysis of variance (ANOVA) using the general linear models (GLM) procedure of JMP Genomics 7 (SAS Institute, Cary, NC). Genotype was considered a fixed effect, and replication was considered a random effect. For comparisons among genotypes, least significant differences (LSDs) were calculated with a Type I (α) error rate of $P = 0.05$. The mean, range, standard deviation (SD), and standard error (SE) were estimated for both oxalate concentrations on fresh weight (mg/100 g) and dry weight (mg/g) basis using 'Tabulate'; the distributions of these two traits was also drawn using 'Distribution'; and the correlation coefficient between the traits for both fresh and dry weight analysis in the 310 spinach genotypes was estimated and drawn using 'Fit Y by X' in JMP Genomics 7.

DNA extraction, GBS, and SNP discovery

Genomic DNA was extracted from leaves of spinach plants using the CTAB (hexadecyltrimethyl ammonium bromide) method (Kisha et al. 1997). A DNA library was prepared using the restriction enzyme *ApeK*I following the GBS protocol described by Elshire et al. (2011). The 90 bp double-end sequencing was performed on each spinach accession using GBS protocol by an Illumina HiSeq 2000 in Genomics Research Institute (BGI), Hong Kong. The GBS data assembly, mapping and SNP discovery were done using SOAP family software (http://soap.genomics.org.cn/) by the bioinformatics term in BGI. The GBS data provided by BGI averaged 3.26 M with 90 bp short-read nucleotides for each spinach sample. The short reads of the GBS data were first aligned to the spinach genome reference Viroflay-1.0.1 with AYZV01 project (http://www.ncbi.nlm.nih.gov/Traces/wgs/?val=AYZV01) using SOAPaligner/soap2 (http://soap.genomics.org.cn/). After the Spinach-1.0.3 spinach genome reference released on July 22, 2015, the AYZV01 series of contig accessions were changed to AYZV02 accessions (http://www.ncbi.nlm.nih.gov/Traces/wgs/?val=AYZV02), all SNP information was updated to the AYZV02 version. The two versions of the reference spinach genome were also published at http://bvseq.molgen.mpg.de/Genome/Download/Spinach/. The SOAPsnp v 1.05 was used for SNP calling (Li 2011; Li et al. 2009). Approximately 0.5 M SNPs were identified from the GBS data among the 310 spinach genotypes. The SNP data was filtered by minor allele frequency (MLF) >2 %, missing data <7 %, and heterozygous genotype <35 %. After filtering, 841 SNPs for 310 spinach accessions were used for genetic diversity and association analysis.

Population structure and genetic diversity

The model-based program STUCTURE 2.3.4 (Pritchard et al. 2000) was used to assess the population structure of the 310 spinach accessions/cultivars based on 841 SNP loci. In order to identify the number of populations (K) making up the structure of the data, the burn-in period was set at 20,000 with the Markov Chain Monte Carlo iterations and the run length set at 20,000 in an admixture model. The analysis then correlated allele frequencies which was independent for each run (Lv et al. 2012). Ten runs were performed for each simulated value of K, which ranged from 1 to 10. For each simulated K, the statistical value delta K was calculated using the formula described by Evanno et al. (2005). The optimal K was determined using STRUCTURE HARVESTER (Earl and von Holdt 2012; http://taylor0.biology.ucla.edu/structureHarvester/). After the optimal K was determined, a Q-matrix was obtained and was used in TASSEL 5 (Bradbury et al. 2007) for association analysis. Each spinach genotype was then

assigned to a cluster (Q) based on the probability determined by the software that the genotype belonged in the cluster. The cut-off probability for assignment to a cluster was 0.525. Based on the optimum K, a Bar plot with 'Sort by Q' was obtained to show the visual of the population structure among the 310 spinach accessions. Genetic diversity was also assessed and the phylogeny trees were drawn using MEGA 6 (Tamura et al. 2013) based on the Maximum Likelihood tree method with the following parameters (Shi et al. 2016).

Association analysis

Association analysis was performed using TASSEL 5 software, in which the single marker regression (SMR) model without structure and without kinship, the regression linear model (GLM), and the mixed linear model (MLM) methods were used as described by Bradbury et al. (Bradbury et al. 2007; http://www.maizegenetics.net/tassel). Two workflows of regression linear models, GLM (Q) and GLM (PCA), and two workflows of mixed linear models, MLM (Q + K) and MLM (PCA + K), were used in Tassel 5 for association analysis of SNP markers. Population structure (Q) was estimated using STUCTURE 2.3.4 (Pritchard et al. 2000); principal component analysis (PCA) was estimated by the tool PCA with covariance and three components; and Kinship (K) was estimated by the tool Kinship with Scald_IBS method in Tassel 5.

