Pak. J. Bot., 49(6): 2255-2262, 2017.

STEPWISE REGRESSION ANALYSIS OF THE CORRELATION BETWEEN LEAF TRAITS AND SSR MARKERS OF *MALUS SIEVERSII*

LIHUI ZUO[1,2,#], JUN ZHANG[1,2,#], WENLIN ZHANG[1,2], RUIXUE WANG[1,2] AND MINSHENG YANG[1,2*]

[1]Institute of Forest Biotechnology, ForestryCollege, Agricultural University of Hebei, 071000 Baoding, P. R. China
[2]Hebei Key Laboratory for Tree Genetic Resources and Forest Protection, 071000 Baoding, P. R. China
[*]Corresponding author's e-mail: yangms100@126.com; Ph.: +86 312 7528715; Fax: +86 312 7528799
[#]Contributed equally to this study

Abstract

In this study, variation of leaf traits of 18 clones selected from 600 *Malus sieversii* clones collected from Tianshan Mountain, Xinjiang (China) was investigated. SSR molecular markers were prepared; correlation between leaf traits and SSR(Simple Sequence Repeats) marker was determined through stepwise regression. The results showed that there were significant differences between the 15 leaf traits. The tip angle α showed the largest average variation coefficient of 25.39%, whereas the leaf index L_1/L_3 yielded the lowest variation of 10.31%. The repeatability of most leaf traits were more than 0.80, which indicated that genetic factors accounted for a high proportion of phenotypic leaf traits. The clones were completely separated with the clustered result of 15 leaf traits and the genetic distance were varied between 1.293 and 7.235. Furthermore, 18 clones with 30 pairs of polymorphic SSR primers, which were evenly distributed among 17 chromosome linkage groups, were subjected to cluster analysis; the clones were completely separated. The genetic distances varied between 0.089 and 0.689, with the average genetic distance was 0.4328. Cluster results of all leaf traits were not correlated with that of SSR markers. Stepwise regression results showed that the same leaf trait was closely related to several sites; likewise, the same site was closely related to several leaf traits. The sites correlation with each single trait varied between 11 and 20, in which 11 sites yielded a leaf index of L_1/A_1, 11 sites exhibited a leaf index of L_1/A_4, and 20 SSR sites were correlated with the width of 3/4 of the leaf (A_4). Therefore, phenotypic traits of the leaf were closely correlated with several SSR sites.

Key words: *Malus sieversii*, Phenotypic traits, SSR, Cluster, Stepwise regression, Correlation analysis.

Introduction

Malus Mill. are characterized with rich genetic resources and diverse varieties, including a considerable number of wild, half-wild, and cultivated varieties. *Malus sieversii* (Ledeb.) Roem. are famous for its economic value in fruit cultivation, production, and breeding (Qian & Tang 2005; Treuren *et al.*, 2010; Liu *et al.*, 2013). For instance, *M. sieversii* (Ledeb.) Roem is an important wild germplasm resource in China. Due to the excellent resistance to cold, disease, drought and grafting affinity, *M. sieversii* were widely used as the rootstock of cultivated apple that had been considered as one of the main rootstocks in the northwest and other apple production areas in China (Zhang *et al.*, 2007; Chen *et al.*, 2007; Zhang *et al*., 2009; Dong *et al*., 2013).

Leaf is one of the most important organs in plants, many physiological and biochemical reactions are completed in the leaves, which are closely related to the growth and development of plants (Mondal *et al.*, 2016). Most of the indexes of leaf traits of plants are quantitative traits, and most quantitative traits are effected by multi gene interaction. Therefore, it is very important to grasp the relationship between quantitative traits and genes in plants, which can further promote the breeding work. There are many studies on the relationship between plant phenotypic traits and molecular markers, Liu *et al*. (2013) reported that morphological characteristics were closely related to molecular markers. Lai *et al*. (2013) studied *Hordeum vulgare* by linkage disequilibrium (LD) analysis of 86 SSR markers and found that sites were closely related to phenotypic traits, such as plant height, spike length, awn length, and grain number. Liu (2012) argued that phenotypic traits were significantly correlated with SSR cluster results of pumpkin. Other studies have also analyzed the correlation between phenotypic traits and molecular markers in crops (Ivandic *et al.*, 2002, 2003; Brantestam *et al.*, 2003); however, in a few studies trees were also evaluated. In the present study, 18 *M. sieversii* clones in Tianshan Mountain of Xinjiang were used as subjects to conduct stepwise regression analysis of leaf data and SSR molecular markers. This study was also performed to determine sites closely correlated with leaf indexes. Furthermore, this study aimed to establish a reference of gene mining, expression, and positioning and to provide technical assistance in molecular breeding and protection of apples. Our results revealed wild genetic resources that could be used in several applications.

Materials and Method

Experimental materials: A total of 18 *M. sieversii* clones were selected from 600 wild plants collected from Xinyuan and Gongliu counties in Tianshan Mountain, Xinjiang. Wild plants were grafted onto a two-year crabapple rootstock, and 21 plants were used for each clone. The clones were planted in the sample garden of the Agricultural University of Hebei by using a randomized block design, with three replicates and seven plants in each block. In the second year, heading was trimmed to 80 cm. In the fourth year, phenotypic traits were investigated (Table 1). In spring of 2014, branches with new leaves were cut, each clone 3-5 branch. The branches were placed in numbered envelopes with silica gel. Silica gel was periodically replaced until the leaves were dried. In our laboratory, dried leaves were detached from the branches and sealed in valve bags, in which two bags were used for each line. The bags were then stored in refrigerator at −70 °C for SSR marker analysis.

