Genetic diversity and population structure of *Gerbera delavayi* Franch. revealed by newly developed genic-SSR markers

Wei Zheng [1] , **Xiaodan Xu** Corresp. [2]

[1] Faculty of Architecture and City Planning, Kunming University of Science and Technology, Kunming, Yunnan, China

[2] Faculty of Art and Communication, Kunming University of Science and Technology, Kunming, Yunnan, China

Corresponding Author: Xiaodan Xu
Email address: 2201322272@qq.com

Background: *Gerbera delavayi* Franch., a perennial herb native to southwest China, is of important and special value in textile and culture. Its abaxial leaves have a layer of soft fibers, which can be manually peeled, twisted into a yarn, and applied in textiles for hand-made clothes. The particular craftsmanship, unique all over the world, is a part of the intangible cultural heritage in Yunnan and Sichuan Province. To date, the wild species has become endangered because its leaves are excessively harvested for textile applications. Estimating the genetic variation of plant species is crucial for plant conservation and breeding. So, the aim of this study was to (1) provide an accurate outline of the genetic diversity and population structure of *G. delavayi* and (2) propose targeted conservation strategies for these wild resources.

Methods: Simple sequence repeats (SSR) markers were developed from the *G. delavayi* transcriptome for its genetic analysis. A total of 5,179 genic-SSR markers were identified and 350 primer pairs were selected for PCR validation. After primary screening , 19 polymorphic and neutral primer pairs were used to analyze the genetic diversity and population structure of nine wild fireweed populations (a total of 204 individuals) collected from Southwest China.

Results: Moderate genetic diversity (allelic richness (*Ar*)=3.436, Shannon's information index (*I*)=0.789 , observed heterozygosity (*Ho*)=0.304 , and expected heterozygosity (*He*)=0.415) and moderate differentiation (F_{ST}=0.110, *P*=0.000) were found among nine populations. The lowest genetic diversity (*Ar*=2.351, *I*=0.497, *Ho*=0.178, and *He*=0.284) and highest inbreeding (inbreeding coefficient (F_{IS})=0.365) were both detected within the isolated population PE. Furthermore, high genetic differentiation, limited gene flow, and clear genetic barrier were detected between PE and other populations, showing the significant effects of habitat fragmentation on the genetic differentiation of these *G. delavayi* populations.

Conclusions: Our results showed that human overexploitation has led to moderate genetic diversity in *G.delavayi* populations. Furthermore, habitat fragmentation has resulted in clear genetic barrier and high differentiation between the isolated and concentrated populations. Upon the genetic features of wild *G. delavayi* populations, we proposed several strategies for protecting and utilizing the current wild resources.

Genetic diversity and population structure of *Gerbera delavayi* Franch. revealed by newly developed genic-SSR markers

Wei Zheng [1], Xiaodan Xu [2, *]

[1] Faculty of Architecture and City Planning, Kunming University of Science and Technology, Kunming 650500, China

[2] Faculty of Art and Communication, Kunming University of Science and Technology, Kunming 650500, China

[*] Corresponding author: Xiaodan Xu

E-mail: 2201322272@qq.com

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

ABSTRACT

Background: *Gerbera delavayi* Franch., a perennial herb native to southwest China, is of important and special value in textile and culture. Its abaxial leaves have a layer of soft fibers, which can be manually peeled, twisted into a yarn, and applied in textiles for hand-made clothes. The particular craftsmanship, unique all over the world, is a part of the intangible cultural heritage in Yunnan and Sichuan Province. To date, the wild species has become endangered because its leaves are excessively harvested for textile applications. Estimating the genetic variation of plant species is crucial for plant conservation and breeding. So, the aim of this study was to (1) provide an accurate outline of the genetic diversity and population structure of *G. delavayi* and (2) propose targeted conservation strategies for these wild resources.

Methods: Simple sequence repeats (SSR) markers were developed from the *G. delavayi* transcriptome for its genetic analysis. A total of 5,179 genic-SSR markers were identified and 350 primer pairs were selected for PCR validation. After primary screening, 19 polymorphic and neutral primer pairs were used to analyze the genetic diversity and population structure of nine wild fireweed populations (a total of 204 individuals) collected from Southwest China.

Results: Moderate genetic diversity (allelic richness (*Ar*)=3.436, Shannon's information index(*I*)=0.789, observed heterozygosity (*Ho*)=0.304, and expected heterozygosity (*He*) =0.415) and moderate differentiation (F_{ST}=0.110, *P*=0.000) were found among nine populations. The lowest genetic diversity (*Ar*=2.351, *I*=0.497,

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

Ho=0.178, and *He*=0.284) and highest inbreeding (inbreeding coefficient (F_{IS})=0.365) were both detected within the isolated population PE. Furthermore, high genetic differentiation, limited gene flow, and clear genetic barrier were detected between PE and other populations, showing the significant effects of habitat fragmentation on the genetic differentiation of these fireweed populations.

Conclusions: The results showed that human overexploitation has led to moderate genetic diversity in *G. delavayi* populations. Furthermore, habitat fragmentation has resulted in clear genetic barrier and high differentiation between the isolated and concentrated populations. Upon the genetic features of wild *G. delavayi* populations, we proposed several strategies for protecting and utilizing the current wild resources.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

INTRODUCTION

Fireweed (*Gerbera delavayi* Franch.) is an endangered perennial herb of the Compositae family and mainly distributed in plains, slopes, and woodlands at an altitude of 1,800m to 3,200m in Southwest China and North Vietnam (Cheng, 1996). To the best of our knowledge, fireweed fiber is a unique natural material that is obtained from the abaxial side of a leaf blade and is used for textile applications (Fig. 1). This material exhibits good characteristics in textile (Xu et al. 2016). The leaf-pasted fibers can be manually peeled, twisted into a yarn, and applied to make traditional clothes without any chemical treatments. This craft began more than 500 years ago and has been included in the intangible cultural heritage in Yunnan and Sichuan Provinces, China (Li and Li 2010). With special cultural symbols, hand made fireweed clothes represent love and wisdom in some ethnic minority groups (e.g. Yi, Bai, and Lisu) in Southwest China. In addition, because the fleshy fibrous root of *G. delavayi* contains coumarin that has bactericidal effects (Liu et al. 2010), it is also collected to make Chinese traditional medicine and to treat dysentery, stomach pain, and indigestion (Xu 2014). What's more, fireweed has strong resistance and hard pedicel; therefore, Chinese breeders consider it as the parent of the distant hybridization of Barberton Daisy (*Gerbera jamesonii* Bolus).