Results and discussion

Phenotyping of oxalate concentration

There were significant differences in fresh and dry weight oxalate concentrations among the 310 spinach genotypes evaluated (Supplement Table S1). There was a wide range in oxalate concentrations based on fresh and dry weight with a near normal distribution (Table S1; Fig. 1). The oxalate concentration on a fresh weight basis ranged from 647.2 to 1286.9 mg/100 g of fresh leaf weight and averaged 984.0 mg/100 g. Oxalate concentration on a dry weight basis ranged from 53.4 to 108.8 mg/g and averaged 80.6 mg/g. The standard deviation of oxalate on a fresh weight basis was 112.7 with the standard error 6.4; and the standard deviation of oxalate on a dry weight basis was 10.7 with the standard error 0.6, indicating there were significant genetic differences of oxalate concentration among the 310 spinach genotypes based on both the fresh and dry weight basis (Supplement Table S1). The oxalate concentration on a dry weight basis was highly correlated with oxalate concentration on a fresh weight basis ($r = 0.803$, $P < 0.01$). A linear regression model between oxalate dry weight (Y, mg/g) and oxalate fresh weight (X, mg/100 g) was postulated as $Y = 5.513 + 0.07636 \times X$ (Fig. 2).

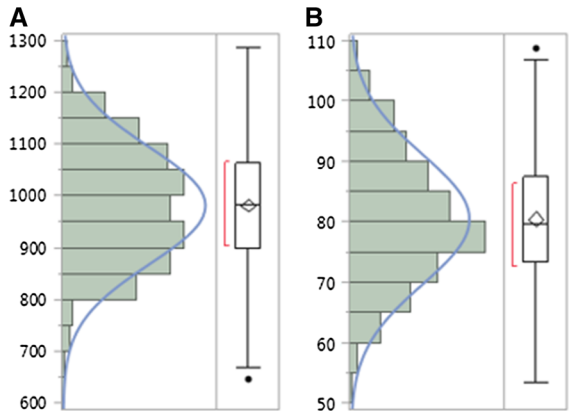

Fig. 1 The distribution of oxalate concentration expressed on fresh and dry weight bases among the 310 spinach genotypes evaluated (**a**: expressed on fresh weight with milligrams (mg) oxalate concentration per 100 gram (g) fresh leave weight, and **b**: on dry weight bases with mg per gram (g) dried leaf weight)

Genetic diversity and population structure

The population structure of the 310 spinach genotypes was initially inferred using STRUCTURE 2.3.4 (Pritchard et al. 2000) and the peak of delta K was observed at K = 2, indicating the presence of two main populations (designated clusters Q1 and Q2) in the spinach panel (Fig. 3a,b). The classification of accessions into populations based on the model-based structure from STRUCTURE 2.3.4 was shown in Fig. 3b and Supplementary Table S1. Based on the software STRUCTURE, each spinach genotype in the tested population of 310 genotypes is assigned a Q-value in each of the structured populations with a Q value as the probability it belongs to. The total of Q-values for each spinach genotype always is 1, and it is then divided into several Q values. Here there are two Q values: Q1- and Q2-value and whereby Q1 + Q2 = 1.0 because we have determined two

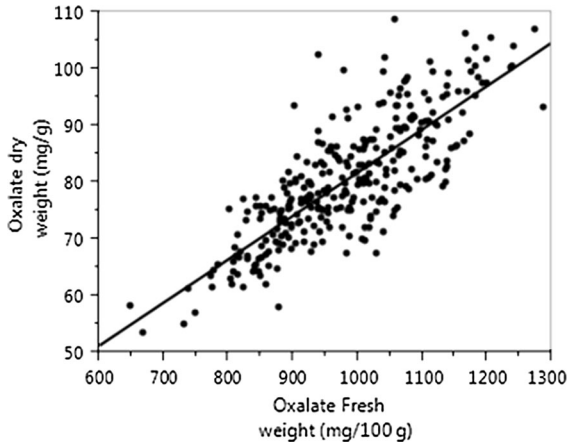

Fig. 2 The correlation between oxalate concentration on a dry weight basis and on a fresh weight basis ($r = 0.803$, $P < 0.01$) and the oxalate dry weight (mg/g) = 5.513 + 0.07636 × Oxalate Fresh weight (mg/100 g)

structured populations are best to fit the whole tested population with the 310 spinach genotypes. A Q-value = 0.525 was used as the value to divide the clusters. For example, if a spinach genotype had a Q1 value ≥ 0.525, i.e. its Q2 value of ≤ 0.475, it would be divided into the Cluster Q1; if a spinach genotype had a Q2 value ≥ 0.525, i.e. its Q1 value of ≤ 4.75, it would be divided into the Cluster Q2; genotypes that had intermediate Q1 and Q2 values (0.475 < Q1 value < 0.525 or 0.475 < Q2 value < 0.525) would be grouped as the Q1Q2 admixture. In total, 295 accessions (95.3 %) were assigned to one of the two populations (Q1 or Q2). Population 1 and 2 (Q1 and Q2) consisted of 189 (61.0 %) and 106 (34.2 %) accessions, respectively. The remaining 15 accessions (4.2 %) were categorized as having admixed ancestry (Supplementary Table S1).