Table 1. Phenotypic trait selection and coding.

No.	Character	Code
1.	Leaf width	A_1
2.	Leaf1/2 width	A_2
3.	Leaf1/3 width	A_3
4.	Leaf 3/4width	A_4
5.	Leaf length	L_1
6.	Length of leaf stalk	L_2
7.	Distance from Leaf width to leaf base	L_3
8.	Shape of leaf tip	α
9.	Shape of leaf base	γ
10.	Shape index of Leaf	L_1/A_1
11.	Shape index of Leaf	L_1/A_2
12.	Shape index of Leaf	L_1/A_3
13.	Shape index of Leaf	L_1/A_4
14.	Shape index of Leaf	L_1/L_2
15.	Shape index of Leaf	L_1/L_3

Experimental Method

Investigation and coding of phenotypic traits: After three years of continual observation (Liu *et al.,* 2005), 15 leaf traits with stable performance and great difference among clones were selected for investigation (Fig. 1). Ten plants were randomly selected from each clone in each block to ensure the accuracy of leaf data were obtained. In addition, 10 fully expanded functional leaves were collected from the same part of each plant and transported to laboratory to measure leaf indexes of each clone by using an electrical ruler, a goniasmometer, and other measuring instruments at low temperature.

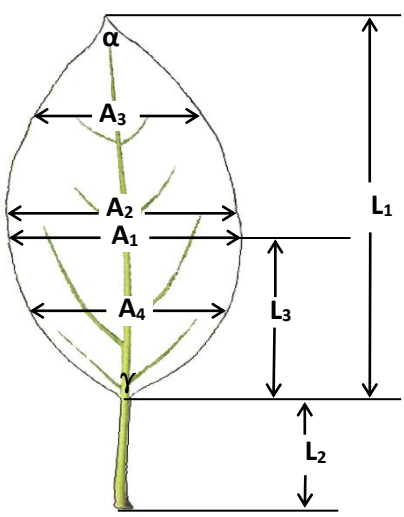

Fig.1. Leaf Trait Scheme of *Malus sieversii*.

DNA extraction: Genomic DNAs of all clones were extracted according to the improved CTAB method (Han *et al.,* 2009; Sun *et al*., 2015). First, tissues and cells were crushed. The soluble substances in the cells were removed through dissolution and centrifugation. The cellnuclei were collected and subjected to further pyrolysis and separation. After the solution was purified, genomic DNAs were detected. After determining the quantity with a NanoDrop 2000, all DNA samples were diluted to 30 $ng \cdot \mu L^{-1}$ and stored at −70°C until the subsequent PCR reaction.

SSR marker primer: SSR primer sequences were prepared according to those described in previous studies (Guilford *et al.,* 1997; Hokanson *et al.,* 1998, 2001; Gianfranceschi *et al.,* 1998) and in NCBI. A total of 80 pairs of SSR primers were selected from 17 linkage groups of the apple genome. Finally, 30 pairs of SSR primers with high polymorphism, clear bands and good repeatability were screened by gel electrophoresis analysis. All primers were synthesized by Shanghai Sangon Bio-Engineering Company (Table 2).

PCR system construction and expansion procedure : A 10 μL PCR system was used in this study. The components and proportions were listed as follows: 4 μL of 2×*Taq* Master mix, 0.5 μL of each of the forward and reverse primer, 1 μL of DNA template, and 4 μL of deionized water. The following PCR conditions were set: pre-denaturation at 95°C for 5 min, denaturation at 95°C for 1 min, annealing at 50–60°C for 1 min, extension at 72°C for 1 min (30 cycles), and further extension at 72°C for 7 min; the reaction was maintained at 4°Cfor 5 min. PCR was performed using a TG-type PCR cycler (Biometer Company).

Gel electrophoresis and silver staining: Electrophoresis was conducted using 8% non-denaturing polyacrylamide gel. Approximately 5 μL of 6× loading buffer was added to PCR products and 2 μL was added to the sample hole. Electrophoretic ink 1×TBE was used. After the sample was added, electrophoresis was performed at 230 V for 30 – 40 min until bromphenol blue strip was 1 cm from the bottom. A 20 bp DNA ladder marker (TaKaRa) was used. Non-denaturing polyacrylamide gel (8%) was immersed in 1% $AgNO_3$ and slightly shaken for 10 min. It was washed two to three times with deionized water and immersed in colored liquid for 2–3 min until the strip was clearly observed. It was washed two to three times with deionized water again and retrieved for photographing and recording.

Statistical analysis: The phenotypic characters were coded and the data were analyzed by variance analysis and cluster analysis, The average value, coefficient of variation and repeatability of each trait were calculated. Repeatability refers to the degree of continuous stability of the phenotype of the organism with the same genotype at different times or different places. Repeatability can also be defined as the stability of a quantitative trait after measurement is performed several times and expressed as follows:

$$R = V_b / (V_b + V_w/K),$$

where V_b is the between-group variable, V_w is the inside-group variable, and K is the number of individuals of each clone for measurement (Duan *et al.,*1995).