For hundreds of years, few varieties of *G. delavayi* were domesticated and cultivated for textile and medical applications. Local farmers often collected raw materials from neighboring wild populations. Unfortunately, because of the increasing price of raw materials (eg., 60 dollars per kg of leaves in 2015) and the growing

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

71 demands for fireweed products, numerous wild *G. delavayi* resources have been cut
72 down to obtain leaves and roots, leading to severe loss of natural resources and an
73 urgent need to breed and conserve this endangered species.

74 Knowledge on genetic diversity and population structure can provide a basis for
75 plant breeding and conservation. However, few research have been conducted to
76 estimate the genetic variation of *G. delavayi*. Genetic molecular markers, such as
77 inter-simple sequence repeat (ISSR)markers, amplified fragment length
78 polymorphisms (AFLP), single nucleotide polymorphisms (SNP), and simple
79 sequence repeats (SSR), have been widely used to analyze the genetic diversity and
80 population structure of many plant species (Xia et al., 2007; Emanuelli et al., 2013;
81 Hou and Luo, 2011; Zhang et al., 2016). Compared with ISSR, RAPD, and AFLP
82 markers, SSR markers have been recognized as the most effective tools for the study
83 of genetic variation because of their high polymorphisms, co-dominant inheritance,
84 high reproducibility, and low costs (Arnaud-Haond et al., 2005). In addition, the
85 developments of sequencing and biological information technologies also provide
86 good opportunities to generate transcriptome data and developing genic-SSRs for *G.*
87 *delavayi*.

88 In this study, 350 new genic-SSR markers developed from the fireweed
89 transcriptome were selected for PCR validation. Finally, 19 high polymorphic and
90 neutral primer pairs were used to estimate the genetic diversity and population
91 structure of nine wild fireweed populations collected from Southwest China. Our
92 goals were to (1) provide an accurate outline of the genetic diversity and population

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

93 structure of *G. delavayi* and (2) propose targeted conservation strategies for these wild
94 resources.
95

96 MATERIALS AND METHODS

97 Plant material, RNA, and DNA extraction

98 Covering the current distribution of wild fireweed populations, 270 healthy
99 plants were collected from nine wild populations in Yunnan and Sichuan provinces
100 (Fig. 2) and subsequently planted in the warming houses of Kunming University of
101 Science and Technology. The tissue plantlets from "Shilin" (SL) were selected for
102 Illumina sequencing. The transcriptome database of "T01" in this population was used
103 to test the genic-SSR contents. Twenty-four samples were randomly selected from
104 nine populations for PCR validation. To analyze the genetic variation of *G. delavayi*, a
105 total of 204 individuals were selected from the nine wild populations (Table 1). The
106 total RNA was extracted using the method of Gong et al. (2012). The total genomic
107 DNA of all studied individuals was extracted from the leaf samples using the method
108 reported by Minerva et al. (2012).

109 Transcriptome sequencing and genic-SSR identification

110 The transcriptome of *G. delavayi* was sequenced on Illumina HiSeq 2500 and
111 assembled de novo with Trinity software (Grabherr et al., 2011). Unigenes that were
112 longer than 1,000 bp were screened for potential microsatellite loci by the MISA
113 (http://pgrc.ipk-gatersleben.de/misa/misa.html). The criteria for SSRs were set as

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

follows: (1) repeat motifs should include di-, tri-, tetra-, and penta-nucleotide repeats; (2) the amplified fragments size should range from 150 bp to 300 bp; (3) unigenes of good quality should be evenly distributed throughout the transcriptome.

Primer design

Sufficient flanking sequences containing microsatellites were selected to design primer pairs using Primer 3 (Rozen and Skaletsky, 1999). The annealing temperature was set to 60°C.The lengths of the designed primers were between 20 and 23 bp with an effective size of Ln=2(G+C) + (A+T). The maximum base repeat number of the 4. 5′ and 3′ regions could not contain over 2 continuous A/T, and 3′ regions could not contain more than 3 G/C. In this study, 3,464 primer pairs were designed, of which 350 pairs were manually selected for further genetic analysis.

PCR amplification and electrophoresis

PCR reaction was performed in 15μL of reaction mixture, consisting of 7.5 μL of Tapmix, 1μL of DNA, 4.5 μL of ddH_2O, and 1μL of each primer labeled with fluorescent FAM, HEX, and TAMRA (Applied Biosystems, Foster City, CA, USA). PCR products were sized on an ABI 3730 Prism Genetic Analyzer (Applied Biosystems). The PCR procedure of all DNA amplification was as follows: an initial denaturation step of 2 min at 98°C was followed by 30 cycles at 98 °C for 10s, 60 °C for 5 s, and 72 °C for 2 s, and a final extension at 72 °C for 10 min.

Data analysis

The raw molecular data of all loci were obtained by GeneMarker version 2.62

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

(SoftGenetics, State College, PA, USA), and all binned peaks were checked for correct assignment to corresponding bands and corrected manually. Dataset editing and formatting were performed in GenAlEx version 6.501 (Peakall and Smouse, 2012). The F_{ST}-outlier approach (Beaumont and Nichols 1996; Beaumont 2005) was used to test for signs of positive and balancing selection on the selected loci in LOSITAN software (Antao et al., 2008). The allelic variation of genic-SSRs and the levels of genetic diversity of the nine populations were tested in GenAlEx version 6.501 and PowerMarker version 3.25(Liu and Muse, 2005). Genetic diversity parameters were as follows: number of observed alleles (*Na*), number of effective alleles (*Ne*), number of private alleles (*Np*), polymorphism information content (*PIC*), Shannon's information index (*I*), observed heterozygosity (*Ho*), expected heterozygosity (*He*), and inbreeding coefficient (F_{IS}). The allelic richness (A_R) was estimated in FSTAT version 2.9.3 (Goudet, 2001). The Hardy–Weinberg equilibrium was tested using GENEPOP version 4.1.4 (Rousset, 2008).