The genetic diversity among spinach accessions was also assessed using the Maximum Likelihood (ML) method by MEGA 6 (Tamura et al. 2013). The populations Q1 and Q2 were defined as the two main clusters and the same colors as the population structure Q1 (red) and Q2 (green) was used from the STRUCTURE 2.3.4 (Fig. 3b) to draw the subtrees of the phylogenetic tree (Fig. 3c) with Q1 (red and round shape), Q2 (green and square shape), and the

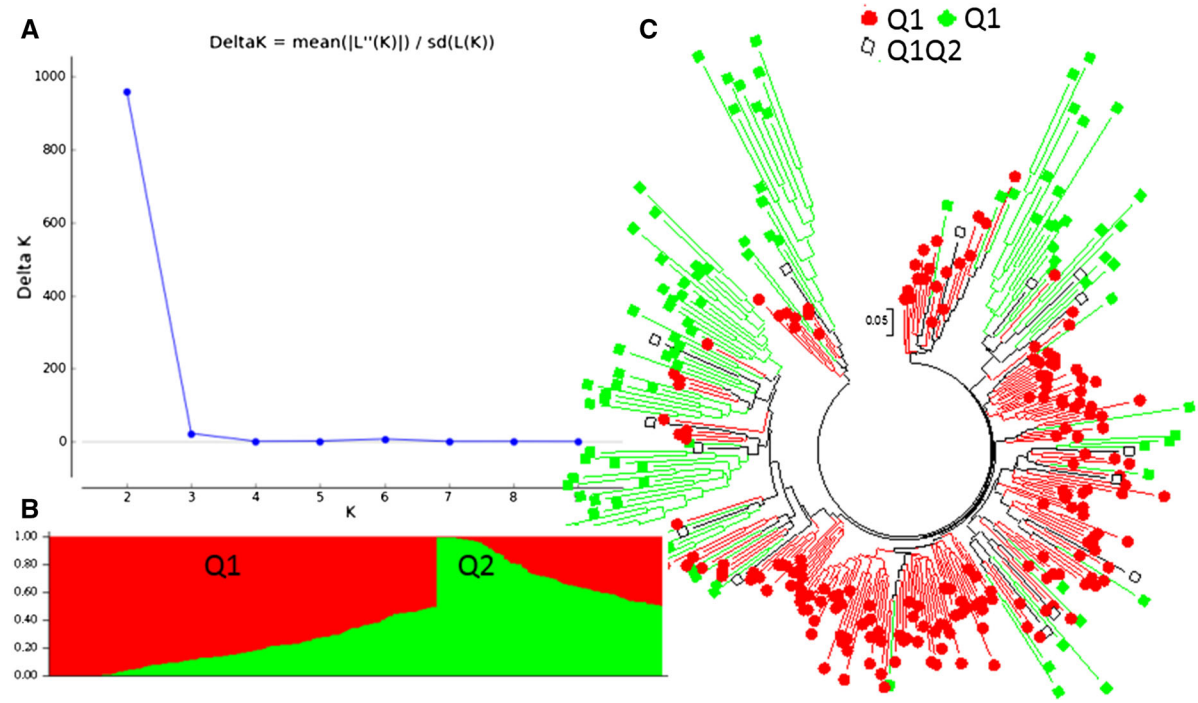

Fig. 3 Model-based populations in the association panel: **a** Delta K values for different numbers of populations assumed (K) in the STRUCTURE analysis. **b** Classification of 310 spinach genotypes into two populations using STRUCTURE 2.3.4. The distribution of the accessions to different populations is indicated by the color code (Q1: *red* and Q2: *blue*). **c** Maximum Likelihood (ML) tree of the 300 accessions drawn by MEGA 6. The color codes for each population are consistent in the figure **b** and **c**, and the empty black square as the admixture Q1Q2. (Color figure online)

admixture Q1Q2 (black empty square). Two phylogenetic trees were included: (1) Fig. 3c, without taxon names in order to compare it to the structure populations from STRUCTURE and view them easily and clearly; (2) Supplementary Fig. S1: the format of the traditional rectangular phylogenetic tree with taxon name. The phylogenetic trees from MEGA 6 (Fig. 3c and Supplementary Fig. S1), were good but not fully consistent with the structure populations (Q1–Q2) from STRUCTURE 2.3.4 (Fig. 2b, 3a), indicating that there were two differentiated genetic populations and admixtures in the spinach panel.

Association analysis

Based on the genetic diversity analysis from STRUCTURE and MEGA, and viewing the phylogenetic trees from Fig. 3 and Supplement Fig S1, the 310 spinach accessions can be organized into two structured populations. Therefore, we used the Q matrix with two structures in the association mapping in TASSEL. In total, five different models in TASSEL were used to do association analysis of oxalate concentration, including one single marker regression (SMR), two regression linear models, GLM (Q) and GLM (PCA), and two mixed linear models, MLM (Q + K) and MLM (PCA + K).