Each ladder of the electrophoresis spectrum corresponded to a molecule marker representing the combination site of a primer. Binary data were calculated according to the ladders: with a adder was recorded as 1 and without a strip was recorded as 0. Polymorphism information content of each primer was

calculated by *PIC* Calc 0.6 software. Based on Nei and Li dissimilarity coefficient, cluster analysis was performed by using the group-average method; a cluster tree-like diagram was established.

The cluster matrix of each trait could be obtained using cluster analysis results and SSR data. Quantitative traits were used as independent variables and all primers were used as dependent variables; the correlation between phenotypic traits and SSR sites could be investigated through stepwise regression. In general, this procedure mainly aims to introduce all considered independent variables (SSR sites) to the regression equation from the largest to smallest according to the corresponding action on Y (leaf traits), significance degree, or contribution. However, variables with no significant action on Y were not introduced to the regression equation. Variables introduced to the regression equation may lose significance after new variables are introduced; as such, such variables should be eliminated. Introduction and elimination of variables from regression equation are steps of stepwise regression. In each step, F should be detected to ensure that only variables with a significant influence on Y are included in the regression equation and non-significant variables are eliminated before new variables are introduced; sites closely related to traits can be determined (Mordechai *et al*., 2014). Statistical analyses were performed in Excel 2007 and DPS© Data Processing System (DPS 7.05) (Tang, 2010).

Results and Analysis

Leaf trait variation analysis: Variance analysis results showed that 15 leaf traits among the 18 clones exhibited very significant variance; these results indicated great variation and high genetic diversity of *M. sieversii* leaf. The variation coefficient of each trait varied between 10.31% and 25.39%, in which the tip angle α showed the largest average variation coefficient; by contrast, the leaf index L_1/L_3 yielded the lowest variation coefficient. The repeatability of all leaf traits were over than 0.80 except the leaf index L_1/L_3 (0.495). All of this indicated that genetic factors accounted for a high proportion of the phenotypic traits of leaves.

Cluster analysis of leaf traits: The Clustering result of 15 leaf traits showed that all the clones were completely separated and could be divided into three categories. The genetic distance varied between 1.293 and 7.235 with the average genetic distance was 3.274 (Fig. 2). At 5.06, all of the clones were divided into three categories. Category 1 included designated as No.1, 2, 4,and other 8 clones; Category 2 included 4 clones, namely, No.5, 6,7,and 12; Category 3 included 3 clones, namely, No.8, 10, and 16.

Table 2. Primers selection.

No.	Primer	Forward primer sequence (5′-3′)	Reverse primer sequence (5′-3′)	Tm (°C)	Linkage group
1.	HI02c07	AGAGCTACGGGGATCCAAAT	GTTTAAGCATCCCGATTGAAAGG	57.8	LG1
2.	CH02b10	CAAGGAAATCATCAAAGATTCAAG	CAAGTGGCTTCGGATAGTTG	55.1	LG2
3.	CH02c02a	CTTCAAGTTCAGCATCAAGACA	TAGGGCACACTTGCTGGTC	56.6	LG2
4.	CH03g07	AATAAGCATTCAAAGCAATCCG	TTTTTCCAAATCGAGTTTCGTT	54.5	LG3
5.	MS14h03	CGCTCACCTCGTAGACGT	ATGCAATGGCTAAGCATA	59.6	LG3
6.	CH02h11a	CGTGGCATGCCTATCATTTG	CTGTTTGAACCGCTTCCTTC	57.8	LG4
7.	CH04e02	GGCGATGACTACCAGGAAAA	ATGTAGCCAAGCCAGAGTAT	57.8	LG4
8.	CH03a09	GCCAGGTGTGACTCCTTCTC	CTGCAGCTGCTGAAACTGG	61.9	LG5
9.	CH05106	TTAGATCCGGTCACTCTCCACT	TGGAGGAAGACGAAGAAGAAAG	60.1	LG5
10.	CH03d02	GCCCAGAAGCAATAAGTAAACC	ATTGCTCCATGCATAAAGGG	58.2	LG6
11.	CH03d07	CAAATCAATGCAAAACTGTCA	GGCTTCTGGCCATGATTTTA	52.2	LG6
12.	CH02a04	GAAACAGGCGCCATTATTTG	AAAGGAGACGTTGCAAGTGG	55.8	LG7
13.	CH04e05	AGGCTAACAGAAATGTGGTTTG	ATGGCTCCTATTGCCATCAT	56.3	LG7
14.	CH01c06	TTCCCCATCATCGATCTCTC	AAACTGAAGCCATGAGGGC	57.8	LG8
15.	CH01f03b	GAGAAGCAAATGCAAAACCC	CCCGGCTCCTATTCTAC	55.8	LG9
16.	CH01h02	AGAGCTTCGAGCTTCGTTTG	ATCTTTTGGTGCTCCCACAC	57.8	LG9
17.	CH02B03	ATAAGGATACAAAAACCCTACACAG	GACATGTTTGGTTGAAAACTTG	57.1	LG10
18.	CH04g07	CCCTAACCTCAATCCCCAAT	ATGAGGCAGGTGAAGAAGGA	57.8	LG11
19.	CH04d07	TGTCCTCCAATCTTAACCCG	CACACAGACGACACATTCACC	57.8	LG11
20.	CH04d02	CGTACGCTGCTTCTTTTGCT	CTATCCACCACCCGTCAACT	57.8	LG12
21.	CH01F02	ACCACATTAGAGCAGTTGAGG	CTGGTTTGTTTTCCTCCAGC	58.0	LG12
22.	CH03a08	TTGGTTTGCTAGGAAAAGAAGG	AAGTTTATCGGGCCTACACG	56.3	LG13
23.	CH05a04	CCTTCGTTATCTTCCTTGCATT	GAGCTTAAGAATAAGAGAAGGGG	56.3	LG13
24.	CH03d08	CATCAGTCTCTTGCACTGGAAA	TAGGGCTAGGGAGTGATGATGA	58.2	LG14
25.	CH02c09	TTATGTACCAACTTTGCTAACCTC	AGAAGCAGCAGAGGAGGATG	56.8	LG15
26.	CH02d11	AGCGTCCAGAGCAACAGC	AACAAAAGCAGATCCGTTGC	59.6	LG15
27.	CH02a03	AGAAGTTTTCAGGGGTGCC	TGGAGACATGCAGAATGGAG	57.6	LG16
28.	CH05a04	GAAGCGAATTTTGCACGAAT	GCTTTTGTTTCATTGAATCCCC	53.7	LG16
29.	CH01a01	GAAAGACTTGCAGTGGGAGC	GGAGTGGGTTTGAGAAGGTT	59.8	LG17
30.	CH05g03	GCTTTGAATGGATACAGGAACC	CCTGTCTCATGGCATTGTTG	58.2	LG17

