Population Structure, Linkage Disequilibrium and Selective Loci in Natural Populations of *Prunus davidiana*

Zhongping Cheng[1,2], Ksenija Gasic[3] & Zhangli Wang[2]

[1] Key Laboratory of plant Germplasm Enhancement and Specialty Agriculture, Chinese Academy of Sciences, Wuhan, Hubei, P.R. China

[2] Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei, P.R. China

[3] Department of Horticulture, Clemson University, Clemson, SC, USA

Correspondence: Zhongping Cheng, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei 430074, P.R. China. Tel: 86-27-8751-0361. E-mail: chenzp2000@hotmail.com

Abstract

Prunus davidiana (Carrie're) Franch is a very important resource for the restoration in dry and arid areas, genetic improvement of peach, and extraction of health-promoting components. To effectively use the resource, we must have a measure of genetic diversity of *P. davidiana* and its population structure. LD (Linkage disequilibrium) provides information for association mapping underlying the phenotypic variation observed. Selective loci reveal adaptive evolution processes resulting from natural selection. A set of 190 genotypes from seven natural populations（SXTB, SIYQ, SXFX, NXXJ, SIJC, GSHT, GSHS) of *P. davidiana* collected from the range of *P. davidiana* in China was fingerprinted with 23 SSR markers, and analyzed with spatial structure, pairwise Fst (differentiation coefficient), PCA (principal coordinate analysis), estimation of groups of populations with STRUCTURE software, selective loci obtained from lnRH tested by standardization distribution and Grubbs. Our results demonstrate that population structure of four groups existed among populations through complementary analyses of the above mentioned methods; significant LD numbers from 22 to 129 between loci within unstructured populations were detected; there were five selective loci in all populations and two common selective loci for local natural selection between populations. We should conserve four populations among seven populations; these selective loci may provide information for disclosing adaption evolution and candidate genes according to selective loci and alleles; LDs inform how to use them for association analysis.

Keywords: *a* species of the *subgenus Amygdalus*, microsatellite, population relation, LD, positive selection

1. Introduction

Prunus davidiana (Carr.) Franch., a wild species and perennial woody plant, is native to China (Yu, 1979; Wang, 1988) and is found in wild areas of China. Wild *P. davidiana* was recorded more than 3000 years ago (Wang & Zhuang, 2001), and Meyer (1915) found wild trees in Shaanxi and Gansu provinces in China which are considered as the center of origin of *P. davidiana*. *P. davidiana* has tolerance to cold temperatures, drought, high soil of pH and resistance to peach aphid (Wang & Zhuang, 2001; Bassi & Monet, 2008). *P. davidiana* is often used as rootstock (Yu, 1979) for fruit production of peach, nectarine and plum in Northwestern and Northern China. Furthermore, it can be crossed with peach and nectarine to improve resistance to plum pox, powdery mildew, leaf curl and etc. (Moing et al., 2003), to breed rootstocks adaptable to marginal soils and more resistant to replant problems (Edin & Garcin, 1994; Pisani & Roselli, 1983; Roselli et al., 1985) or to select for self-incompatibility genes (Foulongne et al., 2003). *P. davidiana* can also be used as landscape trees because of their early flowering and attractive bark, and strength of limbs wood (Bassi & Monet, 2008). *P. davidiana* is widely used as an alternative prior species for ecological restoration in dry and arid areas of northwestern China (Wang & Zhuang, 2001). Seeds of *P. davidiana* have 45% oil that is used in Chinese medicine (Wang & Zhuang, 2001). Stems of *P. davidiana* contain health-promoting components that increase the inhibitory activity on total reactive oxygen species (ROS) and hydroxyl radical (*OH) (Choi, 1991; Jung et al., 2003). Propagation of *P. davidiana* is typically from seeds of isolated natural populations in wild places, so natural populations of *P. davidiana* keep relative stability and integrity of gene pool for genetic analyses. From above mentioned aspects, it is necessary to study on

genetic characterization of natural populations of *P. davidiana*, which provides some information for conservation and utilities.

SSR (Simple sequence repeat or Microsatellite) technology is usually preferred among molecular biology methods due to SSR markers displaying co-dominant inheritance, hypervariability and having high cross-species transferability (Tauraz, 1989; Sosinski et al., 2000; Wünsch, 2009). More than 300 SSR markers have been isolated and characterized in the subgenus *Amygdalus* of *Prunus* (Sook et al., 2008). The markers provide a very reliable and convenient tool for the analyzing genetic diversity of *P. davidiana*. Genetic diversity studies have been performed in peach and other species of the *subgenus Amygdalus* (Aranzana et al., 2002, 2003,2010; Bouhadida et al., 2007; Cheng, 2007a,b; Cheng & Huang, 2009; Cipriani et al., 1999; Dirlewanger et al., 2002; Shiran, 2007; Sosinski et al., 2000; Testolin et al., 2000) however, there are no studies on the population structure, LD and selection of natural populations of *P. davidiana*.

Measuring the population structure of *P. davidiana* using neutral markers is an important first step in association genetic studies in order to avoid false associations between phenotypes and genotypes that may arise from nonselective demographic factors (Krutovsky et al., 2009), and it is more efficient for management and utilization of germplasm (Cho et al., 2008). Softwares or methods such as Cluster analyses using UPGMA (the unweighted pair group method with arithmetic mean method), NJ (Neighbor Joining method) and MP (Maximum parsimony method) (Nei and Kumar, 2000) to identify groups and subgroups according to similarity or distance; STRUCTURE software (Pritchard et al., 2000) for deciding K groups by genetic background analysis; PCA (principal coordinate analysis) implemented by GenALEx6.2 software (Peakall & Smouse, 2006) for characterizing population structure by means of principal coordinates. Some papers have been published about both wild populations (Belaj et al., 2007; Besnard et al., 2007) and cultivars (Inghelandt et al., 2010; Li et al., 2010) in plants and used for guidance of conservation and LD (linage disequilibrium) analyses.

Generally, genetic mapping comes from two basic methods, one is traditional QTL (Quantitative trait loci), and the other is advanced LD (Linage disequilibrium loci). QTL mapping requires segregating populations derived from biparental crosses and has resolution limited (Ecke et al., 2009). LD mapping has higher number of recombination events and a higher resolution in polymorphic populations (Ewens & Spielman, 2001; Jannink et al., 2001). Testing of LD can be calculated with Pairwise LD, Multi-locus LD, Haplotype-specific LD, Model-based LD and recombination (Mueller, 2004). LD has been utilized for genetic mapping of traist or disease loci in humans and model organisms (Mueller, 2004). LD has also successfully been used in plants, and significant LD between loci has been detected and the extent and decay of LD have been observed to vary between expressed species populations and subpopulations(Agrama & Eizenga, 2007; Berloo et al., 2008; Rossi et al., 2009; Comadran et al., 2010; Inghelandt et al., 2010; Brazauskas et al., 2011; Myles et al., 2011). These results provided preconditions for selecting populations or subpopulations for LD mapping, and for evaluating number of markers for use in LD mapping.

Commonly used measures for inferring the present selection are increased linkage disequilibrium between loci; reduced polymorphism; or a skewed allele frequency spectrum at individual loci (Schlotterer et al., 2002). Selection can be detected by methods like lnRV, lnRH, Fay and Wu's H test, the E test, and the joint DH test, the MFDM test (Schlotterer, 2002; Schotterer & Dieringer, 2005; Fay & Wu, 2000; Akey et al., 2002; Sabeti et al., 2002; Zeng et al., 2006; Innan & Kim, 2008; Li, 2010) as well as different models (Excoffier et al., 2009). Selective loci as signature of selection (Kane & Rieseberg, 2007; Chapman et al., 2008); identification of candidate loci (Schlotterer et al., 2002); environmental adaption (Kane & Rieseberg, 2007; Coyer et al., 2011); local Selective Sweeps (Schlotterer, 2002).

There are no published studies on population structure, LD and selection loci of *P. davidiana*. In this study, we investigated seven populations with 23 SSRs that cover the peach genome and appear not tightly linked markers (Aranzana et al., 2003). In this study, we aim to 1) analyze population structure and measure genetic variation among populations and use the information to guide the conservation and use of the germplasm; 2) determine if LD exists between loci among populations and use the information to choose the appropriate strategy for genetic association mapping; 3) search for loci showing evidence of selection in the whole population or between populations to detect genes associated with adaption evolution.

2. Materials and Methods

2.1 Plant Materials

We selected accessions of seven natural populations of *P. davidiana* from the center of origin Shannxi and Gansu provinces and surrounding areas Shanxi and Ningxia provinces in China. Young leaves from more than 30 accessions from each population were collected. The distance between any two accessions collected in a

population were separated by no less than 50 meters. The seven populations sampled in Taibai, Shaanxi; Yangquan, Shanxi; Fuxian, Shaanxi; Xiji, Ningxia; Jiaocheng, Shanxi; Huating, Gansu and Heshui, Gansu are abbreviated as SXTB, SIYQ, SXFX, NXXJ, SIJC, GSHT, GSHS, respectively. Samples were stored in sealed bags containing silica material for DNA extraction, and their geographical information was recorded (Table 1; Figure 1).