Based on the suggestion by Lander and Botstein (1989), a typical LOD threshold should be between 2 and 3 in order to detect the false positive (type I error) rate at 5 % level for a QTL (Churchill and Doerge 1994). In the first marker screening step, candidate SNP markers were screened using an LOD value (LOD = ~(−LOG(P), where P is the *P* value estimated from TASSEL) ≥ 2.0 as the threshold value for all five models. If the LOD was ≥ 2.0 in one of the five models, the SNP was selected as a candidate marker. Based on these criteria, 31 candidate SNP markers were selected from a total of 841 SNPs identified (Table S2).

Among the 31 screened SNPs (Table S2), the different LOD values of each SNP were observed among the five models. For oxalate dry weight (oxalate concentration based on dry weight), there were 11, 7, 14, 5, and 6 SNPs with the LOD > 2.0 from the SMR, GLM (Q), GLM (PCA), MLM (Q + K), and MLM (PCA + K), respectively (Table S2). For oxalate fresh weight, there were 8, 7, 19, 6, and 9 SNPs with the LOD > 2.0 from the SMR, GLM (Q), GLM (PCA), MLM (Q + K), and MLM (PCA + K), respectively (Table S2). For both oxalate weight either based on dry or fresh, the SMR and GLM models including GLM (Q) and GLM (PCA) identified more SNP markers than the MLM model (MLM (Q + K) and MLM (PCA + K)), which was similar to previous studies for the morphological and physiological traits in *Populus simonii* (Wei et al. 2014; Xu et al. 2013). The SMR model had its pitfall without considering population structure and relationship among the individuals. The GLM model only accounts for population structure with a Q-matrix (Pritchard et al. 2000) or PCA matrix (Price et al. 2006), but the MLM model accounts for both the kinship (K-matrix) and the population structure (Q-matrix or PCA matrix) (Price et al. 2006; Yu et al. 2006; Zhang et al. 2010), to prevent the false positive associations due to population or relatedness structure and to increase the power through the application of a correction specific to this structure (Yang et al. 2014) and to correct population stratification (Shin and Lee 2015). Different models have been proposed and used for association mapping studies and it usually recognized the MLM models discovered more robust markers associated with traits of interests in comparison with GLM model (Pritchard et al. 2000; Price et al. 2006; Xu et al. 2013; Yu et al. 2006; Zhang et al. 2010; Zhu et al. 2008), but there are pitfalls with MLM as well such as use of a small subset of markers and effects from case–control ascertainment (Yang et al. 2014).

The 31 candidate SNPs (Table S2) identified in the initial screen with the five models were further evaluated for association with oxalate concentration using additional criteria. It was assumed that if a significant association was identified across multiple models, the SNP marker would be more robust. If it had an LOD value ≥ 2.5 with all models, a SNP was considered to be strongly associated with oxalate content; otherwise, if it had an LOD value ≥ 2.0 with multiple models, the SNP was considered to be weakly associated. Based on these criteria, six SNPs were identified among the original 31 candidate markers (Table 1).

For oxalate dry weight, two of the six SNPs identified, AYZV02283363_2707 and AYZV02296293_852 had LOD scores ≥ 2.5 for all five models, suggesting these two SNPs were the most strongly associated with oxalate concentration based on dry weights; Two other SNPs, AYZV02031464_116 and AYZV02287123_2830 had LOD scores ≥ 2.5 for

Table 1 Six SNP markers associated with oxalate concentration identified from five models using TASSEL in 300 spinach accessions and ten commercial cultivars

Spinach genome Spinach-1.0.3 information			Oxalate dry weight (mg/g)					Oxalate Fresh weight (mg/100 g)					Associated trait
SNP name[a]	SNP Type	Contig at AYZV02 project	SMR	GLM (Q)	GLM (PCA)	MLM (Q + K)	MLM (PCA + K)	SMR	GLM (Q)	GLM (PCA)	MLM (Q + K)	MLM (PCA + K)	
			LOD (-log(P)) value from Tassel 5[b]										
AYZV02031464_95	G/T	AYZV02031464	2.1	1.7	2.4	1.4	1.7	2.4	2.2	2.9	2.3	2.3	Oxalate fresh weight
AYZV02031464_116	T/C	AYZV02031464	3.2	3.2	3.4	2.7	2.4	3.1	3.2	3.6	3.3	2.8	Oxalate dry and fresh weight
AYZV02031464_117	G/T	AYZV02031464	2.3	1.9	2.6	1.4	1.8	2.4	2.3	3.0	2.3	2.3	Oxalate fresh weight
AYZV02283363_2707	A/G	AYZV02283363	3.6	3.3	3.8	2.8	3.2	2.2	2.0	2.6	1.6	2.0	Oxalate dry weight
AYZV02287123_2830	A/G	AYZV02287123	2.6	2.7	2.5	2.1	2.1	2.0	1.9	2.1	1.7	1.9	Oxalate dry weight
AYZV02296293_852	G/C	AYZV02296293	3.0	2.9	3.4	2.5	2.7	2.4	2.1	2.4	2.0	2.0	Oxalate dry and fresh weight
			R-square value from Tassel 5[b]										
AYZV02031464_95	G/T	AYZV02031464	2.4	1.9	2.7	1.5	1.8	2.7	2.6	3.4	2.8	2.6	Oxalate fresh weight
AYZV02031464_116	T/C	AYZV02031464	3.8	3.9	4.1	3.3	2.8	3.6	4.0	4.2	4.3	3.3	Oxalate dry and fresh weight
AYZV02031464_117	G/T	AYZV02031464	2.5	2.2	2.9	1.5	1.9	2.7	2.7	3.5	2.8	2.7	Oxalate fresh weight
AYZV02283363_2707	A/G	AYZV02283363	5.5	5.3	5.8	4.8	5.1	3.4	3.2	3.9	2.8	3.4	Oxalate dry weight
AYZV02287123_2830	A/G	AYZV02287123	3.9	4.2	3.8	3.5	3.3	3.0	3.0	3.2	2.8	3.0	Oxalate dry weight
AYZV02296293_852	G/C	AYZV02296293	3.4	3.5	4.0	3.0	3.2	2.7	2.3	2.6	2.2	2.1	Oxalate dry and fresh weight