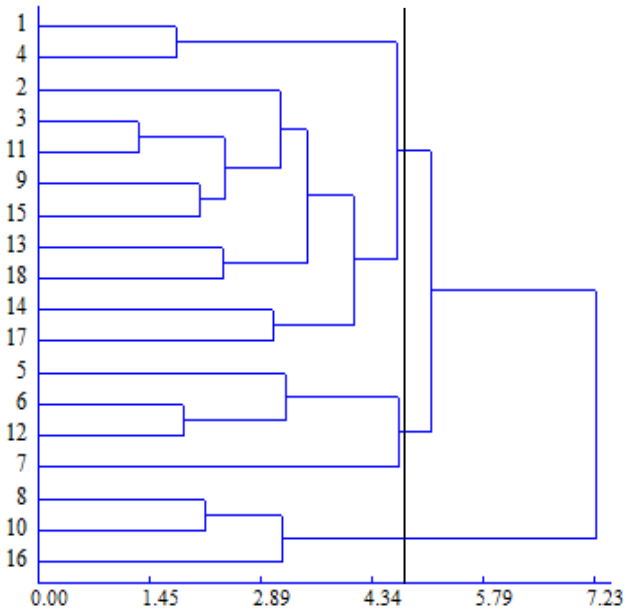

Fig. 2. Cluster results of 15 leaf traits of *Malus sieversii* clones.

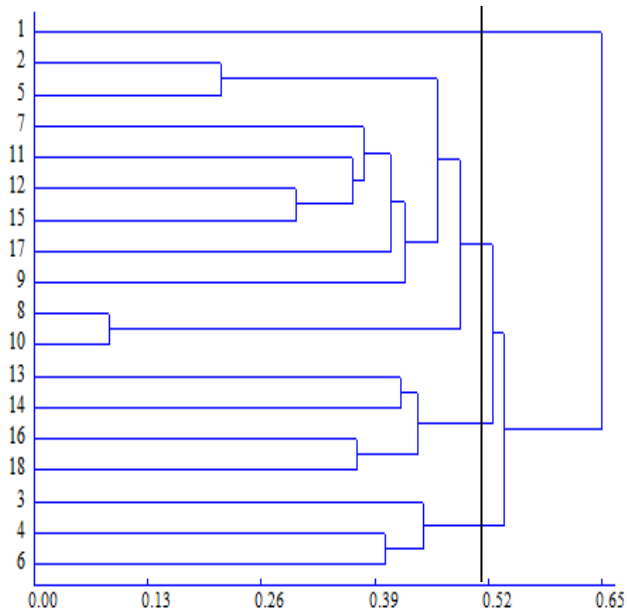

Fig. 3. Clustering chart of 30 pairs of primers of *Malus sieversii*.

SSR cluster analysis: With the expansion data statistical analysis of 30 pairs of primers of 18 *M. sieversii* (Fig.3), 18 clones can be completely distinguished, and the genetic distance among the clones varied between 0.089 and 0.689 with an average genetic distance of 0.433. At 0.52, 18 clones could be divided into four categories, in which No.1 clone belonged to a separate category, Category 2 included No.2, 5, and 8 other clones, Category 3 included No. 13, 14, 16, and 18, and Category 4 included No. 3, 4, and 6.

Correlation analysis between leaf traits and SSR markers: Correlation analysis was performed on the distance matrixes of all leaf trait and SSR marker clusters. The following equation could be fitted: $y = 1.901x + 4.125$. The correlation coefficient was r = 0.105 (n= 153), indicating no significant correlation. In other words, all SSR marker sites were not correlated with the comprehensive leaf parameters of the clones.