Figure 1. Seven populations (NXXJ, GSHS, GSHT, SITB, SXFX, SIYQ and SIJC) of *P. davidiana* collected from the areas

Table 1. Geographic information of collected populations

Name of abbreviation	Place of collection	Latitude	Longitude
SXTB	Taibai county, Shaanxi province	34.098	107.310
SIYQ	Yangquan county, Shanxi province	37.782	113.427
SXFX	Fu county, Shaanxi province	35.973	109.059
NXXJ	Xiji county, Nixia municipality	36.018	106.219
SIJC	Jiacheng county, Shanxi	37.655	111.735
GSHT	Huating county, Gansu province	35.234	106.696
GSHS	Heshui county, Gansu province	36.097	108.524

2.2 Molecular Analysis

Genomic DNA was extracted using a modified CTAB extraction protocol (Doyle & Doyle, 1987). In order to fit the samples into 96 well plates, the number of DNA samples from each population 26 to 28 to conform to a total of 192 samples in two plates. PCR amplifications were run on different two platforms. For fragment separation on 3% high resolution MetaPhor® (Cambrex Charles City Inc, IA) agarose – 1X TBE gels, the amplifications were performed in a total volume of 20 µl with final concentrations of 50 ng of DNA, 0.2 µM of both primers, 200 µM of each dNTP (New England Biolabs, Ipswich, MA) and 0.5 U of New England Biolabs' *Taq* DNA polymerase in 10 mM Tris-HCl pH 8.3, 1.5 mM $MgCl_2$ and 50 mM KCl. DNAs of all accessions was analyzed using 23 published SSR markers which were previously mapped in peach (Table 2). To determine the allelic size amplified by the markers, we selected DNAs of a few accessions from each population amplified by 23 SSRs for fragments analysis ABI 3130 (Applied Biosystems, Foster City, CA). PCR conditions were the same as above with the exception of 0.02 µM of M-13–tagged forward primer, 0.2 µM of reverse primer and 0.2 µM of M-13-tagged dye (6'-FAM, VIC, NED or PET) (ABI). Thermo Scientific MBS Satellite Thermal Cyclers (Thermo Fisher Scientific, Waltham, MA) were used. The PCR program was performed for 3 min of initial denaturation at 94°C, followed by 35 cycles of 1 min at 94°C, 1 min at the annealing temperature (T_a) and 1 min at 72°C, then a final extension step of 5 min at 72°C for each set of primer combinations.

Table 2. Twenty-three SSR markers used for amplification of individuals of seven populations in *P. davidiana*

Marker	Linkage group	Forward sequence	Reverse sequence	Reference
BPPCT 006	G8	GCTTGTGGCATGGAAGC	CCCTGTTTCTCATAGAACTCACAT	Dirlewanger et al., 2003
BPPCT 007	G3	TCATTGCTCGTCATCAGC	CAGATTTCTGAAGTTAGCGGTA	
BPPCT 008	G6	ATGGTGTGTATGGACATGATGA	CCTCAACCTAAGACACCTTCACT	
BPPCT 015	G4	ATGGAAGGGAAGAGAAATCG	GTCATCTCAGTCAACTTTTCCG	
BPPCT 017	G5	TTAAGAGTTTGTGATGGGAACC	AAGCATAATTTAGCATAACCAAGC	
BPPCT 020	G1	CGTGGATGGTCAAGATGC	ATTGACGTGGACTTACAGGTG	
BPPCT 025	G6	TCCTGCGTAGAAGAAGGTAGC	CGACATAAAGTCCAAATGGC	
BPPCT 028	G1	TCAAGTTAGCTGAGGATCGC	GAGCTTGCCTATGAGAAGACC	
CPPCT 002	G3	GGAGCTGCAATATTGCTG	GTTAGGGAAGCATCTCAC	Aranzana et al., 2002
CPPCT 006	G8	AATTAACTCCAACAGCTCCA	ATGGTTGCTTAATTCAATGG	
CPPCT 016	G1	AATTCCCTATGGAAATTAGA	CGCATATTATAGGTAGGAAA	
CPPCT 017	G7	TGACATGCATGCACTAAACAA	TGCAAATGCAATTTCATAAAGG	
CPPCT 022	G7	CAATTAGCTAGAGAGAATTATTG	GACAAGAAGCAAGTAGTTTG	
CPPCT 033	G7	TCAGCAAACTAGAAACAAACC	TTGCAATCTGGTTGATGTT	
UDP96-001	G6	AGTTTGATTTTCTGATGCATCC	TGCCATAAGGACCGGTATGT	Cipriani et al., 1999
UDP96-003	G4	TTGCTCAAAAGTGTCGTTGC	ACACGTAGTGCAACACTGGC	
UDP96-005	G1	GTAACGCTCGCTACCACAAA	CCTGCATATCACCACCCAG	Testolin et al., 2000
UDP96-013	G2	ATTCTTCACTACACGTGCACG	CCCCAGACATACTGTGGCTT	
UDP97-403	G3	CTGGCTTACAACTCGCAAGC	CGTCGACCAACTGAGACTCA	
UDP98-024	G4	CCTTGATGCATAATCAAACAGC	GGACACACTGGCATGTGAAG	
UDP98-025	G2	GGGAGGTTACTATGCCATGAAG	CGCAGACATGTAGTAGGACCTC	
UDP98-409	G8	GCTGATGGGTTTTATGGTTTTC	CGGACTCTTATCCTCTATCAACA	
UDP98-412	G6	AGGGAAAGTTTCTGCTGCAC	GCTGAAGACGACGATGATGA	

When performing PCR for multifluorophore fragment analysis, the conditions above mentioned were followed except for primer pairs with T_a significantly lower than 58ºC (T_a for M-13 forward primer). In such cases, 4 additional cycles were performed at the annealing temperature of the SSR marker followed by 35 cycles at the annealing temperature of the M-13, as described above. PCR amplicons, using 3% MetaPhor® - 1X TBE agarose gels along with New England Biolabs' low molecular weight DNA marker, were visualized with ethidium bromide under UV light, and after pooling the four amplicons together (4 different fluorophores). The samples were cleaned with ExoSAP-IT (USA Scientific or USB) according to manufacturer protocols and run on an ABI 3130 with GeneScan™ 600 LIZ® (Applied Biosystems) internal size standard. PCR products were analyzed by GeneScan with the ABI 3130 and read by Gene Mapper V.4.0 (Applied Biosystem) for multifluorophore fragments.

2.3 Data Analysis

DNAs of accessions from populations were amplified, and their bands with 23 SSR markers, which corresponded to exact sizes detected by Gene Mapper V.4.0, were recorded in Excel.

Genetic variation of SSR markers, including indexes of identifying genotype number, gene diversity, PIC (Polymorphic information content), f (Inbreeding coefficient), was performed using Powermarker software（Liu & Muse, 2005）; Genetic characteristics of including average number of alleles per locus, expected heterozygosity and Theta(H) under the infinite model of populations, were analyzed with Arlequin ver 3.5.1.2 software （Excoffier & Lischer, 2010）.

Population structure was performed with four complementary analyses on genotypic data. First, spatial structure was detected with GenALE x 6.2 software (Peakall & Smouse, 2006) based on genetic distances among populations; second, PCA was implemented with GenALEx6.2 software. Based on the distribution of all accessions along the first three axes, we could detect whether there was any grouping of individuals from populations; third, we used natural populations as a priori groups to test with Wright's Fst index (Weir & Cockerham, 1984) if there was differentiation between populations. The empirical distribution of no differentiation was obtained with Arlequin ver 3.5.1.2 using 10000 permutations; fourth, STRUCTURE 2.3 software, based on the Bayesian model of clustering method (Pritchard et al., 2000), was implemented. We used admixture model assumption to identify K groups of individuals. The assumed K groups varying from 2 to 10 were calculated with thirty replicate runs per K value, a burn in period length of 100000 and a post burn in simulation length of 200000. We decided final K groups through LnP (D) values according to the method recommended by Evanno et al. (2005). Individuals can be allocated into groups with different membership coefficients corresponding to the sum of all being equal to 1.

LD analysis of unstructured populations detected by STRUCTURE was performed using Arlequin ver 3.5.1.2 software under unknown phase between alleles from two heterozygous loci. When allele frequencies for LD were used, those below 5 percent were removed. The number of permutations for LD was 10000, without breaking genotypes to prevent any disequilibrium within loci (Hardy-Weinberg) to affect the significance of disequilibrium between loci. LD between a pair of loci was tested for genotypic data using a likelihood-ratio test, whose empirical distribution was obtained by a permutation procedure (Slatkin & Excoffier, 1996).