[a] SNP name is defined as the contig name plus the SNP position on the contig

[b] Lod (-LOG(P)) value, *P* value from TASSEL, and R-square value calculated from five workflows using TASSEL 5 (Bradbury et al. 2007; http://www.maizegenetics.net/tassel): SMR, GLM (Q), GLM (PCA), MLM (Q + K), and MLM (PCA + K), where SMR = single marker regression without Q or PCA matrix and without kinship (K) matrix; GLM (Q) = regression linear model with Q matrix (Q matrix from STRUCTURE, Pritchard et al. 2000); GLM (PCA) = regression linear model with PCA matrix (PCA matrix from TASSEL); MLM (Q + K) = mixed linear model with Q matrix plus K matrix; and MLM (PCA + K) = using mixed linear model with PCA matrix plus K matrix

four and three out of five models, respectively, suggesting they also are strongly associated with oxalate dry weight; and two other SNPs, AYZV02031464_95 and AYZV02031464_117 had LOD values less $\geq$ 2.0 in two out of the five models for oxalate dry weight and were considered to be weakly associated markers (Table 1).

Among the six SNP markers for oxalate fresh weight, one SNP, AYZV02031464_116 had LOD value $>$ 2.5 for all five models, indicating a strong association; three of the six SNPs (AYZV02031464_117, AYZV020 31464_95, and AYZV02296293_852) had an LOD $\geq$ 2.3, $\geq$ 2.2 and $\geq$ 2.0 across all five models, respectively, suggesting a good association with oxalate fresh weight; and the two other SNPs, AYZV02283363_2707 and AYZV02287123_2830, had relatively low LOD values for most of the models and thus, were considered to only have a weak association for oxalate fresh weight associated markers (Table 1).

Combining the analysis from both fresh and dry weight oxalate concentrations, two SNP marker, AYZV02031464_116 and AYZV02296293_852, were identified as having the strongest association (Table 1). The AYZV02031464_116 SNP was recognized as strongly associated with oxalate content because it had an LOD of 2.5 or higher across all five models for both oxalate concentrations. AYZV0229 6293_852 had an LOD of $\geq$ 2.5 for all five models based on dry weight but an LOD of $\geq$ 2.0 for the five models based on fresh weight (Table 1).

The other four markers, AYZV02031464_117, AYZV02031464_95, AYZV02283363_2707, and AYZV02287123_2830 had an LOD $\geq$ 2.0 in five models for oxalate concentration based on fresh or dry weight, but not both (Table 1), indicating a relatively weak association with oxalate concentration in spinach.

The six SNPs were located on four contigs, which may be located on different chromosomes or different regions of a chromosome, further indicating oxalate concentration was a quantitative trait controlled by multiple genes each with minor effects. Currently, the available spinach genome reference Spinach-1.0.3 (AYZV02) (http://www.ncbi.nlm.nih.gov/Traces/wgs/?val=AYZV02) as released on 7 July 2015, represented approximately one-half of the spinach genome (Dohm et al. 2014; Minoche et al. 2015). A more comprehensive version of the spinach genome assembly may be made available publicly in 2016 (van Deynze, 2014; van Deynze et al. 2015; Allen van Deynze, personal communication), but unfortunately, the entire genome sequence of spinach and physical maps are not currently available. After the whole genome sequences become publicly available, QTLs for oxalate content traits can be mapped to their chromosome location. Although these markers need to be further evaluated, the six SNP markers identified have the potential to be used as good molecular markers to select low oxalate concentration in spinach breeding programs through MAS.

