

# Association Analysis and Identification of SNP Markers for Stemphylium Leaf Spot (*Stemphylium botryosum* f. sp. *spinacia*) Resistance in Spinach (*Spinacia oleracea*)

Ainong Shi<sup>1\*</sup>, Beiquan Mou<sup>2\*</sup>, Jim Correll<sup>3</sup>, Steven T. Koike<sup>4</sup>, Dennis Motes<sup>5</sup>, Jun Qin<sup>1</sup>, Yuejin Weng<sup>1</sup>, Wei Yang<sup>1</sup>

<sup>1</sup>Department of Horticulture, 316 PTSC, University of Arkansas, Fayetteville, USA

<sup>2</sup>US Department of Agriculture, Agricultural Research Service (USDA-ARS), Salinas, USA

<sup>3</sup>Department of Plant Pathology, University of Arkansas, Fayetteville, USA

<sup>4</sup>University of California Cooperative Extension, Salinas, USA

<sup>5</sup>Vegetable Research Center, University of Arkansas, Alma, USA

Email: \*ashi@uark.edu, \*Beiquan.Mou@ARS.USDA.GOV

Received 30 April 2016; accepted 19 August 2016; published 22 August 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

**Stemphylium leaf spot, caused by *Stemphylium botryosum* f. sp. *spinacia*, is an important fungal disease of spinach (*Spinacia oleracea* L.). The aim of this study was to conduct association analysis to identify single nucleotide polymorphism (SNP) markers associated with *Stemphylium* leaf spot resistance in spinach. A total of 273 spinach genotypes, including 265 accessions from the USDA spinach germplasm collection and eight commercial cultivars, were used in this study. Phenotyping for *Stemphylium* leaf spot resistance was evaluated in greenhouse; genotyping was conducted using genotyping by sequencing (GBS) with 787 SNPs; and single marker regression, general linear model, and mixed linear model were used for association analysis of *Stemphylium* leaf spot. Spinach genotypes showed a skewed distribution for *Stemphylium* leaf spot resistance, with a range from 0.2% to 23.5% disease severity, suggesting that *Stemphylium* leaf spot resistance in spinach is a complex, quantitative trait. Association analysis indicated that eight SNP markers, AYZV02052595\_115, AYZV02052595\_122, AYZV02057770\_10404, AYZV02129827\_205, AYZV02152692\_182, AYZV02180153\_337, AYZV02225889\_197, and AYZV02258563\_213 were strongly associated with *Stemphylium* leaf spot resistance, with a Log of the Odds (LOD) of 2.5 or above. The SNP markers may provide a tool to select for *Stemphylium* leaf spot resistance in spinach breeding programs through marker-assisted selection (MAS).**

\*Corresponding author.

**How to cite this paper:** Shi, A.N., Mou, B.Q., Correll, J., Koike, S.T., Motes, D., Qin, J., Weng, Y.J. and Yang, W. (2016) Association Analysis and Identification of SNP Markers for *Stemphylium* Leaf Spot (*Stemphylium botryosum* f. sp. *spinacia*) Resistance in Spinach (*Spinacia oleracea*). *American Journal of Plant Sciences*, 7, 1600-1611.

<http://dx.doi.org/10.4236/ajps.2016.712151>

## Keywords

**Genotyping by Sequencing (GBS), Disease Resistance, Single Nucleotide Polymorphism (SNP), *Spinacia oleracea*, *Stemphylium botryosum*, Stemphylium Leaf Spot**

## 1. Introduction

Spinach (*Spinacia oleracea* L.) is an economically important vegetable crop worldwide [1] [2]. In addition to its economic importance, spinach is one of the faster growing vegetable crops in the US and other regions in terms of per capita consumption, and is considered one of the healthiest vegetables in the human diet due to a high concentration of nutrients and other health-promoting compounds [2] [3]. However, diseases represent a significant constraint in spinach production [4].

In 1997, a new leaf spot disease was discovered in limited spinach acreage in the Salinas Valley of California [5]. Initially, spots on spinach leaves were 2 to 5 mm in diameter, circular to oval, and gray-green in color. These spots expanded and turned tan in color over time. Older leaf spots coalesced, dried, became papery in texture, and sometimes resulted in necrosis of significant portions of leaves. The pathogen was first identified as *Stemphylium botryosum* Wallr., and later defined as *Stemphylium botryosum* f. sp. *spinacia* because isolates of the fungus only infect spinach [5] [6]. The continued occurrence of the disease in spinach crops in the Salinas Valley, especially during rainy seasons, indicates that the pathogen has become established in that state. *S. botryosum* has also been reported as a foliar pathogen in spinach seed crops in Washington [7] and Oregon [8], and in fresh market spinach crops in Delaware and Maryland [9], Florida and Quebec, Canada [10], and Arizona [11]. The percentage of spinach acreage grown for fresh markets vs. processing markets has increased significantly in the US [12]. Quality standards for fresh market spinach are extremely high, so this disease poses an additional threat for growers in California and other states who need to produce defect-free products.

Hernandez-Perez and du Toit (2006) [13] detected *S. botryosum* in 100% of 77 spinach seed lots that were produced in Denmark, the Netherlands, New Zealand, or the US in 2000 to 2003. Hernandez-Perez (2005) [14] demonstrated seed transmission of *S. botryosum* in greenhouse trials. Although seed transmission of the pathogen under field conditions has not been proven, disease outbreaks in baby leaf spinach crops in Arizona, California, and Florida suggest that infected or infested seeds can serve as primary inoculum [10]. Chlorine or hot water seed treatment reduced the incidence of *S. botryosum* significantly but could not eliminate the pathogen from the seed [8].

Development of resistant cultivars is an important part of effective, integrated disease management. Koike *et al.* (2001a) [5] tested 10 flat-leaf, 8 semi-savoy, and 2 savoy spinach cultivars each against four isolates of *S. botryosum* from California. After two weeks, all of the cultivars developed leaf spot symptoms similar to those observed in commercial fields, with only slight differences in disease severity. An additional series of spinach cultivars, including some with resistance to race 5 and race 6 of the downy mildew pathogen, *Peronospora farinosa* f. sp. *spinaciae*, were also susceptible to the disease [15]. Mou *et al.* (2008) [16] screened the USDA spinach collection and commercial cultivars for resistance to *S. botryosum*. No genotype was completely resistant (immune) to the disease. However, there were significant differences in disease incidence (% of plants with leaf spot) and severity (% diseased leaf area) among the genotypes tested. Two accessions from Turkey, PI 169685 and PI 173809, consistently had low disease incidence and severity ratings. In lettuce, for comparison, a source of resistance to *Stemphylium botryosum* f. sp. *lactucum* was found in a wild species, *Lactuca saligna*, for which the resistance is controlled by a dominant gene and a recessive gene [17].