Based on the Bayesian method, STRUCTURE version 2.2.3 (Pritchard et al., 2000) was used to cluster individuals and infer population structure. The combination of admixture and correlated-allele frequency model was used for the analysis. Twenty independent runs were performed for each set, with the *K* values ranging from 1 to 10, a burn-in of 1×10^6 iterations, and 1×10^6 subsequent Markov chain Monte Carlo (MCMC) steps. Based on the method of Evanno et al. (2005), the best *K* value was estimated in the website of the STRUCTURE HARVESTER (http://taylor0.biology.ucla.edu/structureHarvester/) (Earl2012). The CLUMPP

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

version 1.1.2 (Jakobsson and Rosenberg 2007) was used to account for label switching and detect the existence of distinct solutions across replicates. Finally, the graphical displays of the final results were generated in the DISTRUCT version 1.1 (Rosenberg, 2004).

Nei's genetic distance (1983) was calculated in PowerMarker version 3.25 and then used to construct the unrooted UPGMA tree of populations. The bootstrap values were calculated in Phylip version 3.6 (Felsenstein, 2005), and the final files were visualized and edited in MEGA version 5.10 (Tamura et al., 2013). Principal coordinate analysis (PCoA) was performed with NTSYS version 2.1(Rohlf, 2000). The genetic differentiation between pairwise populations was computed using F_{ST} and tested with ARLEQUIN version 3.1 (Excoffier and Lischer, 2010). The gene flow between pairwise populations was estimated using Wright's principle $Nm=(1-F_{ST})/4F_{ST}$ (Wright, 1931).

The correlation analysis between the genetic distance (Rousset 1997), $F_{ST}/(1-F_{ST})$, and geographic distance was carried out with the Mantel test (Mantel 1967) in GenAlEx version 6.501. Finally, based on the maximum difference algorithm of Monmonier, BARRIER version 2.2was used to identify the genetic barriers among the nine populations (Manni 2004).

RESULTS

SSR frequencies and distribution of repeat motif types

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

A total of 15,616 Unigenes were identified in the *G. delavayi* transcriptome, with an average length of 1975 bp (Table 2). A total of 4,133 Unigenes were detected with 5,179 SSRs, of which 839 Unigenes (5.37%) contained more than one SSR. Mononucleotides were the most abundant category, accounting for 54.37% of all SSRs (2816/5179), followed by dinucleotides (22.90%), and trinucleotides (21.70%). Compared with mononucleotides, dinucleotides, and trinucleotides, tetranucleotides, pentanucleotides, and hexanucleotides were significantly less abundant, accounting for 0.07%, 0.01%, and 0.01% of all SSRs, respectively. Without considering mononucleotides repeat, 2,363 SSRs remained. Dinucleotides, trinucleotides, tetranucleotides, pentanucleotides, and hexanucleotides accounted for 50.19% (1186/2363), 47.57% (1124/2363), 1.52% (36/2363), 0.42% (10/2363), and 0.30% (7/2363) of all SSRs, respectively.

In mononucleotide repeats, the most frequent motif (51.27%) was (A)n (not shown). Among the dinucleotide repeats, AG/CT was the most abundant motif, accounting for 60.66% of all dinucleotide SSRs (Fig. 3). Among the trinucleotide repeats, AAT/ATT was the richest motif and accounted for 25.50% (Fig.3).

SSR assessment and selection for genetic analysis

A total of 3,464 SSR primer pairs were designed using Primer 3, and 350 of these SSR primer pairs were randomly selected for PCR validation. Sixty-one primer pairs yielded clear and specific bands, of which 21 were polymorphic for each allele in 24 individuals. Subsequently, the polymorphism level of 21 markers was estimated in 204 individuals (Table 3). A total of 169 alleles were detected across all loci, ranging

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

from 4 (GdS 4) to 14 (GdS 50) alleles, with an average of 8.048 alleles per locus. The effective alleles were 48.305, between 1.068 (GdS 51) and 4.494 (GdS 43), with an average of 2.300 per locus. The mean *PIC* was 0.481, ranging from 0.063 (GdS 51) to 0.750 (GdS 43). The mean *He* and *Ho* were 0.329 and 0.487, respectively. All of these results exhibited the 21 SSR markers' potential in genetic analysis.

The F_{ST} outlier test was conducted to identify the SSR loci potentially under selection pressure (Fig. 4). Results showed that GdS2 and GdS18 were under positive and balancing selections, respectively. Therefore, they were excluded from further genetic analysis, and the remaining 19loci were finally selected to analyze the genetic diversity and population structure ofthenine wild fireweed populations.

Genetic diversity

Moderate genetic diversity (*Ar*=3.436, *I*=0.789, *Ho*=0.304, *He*=0.415) was detected in fireweed populations (Table 4). The lowest *I* was 0.497 and the highest was 1.004. The *Na* was between 2.579 and 4.895, *Ne* was between 1.579 and 2.502, *Ar* was between 2.351 and 4.323, *Ho* was between 0.178 and 0.389 and *He* ranged from 0.284 to 0.496. With a similar trend, all of these indices indicated that the lowest genetic diversity was detected in PE and the highest was observed in WD.

A total of 46 private alleles were identified in 8 out of 9 wild populations (Table 4). A total of 216 rare alleles were detected among the 9 populations, from 12 (PE) to 36 (SL), with an average of 24 per population. All populations significantly deviated from HWE. Furthermore, the F_{IS} was positive and ranged from 0.207 (WD) to 0.365

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

(PE) with an average of 0.262, indicating the moderate inbreeding within populations.