Selective loci were detected with Arlequin ver 3.5.1.2 from the alleles at 23 loci of all accessions according to demand of format in finite model with settings of 30000 permutations. A plot of Fst values against Het/(1-Fst) with permutations and observations was generated. If a locus located out of the plot area, the outlier locus is the most possible selected locus. Schotterer and Dieringer (2005) developed quantitative model-free statistics to identify loci that exhibited the largest reduction in microsatellite diversity ln RH which was more powerful than ln RV (Schlotterer, 2002). Ln RH can be obtained with the expected heterozygosity of compared populations, based on a stepwise mutation mode (Ohta & Kimura, 1973). Ln RH should be approximated by a Gaussian distribution under neutrality. Selective loci as outlier loci were checked between populations by using method of ln RH (Schotterer & Daniel, 2005). For monomorphic loci in a population, we used the method of Kauer et al. (2003) to adjust one additional allele different from the others for avoiding division by zero in the calculation of the ratio between populations. If a locus falls beyond the predetermined confidence bounds (i.e. 95% of a standard normal distribution), it indicates a significant reduction in genetic diversity (Harr et al., 2002). Grubbs' test (Motulsky, 2003), also known as the maximum normed residual test, is a statistical test used to detect outliers in a univariate data set assumed to come from a normally distributed population. We also used Grubbs' test to detect outlier values which should be considered as selective loci. Diagrams of allelic frequencies of selective loci between populations were produced in Excel.

3. Results

3.1 Genetic Diversity of SSR Markers and Populations

The 23 loci amplified by the SSR markers revealed a total of 148 alleles. The number of genotypes identified ranged from 3 with CPPCT002 to 37 with UDP98-412, and the mean of 14.2 for all the markers for 190 accessions. Gene diversity and PIC had similar orders except for a few slight differences, displaying that their values of CPPCT002 and UDP96-005 were the highest and lowest, respectively. Inbreeding coefficient screened by the markers disclosed extreme homozygosity of loci except BPPCT006 (Table 3). For analyses of populations, the mean numbers of alleles per locus ranged from 4.261 in SIJC to 2.826 in SXTB with the mean of 3.7 among seven populations. Expected heterozygosity and Theta (H) under the infinite model of the populations were the same orders as the mean numbers of alleles per locus except SIJC (Table 4).

Table 3. Genetic variation of 23 SSR markers in 190 individuals of seven populations

Marker	Genotype No.	Gene Diversity	PIC	f
BPPCT006	5.0	0.2961	0.2849	-0.1172
BPPCT007	16.0	0.7811	0.7463	0.5840
BPPCT008	22.0	0.7528	0.7219	0.4218
BPPCT015	18.0	0.7816	0.7508	0.7117
BPPCT017	9.0	0.6790	0.6146	0.8149
BPPCT020	5.0	0.4806	0.4054	0.3015
BPPCT025	9.0	0.6961	0.6343	0.5180
BPPCT028	8.0	0.1937	0.1846	0.6754
CPPCT002	3.0	0.0612	0.0593	0.8287
CPPCT006	10.0	0.7189	0.6696	0.9199
CPPCT016	8.0	0.4626	0.4188	0.5127
CPPCT017	3.0	0.3354	0.2792	0.8282
CPPCT022	10.0	0.6989	0.6436	0.6101
CPPCT033	8.0	0.3649	0.3211	0.6841
UDP96-001	13.0	0.3884	0.3680	0.4736
UDP96-003	12.0	0.6944	0.6415	0.7435
UDP96-005	36.0	0.8680	0.8539	0.4926
UDP96-013	10.0	0.2976	0.2825	0.2244
UDP98-024	32.0	0.8048	0.7812	0.7135
UDP98-025	14.0	0.6895	0.6544	0.8405
UDP97-403	21.0	0.7923	0.7638	0.8215
UDP98-409	18.0	0.7134	0.6718	0.6254
UDP98-412	37.0	0.8552	0.8388	0.3377
Mean	14.2	0.5829	0.5474	0.6059

Note: PIC short for polymorphic information content; f for inbreeding coefficient.

Table 4. Average number of alleles per locus, expected heterozygosity and Theta(H) under the infinite model of populations

Population	Number of alleles per locus		Expected heterozygosity		Theta(H)
	Mean	s.d.	Mean	s.d.	Mean
SXTB	2.826	1.497	0.31804	0.26825	0.46636
SIYQ	3.652	1.748	0.45334	0.27104	0.82929
SXFX	4	1.907	0.52194	0.23561	1.09177
NXXJ	3.957	2.325	0.49197	0.29691	0.96837
SIJC	4.261	2.137	0.49912	0.25348	0.99651
GSHT	4.174	1.969	0.53811	0.21719	1.16504
GSHS	3.174	1.557	0.44173	0.26043	0.79124
Mean	3.72	1.877	0.46632	0.25756	0.90123

Note: Mean as average value of the item, s.d as standard deviation.

Expected heterozygosity calculated through $\hat{H} = \frac{n}{n-1}(1 - \sum_{i=1}^{k} p_i^2)$

3.2 Population Structure

According to data of latitude and longitude and amplified bands of accessions of populations, geographic distances and genetic distances between populations were obtained with GenALEx6.2 software. When we analyzed correlation coefficients based on genetic distance and geographic distance between populations, all distance classes displayed no significance at 5% level, which meant there was no spatial structure of populations, that is, the end points were not located beyond the upper or lower red dots lines (Figure 2; Table 5).
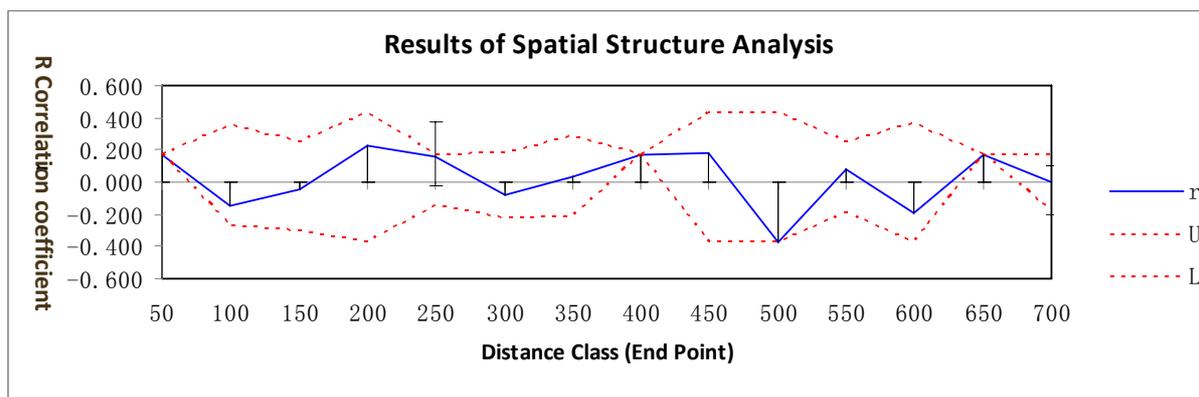
Figure 2. Spatial structure analyses of seven populations

Note: U (Ur error) and L (Lr error) error bars bound the 95% confidence interval about r as determined by bootstrap resampling.

Table 5. Correlation of genetic distance and geographic distance among populations

Distance Class (End Point)	50	100	150	200	250	300	350	400	450	500	550	600	650	700
r	0.167	-0.143	-0.040	0.223	0.153	-0.082	0.035	0.167	0.178	-0.377	0.078	-0.196	0.167	-0.004
U	0.167	0.347	0.251	0.435	0.172	0.179	0.282	0.167	0.435	0.435	0.250	0.361	0.167	0.171
L	0.167	-0.277	-0.311	-0.377	-0.145	-0.225	-0.217	0.167	-0.377	-0.377	-0.188	-0.377	0.167	-0.183
P(r-rand >= r-data)	1.000	0.860	0.640	0.260	0.070	0.690	0.470	1.000	0.290	1.000	0.290	0.870	1.000	0.500
P(r-rand <= r-data)	1.000	0.170	0.370	0.860	0.940	0.330	0.550	1.000	0.780	0.070	0.720	0.230	1.000	0.520
Correction	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167
r uc	0.000	-0.310	-0.207	0.056	-0.013	-0.248	-0.132	0.000	0.011	-0.544	-0.089	-0.363	0.000	-0.170
U uc	0.000	0.180	0.085	0.268	0.006	0.012	0.115	0.000	0.268	0.268	0.084	0.194	0.000	0.004
L uc	0.000	-0.444	-0.477	-0.544	-0.312	-0.392	-0.383	0.000	-0.544	-0.544	-0.354	-0.544	0.000	-0.350
Mean Bootstrap r	0.000	0.000	0.000	0.000	0.167	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.008
Ur error	-0.167	0.143	0.040	-0.223	0.220	0.082	-0.035	-0.167	-0.178	0.377	-0.078	0.196	-0.167	0.104
Lr error	0.167	-0.143	-0.040	0.223	0.181	-0.082	0.035	0.167	0.178	-0.377	0.078	-0.196	0.167	0.196
Ur	0.000	0.000	0.000	0.000	0.374	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.100
Lr	0.000	0.000	0.000	0.000	-0.028	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.199
Intercept	76.871													

Note: r, U and L values are adjusted by the correction factor. Uncorrected values are shown as r uc, U uc, L uc. Bootstrap mean, Ur, Lr are also adjusted by the correction factor.