*Stemphylium* leaf spot resistance in spinach appears to be a complex (quantitative) trait with different responses to different isolates of the pathogen [5] [6] [15] [16]. It is time-consuming to transfer quantitative traits through classic plant breeding approaches. However, molecular plant breeding methods can provide efficient ways to select quantitative traits through marker assisted selection (MAS). Single nucleotide polymorphism (SNP) analysis, with the abundance of SNPs, cost-efficiency and high-throughput scoring, has become a powerful tool in genome mapping, association studies, diversity analysis, and tagging of important genes in plant genomics [18] [19]. Therefore, identification of SNP markers associated with *Stemphylium* leaf spot resistance should provide breeders with a useful tool to assist in selecting for resistance to this disease in spinach breeding

programs. Genotyping by sequencing (GBS) is a next-generation sequencing platform used to discover SNPs without prior knowledge of the genome [20]–[22]. The spinach genome sequences AYZV01 and AYZV02 are available to the public at (<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 [23] [24]. In addition, a more comprehensive version of the spinach genome assembly will be made available publicly in 2016 [25] [26] (Allen van Deynze, personal communication). These resources provide a reference for SNP discovery and association analysis in spinach. The objective of this research was to conduct association analysis to identify SNP markers associated with *Stemphylium* leaf spot resistance.

## 2. Materials and Methods

### 2.1. Plant Material

A total of 273 spinach genotypes, including 265 accessions of the USDA spinach germplasm collection and 8 commercial cultivars, were used for the association analysis of leaf spot resistance (**Supplementary Table S1**). The 273 spinach genotypes were collected originally from 30 countries, with a majority from eight countries: Turkey, United States (US), Serbia, Afghanistan, England, Iran, China, and Belgium for 80, 32, 22, 22, 16, 15, 12, and 11 accessions or cultivars, respectively. These comprised 76.6% of the spinach genotypes tested (**Supplementary Table S1**). All seeds of USDA germplasm accessions were kindly provided by the USDA-ARS North Central Regional Plant Introduction Station at Iowa State University, Ames, IA. The seeds of “Symphonie” was from Gautier.Graines; “Space” and “Springfield” from Gowan Seed Co; “Whale” from Rijk. Zwaan; and “Polka”, “Seven R”, “Unipack 12”, and “Unipack 277” from Seminis.

### 2.2. Leaf Spot Phenotyping

Leaf spot disease evaluation was conducted at the Agricultural Research Station of the USDA in Salinas, CA [16]. The 265 accessions of the USDA spinach germplasm collection plus 8 commercial cultivars were screened for *Stemphylium* leaf spot resistance in a greenhouse. The experimental design was a randomized complete block (RCB) with two replications/genotype. In each replication, eight seeds from each genotype were planted in a plastic pot (10 cm × 10 cm × 10 cm) filled with Sunshine Mix #1 (Sun Gro Horticulture Canada Ltd., Seba Beach, Canada) and seedlings were thinned to 5 plants per pot after emergence.

An isolate of *S. botryosum* obtained from infected plants of the spinach cultivar Cheetah grown in Arizona was maintained for 4 weeks on plates of V8 juice agar under a 12 h/12 h light/dark regime. A conidial suspension (approximately 10<sup>5</sup> conidia per ml) was prepared in water and applied using a hand-held mister onto leaves of each genotype five weeks after seeding (four- to six-true-leaf stage). Inoculated plants were incubated in a humidity chamber maintained at 100% relative humidity for 72 h, and then maintained in a greenhouse at 14°C /27°C (night/day cycle). Three weeks after inoculation, the plants in each pot were evaluated for severity (percent of leaf area with symptoms) of *Stemphylium* leaf spot symptoms. Re-isolations for *S. botryosum* were conducted to confirm the presence of the pathogen in association with the leaf spots.

Disease severity 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 ( $\alpha$ ) error rate of  $P = 0.05$ . The disease severity data also were analyzed using Microsoft (MS) Excel 2013 for the average, range, standard deviation (SD), standard error (SE), and coefficient of variation (CV).

### 2.3. DNA Extraction, GBS, and SNP Discovery

Genomic DNA was extracted from leaves of spinach plants using the CTAB (hexadecyltrimethyl ammonium bromide) method [27]. A DNA library was prepared using the restriction enzyme *ApeKI* following the GBS protocol described by Elshire *et al.* (2011) [20]. The 90 bp, double-end sequencing was performed on each spinach accession/cultivar using the GBS protocol by an Illumina HiSeq 2000 machine at the Beijing Genomics Institute (BGI) in Hong Kong. GBS data assembly, mapping and SNP discovery were done using SOAP family software (<http://soap.genomics.org.cn/>) by the bioinformatics team at 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 reference genome Viroflay-1.0.1 with the 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 was released on July 22, 2015, the AYZV01 series of contig accessions were changed to the AYZV02 accessions (<http://www.ncbi.nlm.nih.gov/Traces/wgs/?val=AYZV02>), and all SNP information was updated to the AYZV02 version. The SOAPsnp v 1.05 was used for SNP calling [28] [29]. Approximately a half-million SNPs were discovered from the GBS data among the 273 spinach germplasm accessions, as provided by BGI. The spinach accessions and SNPs were filtered before conducting genetic diversity and association analyses. If a spinach accession had >20% missing SNP data and the heterozygous SNP genotype was >30%, the spinach genotype was removed from the panel. The SNP data was filtered for minor allele frequency (MAF) >2%, missing data <7%, and heterozygous genotype <25%. After filtering, 787 SNPs for 273 spinach accessions were used for genetic diversity and association analysis.

## 2.4. Population Structure and Genetic Diversity

The model-based program STRUCTURE 2.3.4 [30] was used to assess the population structure of the 273 spinach accessions/cultivars based on 787 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 10,000 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 independently for each run [31]. Ten runs were performed for each simulated value of K, which ranged from 1 to 11. For each simulated K, the statistical value delta K was calculated using the formula described by Evanno *et al.* (2005) [32]. The optimal K was determined using STRUCTURE HARVESTER [33] (<http://taylor0.biology.ucla.edu/structureHarvester/>). After the optimal K was determined, a Q-matrix was obtained and used in Tassel 5 for association analysis. Each spinach accession 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 0.55 for only two clusters (structure populations), or 0.50 for three or more clusters. Based on the optimum K, a Bar plot with 'Sort by Q' was obtained to show the population structure among the 273 spinach accessions.