Use of spinach accessions with low oxalate concentration

From this research, it was evident that there was a wide range in oxalate concentrations among the genotypes tested with a near normal distribution indicating that the oxalate concentration in spinach was a complex trait, which may be a quantitative trait controlled by minor genes in the tested spinach association panel. However, it has not been definitively determined whether the oxalate concentration in spinach is a quantitative or qualitative trait controlled by major genes or minor genes. No major QTLs with large effect for oxalate concentration in spinach were identified in this study. All six SNP markers identified had very low R-square values, further indicating there were multiple genes with minor effect in the spinach genotypes tested for oxalate concentration. However, we cannot exclude the possible existence of major genes that control oxalate concentration in spinach. Further QTL mapping using bi-parent crosses derived from high and low-oxalate spinach lines are underway to further determine the genetics of oxalate concentration in spinach.

Although it is not clear whether the oxalate concentration in spinach is a quantitative or qualitative trait, the genetic variation of oxalate content in spinach has been identified and confirmed from several reports and also several low-content spinach genotypes have been reported (Kitchen et al. 1964; Kaminishi and Kita 2006; Kohman 1939; Moir 1953; Mou 2008; Murakami et al. 2009; Solberg et al. 2015). These studies indicate that there is potential to use the genetic diversity of spinach to reduce oxalate in spinach.

Among the set of 300 spinach USDA accessions evaluated in this study, eight (PI165710, PI181923,

Table 2 Lists of nine low oxalate concentration spinach accessions plus ten commercial spinach hybrids with their accession number, name, origin region (country), oxalate concentration expressed on fresh weight and dry weight bases (mg per unit leaf weight)

Accession number/name	Plant name	Original (country)	Oxalate fresh weight (mg/100 g)	Oxalate dry weight (mg/g)	Q_cluster[a]
PI165710	Cornell ID #247	Japan	731.6	55.1	Q1
PI181923	Hama no. 20	Syria	747.8	57.0	Q2
PI339548	101–25	Turkey	777.7	64.6	Q1
PI358252	Edrolisten	Macedonia	774.7	61.4	Q1
PI445782	Shami	Syria	772.0	63.4	Q1
PI445784	Baladi	Syria	668.8	53.4	Q1
PI531457	SZEKESFEHERVARI	Hungary	738.3	61.3	Q1
PI608762	K-17068	Thailand	647.2	58.3	Q1
Alrite.F1	Alrite.F1	Japan	839.5	64.2	Q2
Bolero.F1	Bolero F1	Netherlands	1056.2	78.4	Q2
Bordeaux.F1	Bordeaux.F1	Netherlands	876.8	58.1	Q1
Hellcat.F1	Hellcat.F1	Netherlands	943.6	75.2	Q1
Indian.Summer.F1	Indian.Summer.F1	United States	1029.2	79.9	Q2
Lion.F1	Lion.F1	Netherlands	940.5	71.3	Q1Q2
Melody.F1	Melody.F1	Netherlands	894.1	75.2	Q2
Nordic.IV.F1	Nordic.IV.F1	United States	807.7	66.1	Q1
Unipack.151.F1	Unipack.151.F1	Netherlands	840.2	64.4	Q2
Whale.F1	Whale.F1	Netherlands	936.2	76.5	Q1

[a] Two population structures (Clusters) identified from Mega 6: Q1 means the spinach accession belongs to the population group 1; Q2 to group 2; and Q1 Q2 to group 1 or 2

PI339548, PI358252, PI445782, PI445784, PI531457, and PI608762) had low oxalate concentrations of less than 780.0 mg/100 g based on fresh weight tissue and less than 65.0 mg/g based on dry weight (Table 2). These eight accessions may be used as initial sources of low oxalate concentration in spinach breeding efforts.

Breeding utilizes the genetic diversity available. It is anticipated that there would be a wider range in variation in a given phenotype among progeny derived from two parents with a broader genetic background or larger genetic distance. As an approach to using the accessions, the genetic distance between the eight accessions with low oxalate concentrations and the commercial cultivars was compared (Table 2). A phylogenetic tree among the 18 genotypes was built using SNP alleles by MEGA 6 (Fig. 4), which also showed two structured, and well-divided populations (Q1 and Q2) similar to those among the 310 tested spinach genotypes from this study. Seven of the eight spinach accessions (PI81923 being the exception) were divided into cluster 1 (Q1) and were genetically close to each other, indicating that each of them could be effective parents in spinach breeding programs to

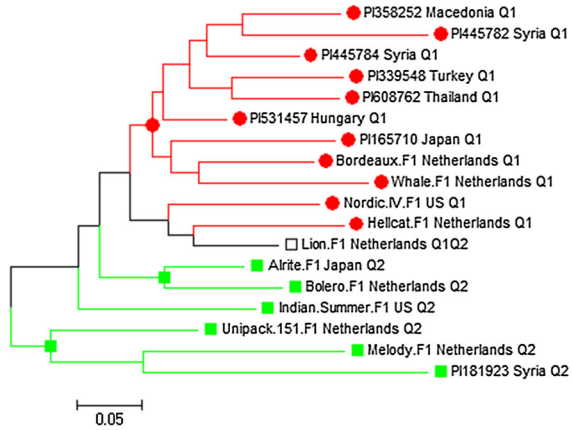

Fig. 4 A phylogenetic tree drawn by MEGA 6 among eight spinach germplasm accessions with low oxalate concentrations plus ten spinach commercial cultivars