Stepwise regression analysis of single leaf trait and SSR sites: Each leaf trait and primer were separately clustered to further explore the correlation between leaf traits and SSR molecular markers of *M. sieversii* clones. All of the testing SSR sites were screened using regression equation in which phenotypic traits were used as independent variables and all of the primers were used as dependent variables to determine the SSR sites closely related to the phenotypic traits. The importance of the sites could be determined according to the parameters of regression equation and expressed as a ratio (correlation degree, %) of one coefficient to all coefficients (absolute value). Table 4 showed that the sites correlated with each single trait varied between 11 and 20, in which 11 sites yielded the leaf index of L_1/A_1, 11 sites exhibited the leaf index of L_1/A_4, and 20 sites correlation with the width of 3/4 of the leaf A_4. In addition, the site of No.17 primer yielded the highest correlation degree of L_1/A_4 and reached 29.97%. The cumulative value of the first six sites with the highest correlation degree showed that the traits varied between 50.72% and 71.80%. All of the parameters reached a very significant correlation level. Multiple correlation coefficients varied between 0.5246 and 0.7164, in which the correlation coefficient of L_1/L_3 was the smallest and that of A_4 was the largest.

Plant growth and development, as well as gene regulation, expression, and translation, are affected by various conditions. The same locus was associated with multiple traits, but the importance of the association was different (Table 5). The number of traits correlated with the primer sites varied between 3 and 13, in which the number of traits that correlated with No. 11, 15, and 30 sites were the least and the number of traits that correlated with No.9 was the highest. The average contribution rate of the sites to all leaf traits varied from 2.73% to 14.24%. In particular, the average contribution rate of No. 1 site was the greatest, indicating that this site was closely correlated with leaf traits. By contrast, the contribution rate of No.15 site was the smallest, indicating that this site was little correlated with leaf traits.

Discussion

Apple has been cultivated in China for over 2000 years. During evolution, *M. sieversii* has been affected by drastic changes in geological and climatic environments. Hybridization of various types has produced diverse hybrid seedlings. In addition, different types compete with one another in warm and wet environments. In wet valleys, basins, and high-altitude inversion layers, as well as under changing slope aspect and soil moisture, wild apple shows diverse ecological types with different plant types, fruit types, colors, maturities, and flavors. After a long time evolution, it formed rich and stable genetic resources of *M. sieversii* (Zhang, 1973). In our study, *M. sieversii* clones exhibited rich genetic diversities in leaf traits. The

variation coefficients of the traits varied between 10.31% and 25.39%. These traits were significantly different among the clones. The quantitative traits of plants are greatly affected by the environment; as such, accurate measurement of the genetic variation from the phenotype is difficult. In this study, all clones were planted in the nurseries with consistent environmental conditions in a randomized block design, which eliminates individual phenotypic difference caused by environmental factors to a large extent. The repeatabilities of all traits were over than 0.8. Thus, genetic difference of traits was accurately revealed and reliability of the experiment was verified.

Studies have been conducted regarding the correlation between molecular markers and phenotypic traits through traditional linkage and LD analyses. Based on family research, linkage analysis require large numbers of samples and long breeding cycles. Moreover, the genetic diversity of filial generation is reduced because of several backcrosses. As a result, the reliability of results is reduced (Salvi & Tuberosa, 2005). In contrast to linkage analysis, LD analysis is commonly used because of the following advantages: (1) existing natural group can be used as material rather than a specially constructed mapping population; (2) several alleles can be simultaneously detected at the same locus; (3) high positioning accuracy can be obtained and single gene level can be reached. However, LD analysis poses some problems. For example, mating system (LD decreases more slowly than that of sibling species by consanguineous mating) and migration among populations can increase LD. Gene flow in subpopulations can greatly increase LD level of a whole population, leading to false correlation (Salvi & Tuberosa , 2005). Researchers also determined the correlation between SSR markers and phenotypic traits based on the correlation among matrixes. For instance, Mamunur *et al*. (2011) used 21 rice as materials, and studied their correlation of genetic matrix with 34 SSR primers and 13 phenotypic traits, and they found that r was 0.321, indicating a very significant correlation level. Yoseph *et al*. (2005) analyzed 15 phenotypic traits of 62 corn materials from Ethiopia by using genetic correlation matrix method in which eight pairs of AFLP primers and 20 pairs of SSR primers were considered; Yoseph (2005) found that the correlation of SSR and AFLP with the genetic matrix of phenotypic traits is significant, with coefficients of 0.43 and 0.39, respectively. Liu (2012) studied the correlation between 33 pairs of SSR primers and 44 phenotypic traits of 76 pumpkin samples by using the correlation matrix method and obtained r of 0.7295, indicating an extremely significant correlation. These studies have shown significant correlation between phenotypic traits and SSR cluster results. However, Chu (2007) investigated 76 Chinese pumpkin samples and found that phenotypic traits were not significantly correlated with molecular markers, with r of 0.105. The following reasons may be accounted for these varying results: (1) Phenotypic traits are controlled by multiple genes; therefore, regulatory mechanism is very complex and detection sites are insufficient to cover all genes. (2) The selected primers are randomly distributed on a chromosome, leading to great randomness; the variation reflected by the detection sites is inconsistent with the investigated phenotypic traits. (3) The difference in DNA structure is unnecessarily expressed in the phenotype or diverse expression may be a result the interaction or regulation among genes. (4) Phenotypic traits are greatly affected by the environment, and its phenotypic value is the result of the combined action of the genotype and the environment. In some cases, phenotypic traits cannot truly reflect the genetic diversity of plants.

Table 3. Leaf trait variation of each clone.