Genetic diversity of the 273 spinach accessions was also assessed, and the phylogeny trees were drawn using MEGA 6 [34] based on the Maximum Likelihood tree method with the following parameters. Test of phylogeny: bootstrap method with 500 bootstrap replications. Model/Method: General time reversible model. Rates among sites: Gamma distributed with invariant sites (G + I). Number of discrete gamma categories: 4. Gaps/missing data treatment: use of all sites. ML heuristic method: Subtree-pruning-regrafting-extensive (SPR level 5). Initial tree for ML: make initial tree automatically (Neighbor Joining). Branch swap filter: moderate. In order to compare results from the two software programs during the drawing of phylogeny trees using MEGA, the color, shape and branch of each spinach genotype was drawn using the same color located at the cluster (Q) from STRUCTURE. For the sub-tree of each Q (cluster), the shape of 'Node/Subtree Marker' and the 'Branch Line' was drawn with the same color as the bar plot of population clusters from the STRUCTURE analysis.

## 2.5. 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.* (2007) [35] (<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 STRUCTURE 2.3.4 [30]; 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.

## 3. Results

### 3.1. Phenotyping of *Stemphylium* Leaf Spot Resistance

*Stemphylium* leaf spot severity varied among the 273 spinach genotypes tested, ranging from 0.2% to 23.5%, and averaged 4.6% with a standard deviation of 3.8, and with a standard error of 0.23%, indicating significant

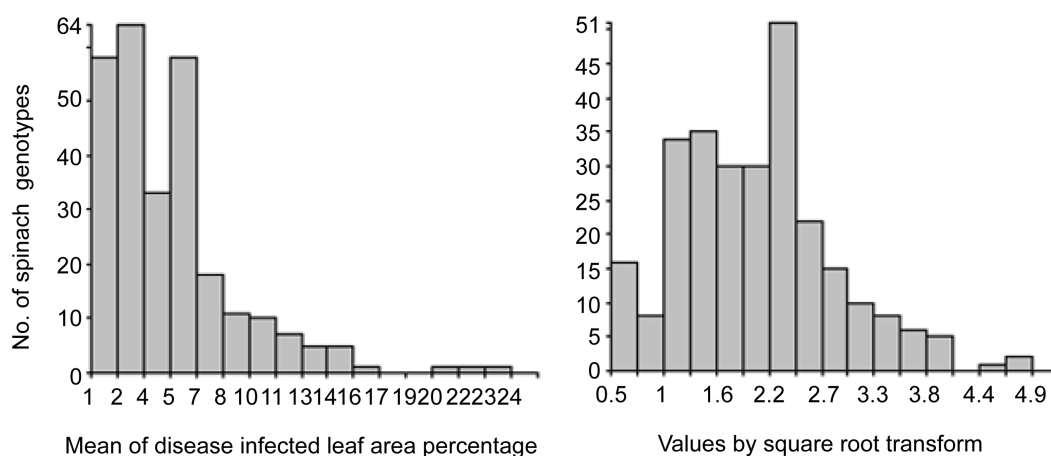
genetic differences in *Stemphylium* leaf spot resistance among the accessions (**Supplementary Table S1**). The distribution of *Stemphylium* leaf spot resistance was skewed to low disease severity and near normal after transformation by square root (**Figure 1**), indicating that *Stemphylium* leaf spot resistance is quantitative and probably controlled by multiple minor genes.

### 3.2. Genetic Diversity and Population Structure

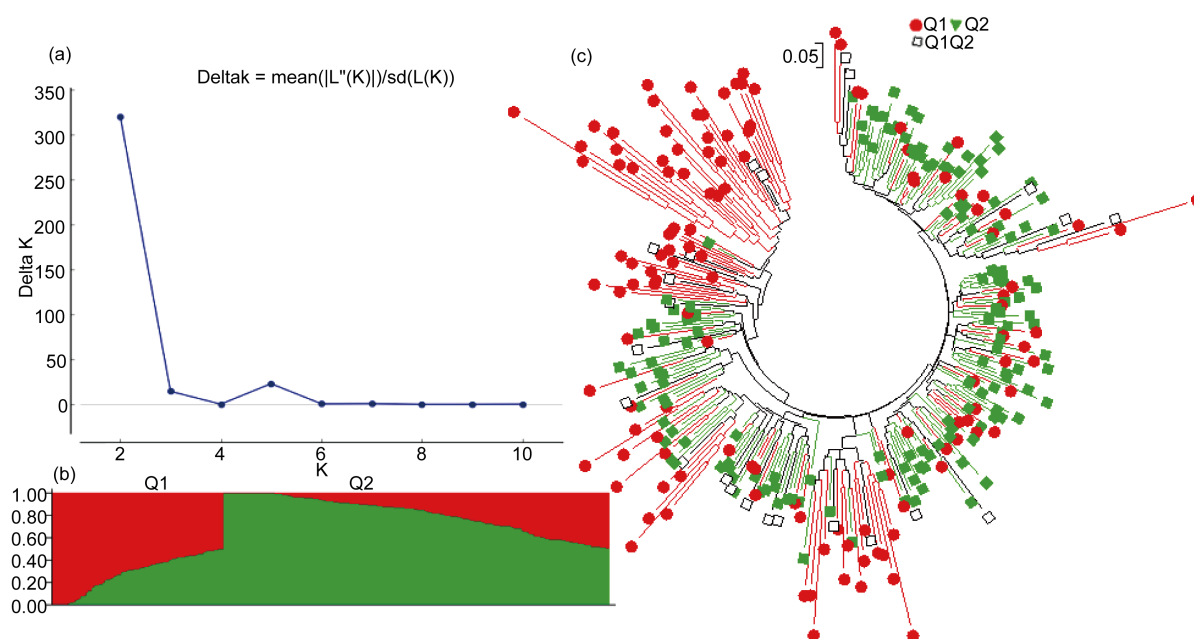
The population structure of the 273 spinach accessions and cultivars was initially inferred using STRUCTURE 2.3.4 [30]. The peak delta K was observed at  $K = 2$ , indicating the presence of two main population clusters, Q1 and Q2, in the spinach panel (**Figure 2(a)** and **Figure 2(b)**). The classification of accessions into populations or clusters based on the model-based structure from STRUCTURE 2.3.4 is shown in **Figure 2(b)** and **Supplementary Table S1**. A Q-value = 0.55 was used to divide the clusters, *i.e.*, if a spinach genotype had a  $Q1 \geq 0.55$ , the genotype was placed into Cluster Q1; if a spinach had a  $Q2 \geq 0.55$ , it was placed into Cluster Q2; and the remaining genotypes ( $0.45 < Q1 < 0.55$  or  $0.45 < Q2 < 0.55$ ) were placed into Q1Q2 of the admixture. In total, 252 of the 273 accessions or cultivars (92.3%) were assigned to one of the two populations or clusters. Q1 and Q2 consisted of 115 (42.1%) and 137 (50.2%) accessions and cultivars, respectively. The remaining 21 accessions (7.7%) were categorized as having admixed ancestry between Q1 and Q2, called Q1Q2 (**Supplementary Table S1**).