reduce oxalate content. Among them, two accessions from Syria, PI445782 ('Shami') and PI 445784 ('Baladi'), were recommended earlier by Mou (2008) as sources of low oxalate concentration in spinach because they had the lowest oxalate concentration. From the diversity analysis (Fig. 4), the two accessions were more genetically distant from the ten commercial cultivars evaluated. Another accession, PI81923 ('Hama no. 20') from Syria, was also a good source for low oxalate because it belonged to population 2 (Q2), it also had a further genetic distances from others, and it was grouped separately from all other accessions and cultivars except Melody F1 (Fig. 4). Thus, these spinach accessions may provide good sources of low oxalate concentration genotypes as parents in spinach breeding.

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

References

Bohn T, Davidsson L, Walczyk T, Hurrell RF (2004) Fractional magnesium absorption is significantly lower in human subjects from a meal served with an oxalate-rich vegetable, spinach, as compared with a meal served with kale, a vegetable with a low oxalate content. Br J Nutr 91:601–606

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Chan-Navarrete R, Dolstra O, van Kaauwen M, van Bueren ETL, van der Linden CG (2016) Genetic map construction and QTL analysis of nitrogen use efficiency in spinach (*Spinacia oleracea* L). Euphytica 208:621–636

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963–971

Collard BCY, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos. Trans. R. Soc. B12(363):557–572

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 142:169–196

Correll JC, Bluhm BH, Feng C, Lamour K, du Toit LJ, Koike ST (2011) Spinach: better management of downy mildew and white rust through genomics. Eur J Plant Pathol 129:193–205

Dicoteau DR (2000) Vegetable crops. Printice Hall, New Jersey

Dohm JC, Minoche AE, Holtgräwe D, Capella-Gutierrez S, Zakrzewski F, Tafer H, Rupp O, Sörensen TR, Stracke R, Reinhardt R, Goesmann A, Kraft T, Schulz B, Stadler PF, Schmidt T, Gabaldon T, Lehrach H, Weisshaar B, Himmelbauer H (2014) The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). Nature 505:546–549

Earl DA, von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361

Elshire RJ, Glaubitz JC, Sun Q (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6(5):19379

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

Franceschi VR, Horner HT (1980) Calcium oxalate crystals in plants. Bot Rev 46:361–427

He J, Zhao X, Laroche X, Lu Z, Liu H, Li Z (2014) Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front Plant Sci 5:484

Holmes RP, Kennedy M (2000) Estimation of the oxalate content of foods and daily oxalate intake. Kidney Int 57:1662–1667

Iquira E, Humira S, Francois B (2015) Association mapping of QTLs for Sclerotinia stem rot resistance in a collection of soybean plant introductions using a genotyping by sequencing (GBS) approach. BMC Plant Biol 15:5. doi:10.1186/s12870-014-0408-y

Kaminishi A, Kita V (2006) Seasonal change of nitrate and oxalate concentration in relation to the growth rate of spinach cultivars. HortSceince 41:1589–1595

Kawazu Y, Okimura M, Ishii T, Yui S (2003) Varietal and seasonal differences in oxalate content of spinach. Sci Hortic 97:203–210

Khattak J, Torp AM, Andersen SB (2006) A genetic linkage map of *Spinacia oleracea* and localization of a sex determination locus. Euphytica 148:311–318

Kisha T, Sneller CH, Diers BW (1997) Relationship between genetic distance among parents and genetic variance in populations of soybean. Crop Sci 37:1317–1325

Kitchen JW, Burns EE, Perry BA (1964) Calcium oxalate content of spinach (*Spinacia oleracea* L). Proc Am Soc Hortic Sci 84:441–445

Koh E, Charoenprasert S, Mitchell AE (2012) Effect of organic and conventional cropping systems on ascorbic acid, vitamin C, flavonoids, nitrate, and oxalate in 27 varieties of spinach (*Spinacia oleracea* L). J Agric Food Chem 60:3144–3150

Kohman EF (1939) Oxalic acid in foods and its behavior and fate in diet. J Nutr 18:233–246

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Lester GE, Makus DJ, Hodges DM, Jifon JL (2013) Summer (Subarctic) versus winter (Subtropic) production affects spinach (*Spinacia oleracea* L) Leaf bionutrients: vitamins (C, E, Folate, K1, provitamin A), lutein, phenolics, and antioxidants. J Agric Food Chem 61:7019–7027

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987–2993

Li R, Yu C, Li Y, Lam T, Yiu S, Kristiansen K, Wang J (2009) SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25:1966–1967

Liu H, Bayer M, Druka A, Russell JR, Hackett CA, Poland J, Ramsay L, Hedley PE, Waugh R (2014) An evaluation of genotyping by sequencing (GBS) to map the Breviaristatum-e (ari-e) locus in cultivated barley. BMC Genom 15:104