No.	Character	Mean	Max.	Min.	Standard deviation	Coefficient of variation(%)	F-value	Repeatability
1.	A_1	5.623	8.20	3.99	1.10	19.50	25.832**	0.961
2.	A_2	5.52	8.13	3.90	1.13	20.43	21.006**	0.952
3.	A_3	4.53	7.06	3.21	1.03	22.63	8.889**	0.888
4.	A_4	5.04	7.27	3.68	0.99	19.65	10.322**	0.903
5.	L_1	9.772	12.55	7.65	1.45	14.82	30.632**	0.967
6.	L_2	2.51	3.06	1.56	0.46	18.28	7.203**	0.861
7.	L_3	3.95	5.50	2.86	0.73	18.49	5.666**	0.824
8.	α	60.87	90.00	37.33	15.46	25.39	5.797**	0.827
9.	γ	138.09	181.67	100.67	24.87	18.01	8.708**	0.885
10.	L_1/A_1	1.77	2.18	1.35	0.25	13.97	14.742**	0.932
11.	L_1/A_2	1.81	2.29	1.36	0.27	14.71	13.086**	0.924
12.	L_1/A_3	2.23	3.07	1.58	0.38	16.92	8.225**	0.878
13.	L_1/A_4	1.98	2.43	1.49	0.29	14.70	12.946**	0.923
14.	L_1/L_2	3.98	4.90	3.09	0.57	14.29	6.715**	0.851
15.	L_1/L_3	2.53	2.93	2.09	0.26	10.31	1.982**	0.495

Note: ** Refers to extremely significant difference ($p<0.01$) of the discrepancy level of all clones

Table 4. Stepwise regression statistics table.

Character	Number of correlated sites	First 6 sites which had the highest correlation with the traits (correlation degree)	Accumulative correlation degree of first 6 sites	Multiple correlation coefficients
A_1	15	16 (11.31%) 21(9.61%) 13(8.30%) 27(7.65%) 9(7.54%) 29(7.44%)	51.86%	0.6955**
A_2	17	16(11.74%) 21(10.13%) 13(7.83%) 27(7.29%) 9(7.06%) 19(6.67%)	50.72%	0.7096**
A_3	18	16(12.39%) 27(11.00%) 21(9.18%) 23(7.41%) 18(6.61%) 9(6.24%)	52.83%	0.6744**
A_4	20	13(10.37%) 16(9.00%) 27(8.89%) 23(8.13%) 9(7.44%) 21(7.03%)	50.86%	0.7164**
L_1	12	21(11.64%) 27(11.18%) 29(10.86%) 19(9.76%) 17(8.81%) 20(8.61%)	60.85%	0.5330**
L_2	18	21(12.93%) 4(10.25%) 13(7.99%) 24(7.63%) 18(6.26%) 8(6.19%)	51.24%	0.6610**
L_3	16	1(27.31%) 20(9.87%) 25(9.54%) 17(8.97%) 4(7.93%) 13(4.50%)	68.13%	0.6284**
α	14	27(13.37%) 25(10.85%) 16(9.52%) 29(9.48%) 7(8.38%) 28(7.41%)	59.00%	0.6382**
γ	12	4(16.21%) 26(14.53%) 20(12.25%) 1(10.54%) 16(9.55%) 18(7.71%)	70.79%	0.6035**
L_1/A_1	11	27(21.90%) 16(12.61%) 24(11.64%) 5(11.18%) 7(7.51%) 9(6.97%)	71.80%	0.6086**
L_1/A_2	12	27(21.86%) 16(13.26%) 24(12.49%) 9(9.16%) 5(9.03%) 2(5.91%)	71.71%	0.6123**
L_1/A_3	13	27(15.83%) 29(11.23%) 24(8.73%) 5(8.61%) 16(8.56%) 18(8.04%)	61.01%	0.6038**
L_1/A_4	11	27(29.97%) 7(11.01%) 16(9.69%) 9(9.68%) 4(8.21%) 13(7.66%)	76.21%	0.5757**
L_1/L_2	13	24(17.06%) 18(14.23%) 4(7.97%) 8(7.92%) 17(7.04%) 25(6.80%)	61.02%	0.5883**
L_1/L_3	14	1(21.30%) 17(9.49%) 4(9.08%) 29(7.92%) 25(7.81%) 20(7.47%)	63.07%	0.5246**

Note: ** Refers to extremely significant correlation ($p<0.01$)

Table 5. Primer site correlation analysis.