The genetic diversity among spinach accessions or cultivars was also assessed using the Maximum Likelihood (ML) method in MEGA 6 [34], with phylogenetic trees drawn based on the results. Q1 and Q2 were defined as the two main clusters or populations (see above), with the same colors as the population structure Q1 (red) and Q2 (green) from the STRUCTURE 2.3.4 analysis (**Figure 2(b)**) to draw subtrees of the phylogenetic tree (**Figure 2(c)**). Q1 is denoted with a red color and round shape, Q2 with a green color and square shape, and the admixture Q1Q2 with black, empty squares. Four phylogenetic trees were included: (1) without taxon names assigned in order to compare the populations from STRUCTURE (**Figure 2(c)**); (2) with two formats of the traditional rectangular phylogenetic tree (**Supplementary Figure S1-1** and **Figure S1-2**); and (3) the circular phylogenetic trees (**Supplementary Figure S2-1** and **Figure S2-2**, which represent a different format of **Figure S1-1**). The phylogenetic trees from MEGA 6 (**Figure 2(c)** and **Supplementary Figure S1-1**, **Figure S1-2**, **Figure S2-1**, and **Figure S2-2**), were not fully consistent with the structure populations Q1 and Q2 developing in STRUCTURE 2.3.4 (**Figure 2(a)** and **Figure 2(b)**), indicating there were two differentiated genetic populations and admixtures in the spinach panel, which was not divided distinctly into two clusters.

Besides the two structured populations inferred using the STRUCTURE analysis, the second highest peak of delta K was observed at  $K = 5$  using Structure Harvester, indicating the presence of five sub-population clusters



**Figure 1.** The distribution of disease severity for 273 spinach genotypes evaluated for resistance to *Stemphylium* leaf spot caused by *Stemphylium botryosum*, including 265 USDA spinach germplasm accessions and 8 commercial spinach cultivars. For each accession or cultivar, *Stemphylium* leaf spot severity was rated for two replications of five plants. The distribution on left was based on the mean of disease infected leaf area percentage and the one on right based on values by square root transform.



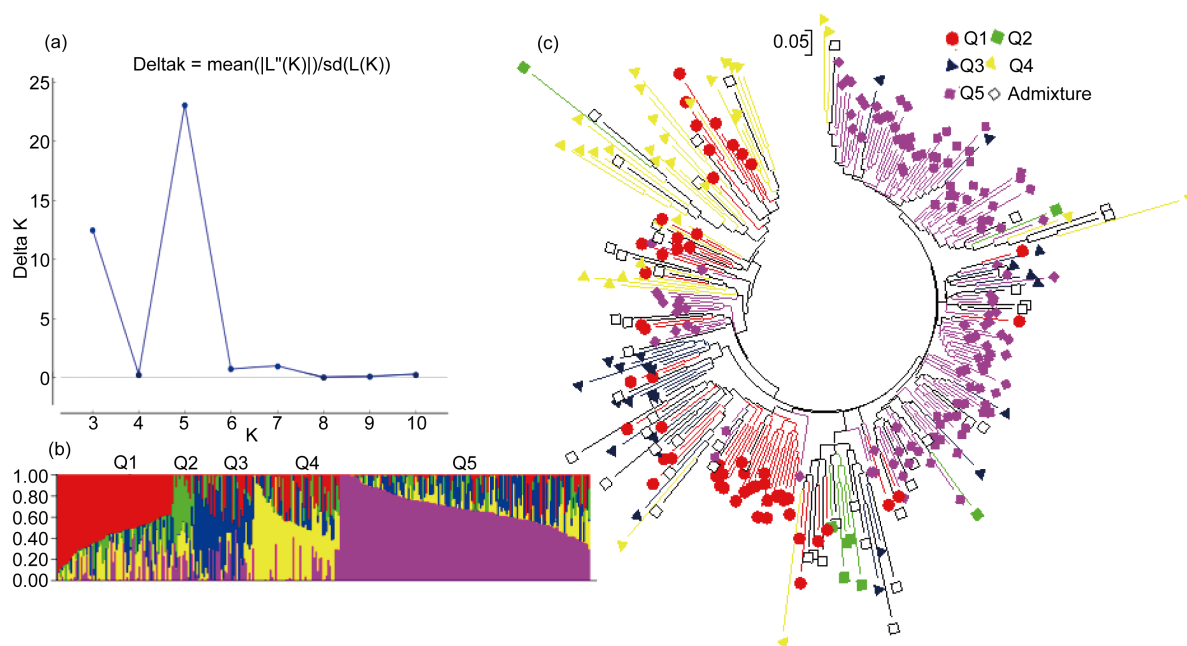
**Figure 2.** Model-based populations in association panels for spinach accessions and cultivars evaluated for resistance to *Stemphylium* leaf spot: (a) Delta K values for different numbers of populations (K) assumed in analysis completed with the STRUCTURE software. (b) Classification of 273 spinach genotypes, including 265 USDA spinach accessions and 8 commercial cultivars, into two populations using STRUCTURE Version 2.3.4, where the numbers on the y-axis show the subgroup membership, and the x-axis shows the different accession. The distribution of accessions into different populations is indicated by the color coding (Cluster 1, Q1, is red; and Cluster 2, Q2, is green). (c) Maximum Likelihood (ML) tree of the 273 spinach accessions or cultivars drawn in MEGA 6. The color code for each population is consistent in the (b) and (c), and the empty black square represents accessions or cultivars aligned with the admixture cluster or population, Q1Q2.