Population structure

The genetic distance (Nei, 1983) had a range of 0.063 to 0.184 (Table 5); the lowest was generated between SL and XP, and the highest was detected between KY and PE. Most notably, the genetic distance between PE and other populations was higher than that between the remaining pairwise populations. We constructed the unrooted UPGMA tree for populations based on Nei's genetic distance (Fig. 5 A). In the population tree, PE (cluster Ⅰ) from South Yunnan was separated from the rest of the populations with 100% bootstrap support, followed by EY and LJ (cluster Ⅱ) from Northwest Yunnan with 85% bootstrap support. Subsequently, KY from Southeast Yunnan was assigned to cluster III with a bootstrap value of 43%, and the 5 populations from central Yunnan and South Sichuan were grouped in cluster IV with a bootstrap value of 69%. These four clusters were highly relevant to the geographic distribution of the nine studied populations.

In this study, model-based STRUCTURE analysis failed to detect any structure among the nine populations (Fig. 5 B). Referring to the suggestion of Pritchard et al. (2010), with the population location as prior information, we also used LOCPRIOR to assist in the clustering, but still detected the absence of population structure. The maximum LnP (D) value was detected at K=5, indicating that 204 individuals could be clustered into five genetic clusters (Fig. S1). Interestingly, at Q>0.800, population PE could be assigned into a distinct cluster. Frequent gene introgression was detected between EY and LJ, and among KY, FY, SL, SDC, XP, and WD.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

Consistent with STRUCTURE, genetic distance, and UPGMA tree, PCoA also showed that PE was significantly separated from other populations with a small admix (Fig. S2).

Relatedness between genetic and geographic distance and Genetic Barrier

A data analysis method, which involves the regression of $F_{ST}/(1- F_{ST})$ estimate for pairs of subpopulations on geographic distance for populations along linear habitats or logarithm of distance for populations in two-dimensional habitats, was used in the Mantel test (Mantel, 1967). Results showed a significant and positive correlation between genetic and geographical distances (r=0.573, P=0.001) (Fig. S3), indicating the isolation-by-distance (IBD) among the nine populations. Based on the maximum difference algorithm of Monmonier, our genetic barrier prediction analysis identified three putative barriers among the nine populations (Fig. 6). The first barrier separated PE from the rest of the populations. The second one separated KY from the remaining seven populations. The third one separated EY and LJ from the rest of the populations.

Genetic differentiation and gene flow

In the analysis of molecular variance (AMOVA), moderate differentiation (F_{ST}=0.110, P=0.000) was detected among the nine populations, and 89.03% of variation occurred within populations (Table 6). The pairwise F_{ST} and gene flow among the nine populations showed higher differentiation and limited gene flow between PE and the other populations when compared with those between the

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

remaining pairwise populations (Table 5 and S1). According to the results of UPGMA and genetic barrier analysis, the nine populations were assigned to four clusters, cluster I (PE), cluster II (EY and LJ), cluster III (KY), and cluster IV (FY, SL, SDC, XP, and WD) for further analyses. Moderate genetic differentiation (F_{ST}=0.131, P=0.000) and low genetic variation (7.65%) were detected among the four clusters (Table 6). The F_{ST} values between cluster I and the other clusters were significantly higher than those between the remaining pairwise clusters (Table S2). The gene flow values between cluster I and the other clusters were much lower than those between the remaining pairwise clusters (Table S2).

DISCUSSION

Genic-SSR markers

A total of 5,179 genic-SSR markers were characterized in the *G. delavayi* transcriptome. The frequencies and distribution of the identified repeat motif types were analyzed in this study. Results showed that the frequency of a repeat decreased as its length increased. Three most abundant categories, mononucleotides, dinucleotides, and trinucleotides, comprised 98.97% of all identified SSRs. Excluding mononucleotide, dinucleotide was the most abundant category, followed by trinucleotide. This result was consistent with those of other studies (Wu et al., 2014; Dutta et al., 2011) but different from the report on *Gerbera hybrid* (Benemann et al. 2012), which indicated that the richest motif was trinucleotide. The most abundant

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

dinucleotide motif was AG/CT, which was consistent with many previous reports (Zhang et al., 2012; Moccia et al., 2009). However, the richest trinucleotide motif was AAT/ATT, which was different from that (AAG/CTT) in many dicotyledons (Morgante et al., 2002).

of the 350 primer pairs, 21 genic-SSR markers were selected to genotype 204 fireweed individuals. The mean *Ho*, *He,* and *PIC* of 21 loci were 0.329, 0.481, and 0.487, respectively, showing their high potential in the genetic analysis of fireweed populations. We conducted the F_{ST} outlier test across 21 loci because most of the genetic parameters were calculated under the hypothesis that all the loci are neutral (Shen et al., 2013). Results showed that GdS2 and GdS18 were under significant selection. Therefore, we excluded them from further genetic analyses.

Moderate genetic diversity

Moderate genetic diversity was detected in fireweed populations. The mean *Ar*, *I*, *Ho,* and *He* of fireweed populations were 3.436, 0.789, 0.304, and 0.415, respectively, which were significantly lower than those of some plant species from the same regions (*Ar*=6.778, *I*=0.789, *Ho*=0.502, *He*=0.597) (Zhao et al., 2014) but similar to those of many partially endangered species (*Ar*: 2.016–4.000, *I*: 0.789–0.885, *Ho*: 0.260–0.441, *He*: 0.334–0.457) (Xu et al., 2016; Feng et al., 2014; Colling et al., 2010). These results indicated the existing grim situation of these fireweed populations. Human overexploitation may be the most possible reason for this moderate genetic diversity. Our latest field investigation found that because of human overexploitation, the natural distribution and population size of *G. delavayi* has

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

drastically decreased in recent years, and the partial natural populations reported by previous research have disappeared. Collecting the leaves and roots of wild fireweed to make related products has already caused serious, even catastrophic damage to these wild resources, and has led to a moderate level of genetic diversity.