Primer sites	Number of correlated traits	Maximum correlation degree	Minimum correlation degree	Average of correlation degree
1.	5	L_3(27.31%)	A_3(4.67%)	14.24%
2.	5	L_1/A_2(5.91)	L_2(3.46%)	4.91%
3.	5	L_1/A_3(7.60%)	A_3(3.40%)	5.55%
4.	6	γ(16.21%)	L_3(7.93%)	9.94%
5.	11	L_1/A_1(11.18%)	α(3.44%)	5.85%
6.	8	L_1/A_2(5.15%)	A_4(1.98%)	3.63%
7.	7	L_1/A_4(11.01%)	A_2(4.02%)	6.39%
8.	4	L_1/L_2(7.92%)	L_3(2.46%)	4.97%
9.	13	L_1/A_4(9.68%)	L_2(3.62%)	6.60%
10.	6	L_1/L_2(6.13%)	L_2(3.17%)	4.38%
11.	3	L_1(5.15%)	A_4(2.26%)	3.58%
12.	6	α(7.07%)	L_3(3.29%)	4.57%
13.	9	A_4(10.37%)	A_3(2.67%)	6.88%
14.	6	L_1(7.25%)	L_2(2.71%)	5.26%
15.	3	L_1/L_3(3.03%)	A_4(2.45%)	2.73%
16.	10	L_1/A_2(13.26%)	L_1/A_3(8.56%)	10.76%
17.	5	L_1/L_3(9.49%)	A_4(3.75%)	7.61%
18.	10	L_1/L_2(14.23%)	L_1/L_3(3.21%)	6.59%
19.	10	L_1(9.76%)	L_2(2.63%)	5.12%
20.	6	γ(12.25%)	A_3(4.00%)	7.81%
21.	8	L_2(12.93%)	γ(5.08%)	8.86%
22.	7	γ(5.69%)	L_1/A_3(2.97%)	4.27%
23.	5	A_4(8.13%)	L_2(5.03%)	6.60%
24.	9	L_1/L_2(17.06%)	L_1/A_4(3.91%)	8.78%
25.	9	α(10.85%)	L_1/A_4(4.14%)	7.15%
26.	6	γ(14.53%)	α(3.19%)	5.70%
27.	10	L_1/A_4(29.97%)	L_3(2.89%)	13.17%
28.	9	α(7.41%)	L_3(3.25%)	4.65%
29.	9	L_1/A_3(11.23%)	L_1/L_2(4.55%)	7.65%
30.	3	L_1/L_2(6.35%)	L_3(1.93%)	3.55%

This study is the first to investigate the correlation between single trait and a SSR marker through stepwise regression analysis based on genetic matrix. In stepwise regression, irrelevant sites, can be eliminated and SSR sites that are correlated with leaf traits can be identified; furthermore, the correlation degree between each site and a leaf trait can be analyzed. In our study, the same leaf trait was closely correlated with several sites; likewise, the same site was closely correlated with several leaf traits. This finding is consistent with QTL results obtained by analyzing the correlation between phenotypic traits and SSR markers of 24 tobacco samples by LD analysis (Zheng *et al.*, 2014). Our results also showed that single leaf trait was correlated with several SSR sites (11 sites to 20 sites), indicating that leaf traits were determined by the interaction of multiple genes rather than controlled by single gene (site). The degrees of correlation of different sites with a trait were also different, in which the correlation degree between No.17 primer site and L_1/A_4 (29.97%) was the highest. The accumulative correlation degree of the first six sites with the highest correlation degrees with the traits over 50%, which could explain most traits. Site analysis results showed that the same site was correlated with several different phenotypic traits, but the correlation degree was different. Three traits, which were the least, were correlated with No. 11, 15, and 30 sites; by contrast, 13 traits were correlated with No.9 site, and the number of these traits was the highest. This finding was observed possibly because expression and regulation of genes correlated with different sites were related with those of the genes of some intermediate products, such as synthetase, or those under different regulatory mechanisms; thus, different traits were influenced. The average contribution rate of the sites to all leaf traits varied between 2.73% and 14.24%. In particular, the contribution rate of No.1 site was the largest, indicating the close correlation between No.1 site and the leaf traits. The contribution rate of No.15 site was the smallest, indicating this site was poorly correlated with leaf traits. Furthermore, SSR molecular markers were possibly correlated with phenotypic traits. Our study is the first to determine the correlation between traits and SSR markers through stepwise regression analysis; however, theoretical foundation and scientific explanation of this method should be further investigated because this method can be of great significance to functional gene mining, expression, and positioning.

Acknowledgments

This research was supported by the Special Research Program for Social and Public Interests of State Forestry Administration of China (Grant No. 201104039).

References

Brantestam, A.K., R. Bothmer, C. Dayteg, I. Rashal, S. Tuvesson and J. Weibull. 2007. Genetic diversity changes and relationships in spring barley (*Hordeum vulgare* L.) germplasm of Nordic and Baltic areas as shown by SSR markers. *Genet. Resour. Crop. EV.*, 54(4): 749-758.

Chen, X.S., T. Feng, Y.M. Zhang, T.M. He, J.R. Feng and C.Y. Zhang. 2007. Genetic Diversity of volatile components in Xinjiang wild apple (*Malus sieversii*). *J. Genet. Genomıcs.*, 34(2): 171-179.

Chu, P.P. 2007. Studies on genetic diversity of *cucurbita moschat*a germplasmresource. Wuhan: Huazhong Agricultural University.

Dong, Y., J. Zhang, Y.C Ren and Z.X. Han. 2013. Study on genetic diversity of natural populationin *Malus sieversii* with microsatellite. *J. Plant Genetic Res.*, 14(5): 771-777.

Duan, A.A., C.X. Zhang and C.X. Bi. 1995. Research on the methods of repeatability estimation in clonal trial of trees. *J. Northwest Forestry College*, (3): 21-24.

Gao, Y., F.Z. Liu, Y.F. Cao and K. Wang. 2007. Analysis of genetic relationship for *Malus* germplasm resources by SSR markers. *J. Fruit Sci.*, 24(2): 129-134.

Gianfranceschi, L., N. Seglias, R. Tarchini, M. Komjanc and C. Gessler. 1998. Simple sequence repeats for the genetic analysis of apple. *Theor. Appl. Genet.*, 96(8): 1069-1076.

Guilford, P., S. Prakash, J.M. Zhu, E. Rikkerink, S. Gardiner, H. Bassett and R. Forster. 1997. Microsatellites in *Malus*×*damestica* (apple): abundance, polymorphism and cultivar identification. *Theor. Appl. Genet.*, 94(2): 249-254.