(Q1 to Q5) within the 273 spinach accessions or cultivars (Figure 3(a)). Figure 3(b) shows the bar plot drawn in STRUCTURE to visualize the five populations, where Q1 is red; Q2 is green; Q3 is blue; Q4 is yellow; Q5 is purple; and the admixture of the five populations is represented by black empty squares. The classification of accessions into populations based on the model-based structure developed in STRUCTURE 2.3.4 is shown in Figure 3(b) and Supplementary Table S1. Each spinach accession was also assigned to one of the five populations based on probabilities calculated in STRUCTURE (Supplementary Table S1). A Q value = 0.5 was used to divide the five clusters and the admixture. In total, 230 accessions or cultivars (84.2%) were assigned to one of the five populations, Q1 to Q5. Q1 to Q5 consisted of 49 (17.9%), 7 (2.6%), 24 (8.8%), 32 (11.7%), and 118 (43.2%) accessions/cultivars, respectively. The remaining 43 accessions (15.8%) were categorized as having admixed ancestry between Q1 to Q5 (Supplementary Table S1).

The genetic diversity of the 273 spinach accessions also was analyzed using the ML method in MEGA 6 by combining the five structured populations, Q1 to Q5, from STRUCTURE as done for the two structured populations above. The five clusters shown in Figure 3(c) were divided according to the five structured populations, Q1 to Q5, with same colors as in Figure 3(b), indicating five differentiated genetic populations and admixtures among the 273 accessions. The same approaches and methods were also used to analyze genetic diversity of the five structured populations using MEGA 6. The four phylogenetic trees drawn were consistent with the structure populations Q1 to Q5 from STRUCTURE 2.3.4, indicating that there were five differentiated genetic subpopulations and admixtures in the spinach panel. However, the five structured subpopulations were not very clear (Supplementary Figure S3-1, Figure S3-2, Figure S4-1 and Figure S4-2).

### 3.3. Association Analysis

Based on the genetic diversity analysis from STRUCTURE and MEGA, and the phylogenetic trees from Figure 2 and Figure 3, as well as Supplementary Figure S1-1, Figure S1-2, Figure S2-1, Figure S2-2, Figure S3-1, Figure S3-2, Figure S4-1, and Figure S4-2, the 273 spinach accessions/cultivars were organized into two



**Figure 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 spinach accessions into five populations using STRUCTURE 2.3.4, where the numbers on the y-axis show the subgroup membership, and the x-axis shows the different accession. The distribution of the accessions to different populations is indicated by the color code (Q1: red, Q2: green, Q3: blue, Q4: yellow, and Q5: purple). (c) Maximum Likelihood (ML) tree of the 273 accessions drawn by MEGA 6. The color codes for each population are consistent in the figure (b) and (c), but the empty black square as all admixtures among Q1, Q2, Q3, Q4, and Q5.

or five structured populations, as described above. Therefore, we used both the Q matrix with two structures and the matrix with five structures for association mapping in TASSEL. In total, nine models were used in TASSEL to do association analysis of *Stemphylium* leaf spot resistance: a single marker regression (SMR) model; two workflows with linear regression models, GLM (Q) and GLM (PCA); two workflows with mixed linear models, MLM (Q + K) and MLM (PCA + K); two Q-matrices, with 2 vectors and 5 vectors; and two PCA-matrices, with 2 and 5 vectors. First, we screen SNP markers using a Log of the Odds (LOD) (or likely  $\text{LOD} = -\log(P)$ , where P is the P value). Based on the suggestion by Lander and Botstein (1989) [36], a typical LOD threshold should be between 2 and 3 in order to detect at 5% level for a QTL [37]. The LOD value of 2.5 was used as the threshold to screen associated SNP markers. For each SNP, if a LOD value  $\geq 2.5$  was observed in one of the nine models, the SNP was selected for further identification, as listed in **Supplementary Table S2**.

Among the 16 SNPs selected (**Table S2**), different LOD values were observed from different models, SMR gave the largest LOD values, GLM (GLM (Q) and GLM (PCA)) second, and the MLM (MLM (Q + K) and MLM (PCA + K)) the smallest. Of the SNPs with an  $\text{LOD} \geq 2.5$ , there were 13 SNPs in the SMR model; 14 SNPs in GLM (Q) under both two-structured and five-structured populations; 13 and 8 SNPs in the GLM (PCA); 7 and 4 SNPs in the MLM (Q + K), 6 and 5 SNPs in the MLM (PCA + K) under the two-structured and five-structured populations, respectively (**Table S2**). The two-structured populations based Q-matrix and PCA provided more SNP markers with significant association with *Stemphylium* leaf spot resistance, which was consistent with the population structure analysis by STRUCTURE where the peak delta K was observed at K = 2.

The SNP markers listed in **Table S2** were further identified based on five models from TASSEL under the two-structured populations in order to identify and confirm the SNP markers with stable results because different models can give different results. It was assumed that if significant associations were calculated using models for a particular SNP, the SNP marker should be robust for association mapping. If a SNP had an  $\text{LOD} \geq 2.5$  or close to 2.5 in the five models, SMR, GLM (Q), GLM (PCA), MLM (Q + K), and MLM (Q + PCA), the SNP marker was identified as relevant for association mapping. Using these selection criteria, eight SNP markers were identified as being strongly associated with *Stemphylium* leaf spot resistance (**Table 1**).

## 4. Discussion

The distribution of *Stemphylium* leaf spot resistance in the 273 spinach genotypes tested in this study showed a skewed distribution of disease severity that ranged from 0.2% to 23.5% and a near normal distribution was observed after transformation by square root (Figure 1), indicating that *Stemphylium* leaf spot resistance in spinach is a complex trait which is probably controlled by multiple minor genes. Mou *et al.* (2008) [16] screened most of the USDA spinach germplasm collection and commercial cultivars for resistance to *S. botryosum*, and found that no genotype was completely resistant (immune) to the disease but there were significant differences in disease incidence (percentage of plants with leaf spot) and severity (percentage of leaf area diseased) among the genotypes tested. So far, it is not evident whether *Stemphylium* leaf spot resistance in spinach is a quantitative or qualitative trait, or controlled by major genes or minor genes. In this study, several QTLs for *Stemphylium* leaf spot resistance appear to have been identified. All identified SNP markers had low  $R^2$  values, indicating that the resistance identified was associated with minor genes in these spinach genotypes. However, there is no evidence to indicate that major genes for resistance to *Stemphylium* leaf spot could not be responsible for resistance in spinach germplasm. For resistance to similar diseases caused by *S. botryosum* in other crops, major genes have been identified, e.g., Netzer *et al.* (1985) [17] reported two genes, one dominant (*Sm1*) and one recessive (*sm2*), for resistance to *S. botryosum* isolates that infect lettuce. Behare *et al.* (1991) [38] reported that resistance of tomato to gray leaf spot caused by four *Stemphylium* species, *S. solani*, *S. floridanum*, *S. botryosum*, and *S. vesicarium*, was controlled by a single dominant gene, *Sm*, which was mapped onto tomato chromosome 11 by restriction fragment length polymorphism (RFLP) markers. In contrast, minor genes (QTLs) have also been reported, e.g., Kumar (2007) [39] reported that genetic resistance to *Stemphylium* blight of lentil, caused by *S. botryosum*, was quantitatively inherited in progeny of the cross Barimasur-4 × CDC Milestone; and three QTLs for resistance were mapped in lentil [40]. Further QTL mapping using bi-parent populations derived from highly susceptible and highly resistant spinach lines will be needed to fully validate the genetics of resistance to *Stemphylium* leaf spot in spinach.