The lowest genetic diversity and highest inbreeding were both detected in PE, a relatively isolated population compared with the remaining eight populations. For PE, marginal and isolated distribution may reduce the chance to communicate with other populations and lead to low genetic diversity and high inbreeding. This observation can be supported by the limited gene flow between PE and other populations and the significantly higher genetic diversity in other concentrated populations, such as WD, LJ, EY, and FY.

Genetic barrier and population structure

STRUCTURE, UPGMA tree, PCoA, and pairwise F_{ST} all showed that PE was significantly separated and differentiated from other populations. According to Manni et al. (2004), the most important genetic barrier was detected between PE and other populations in the genetic barrier analysis. Population isolation is probably responsible for the substantial genetic barrier and high differentiation between PE and other populations because population isolation caused by habitat fragmentation may lead to increased genetic differentiation among populations (Young et al., 1993). PE is the only sampling population in South Yunnan. The Ailao Mountains and Honghe rift valley, where few fireweeds inhabit, isolate PE from other populations. The high mountains and wide rift valley hinder the gene flow through seeds and pollen, finally

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

329 leading to the clear genetic barrier and high differentiation between PE and the other
330 populations.

331 In addition to PE, our study detected moderate genetic differentiation among the
332 remaining eight populations. Results of STRUCTURE, UPGMA tree, and Mantel test
333 showed the absence of genetic structure and close genetic relationship among
334 populations from the same or adjacent regions. For plant species, mating system plays
335 an important role in shaping population structure and genetic differentiation (Petit and
336 Hampe, 2006). Fireweed is a strictly outcrossing species. In a certain geographic
337 range, seed and pollen can be effectively dispersed by insects and animals, which
338 improves the gene flow and weaken the genetic differentiation among adjacent
339 populations. Perennial characteristics can also maintain shared ancestral information
340 and limit the differentiation among populations (Petit and Hampe, 2006). Therefore,
341 these combined factors contribute to the absence of population structure and the
342 moderate differentiation among the remaining eight populations.

343 **Conservation implications**

344 This study detected moderate genetic diversity in the nine wild fireweed
345 populations. Human activities and habitat fragmentation have led to significant
346 genetic diversity reduction, high genetic differentiation, and substantial genetic barrier
347 in *G. delavayi*. The abundant rare alleles within populations also indicated the
348 potential loss of genetic diversity. Therefore, resource protection and public education
349 are necessary. As for *G. delavayi*, the breeding varieties used for medicine and textile
350 may be the most basic method to reduce the exhaustive exploitation of wild resources.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

Therefore, we also suggest *ex situ* conservation for individuals with specific excellent characteristics for future plant breeding. Moreover, given the low genetic diversity in the isolated PE population, we propose seed and pollen dispersals among populations to improve the genetic diversity within populations and strengthen the gene flow among the nine populations.

CONCLUSIONS

5179 Genic-SSRs were identified from 15616 Uni-scaffolds in *G. delavayi* transcriptome. Twenty one primer pairs were polymorphic and selected to yield PCR amplicons across 204 individuals. After removing two outlier loci, nineteen primer pairs were used to study the genetic diversity and population structure of nine wild fireweed populations. Our results showed that the human over-exploitation has led to the moderate genetic diversity in these populations, and habitat fragment has resulted in clear genetic barrier and high differentiation between the isolated and concentrated populations. Finally, we proposed several strategies for protecting and utilizing the current wild resources.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

Acknowledgment

We would like to extend our gratitude to Xingfu Zhang and Chengchuan Ou for the material collection, and Jing Zhang for the data analysis.

REFERENCES

Antao T, Lopes A, Lopes R, Beja-Pereira A, Luikart G. 2008. LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics* **9:**323 DOI 10.1186/1471-2105-9-323.

Arnaud-Haond S, Alberto F, Teixeira S, Procaccini G, Serrão EA, Duarte CM. 2005. Assessing genetic diversity in clonal organisms: low diversity or low resolution? Combining power and cost efficiency in selecting markers. *Journal of Heredity* **96(4):**434–440 DOI 10.1093/jhered/esi043.

Beaumont MA, Nichols RA. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of Royal Society of London* **263(1377):**1619–1626.

Beaumont MA. 2005. Adaptation and speciation: what can F-st tell us? *Trends in Ecology and Evolution* **20(8):**435–440 DOI 10.1016/j.tree.2005.05.017.

Benemann DDP, Machado LN, Arge LWP, Bianchi VJ, Oliveira ACD, Maia LCD, Peters JA. 2012. Identification, characterization and validation of SSR markers from the gerbera EST database. *Plant Omics* **5(2):**159–166.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

388 Colling G, Hemmer P, Bonniot A, Hermant S, Matthies D. 2010. Population genetic
389 structure of wild daffodils (*Narcissus pseudonarcissus* L.) at different spatial
390 scales. *Plant Systematics and Evolution* **287(3−4):**99−111 DOI
391 10.1007/s00606-010-0298-x.
392 Cheng YQ. 1996. China flora, Vol 79, Science Press, Beijing, China.
393 Dutta S, Kumawat G, Singh BP, Gupta DK, Singh S, Dogra V, Gaikwad K, Sharma
394 TR, Raje RS, Bandhopadhya TK, Datta S, Singh MN, Bashasab F, Kulwal P,
395 Wanjari KB, Varshney RK, Cook DR, Singh NK. 2011. Development of
396 genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus*
397 *cajan* (L.) Millspaugh]. *BMC Plant Biology* **11:** 17 DOI
398 10.1186/1471-2229-11-17.
399 Earl DA. 2012. STRUCTURE HARVESTER: a website and program for visualizing
400 STRUCTURE output and implementing the Evanno method. *Conservation*
401 *Genetics Resources* **4(2):**359−361 DOI 10.1007/s12686-011-9548-7.
402 Emanuelli F, Lorenzi S, Grzeskowiak L, Catalano V, Stefanini M, Troggio M, Myles
403 S, Martinez-Zapater JM, Zyprian E, Moreira FM, Grando MS. 2013. Genetic
404 diversity and population structure assessed by SSR and SNP markers in a large
405 germplasm collection of grape. *BMC Plant Biology* **13:**39 DOI
406 10.1186/1471-2229-13-39.
407 Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals
408 using the software STRUCTURE: a simulation study. *Molecular Ecology*
409 **14(8):**2611−2620 DOI 10.1111/j.1365-294X.2005.02553.x.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