Upper (Ur error) and lower (Lr error) error bars bound the 95% confidence interval about r as determined by bootstrap resampling.

Upper (U) and lower (L) confidence limits bound the 95% confidence interval about the null hypothesis of No spatial structure for the combined data set as determined by permutation.

Fst was calculated for all pairs of populations, and all pairwise differences between populations appeared significant at the 5% level (Table 6). It meant there was genetic differentiation between the populations. When we used all accessions from the populations for PCA, the first three axes explained 24.81%, 23.08% and 16.39% of total variation, respectively. Accessions from SXTB or SIYQ almost clustered together according to their originations, whereas accessions from the other five populations distributed in overlapping (Figure 3).
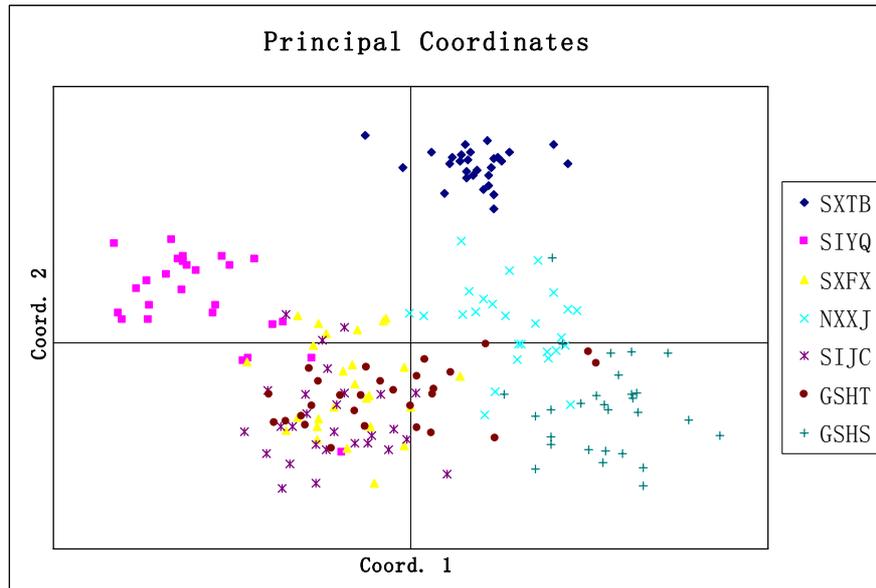
Figure 3. Principal coordinates analysis of 190 individuals from seven populations

We used STRUCTURE2.3 software in admixture model to analyze genetic structure of populations. When performed with assumed K = 2 to 10, there were no distinct groups to be decided because of the values of LnP(D) slightly increasing with values of K, so we used the method suggested by Evanno et al. (2005) to calculate K groups through values of LnP(D), and the highest peak of the curve line was found at K = 4 (groups) (Figure 4). Accessions' membership probabilities of seven populations allocated in four groups were more than 0.83 just except 0.77 from GSHT. Accessions from populations were distributed among assumed four groups similar to the groups with PCA analysis (Figure 3).
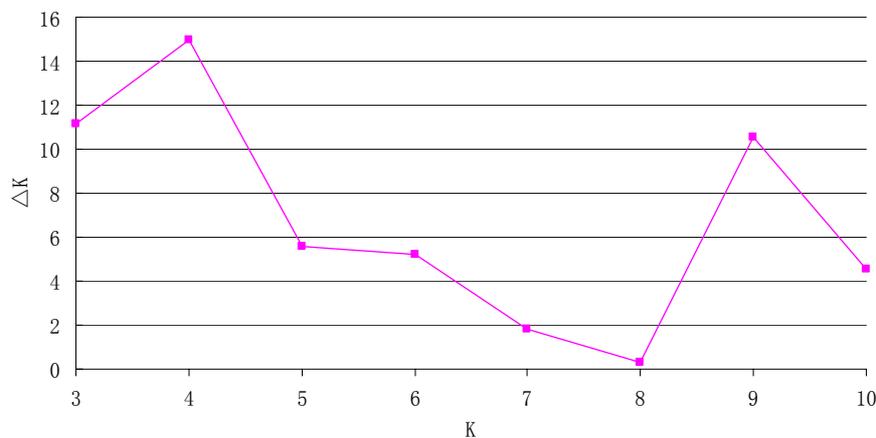
Figure 4. Detecting the number of cluster of 190 individuals from seven populations. ΔK calculated as ΔK = m|L''(K)|/s[L(K)]. The modal value of this distribution is the true K(*) or the uppermost level of structure, here four clusters (Evanno et al., 2005)

3.3 LD

Because population structure can cause spurious LD, we combined the four methods discussed above to identify unstructured populations (almost the same color in Figure 5) including SXTB; SIYQ;SXFX,SIJC and GSHT; NXXJ and GSHS for further studying, meanwhile, the whole population including seven populations was used as control to compare with unstructured populations. We performed the aforementioned populations to investigate significance of genotypic disequilibrium between all loci after removing low frequent alleles (only considering MAF>=0.05). There were some significant LDs between loci of 23 SSRs, displaying that 199 and172 in the whole seven populations; 22 and 8 in SXTB; 37 and 20 in SIYQ; 129 and 95 in SXFX, SIJC and GSHT; 72 and 50 in NXXJ and GSHS at 5% and 1% significant level were found (Figure 6, 7, 8, 9, 10).
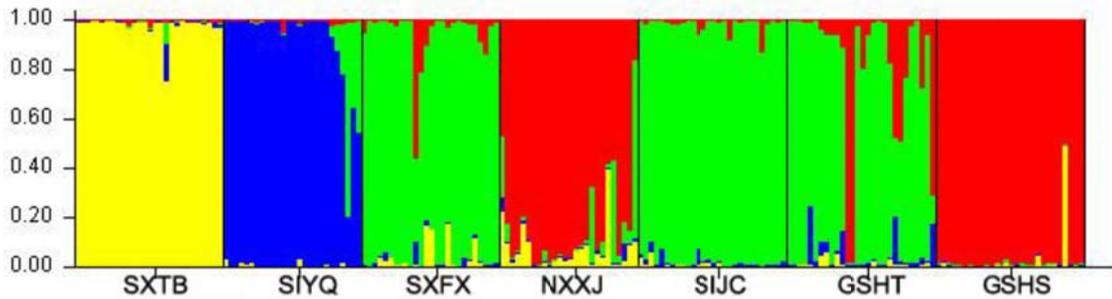
Figure 5. Population structure of 190 individuals from seven populations using 23 loci (k=4)
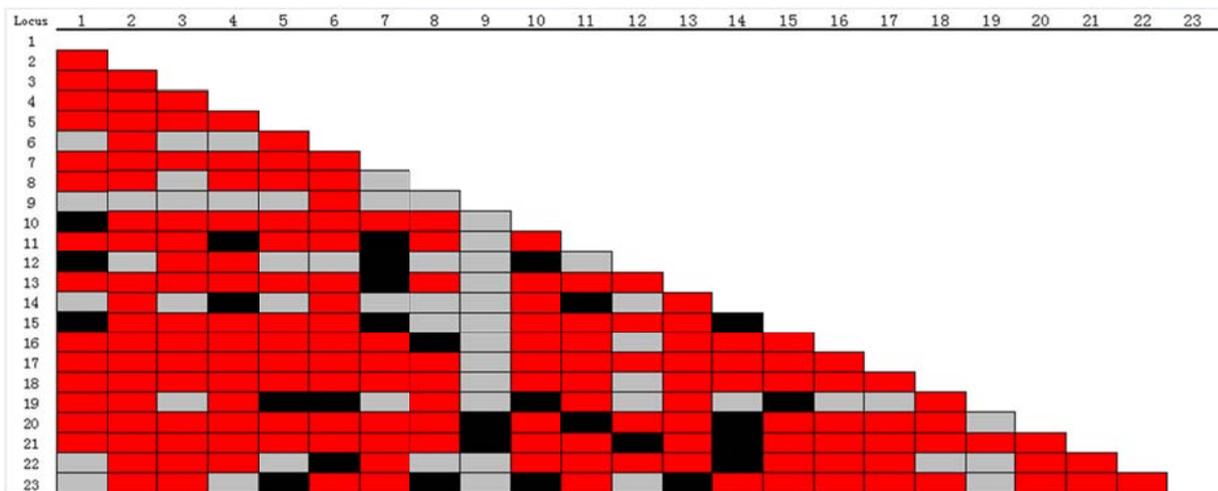
Figure 6. Test of linkage disequilibrium between all pairs of loci in seven populations