In this study, five models, and two population structured matrices were used to conduct association analysis of *Stemphylium* leaf spot resistance in this collection of spinach genotypes. Many SNPs showed different results with the different models evaluated. It was assumed that if a particular SNP had significant associations in different models, the SNP marker should be effective for marker associate mapping for resistance to *Stemphylium* leaf spot. Based on LOD ( $-\log(P)$ ) values from all nine models evaluated in this study, eight SNP markers were

**Table 1.** Eight SNP markers associated with *Stemphylium* Leaf spot disease resistance in spinach.

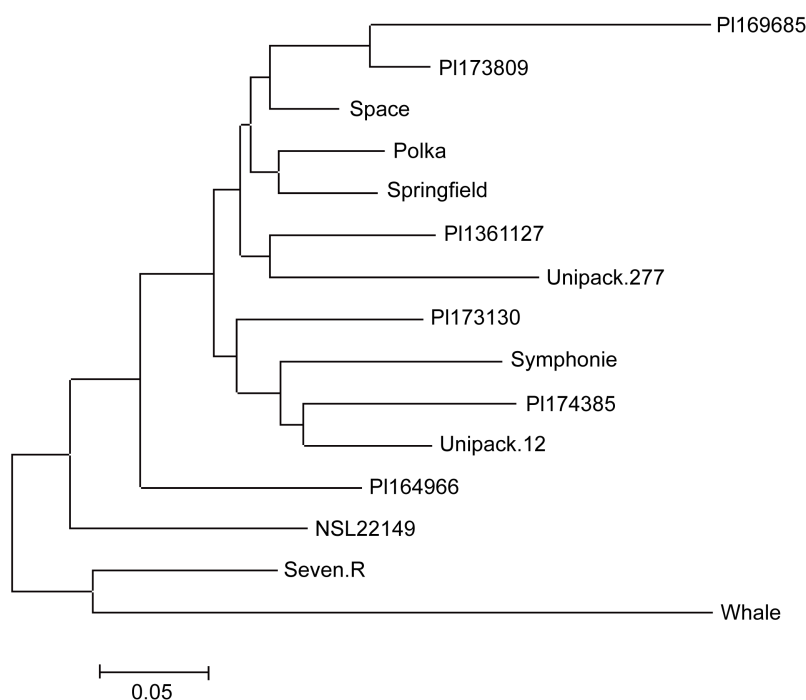
SNP name*	SNP Type	Contig at AYZV02 project	SNP Position	LOD ( $-\log(P)$ ) value from TASSEL 5 <sup>#</sup>					R-sqaure (%) value from TASSEL 5 <sup>#</sup>				
				SMR	GLM (Q)	GLM (PCA)	MLM (Q + K)	MLM (PCA + K)	SMR	GLM (Q)	GLM (PCA)	MLM (Q + K)	MLM (PCA + K)
AYZV02052595_115	A/T	AYZV02052595	115	4.20	4.12	4.02	3.05	3.24	5.90	5.78	5.63	4.31	4.59
AYZV02052595_122	A/C	AYZV02052595	122	3.61	3.48	3.47	2.51	2.62	4.98	4.78	4.76	3.40	3.54
AYZV02057770_10404	A/T	AYZV02057770	10404	2.95	2.84	2.70	2.24	2.26	5.16	4.98	4.74	4.05	4.05
AYZV02129827_205	T/C	AYZV02129827	205	2.62	2.79	2.86	2.98	2.26	4.59	4.88	5.01	5.42	4.04
AYZV02152692_182	A/C	AYZV02152692	182	3.53	3.54	3.56	2.78	2.23	6.00	6.03	6.05	5.12	4.03
AYZV02180153_337	G/T	AYZV02180153	337	4.78	4.76	4.68	4.00	4.05	8.01	8.00	7.89	7.45	7.68
AYZV02225889_197	G/A	AYZV02225889	197	5.98	5.97	5.99	4.70	4.19	10.19	10.21	10.25	8.88	7.83
AYZV02258563_213	C/G	AYZV02258563	213	3.65	3.58	3.56	3.34	3.27	5.00	4.91	4.87	4.71	4.55

\*SNP name defined as contig name plus the SNP position on the contig. <sup>#</sup>Lod ( $-\log(P)$ ) value (where P is the 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.



identified as being strongly associated with *Stemphylium* leaf spot resistance (**Table 1**). Among the eight SNP markers, two markers, AYZV02052595\_115 and AYZV02052595\_122, were located on the same contig, AYZV02052595 at only a 7 bp distance, and the other six markers were located on different contigs. The eight markers had the LOD values  $\geq 2.5$  in all five models except AYZV02057770\_10404 at MLM (Q + K) with LOD value of 2.24, AYZV02057770\_10404, AYZV02129827\_205, and AYZV 02152692\_182 with LOD values, 2.26, 2.26, and 2.23, respectively (**Table 1**). The two markers, AYZV 02180153\_337 and AYZV02225889\_197 showed strongest associated with *Stemphylium* leaf spot resistance with LOD value equaled to or greater than 4.0 and  $R^2$  values greater than 7.4 and 7.8 respectively in all five models. AYZV02052595\_115 and AYZV02258563\_213 had LOD value greater than 3.0 in all five models, showing a strong association. The other four markers, AYZV02052595\_122, AYZV02057770\_10404, AYZV 02129827\_205, and AYZV02152692\_182 also showed significant associations with *Stemphylium* leaf spot resistance with LOD  $\geq 2.5$  or close to 2.5 in the five models (**Table 1**). However, the  $R^2$  values were low or say, not high with  $<10.3\%$  in all the models, providing further evidence that *Stemphylium* leaf spot resistance is a quantitative trait controlled by minor genes.