410 Excoffier L, Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to
411 perform population genetics analyses under Linux and Windows. *Molecular*
412 *Ecology Resources* **10(3):**564–567 DOI 10.1111/j.1755-0998.2010.02847.x.
413 Felsenstein J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed
414 by the author. Seattle, WA: Department of Genome Sciences, University of
415 Washington.
416 Feng XY, Wang YH, Gong X. 2014. Genetic diversity, genetic structure and
417 demographic history of *Cycas simplicipinna* (Cycadaceae) assessed by DNA
418 sequences and SSR markers. *BMC Plant Biology* **14:**187 DOI
419 10.1186/1471-2229-14-187.
420 Gong L, Deng Z. 2012. Selection and application of SSR markers for variety
421 discrimination, genetic similarity and relation analysis in gerbera (*Gerbera*
422 *hybrida*). *Scientia Horticulturae* **138:**120–127 DOI
423 10.1016/j.scienta.2012.02.020.
424 Goudet J. 2001. FSTAT, a program to estimate and test gen diversities and fixation
425 indices (version 2.9.3).
426 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan
427 L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind
428 N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A.
429 2011. Full-length transcriptome assembly from RNA-Seq data without a
430 reference genome. *Nature Biotechnology* **29:**644–652 DOI 10.1038/nbt.1883.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

Hou Y, Luo A. 2011. Population genetic diversity and structure of a naturally isolated plant species, *Rhodiola dumulosa* (Crassulaceae). *PLoS One* **6(9):**e24497 DOI 10.1371/journal.pone.0024497.

Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23(14):**1801−1806.

Li XC, Li Y. 2010. Fireweed cloth spinned and weaved by the minority nationalities in southwest Chinese. *Socical Science of Yunnan* **20(2):**64−67.

Liu K, and Muse, SV. 2005. PowerMarker: An integrated analysis environment for genetic marker data. *Bioinformatics* **21(9):**2128−2129 DOI 10.1093/bioinformatics/bti282.

Liu SZ, Feng JQ, Wu J. 2010. A new monoterpene−coumarin and a new monoterpene-chromone from *Gerbera delavayi*. *Helvetica Chimica Acta* **93(10):**2026−2029.

Manni F, Guerard E, Heyer E. 2004. Geographic patterns of (genetic, morphologic, linguistic) variation: how battiers can be detected by using Monmonier's algorithm. *Human Biology* **76(2):**173−190 DOS 10.1353/hub.2004.0034.

Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research* **27(2)**:209−220.

Minerva G, Surinder K, Manisha T. 2012. *In vitro* propagation and assessment of plant genetic fidelity by RAPD and SSR markers in gerbera (*Gerbera jamesonii*). *Research Journal of Biotechnology* **7(4):**53−57.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

453 Moccia MD, Oger-Desfeux C, Marais GAB, Widmer A. 2009. A white campion
454 (*Silene latifolia*) floral expressed sequence tag (EST) library: annotation,
455 EST-SSR characterization, transferability, and utility for comparative mapping.
456 *BMC Genomics* **10:**243 DOI 10.1186/1471-2164-10-243.
457 Morgante M, Hanafey M, Powell W. 2002. Microsatellites are preferentially
458 associated with nonrepetitive DNA in plant genomes. *Nature Genetics*
459 **30(2):**194–200 DOI 10.1038/ng822.
460 Nei M, Tajima F, Tateno Y. 1983. Accuracy of estimated phylogenetic trees from
461 molecular data. II. Gene frequency data. *Journal of Molecular Evolution*
462 **19(2):**153–170 DOI 10.1007/BF02300753.
463 Peakall R, Smouse PE. 2012. GenAlEx 65: genetic analysis in Excel. Population
464 genetic software for teaching and research–an update. *Bioinformatics*
465 **28(19):**2537–2539.
466 Petit RJ and Hampe A. 2006. Some evolutionary consequences of being a tree. *Annual*
467 *Review Ecology Evolution and Systematics* **37:**187–214 DOI
468 10.1146/annurev.ecolsys.37.091305.110215.
469 Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using
470 multilocus genotype data. *Genetics* **155:**945–959.
471 Rohlf FJ. 2000. NTSYS 2.1: Numerical Taxonomic and Multivariate Analysis System.
472 Exeter Software. New York.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

473 Rosenberg NA. 2004. DISTRUCT: a program for the graphical display of population
474 structure. *Molecular Ecology Notes* **4(1):**137–138 DOI
475 10.1046/j.1471-8286.2003.00566.x.
476 Rousset F. 2008. genepop'007: a complete re-implementation of the genepop software
477 for Windows and Linux. *Molecular Ecology Resources* **8(1):**103–106 DOI
478 10.1111/j.1471-8286.2007.01931.x.
479 Rousset F. 1997. Genetic differentiation and estimation of gene flow from Fstatistics
480 under isolation by distance. *Genetics* **145(4):**1219–1228.
481 Rozen S, Skaletsky H. 1999. Primer3 on the WWW for general users and for biologist
482 programmers. *Methods in Molecular Biology* **132:**365–385.
483 Shen DF, Bo WH, Xu F, Wu RL. 2013. Genetic diversity and population structure of
484 the Tibetan poplar (*Populus szechuanica* var. *tibetica*) along an altitude gradient.
485 *BMC Genetics* **15(S1):**S11 DOI 10.1186/1471-2156-15-S1-S11.
486 Tamura K, Stecher G, Peterson D, Filipski A, Kumar. 2013. MEGA6: molecular
487 evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*
488 **30(12):**2725–2729 DOI: 10.1093/molbev/mst197.
489 Wright S. 1931. Evolution in Mendelian populations. *Genetics* **16(2):**97–159.
490 Wu J, Cai CF, Cheng FY, Cui HL, Zhou H. 2014. Characterisation and development
491 of EST-SSR markers in tree peony using transcriptome sequences. *Molecular*
492 *Breeding* **34(4):**1853–1866 DOI 10.1007/s11032-014-0144-x.
493 Xia T, Chen SL, Chen SY, Zhang DF, Zhang DJ, Gao QB, Ge XJ. 2007. ISSR analysis
494 of genetic diversity of the Qinghai-Tibet Plateau endemic *Rhodiola*

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

chrysanthemifolia (Crassulaceae). *Biochemical Systematics and Ecology* **35:**209–214 DOI 10.1016/j.bse.2006.09.016.