Note: Red and black filled cells mark indicates significant linkage disequilibrium of pairwise loci at the 1% or 5% significance level. The numbers 1, 2, ……, 23 represent the SSR loci amplified in Table 2 in order. Below is as the same.
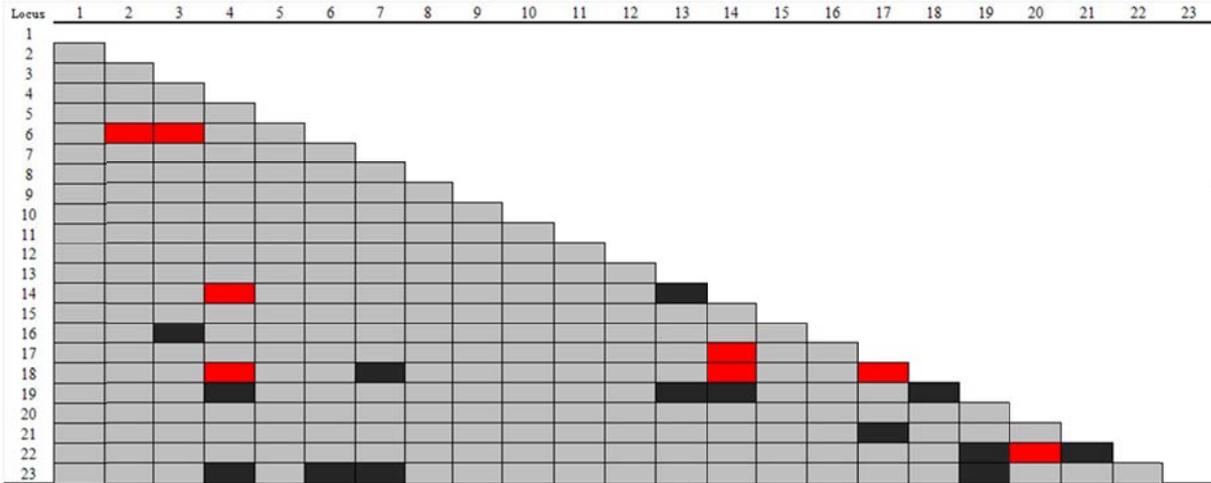
Figure 7. Test of linkage disequilibrium between all pairs of loci in the SXTB population
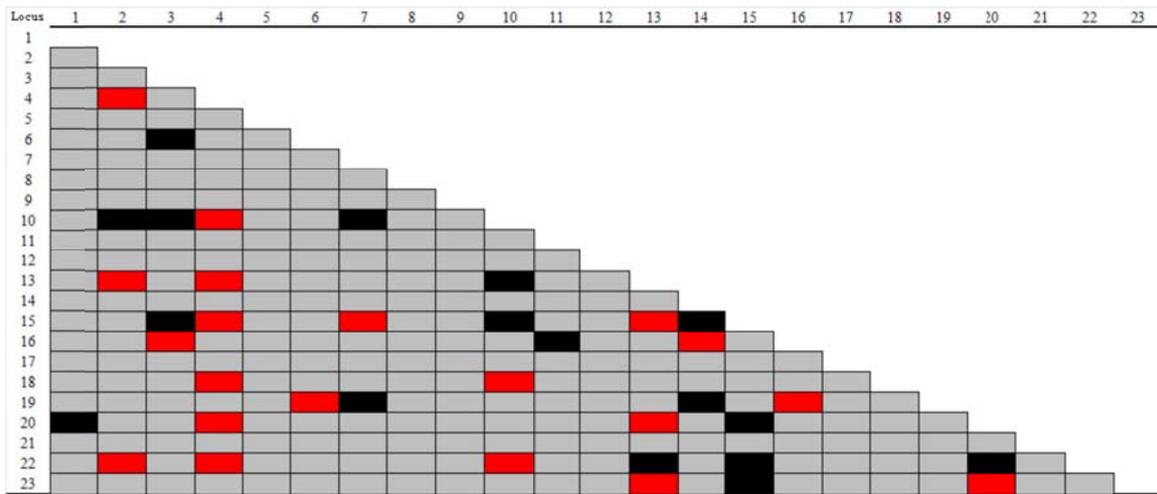
Figure 8. Test of linkage disequilibrium between all pairs of loci in the SIYQ population
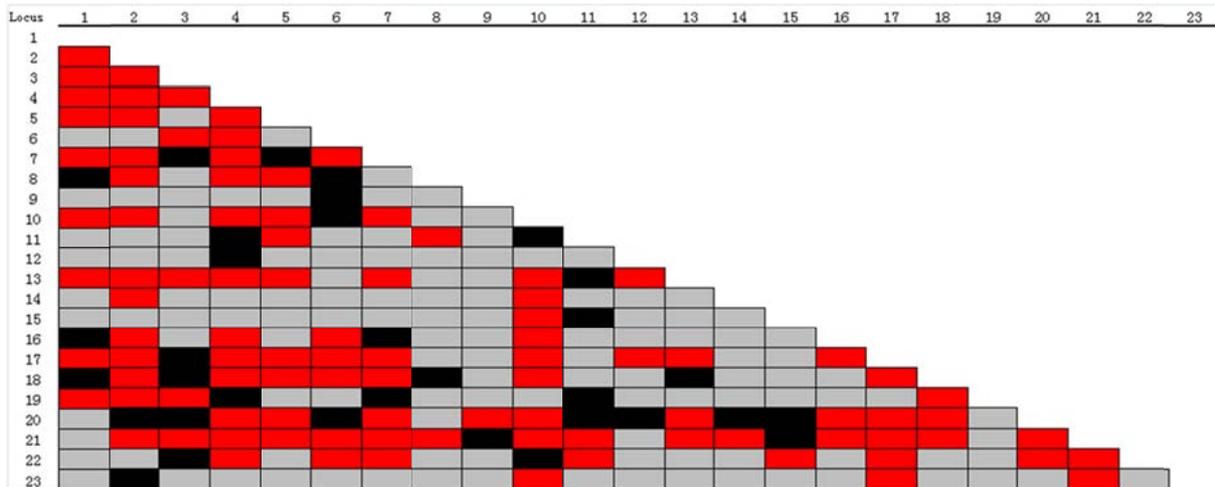
Figure 9. Test of linkage disequilibrium between all pairs of loci in the SXFX, SIJC and GSHS populations
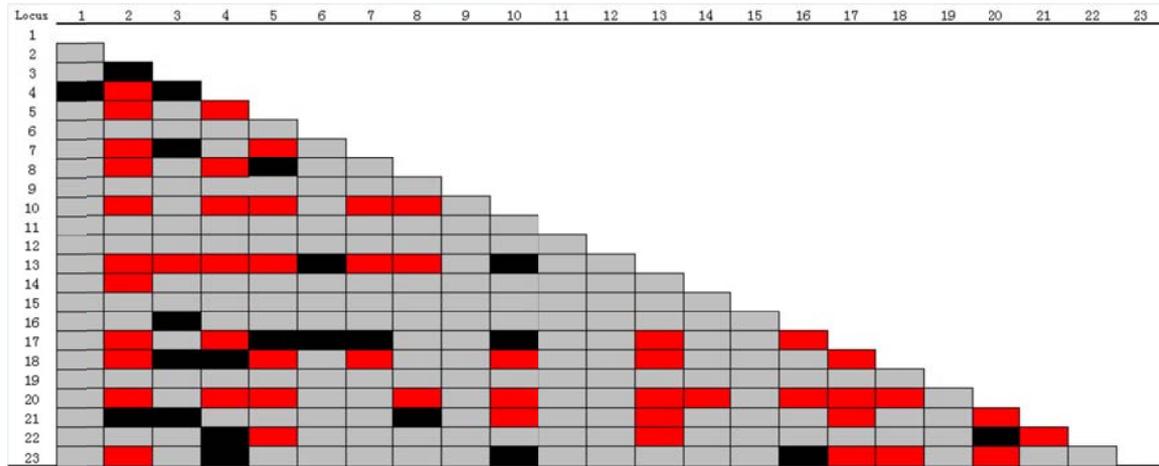
Figure 10. Test of linkage disequilibrium between all pairs of loci in the NXXJ, GSHS populations