Among the 265 genotypes of spinach germplasm, seven accessions, NSL22149, PI164966, PI169685, PI173130, PI173809, PI174385, and PI361127 had very low disease severity of *Stemphylium* leaf spot resistance and also showed stable low disease severity ratings evaluated over two years [16]. Two genotypes, PI169685 and PI173809, also consistently had low disease incidence [16]. The genetic associations among the seven accessions and eight spinach cultivars, Polka, Seven R, Space, Springfield, Symphonie, Unipack 12, Unipack 277, and Whale were further analyzed and a phylogenetic tree was generated by MEGA 7 (**Figure 4**). From the phylogenetic tree consisted of 15 spinach genotypes, PI169685 and PI173809 merged together, closest to Space, and then to Poka and Springfield; PI361127 was closest to Unipack 277; and the seven accessions and cultivars merged together as a group 1. PI174385 was closest to Unipack 12; and then to Symphonie; the three with PI173130 merged together as a group 2. Group 1 and 2 merged as larger group 12. PI164966 and NSL22149 were two outliers, close to the group 12. Seven R and Whale merged together, located at outside of the group 12. The phylogenetic analysis of the genetic associations among the 15 spinach accessions and cultivars that had the highest level of resistance can provide the information on how to use the seven accessions with low disease severity of *Stemphylium* leaf spot resistance.



**Figure 4.** Phylogenetic tree among 15 spinach accessions and cultivars drawn by MEGA 7.

## 5. Conclusion

Eight SNP markers, AYZV02052595\_115, AYZV02052595\_122, AYZV02057770\_10404, AYZV02129827\_205, AYZV02152692\_182, AYZV02180153\_337, AYZV02225889\_197, and AYZV02258563\_213 were strongly associated with *Stemphylium* leaf spot resistance. The SNP markers may be useful to select for *Stemphylium* leaf spot resistance in spinach breeding programs through marker-assisted selection.

## Acknowledgements

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

## References

- [1] Morelock, T.E. and Correll, J.C. (2008) Spinach. In: Prohens, J. and Nuez, F., Eds., *Vegetables I: Asteraceae, Brassicaceae, Chenopodiaceae, and Cucurbitaceae*, Springer, New York, 189-218. [http://dx.doi.org/10.1007/978-0-387-30443-4\\_6](http://dx.doi.org/10.1007/978-0-387-30443-4_6)
- [2] Correll, J.C., Bluhm, B.H., Feng, C., Lamour, K., du Toit, L.J. and Koike, S.T. (2011) Spinach: Better Management of Downy Mildew and White Rust through Genomics. *European Foundation for Plant Pathology*, **129**, 193-205. <http://dx.doi.org/10.1007/s10658-010-9713-y>
- [3] Dicoteau, D.R. (2000) *Vegetable Crops*. Prentice Hall, Upper Saddle River.
- [4] Correll, J.C., Morelock, T.E., Black, M.C., Koike, S.T., Brandenberger, L.P. and Dainello F.J. (1994) Economically Important Diseases of Spinach. *Plant Disease*, **78**, 653-660. <http://dx.doi.org/10.1094/PD-78-0653>
- [5] Koike, S.T., Henderson, D.M. and Butler, E.E. (2001) Leaf Spot Disease of Spinach in California Caused by *Stemphylium botryosum*. *Plant Disease*, **85**, 126-130. <http://dx.doi.org/10.1094/PDIS.2001.85.2.126>
- [6] Koike, S.T., Henderson, D.M. and Butler, E.E. (2001) Host-Specific Strain of *Stemphylium* Causes Leaf Spot Disease of California Spinach. *California Agriculture*, **55**, 31-34. <http://dx.doi.org/10.3733/ca.v055n05p31>
- [7] du Toit, L.J. and Derie, M.L. (2001) *Stemphylium botryosum* Pathogenic on Spinach Seed Crops in Washington. *Plant Disease*, **85**, 920. <http://dx.doi.org/10.1094/PDIS.2001.85.8.920B>
- [8] du Toit, L.J. and Hernandez-Perez, P. (2005) Efficacy of Hot Water and Chlorine for Eradication of *Cladosporium variable*, *Stemphylium botryosum*, and *Verticillium dahliae* from Spinach Seed. *Plant Disease*, **89**, 1305-1312. <http://dx.doi.org/10.1094/PD-89-1305>
- [9] Everts, K.L. and Armentrout, D.K. (2001) Report of Leaf Spot of Spinach Caused by *Stemphylium botryosum* in Maryland and Delaware. *Plant Disease*, **85**, 1209. <http://dx.doi.org/10.1094/PDIS.2001.85.11.1209B>
- [10] Raid, R.N. (2004) Management of *Stemphylium* Leaf Spot on Baby Spinach Harvested for Spring Mix Production in the Everglades Agricultural Area of Florida. *Summaries and Program of the 17th International Lettuce and Leafy Vegetable Conference*, Quebec, 28-31 August 2004, 24.
- [11] Koike, S.T., Matheron, M.E. and du Toit, L.J. (2005) First Report of Leaf Spot of Spinach Caused by *Stemphylium botryosum* in Arizona. *Plant Disease*, **89**, 1359. <http://dx.doi.org/10.1094/PD-89-1359A>
- [12] Morelock, T.E. and Correll, J.C. (2006) History of Spinach Production in the United States. *Processing of the International Spinach Conference*, La Conner, WA, 13-14 July 2006, 16.
- [13] Hernandez-Perez, P. and du Toit, L.J. (2006) Seedborne *Cladosporium variable* and *Stemphylium botryosum* in spinach. *Plant Disease*, **90**, 137-145. <http://dx.doi.org/10.1094/PD-90-0137>
- [14] Hernandez-Perez, P. (2005) Management of Seedborne *Stemphylium botryosum* and *Cladosporium variable* Causing Leaf Spot of Spinach Seed Crops in Western Washington. M.S. Thesis, Washington State University, Pullman.
- [15] Koike, S.T., O'Neill, N.R. and Camara, M.P.S. (2001) Further Characterization of the Pathogen That Causes *Stemphylium* Leaf Spot of Spinach in California. *Abstracts of the National Spinach Conference*, Fayetteville, AR, 14-15 November 2001, 7.
- [16] Mou, B., Koike, S.T. and du Toit, L.J. (2008) Screening for Resistance to Leaf Spot Diseases of Spinach. *HortScience*, **43**, 1706-1710.
- [17] Netzer, D., Globerson, D., Weintal, C.H. and Elyassi, R. (1985) Sources and Inheritance of Resistance to *Stemphylium* Leaf Spot of Lettuce. *Euphytica*, **34**, 393-396. <http://dx.doi.org/10.1007/BF00022934>
- [18] Collard, B.C.Y. and Mackill D.J. (2008) Marker-Assisted Selection: An Approach for Precision Plant Breeding in the Twenty-First Century. *Philosophical Transactions of the Royal Society B*, **363**, 557-572. <http://dx.doi.org/10.1098/rstb.2007.2170>