Xu BY. 2014. Identification of *Gerbera delavayi* by thin layer chromatography (TLC) method. *Yunnan Journal of Traditional Chinese Medicine* **35:**60–61.

Xu XD, Zheng W, Xu JG, Chen LQ. 2016. Properties of natural fibers from the abaxial side of fireweed (*Gerbera delavayi*) leaf blade for manual spinning. *Journal of Natural Fibers* DOI 10.1080/15440478.2016.1146644.

Xu XX, Cheng FY, Xian HL, Peng LP. 2016. Genetic diversity and population structure of endangered endemic *Paeonia jishanensis* in China and conservation implications. *Biochemical Systematics and Ecology* **66:**319–325 DOI: 10.1016/j.bse.2016.05.003.

Young AG, Merriam HG, Warwick SI. 1993. The effects of forest fragmentation on genetic variation in *Acer saccharum* Marsh. (sugar maple) populations. *Heredity* **71(3):**277–289.

Zhang HY, Wen LB, Miao HM, Zhang T, Wang CY. 2012. Development and validation of genic-SSR markers in sesame by RNA-seq. *BMC Genetics* **13:**316 DOI 10.1186/1471-2164-13-316.

Zhang L, Dietrich CH, Qin DZ. 2016. Microsatellite markers from tea green leafhopper *Empoasca* (*Matsumurasca*) *onukii*: a powerful tool for studying genetic structure in tea plantations. *BMC Genetics* **17:**112 DOI 10.1186/s12863-016-0420-3.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

516 Zhao DW, Yang JB, Yang SX, Kato K, Luo JP. 2014. Genetic diversity and
517 domestication origin of tea plant *Camellia taliensis* (Theaceae) as revealed by
518 microsatellite markers. *BMC Plant Biology* **14:**14 DOI
519 10.1186/1471-2229-14-14.

520

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

Table 1 Sample sizes and locations of nine wild *Gerbera delavayi* populations.

Population	Locality	Sample size	Latitude	Longitude	Altitude
WD	Wuding, Yunnan	22	25°28′12.39″N	102°19′32.92″E	2026m
FY	Fuyuan, Yunnan	24	25°50′30.84″N	103°17′18.83″E	2468m
LJ	Lijiang, Yunnan	23	26°53′59.81″N	100°10′47.33″E	2564m
EY	Eryuan, Yunnan	17	26°09′38.82″N	99°54′20.12″E	2312m
KY	Kaiyuan, Yunnan	23	23°48′19.36″N	103°37′52.97″E	1456m
PE	Puer, Yunnan	24	22°46′20.11″N	101°00′56.81″E	1466m
SL	Shilin, Yunnan	22	24°54′31.20″N	103°20′42.31″E	1862m
XP	Xinping, Yunnan	24	23°55′55.32″N	102°08′21.12″E	1528m
SDC	Dechang, Sichuan	25	27°19′20.92″N	102°17′27.08″E	2180m

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

Table 2 Detailed characteristics of Genic-SSRs in *Gerbera delavayi* transcriptome.

Searching Item	Number
Total number of sequences examined	15,616
Total size of examined sequences (bp)	30,835,470
Total number of identified SSRs	5,179
Number of SSR containing sequences	4,133
Number of sequences containing more than 1 SSR	839
Number of SSRs present in compound formation	271
Mononucleotide	2,816
Dinucleotide	1,186
Trinucleotide	1,124
Tetranucleotide	36
Pentanucleotide	10
Hexanucleotide	7

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

Table 3 Characteristics of 21 selected genic-SSR markers in 204 individuals.

Locus	Motif	Primer sequence(5'-3')	*Na*	*Ne*	*PIC*	*Ho*	*He*	Genbank numbers
GdS2	TCT	F: TGAACCACACCGTTACGAAG	6	3.340	0.469	0.271	0.701	KY050788
		R: CAGAACCAACTGGAAGAGGG						
GdS4	TCG	F: GTCGAGGAGGTTGACGAAAG	4	1.133	0.114	0.084	0.118	KY050789
		R: GCGGCACTCTCCTTCTATGA						
GdS5	GGC	F: GTTATAAGGCGGTGGAGGGT	9	1.686	0.380	0.257	0.407	KY050790
		R: GGTGGTTCTAGCAGCTCAGG						
GdS17	CTG	F: GGAGTTGCTACTGCAAAGGC	5	2.995	0.600	0.643	0.666	KY050791
		R: ATGGTAGAGCATCGGAATGG						
GdS18	CT	F: TGCAGCTTGTTAATGGAGGA	9	2.470	0.615	0.837	0.595	KY050792
		R: GTCGGTGAGAGGAGAGTTGC						
GdS20	CT	F: CGCTGCTACAGAACTCATCG	5	2.100	0.436	0.434	0.524	KY050793
		R: GGAATCGAAGCAGTCTCCAC						
GdS22	ATG	F: GTACTCCGGATCTCCCAACA	5	1.139	0.120	0.066	0.122	KY050794
		R: GGGACTCAGACTAACCGACG						
GdS27	GAC	F: TGGGGTTGTGTTGAAGATGA	6	2.376	0.493	0.040	0.579	KY050795
		R: CCAAACTTGATGCCCAATTC						
GdS29	TG	F: GATGGACCTTCTGCACCAGT	8	2.600	0.570	0.429	0.615	KY050796
		R: CGGCATCAGCAGAATCTACA						
GdS31	TGA	F: AGCATTCACAACCACAACCA	7	1.253	0.195	0.105	0.202	KY050797
		R: CTAGCGCTTGTCAGAACACG						
GdS32	CGC	F: GACAAGTGAAGCAGGGGAAG	10	3.674	0.682	0.295	0.728	KY050798
		R: AGTAATCCCAACCCCACACA						