Lv J, Qi J, Shi Q, Shen D, Zhang S, Zhang A, Shao G, Li H, Sun Z, Weng Y, Shang Y, Gu X, Li X, Zhu X, Zhang J, van Treuren R, van Dooijeweert W, Zhang Z, Huang S (2012) Genetic diversity and population STRUCTURE of cucumber (*Cucumis sativus* L). PLoS ONE 7:46919

Massey LK, Roman-Smith H, Sutton RAL (1993) Effect of dietary oxalate and calcium on urinary oxalate and risk of formation of calcium oxalate kidney stones. J Am Diet Assoc 93:901–906

Minoche AE, Dohm JC, Schneider J, Holtgrawe D, Viehover P, Montfort M, Sorensen TR, Weisshaar B, Himmelbauer H (2015) Exploiting single-molecule transcript sequencing for eukaryotic gene prediction. Genome Biol 16:184

Moir KW (1953) The determination of oxalic acid in plants. Qld J Agric Sci 10:1–3

Morelock TE, Correll JC (2008) Spinach. In: Prohens J, Nuez F (eds) Vegetables I: asteraceae, brassicaceae, chenopodiaceae, and cucurbitaceae. Springer, New York, pp 189–218

Mou B (2008) Evaluation of oxalate concentration in the US spinach germplasm collection. HortScience 43:1690–1693

Murakami K, Edamoto M, Hata N, Itami Y, Masuda M (2009) Low-oxalate spinach mutant induced by chemical mutagenesis. J Jpn Soc Hortic Sci 78:180–184

Nimmakayala P, Levi A, Abburi L, Abburi VL, Tomason YR, Saminathan T, Vajja VG, Malkaram S, Reddy R, Wehner TC, Mitchell SE (1011) Reddy UK (2014) Single nucleotide polymorphisms generated by genotyping by sequencing to characterize genome-wide diversity, linkage disequilibrium, and selective sweeps in cultivated watermelon. BMC Genom 15:767. doi:10.1186/1471-2164-15-767

Noonan SC, Savage GP (1999) Oxalate content of foods and its effect on humans. Asia Pac J Clin Nutr 8:64–74

Oguchi Y, Weerakkody WAP, Tanaka A, Nakazawa S, Ando T (1996) Varietal differences of quality-related compounds in leaves and petioles of spinach grown at two locations. Bull. Hiroshima Prefect Agric. Res. Cent. 64:1–9

Oke OL (1969) Oxalic acid and plants and in nutrition. World Rev Nutr Diet 10:262–303

Okutani I, Sugiyama N (1994) Relationship between oxalate concentration and leaf position in various spinach cultivars. HortScience 29:1019–1021

Palaniswamy UR, Bible BB, McAvoy RJ (2004) Oxalic acid concentrations in purslane (*Portulaca oleraceae* L.) is altered by the stage of harvest and the nitrate to ammonium ratios in hydroponics. Sci. Hort. 102:267–275

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Shi A, Buckley B, Mou B, Motes D, Morris JB, Ma J, Xiong H, Qin J, Yang W, Chitwood J, Weng Y, Lu W (2016) Association analysis of cowpea bacterial blight resistance in USDA cowpea germplasm. Euphytica 208:143–155

Shin J, Lee C (2015) A mixed model reduces spurious genetic associations produced by population stratification in genome-wide association studies. Genomics 105:191–196

Solberg SO, Yndgaard F, Axelsson J (2015) Nitrate and oxalate in germplasm collections of spinach and other leafy vegetables. Emirates J Food Agric 27(434):698–705

Sonah H, Bastien M, Iquira E, Tardivel A, Legare G (2013) An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PLoS ONE 8:54603

Sonah H, O'Donoughue L, Cober E, Rajcan I, Belzile F (2015) Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. Plant Biotechnol J 13(2):211–221

Tamura K, Stecher G, Peterson D, Ailipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

van Deynze A (2014) A de novo draft assembly of spinach using Pacific Biosciences technology. Plant & Animal Genomics XXII Conference, January 10-15, 2014, San Diego, CA (http://aa314.gondor.co/webinar/a-de-novo-draft-assembly-of-spinach-using-pacific-biosciences-technology/, accessed on May 3, 2016)

van Deynze A, Ashrafi H, Hickey L, Peluso P, Rank D, Chin J, Rapicavoli N, Drake J, Garvin T, Schatz M (2015) Using spinach to compare technologies for whole genome assemblies. Plant & Animal Genomics XXIII Conference, January 10-14, 2015, San Diego, CA

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

Xu Y, Crouch JH (2008) Marker-assisted selection in plant breeding: from publications to practice. Crop Sci 48:391–407

Xu J, Ranc N, Munos S, Rolland S, Bouchet JP, Desplat N, Le Paslier MC, Liang Y, Brunel D, Causse M (2013) Phenotypic diversity and association mapping for fruit quality traits in cultivated tomato and related species. Theor Appl Genet 126(3):567–581

Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL (2014) Advantages and pitfalls in the application of mixed-model association methods. Nat Genet 46:100–106

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DK, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38(2):203–208

Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42:355–360

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. The Plant Genome 1:5–20