3.4 Selective Loci

Wild populations under different environments experienced natural selection, so selective loci maybe appeared in whole populations or between populations. We used Arlequin ver 3.5.1 in finite model to detect selective loci (Table 7). There were five loci with significance at 5% level among populations, of which two with 1% level were apparently located outside of scattered dots plot of 30000 of coalescent simulations performed for all populations (Figure 11). When data of lnRH values were tested using beyond 95% confidence scale, 18 outlier loci or selective loci were found between populations (Electronic supplementary material S1), while data of lnRH values were tested by Grubbs, there were only three outlier loci between populations (Electronic supplementary material S2). Although the two testing methods detected different number of selective loci, there still were two similar loci, BPPCT 025 (between SXFX and SIJC) and CPPCT022 (between SIYQ and NXXJ). Allele frequencies in loci BPPCT 025 or CPPCT022 between populations apparently were different from each other, and variation of allele number became narrow (Figure 12a, b).
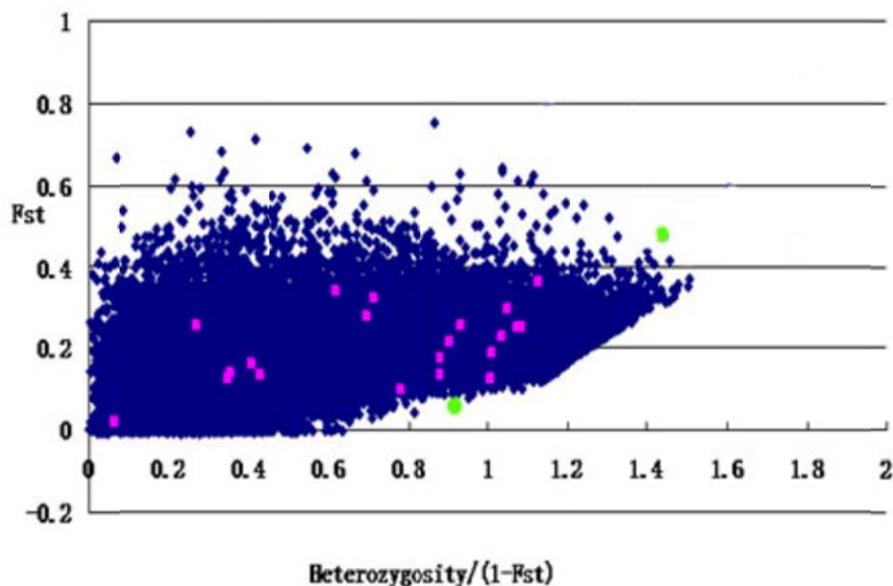

Figure 11. Joint distribution of Fst against heterozygosity. Pink diamonds correspond to observed microsallite loci, whereas blue dots are simulated loci under finite model with 30000 permutations. Significant loci (P< 0.01) are shown as large green dots
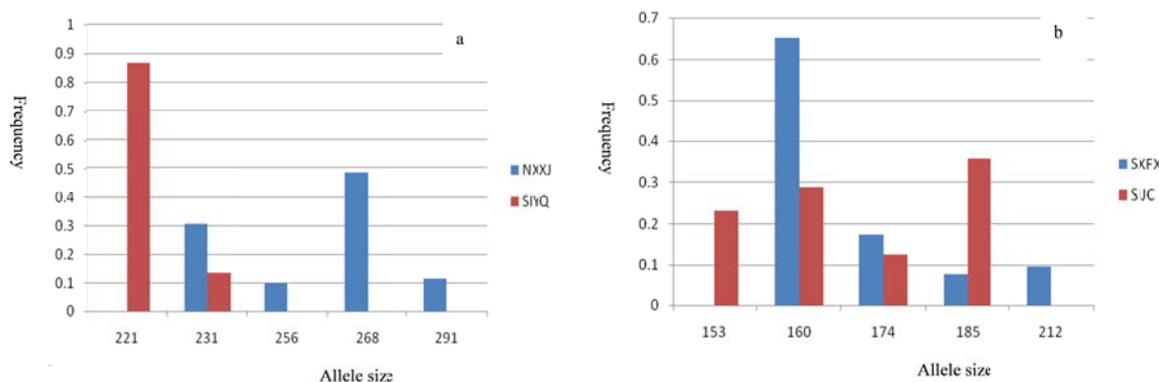
Figure 12. Frequency distribution of allele sizes with significance(P<0.01) under mutual comparisons of all populations. a) Frequency distribution of alleles amplified by BPPCT 025 between NXXJ and SIYQ populations. a) Frequency distribution of alleles amplified by CPPCT 022 between SXFX and SIJC populations

4. Discussion

Information of genetic diversity from both markers and populations in this study appeared very affluent. Inbreeding coefficients with positive values inferred to self-compatible type as main reproduction, justifying that isolated populations has very few chance to exchange genes from outside populations, and conforming to population differentiation among populations (Table 6).

Table 6. Fst values between populations calculated with distance method of pairwise difference

	SXTB	SIYQ	SXFX	NXXJ	SIJC	GSHT	GSHS
SXTB		*	*	*	*	*	*
SIYQ	0.323		*	*	*	*	*
SXFX	0.301	0.212		*	*	*	*
NXXJ	0.207	0.248	0.163		*	*	*
SIJC	0.317	0.239	0.179	0.190		*	*
GSHT	0.277	0.204	0.174	0.170	0.147		*
GSHS	0.320	0.334	0.236	0.141	0.243	0.173	

Note: Fst Values listed below diagonal and "*" with significance at 5% level shown above diagonal.

Seven populations were analyzed with population spatial structure, displaying that no correlation between genetic distance and geographic distance (Figure 2). There is not any spatial structure, at least two reasons are assumed as follows: first, the populations separates with long distance, and pollination is very difficult for individuals from different population; second, fruits of *P. davidiana* cannot be eaten because of no flesh, so fruits and seeds seldom have been brought to other places for growing as natural populations. The fact conformed to inbreeding coefficients with most positive values detected by the markers. Just as through analysis of spatial genetic structure of the Laperrine's olive which was very isolated endangered populations (Besnard et al., 2007), evident genetic particularities have to be urgently considered for their endemism. STRUCTURE and PCA analysis can identify population structure and be effectively applied for deciding groups. Population structure of wild olives and commercial maize (Belaj et al., 2007; Inghelandt et al., 2010) built four gene pools for gemplasm, respectively. By using successful methods, Population structure indicated there were four groups in this study (Figure 3, 4, 5). Fst between all studied populations were significant (Table 6). All information from above analyses for population structure of *P. davidiana* infers that we try to conserve individuals from four groups of these wild populations.

LDs in the whole population and unstructured populations expressed great differences. What reason is for highest value in the whole population is that there is population structure existed in it, which causes spurious LD. For example, Wang et al. (2008) found that 63.89% LDs of loci pairs at a 1% level were in the entire sample, but a range of 18.75–40.28% was in the subgroups. We selected unstructured populations for further analyzing of LD, Many LDs of loci pairs in unstructured populations were detected, which explained that natural populations might have experienced genetic bottleneck from their progenitor and natural selection for a long time, and self-compatible individuals generated genetic drift because some deadly genes became homogeneous. LDs screened the populations creates precondition for association mapping and marker assisted selection (MAS). In this study, mean value of 25.7% of loci pairs (SXTB; SIYQ; SXFX SIJC and GSHT; NXXJ and GSHS were 8.6%, 14.6%, 51.0% and 28.5% at 5% significant level, respectively) in *P. davidiana* (Figure 7, 8, 9, 10) was higher than that of 15.1% of the three subpopulations of cultivars in related *P. persica* (melting peaches, nectarines and non-melting peaches were 13.9%,13.4% and 18%, respectively) (Aranzana et al., 2010). The two species belonged to the same genus had more difference of LD, while the latter maybe came from more recombination due to cultivars bred from crossing. Other studies (Barnaud et al., 2009; Rossi et al. 2009) also justified that domestication bottlenecks and vegetative propagation are the primary factors responsible for this difference between cultivated and wild grapevine. Differences of LDs among unstructured populations may be explained that they had different number of accessions, membership of accessions and differential selection for adaptation to complicated environments or for special traits in these populations, but the information still can help us to select ideal populations for association mapping.

Table 7. Selective loci detected in finite model for all individuals from seven populations

Locus	Obs.Het. BP	Het./(1-Fst)	Obs. Fst	Fst P-value
BPPCT 006	0.304	0.354	0.140	0.304
BPPCT 007	0.805	1.074	0.250	0.298
BPPCT 008	0.765	0.883	0.134	0.066
BPPCT 015	0.801	1.035	0.226	0.436
BPPCT 017	0.697	0.934	0.254	0.318
BPPCT 020	0.501	0.695	0.279	0.248
BPPCT 025	0.702	0.780	0.100	0.018*
BPPCT 028	0.199	0.268	0.256	0.243
CPPCT 002	0.061	0.062	0.020	0.065
CPPCT 006	0.741	1.050	0.295	0.142
CPPCT 016	0.483	0.713	0.322	0.137
CPPCT 017	0.340	0.406	0.163	0.368
CPPCT 022	0.748	1.438	0.480	0.000**
CPPCT 033	0.369	0.425	0.133	0.225
UDP96-001	0.408	0.618	0.341	0.123
UDP96-003	0.710	0.907	0.217	0.494
UDP96-005	0.880	1.006	0.125	0.018*
UDP96-013	0.303	0.346	0.125	0.246
UDP97-403	0.820	1.011	0.189	0.319
UDP98-024	0.719	1.127	0.362	0.032*
UDP98-025	0.813	1.085	0.251	0.289
UDP 98-409	0.727	0.882	0.176	0.265
UDP98-412	0.861	0.921	0.065	0.000**

Note : '*'and '**' represent significant level at 1% and 5%, respectively

To adapt to various environments, natural populations through selection have caused variation of alleles. Generally, as long as favorable genes were fixed for positive selection, usually as expression of outlier loci, variation of gene frequencies became low. We used Arlequin ver 3.5.1.2 in finite model to detect outlier loci for individuals from all populations. Five loci were significant for selection in all populations, inferring that genes experienced coinciding evolution with history of demography (Table 7). Two of these loci with 1% significant level, located beyond 30000 permutations plot (Figure 11), demonstrated two positive selective types, one with low Fst at the bottom of the plot was balancing selection; the other with high Fst at the upper of the plot and high heterozygosity was directional selection (Excoffier et al., 2009). On the other hand, we detected selective loci appeared between populations. Selective loci using lnRH tested by both standard distribution scale and Grubbs appeared different number (Electronic supplementary material S1, Electronic supplementary material S2). From statistics, Grubbs test, used for detecting outliers, is more strictly than test of standard deviation. Both methods detected common selective loci: BPPCT 025 loci between SXFX and SIJC, CPPCT022 loci between SIYQ and NXXJ. Allele frequencies displayed apparently differences between populations (Figure 12a,b) as local selective sweep found in human populations (Kayser et al., 2003; Schlotterer, 2002). Some genes of the two loci are very possible responsibility for disclosing adaption evolution and digging out candidate genes.