- [19] Xu, Y. and Crouch, J.H. (2008) Marker-Assisted Selection in Plant Breeding: From Publications to Practice. *Crop Science*, **48**, 391-407. <http://dx.doi.org/10.2135/cropsci2007.04.0191>
- [20] Elshire, R.J., Glaubitz, J.C. and Sun, Q. (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE*, **6**, e19379. <http://dx.doi.org/10.1371/journal.pone.0019379>
- [21] He, J., Zhao, X., Laroche, A., Lu, Z., Liu, H. and Li, Z. (2014) Genotyping-by-Sequencing (GBS), an Ultimate Marker-Assisted Selection (MAS) Tool to Accelerate Plant Breeding. *Frontiers in Plant Science*, **5**, 484. <http://dx.doi.org/10.3389/fpls.2014.00484>
- [22] Sonah, H., Bastien, M., Iquira, E., Tardivel, A. and Legare, G. (2013) An Improved Genotyping by Sequencing (GBS) Approach Offering Increased Versatility and Efficiency of SNP Discovery and Genotyping. *PLoS ONE*, **8**, e54603. <http://dx.doi.org/10.1371/journal.pone.0054603>
- [23] Dohm, J.C., *et al.* (2014) The Genome of the Recently Domesticated Crop Plant Sugar Beet (*Beta vulgaris*). *Nature*, **505**, 546-549. <http://dx.doi.org/10.1038/nature12817>
- [24] Minoche, A.E., Dohm, J.C., Schneider, J., Holtgrawe, D., Viehover, P., Montfort, M., Sørensen, T.R., Weisshaar, B. and Himmelbauer, H. (2015) Exploiting Single-Molecule Transcript Sequencing for Eukaryotic Gene Prediction. *Genome Biology*, **16**, 184. <http://dx.doi.org/10.1186/s13059-015-0729-7>
- [25] Van Deynze, A. (2014) A de Novo Draft Assembly of Spinach Using Pacific Biosciences Technology. *Plant & Animal Genomics XXII Conference*, San Diego, 10-15 January 2014. <http://aa314.gondor.co/webinar/a-de-novo-draft-assembly-of-spinach-using-pacific-biosciences-technology/>
- [26] van Deynze, A., Ashrafi, H., Hickey, L., Peluso, P., Rank, D., Chin, J., Rapicavoli, N., Drake, J., Garvin, T. and Schatz, M. (2015) Using Spinach to Compare Technologies for Whole Genome Assemblies. *Plant & Animal Genomics XXIII Conference*, San Diego, 10-14 January 2015.
- [27] Kisha, T., Sneller, C.H. and Diers B.W. (1997) Relationship between Genetic Distance among Parents and Genetic Variance in Populations of Soybean. *Crop Science*, **37**, 1317-1325. <http://dx.doi.org/10.2135/cropsci1997.0011183X003700040048x>
- [28] Li, H. (2011) A Statistical Framework for SNP Calling, Mutation Discovery, Association Mapping and Population Genetic Parameter Estimation from Sequencing Data. *Bioinformatics*, **27**, 2987-2993. <http://dx.doi.org/10.1093/bioinformatics/btr509>
- [29] Li, R., Yu, C., Li, Y., Lam, T., Yiu, S., Kristiansen, K. and Wang, J. (2009) SOAP2: An Improved Ultrafast Tool for Short Read Alignment. *Bioinformatics*, **25**, 1966-1967. <http://dx.doi.org/10.1093/bioinformatics/btp336>
- [30] Pritchard J.K., Stephens, M. and Donnelly, P. (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, **155**, 945-959.
- [31] 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. and Huang, S. (2012) Genetic Diversity and Population STRUCTURE of Cucumber (*Cucumis sativus* L). *PLoS ONE*, **7**, e46919. <http://dx.doi.org/10.1371/journal.pone.0046919>
- [32] Evanno, G., Regnaut, S. and Goudet, J. (2005) Detecting the Number of Clusters of Individuals Using the Software STRUCTURE: A Simulation Study. *Molecular Ecology*, **14**, 2611-2620. <http://dx.doi.org/10.1111/j.1365-294X.2005.02553.x>
- [33] Earl, D.A. and von Holdt, B.M. (2012) STRUCTURE HARVESTER: A Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Method. *Conservation Genetics Research*, **4**, 359-361. <http://dx.doi.org/10.1007/s12686-011-9548-7>
- [34] Tamura, K., Stecher, G., Peterson, D., Ailipksi, A. and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, **30**, 2725-2729. <http://dx.doi.org/10.1093/molbev/mst197>
- [35] Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S. (2007) TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples. *Bioinformatics*, **23**, 2633-2635. <http://dx.doi.org/10.1093/bioinformatics/btm308>
- [36] Lander, E.S. and Botstein, D. (1989) Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps. *Genetics*, **121**, 185-199
- [37] Churchill, G.A. and Doerge, R.W. (1994) Empirical Threshold Values for Quantitative Trait Mapping. *Genetics*, **138**, 963-971
- [38] Behare, J., Laterrot, H., Sarfatti, M. and Zamir, D. (1991) Restriction Fragment Length Polymorphism Mapping of the *Stemphylium* Resistance Gene in Tomato. *Molecular Plant-Microbe Interaction*, **4**, 489-492. <http://dx.doi.org/10.1094/MPMI-4-489>
- [39] Kumar, P. (2007) Genetics of Resistance to *Stemphylium* Leaf Blight of Lentil (*Lens culinaris*) in the Cross Barimaur-4 × CDC Milestone. MSc Thesis, University of Saskatchewan, Saskatoon.

- 
- [40] Saha, G.C., Sarker, A., Chen, W., Vandemark, G.J. and Muehlbauer, F.J. (2010) Inheritance and Linkage Map Positions of Genes Conferring Resistance to *Stemphylium* Blight in Lentil. *Crop Science*, **50**, 1831-1839.  
<http://dx.doi.org/10.2135/cropsci2009.12.0709>

## Supplementary

All Supplementary figures and tables in this part can be found in this link:

[https://agritech.uark.edu/Ainong\\_Shi/Supplementary\\_Table\\_Figure\\_Spinach\\_LeafSpotAssociation.xlsx](https://agritech.uark.edu/Ainong_Shi/Supplementary_Table_Figure_Spinach_LeafSpotAssociation.xlsx)



**Submit or recommend next manuscript to SCIRP and we will provide best service for you:**

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>