Han, Q., H. Wang, J.M. Wang, J.R. Zhao and M.S. Yang. 2009. Genertic diversity analysis of salicaceae cultivars by transferablity of SSR primers. *Mol. Plant Breed.*, 7(5): 904-911.

Hokanson, S.C., W.F. Lambcy, A.K. Szewc-McFadden and J.R. McFerson. 2001. Microsatellite (SSR) variation in a collection of *Malus* (apple) speciesand hybrids. *Euphytica*, 11(8): 1294.

Hokanson, S.C., A.K. Szewc-McFadden, W.F. Lamboy and J.R. McFerson. 1998. Microsatellite (SSR) markers reveal genetic identities diversity and relationship in *Malus* ×*domestic* Borkh.core subset collection. *Theor. Appl. Genet.*, 97(5-6): 671-683.

Ivandic, V., C.A. Hackett, E. Nevo, R. Keith, W.T.B. Thomas and B.P. Forster. 2002. Analysis of simple sequence repeats (SSRs) in wild barley from the Fertile Crescent: Associations with ecology, geography and floweringtime. *Plant Mol. Biol.*, 48: 511-527.

Ivandic, V., W.T.B. Thomas, E. Nevo, Z. Zhang and B.P. Forster. 2003. Associations of simple sequence repeats with quantitative trait variation including biotic and abiotic stress tolerance in *Hordeum spontaneum*. *Plant Breeding*, 122(4): 300-304.

Lai, Y., P.X Wang, G.Q. Fan, E.J. Si, J. Wang, K. Yang, Y.X. Meng, B.C. Li, X.L. Ma, X.W. Shang and H.J. Wang. 2013. Genetic diversity and association analysis using SSR markers in barley. *Scientia Agricultura Sinica*, 46(2): 233-242.

Liu, C. 2012. Studies on genetic diversity of seed-used pumpkin germplasm resources [M. D. Dissertation]. Harbin,China: Northeast Agricultural University.

Liu, C., Y. Ge, D.J. Wang, X. Li, X.X. Yang, C.S. Cui and S.P. Qu. 2013. Morphological and molecular diversity in a germplasm collection of seed pumpkin. *Scientia Horticulturae*, 154: 8-16.

Liu, F.Z., Y.F. Cao and K. Wang. 2005. *Apple germplasm resource description specification and standard data*. China Agriculture Press, Beijing.

Mamunur, R.M., R.M. Azizur, A. Hossain and G. Rasul. 2011. Comparative study on morphological, physiological and molecular genetic diversity analysis in rice (*Oryza sativa* L.). *Libyan Agri. Res. Center J. Int.*, 2 (2): 85-93.

Mondal, M.M.A., A.H.M. Razzaque and A.B. Puteh. 2016. Debranching improves morpho-physiological characters, fruit quality and yield of tomato. *Pak. J. Bot.*, 48(6): 2249-2253.

Mordechai, S. and B. Neima. 2014. Application of stepwise regression for dynamic parameter estimation. *Comput. Chem. Eng.*, 69: 26-38.

Qian, G.Z. and G.G. Tang. 2005. A review on the plant taxonomic study on the genus *Malus Miller*. *J. Nanjing Forestry University*, 29(3): 94-98.

Salvi, S., and R. Tuberosa. 2005. To clone or not to clone plant QTLs: present and future challenges. *Trends Plant Sci.*, 10(6): 297-304.

Sun, Y.L., H.M. Kang, S.H. Han, Y.C. Park and S.K. Hong. 2015. Taxonomy and phylogeny of the genus Citrus based on the nuclear ribosomal DNA ITS region sequence. *Pak. J. Bot.* , 47(1): 95-101.

Tang, Q.Y. 2010. *DPS data processing system*. Science Press, Beijing.

Treuren, R., H. Kemp, G. Ernsting, B. Jongejans, H. Houtman and L. Visser. 2010. Microsatellite genotyping of apple (*Malus*×*domestic*a Borkh.) genetic resources in the Netherlands: application in collection management and variety identification. *Genet. Resour. Crop Evol*., 57(6): 853-865.

Yoseph, B., B. Anna-Maria and A.M. Alexander. 2005. A comparative study of molecular and morphological methods of describing genetic relationships in traditional Ethiopian highland maize. *Afr. J. Biotechnol*., 4 (7): 586-595.

Zhang, C.Y., X.S. Chen, Y.M. Zhang, Z.H. Yuan, Z.C. Liu, Y.L. Wang and Q. Lin. 2009. Method of constructing core collection for*Malus sieversii* in Xinjiang, China using molecular markers. *Agr. Sci. China.*, 8(3): 276-284.

Zhang, C., X. Chen, T. He, X. Liu, T. Feng and Z. Yuan. 2007. Genetic structure of *Malus sieversii* population from Xinjiang, China, revealed by SSR markers. *J. Genet. Genomics*, 34(10): 947-955.

Zhang, S.X. 1973. On the ecogeographical characters and the problems of classification of the wild fruit-tree forest in the ilivalley of sinkiang. *J. Integ. Plant Biol.*, 15(2): 239-253.

Zheng, J.Y., C.J. Zhang, A.G. Yang, Y. Feng, Q.F. Feng and M. Ren. 2014. Study on SSR loci associated with some chemical component of tobacco. *Chinese Agri. Sci. Bull.*, 30(13): 102-106.

(Received for publiation 2 August 2016)