Acknowledgement

References

Agrama, H. A., & Eizenga, G. C. (2007). Molecular diversity and genome-wide linkage disequilibrium patterns in a worldwide collection of *Oryza sativa* and its wild relatives. *Euphytica, 160*, 339-355. http://dx.doi.org/10.1007/s10681-007-9535-y

Akey, J. M., Zhang, G., Zhang, K., Jin, L., & Shriver, M. D. (2002). Interrogating a high-density SNP map for signatures of natural selection. *Genome Res., 12*, 1805-1814. http://dx.doi.org/10.1101/gr.631202

Aranzana, M. J., Abbassi, E. K., Howad, W., & Arús, P. (2010). Genetic variation, population structure and linkage disequilibrium in peach commercial varieties. *BMC Genetics, 11*, 69

Aranzana, M. J., Carbo, J., & Arus, P. (2003). Microsatellite variability in peach [*Prunus persica* (L.) Batsch]: cultivar identification, marker mutation, pedigree inferences and population structure. *Theor Appl Genet, 106*, 1341-1352

Aranzana, M. J., Garcia-Mas, J., Carbo, J., & Arus, P. (2002). Development and variability analysis of microsatellite markers in peach. *Plant Breeding, 121*, 87-92. http://dx.doi.org/10.1046/j.1439-0523.2002.00656.x

Aranzana, M. J., Pineda, P., Cosson, E., Dirlewanger, J., Ascasibar, G., Cipriani, C. D., … Arús, P. (2003). A set of simple-sequence repeat (SSR) markers covering the *Prunus* genome. *Theor Appl Genet, 106*, 819-825. http://dx.doi.org/10.1186/1471-2156-11-69

Barnaud, A., Laucou, V., This, P., Lacombe, T., & Doligez, A. (2009). Linkage disequilibrium in wild French grapevine, *Vitis vinifera* L. subsp. *silvestris*. *Heredity, 104*, 431-437. http://dx.doi.org/10.1038/hdy.2009.143

Bassi, D., & Monet, R. (2008). Botany and taxonomy. In D. R. Layne, & D. Bassi (Eds.), *The Peach: Botany, Production and Uses (chapter 1)*. CABI Wallingford Oxfordshire UK. http://dx.doi.org/10.1079/9781845933869.0001

Belaj, A., Munoz-Diez, C., Baldoni, L., Porceddu, A., Barranco, D., & Satovic, Z. (2007). Genetic Diversity and Population Structure of Wild Olives from the North-western Mediterranean Assessed by SSR Markers. *Annals of Botany, 100*, 449-458. http://dx.doi.org/10.1093/aob/mcm132

Berloo, R., Zhu, A., Ursem, R., Verbakel, H., Gort, G., & Eeuwijk, F. A. (2008). Diversity and linkage disequilibrium analysis within a selected set of cultivated tomatoes. *Theoretical and Applied Genetics, 117*, 89-101. http://dx.doi.org/10.1007/s00122-008-0755-x

Besnard, G., Christin, P. A., Baali-Cherif, D., Bouguedoura, N., & Anthelme, F. (2007). Spatial genetic structure in the Laperrine's olive (*Olea europaea* subsp. *laperrinei*), a long-living tree from the central Saharan mountains. *Heredity, 99*, 649-657. http://dx.doi.org/10.1038/sj.hdy.6801051

Bouhadida, M., Casas, A.M., Moreno, M. A., & Gogorcena, Y. (2007). Molecular characterization of Miraflores peach variety and relatives using SSRs. *Sci Hortic, 111*, 140-145. http://dx.doi.org/10.1016/j.scienta.2006.10.018

Brazauskas, G., Lenk, I., Pedersen, M. G., Studer, B., & Lübberstedt, T. (2011). Genetic variation, population structure, and linkage disequilibrium in European elite germplasm of perennial ryegrass. *Plant Science.* http://dx.doi.org/10.1016/j.plantsci.2011.06.013

Chapman, M. A., Pashley, C. H., Wenzler, J., Hvala, J., Tang, S., Knapp, S. J., & Burke, J. M. (2008). A Genomic Scan for Selection Reveals Candidates for Genes Involved in the Evolution of Cultivated Sunflower (*Helianthus annuus*). *The Plant Cell Online, 20*, 2931-2945. http://dx.doi.org/10.1105/tpc.108.059808

Cheng, Z. P. (2007a). *Molecular biological study on phylogeny of subgenus Amygdalus and genetic diversity of Prunus persica*. PhD Dissertation (in Chinese). http://dx.doi.org/10.1016/j.scienta.2006.10.020

Cheng, Z. P. (2007b). Genetic characterization of different demes in *Prunus persica* revealed by RAPD markers. *Sci. Hortic, 111*, 242-247. http://dx.doi.org/10.1016/j.scienta.2008.10.008

Cheng, Z., Huang, P., & H. W. (2009). SSR fingerprinting Chinese peach cultivars and landraces (*Prunus persica*) and analysis of their genetic relationships. *Sci Hortic, 120*, 188-193.

Cho, G. T., Lee, J., Moon, J. K., Yoon, M. S., Baek, H. J., Kang, J. H., Kim, T. S., & Paek, N. C. (2008). Genetic Diversity and Population Structure of Korean Soybean Landrace [Glycine max (L.) Merr.]. *J Crop Sci Biotech, 11,* 83-90. http://dx.doi.org/10.1007/s001220051209

Choi, J. S. (1991). Antihypeiuipidemic effect of flavonoids from *Prunus davidiana*. *J Natl Prod, 54*, 218-224. http://dx.doi.org/10.1021/np50073a022

Cipriani, G., Lot, G., Huang, W. G., Peterlunger, E., & Testolin, R. (1999). AC/GT and AG/CT microsatellite repeats in peach [*Prunus persica* (L.) Batsch]: isolation characterization and cross-species amplification in *Prunus. Theor Appl Genet, 99*, 65.

Comadran, J., Ramsay, L., MacKenzie, K., Hayes, P., Close, T. J., Muehlbauer, G., Stein, N., & Waugh, R. (2010). Patterns of polymorphism and linkage disequilibrium in cultivated barley. *Theoretical and Applied Genetics, 122*, 523-531. http://dx.doi.org/10.1007/s00122-010-1466-7

Coyer, J. A., Hoarau, G., Pearson, G., Mota, C., Jüterbock, A., Alpermann, T., John, U., & Olsen, J. L. (2011). Genomic scans detect signatures of selection along a salinity gradient in populations of the intertidal seaweed Fucus serratus on a 12 km scale. *Marine Genomics, 4*, 41-49. http://dx.doi.org/10.1016/j.margen.2010.12.003

Dirlewanger, E., Cosson, P., Tavaud, M., Aranzana, M. J., Poizat, C., Zanetto, A., … Laigret, F. (2002). Development of microsatellite markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry. *Theor App Genet, 105*, 27-138.

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull, 19*, 11-15.

Ecke, W., Clemens, R., Honsdorf, N., & Becker, H. C. (2009). Extent and structure of linkage disequilibrium in canola quality winter rapeseed (*Brassica napus* L.). *Theoretical and Applied Genetics, 120*, 921-931. http://dx.doi.org/10.1007/s00122-009-1221-0

Ewens, W. J., & Spielman, R. S. (2001). Locating genes by linkage and association. *Theor Popul Biol, 60*, 135-139. http://dx.doi.org/10.1006/tpbi.2001.1547

Excoffier, L., & Hofer, T. (2009). Detecting loci under selection in a hierarchically structured population." *Heredity, 103*(4), 285-298. http://dx.doi.org/10.1038/hdy.2009.74

Excoffier, L., & Lischer H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources, 10*, 564-567. http://dx.doi.org/10.1111/j.1755-0998.2010.02847.x

Excoffier, L., Hofer, T., & Foll, M., (2009). Detecting loci under selection in a hierarchically structured population. *Heredity, 103*, 285-298. http://dx.doi.org/10.1038/hdy.2009.74

Fay, J. C., & Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. *Genetics, 155*, 1405-1413.

Foulongne, M., Pascal, T., Arús, P., & Kervella, J., (2003). The potential of *Prunus davidiana* for introgression into peach [*Prunus persica* (L.) Batsch] assessed by comparative mapping. *Theor Appl Genet, 107*, 227-238. http://dx.doi.org/10.1073/pnas.202336899

Harr, B., Kauer, M., & Schlotterer, C. (2002). Hitchhiking mapping—a population-based fine mapping strategy for adaptive mutations in *Drosophila melanogaster. Proc Natl Acad Sci USA, 99*, 12949-12954

Inghelandt, D., Melchinger, A. E., Lebreton, C., & Stich, B. (2010). Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theor Appl Genet, 120*, 1289-1299. http://dx.doi.org/10.1007/s00122-009-1256-2

Innan, H., & Kim, Y. (2008). Detecting local adaptation using the joint sampling of polymorphism data in the parental and derived populations. *Genetics, 179*, 1713-1720. http://dx.doi.org/10.1534/genetics.108.086835

Jannink, J. L., Bink, M. C., & Jansen, R. C. (2001). Using complex plant pedigrees to map valuable genes. *Trends Plant Sci, 6*, 337-342. http://dx.doi.org/10.1016/S1360-1385(01)02017-9

Jung, H. A., Jung, M. J., Kim, J. Y., Chung, H. Y., & Choi, J. S. (2003). Inhibitory activity of flavonoids from *Prunus davidiana* and other flavonoids on total ROS and hydroxyl radical generation. *Arch Pharm Res, 26*, 809-815. http://dx.doi.org/10.1007/BF02980025

Kane, N. C., & Rieseberg, L. H. (2007). Selective Sweeps Reveal Candidate Genes for Adaptation to Drought and Salt Tolerance in Common Sunflower, *Helianthus annuus. Genetics, 175*, 1823-1834. http://dx.doi.org/10.1534/genetics.106.067728

Kauer, M. O., Dieringer, D., & Schlotterer, C. (2003). A Microsatellite Variability Screen for Positive Selection Associated With the "Out of Africa" Habitat Expansion of *Drosophila melanogaster. Genetics, 165*, 1137-1148.

Kayser, M., Brauer, S., & Stoneking, M. (2003). A Genome Scan to Detect Candidate Regions Influenced by Local Natural Selection in Human Populations. *Mol Biol Evol, 20*, 893-900. http://dx.doi.org/10.1093/molbev/msg092

Krutovsky, K. V., Clair, J. B., Saich, R., Hlpldns, V. D., & Neale, D. B. (2009). Estimation of population structure in coastal Douglas-fir (*pseudotsuga menziesii* (Mirb.) Franco var. *menziesii*) using allozyme and microsatellite markers. *Tree Genetics & Genomes, 5*, 641-658. http://dx.doi.org/10.1007/s11295-009-0216-y

Li, H. (2010). A New Test for Detecting Recent Positive Selection that is Free from the Confounding Impacts of Demography. *Molecular Biology and Evolution, 28*, 365-375. http://dx.doi.org/10.1093/molbev/msq211

Li, J., Schulz, B., & Stich, B. (2010). Population structure and genetic diversity in elite sugar beet germplasm investigated with SSR markers. *Euphytica, 175*, 35-42. http://dx.doi.org/10.1007/s10681-010-0161-8

Liu, K., & Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics, 21*, 2128-2129. http://dx.doi.org/10.1093/bioinformatics/bti282

Meyer, F. N. (1915). *China, A fruitful field for plant exploration*. Yearbook of the Department of Agriculture.

Moing, A., Poessel, J. L., Svanella-Dumas, L., Loonis, M., & Kervella, J. (2003). Biochemical basis of low fruit quality of *Prunus davidiana,* a pest and disease donor for peach breeding. *J Am Soc Hortic Sci, 128*, 55-62.

Motulsky, H. J. (2003). *Prism 4 Statistics Guide −Statistical analyses for laboratory and clinical researchers*. GraphPad Software Inc., San Diego CA (www.graphpad.com).

Mueller, J. C. (2004). Linkage disequilibrium for different scales and applications. *Briefings in Bioinformatics, 5*, 355-364. http://dx.doi.org/10.1093/bib/5.4.355

Myles, S., Boyko, A. R., Owens, C. L., Brown, P. J., Grassi, F., Aradhya, M. K., … Buckler, E. S. (2011). From the Cover: Genetic structure and domestication history of the grape. *Proceedings of the National Academy of Sciences, 108*, 3530-3535. http://dx.doi.org/10.1073/pnas.1009363108

Nei, M., & Kumar, S. (2000). Molecular evolution and phylogenetics. Oxford, UK: Oxford University Press.

Ohta, T., Kimura, M. (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a genetic population. *Genet Res, 22*, 201-204. http://dx.doi.org/10.1017/S0016672300012994

Peakall, R., & Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes, 6*, 288-295. http://dx.doi.org/10.1111/j.1471-8286.2005.01155.x

Pisani, P. L., & Rosselli, G. (1983). Interspecific hybridization of *Prunus persica* × *P. davidiana* to obtain new peach rootstocks. *Genetica Agraria, 112*, 197-198.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155*, 945-959.

Rossi, M., Bitocchi, E., Bellucci, E., Nanni, L., Rau, D., Attene, G., … Papa, R. (2009). Linkage disequilibrium and population structure in wild and domesticated populations of *Phaseolus vulgaris* L. *Evolutionary Applications, 2*, 504-522. http://dx.doi.org/10.1111/j.1752-4571.2009.00082.x

Sabeti, P. C., Reich, D. E., & Higgins, J. M. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature, 419*, 832-837. http://dx.doi.org/10.1038/nature01140

Schlotterer, C. (2002). A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics, 160*, 753-763. http://dx.doi.org/10.1007/0-387-27651-3_5

Schlotterer, C., & Dieringer, D. (2005). A noevel test statistic for the identification of local selective sweeps based on microsatellite gene diversity. In D. Nurminsky (Eds.), *Selective Sweep* (chapter 5). Kluwer Academic/Plenum Publishers.

Shiran, B., Amirbakhtiar, N., Kiani, S., Mohammadi, S., Sayed-Tabatabaei, B. E., & Moradi, H. (2007). Molecular characterization and genetic relationship among almond cultivars assessed by RAPD and SSR markers. *Sci Hortic, 111*, 280-292. http://dx.doi.org/10.1016/j.scienta.2006.10.024

Slatkin, M., & Excoffier, L. (1996). Testing for linkage disequilibrium in genotypic data using the EM algorithm. *Heredity, 76*, 377-383. http://dx.doi.org/10.1038/hdy.1996.55

Sook J., Margaret, S., Taein, L., Anna, B., Randall, S., Albert, A., & Dorrie, M. (2008). GDR (Genome Database for *Rosaceae*): integrated web-database for *Rosaceae* genomics and genetics data. *Nucleic Acids Research, 2008 January; 36(Database issue):* D1034-D1040.

Sosinski, B., Gannavarapu, M., Hager, L. D., Beck, L. E., King, G. J., Ryder, C. D., … Abbott, A. G. (2000). Characterization of microsatellite markers in peach [*Prunus persica* (L.) Batsch]. *Theor Appl Genet, 101*, 421-428. http://dx.doi.org/10.1007/s001220051499

Tauraz, D. (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res, 17*, 6463-6471. http://dx.doi.org/10.1093/nar/17.16.6463

Testolin, R., Marrazzo, T., & Cipriani, G. (2000). Microsatellite DNA in peach [*Prunus persica* (L.) Batsch] and its use in fingerprinting and testing the genetic origin of cultivars. *Genome, 43*, 512-520.

Wang, R., Yu, Y., Zhao, J., Shi, Y., Song, Y., Wang, T., & Li, Y. (2008). Population structure and linkage disequilibrium of a mini core set of maize inbred lines in China. *Theoretical and Applied Genetics, 117*, 1141-1153. http://dx.doi.org/10.1007/s00122-008-0852-x

Wang, Y. L. (1988). *Taxonomy of deciduous fruit trees*. China Agricultural Publishing House, Beijing(in Chinese)

Wang, Z. H., & Zhuang, E. J. (2001). *Fruit annals in China-Scroll of peach and nectarine*. China Forestry Publishing House, Beijing(in Chinese)

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution, 38*,1358-1370. http://dx.doi.org/10.2307/2408641

Wünsch, A. (2009). Cross-transferable polymorphic SSR loci in *Prunus* species. *Sci Hortic, 120*, 348-352. http://dx.doi.org/10.1016/j.scienta.2008.11.012

Yu, D. J. (1979). *Taxonomy of China fruit trees*. China Agricultural Publishing House, Beijing(in Chinese).

Zeng, K., Fu, Y. X., Shi, S., & Wu, C. I. (2006). Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics, 174*, 1430-1439. http://dx.doi.org/10.1534/genetics.106.061432