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

GdS40	TG	F: CTGCTAAAGCTGTTGCTTGG R: AGGTTAAGTTCGGTTCCGGT	11	3.013	0.628	0.407	0.668	KY050799
GdS43	TAT	F: CACCCAAGTTTGACGATGTG R: TCGTTTCCCAACCAATTCTC	8	4.494	0.750	0.463	0.778	KY050800
GdS45	AAG	F: AGGTGACCATGGTTTCTTGG R: ACCTATAACCCATGGGAGGG	7	1.689	0.382	0.320	0.408	KY050808
GdS48	GA	F: AAGAGCCTTGTTAGGCGTTG R: CCAGAACCCACAGGTACCAC	13	3.273	0.677	0.465	0.694	KY050807
GdS50	GA	F: ACCGCAGTAGGTTGAAGGAA R: AGGTGAAGCAACAGGACCAC	14	2.476	0.559	0.443	0.596	KY050806
GdS51	GT	F: ATGCATCATGTCAAAGCAGC R: GGTTCAGATGGAGGTCTGGA	10	1.068	0.063	0.035	0.064	KY050805
GdS56	TCT	F: GCCATTAGGGTTTCCTCTCA R: ACTTTAGGGTTTGCTCCGGT	10	1.229	0.180	0.155	0.186	KY050804
GdS58	ACA	F: GGTGGTTGCTGTAATGGAGG R: TGATCGGGTGGTTAGAGAGG	10	2.076	0.431	0.399	0.518	KY050803
GdS60	CGG	F: TGCAATGACCGAAGCTATCA R: AATCATCCAACATTCCCACC	7	1.991	0.466	0.350	0.498	KY050802
GdS61	AG	F: TACCCAAAGGAACCCAAGTG R: GTGGAATGCCAGGTAAGGAG	5	2.228	0.475	0.404	0.551	KY050801
Mean			8.048	2.300	0.481	0.329	0.487	
Total			169	48.305				

Na: number of observed alleles; *Ne*: number of effective alleles; *PIC*: polymorphism information content; *Ho*: observed heterozygosity; *He*: expected heterozygosity.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

Table 4 Genetic diversity of nine *Gerbera delavayi* populations based on nineteen genic-SSRs loci.

Pop	*Na*	*Ne*	*Ar*	*Np*	*Nr*	*I*	*Ho*	*He*	*FIS*
PE	2.579	1.57	2.351	1	12	0.497	0.178	0.284	0.365
EY	3.947	2.10	3.702	9	22	0.810	0.265	0.426	0.345
LJ	4.263	2.16	3.787	13	27	0.882	0.356	0.457	0.216
KY	3.632	1.96	3.162	4	24	0.744	0.291	0.402	0.295
SDC	4.263	2.12	3.668	6	33	0.830	0.298	0.424	0.247
FY	3.895	2.17	3.442	6	20	0.854	0.362	0.458	0.261
SL	4.000	1.93	3.485	2	36	0.766	0.292	0.400	0.213
WD	4.895	2.50	4.323	5	29	1.004	0.389	0.496	0.207
XP	3.263	1.95	3.000	0	13	0.718	0.307	0.391	0.210
Mean	3.860	2.05	3.436	5.111	24	0.789	0.304	0.415	0.262
Total				46	216				

Na: number of observed alleles; *Ne*: number of effective alleles; *Ar*: allelic richness; *Np*: number of private alleles; *Nr*: number of rare alleles; *I*:Shannon's information index; *Ho*: observed heterozygosity; *He*: expected heterozygosity; F_{IS}: inbreeding coefficient. Populations that deviate from the Hardy-Weinberg equilibrium are in bold format.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

Table 5 Pairwise Nei's genetic distances (1983) (below diagonal) and pairwise F_{ST} values (above diagonal) between nine *Gerbera delavayi* populations.

	PE	EY	LJ	KY	FY	WD	SDC	SL	XP
PE	0.000	0.261	0.217	0.224	0.194	0.194	0.229	0.190	0.130
EY	0.181	0.000	0.031	0.082	0.086	0.090	0.069	0.107	0.169
LJ	0.162	0.133	0.000	0.087	0.068	0.081	0.093	0.113	0.156
KY	0.184	1.194	0.130	0.000	0.102	0.085	0.081	0.093	0.124
FY	0.161	0.119	0.094	0.130	0.000	0.021	0.067	0.048	0.083
WD	0.171	0.115	0.119	0.113	0.064	0.000	0.041	0.031	0.087
SDC	0.171	0.116	0.116	0.123	0.094	0.076	0.000	0.052	0.097
SL	0.141	0.112	0.119	0.091	0.084	0.077	0.073	0.000	0.068
XP	0.112	0.140	0.136	0.113	0.099	0.104	0.090	0.063	0.000

*P<0.05, **P<0.01.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

Table 6 Analysis of molecular variance (AMOVA) from genic-SSR data.

Source of variation	d.f.	Sun of	Variance	Percentage of	Fixtion
Among nine populations					
Among populations	8	203.557	0.47653Va	10.97	F_{ST}=0.110**
Within populations	399	1543.162	3.86757Vb	89.03	Nm=2.029
Total	407	1746.718	4.3441		
Among four genetic clusters					
Among clusters	3	129.024	0.34059 Va	7.65	F_{ST} =0.131**
Among populations	5	74.532	0.24403 Vb	5.48	F_{SC}=0.059**
Within populations	339	1543.162	3.86757 Vc	86.87	F_{CT}=0.077**
Total	407	1746.718	4.45219		Nm=1.658

F_{ST}: variance among coefficient of individual relative to the total variance; *Fsc*: variance among subpopulations within groups; F_{CT}: variance among groups relative to the total variance; Nm: gene flow. *P<0.05, **P<0.01.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

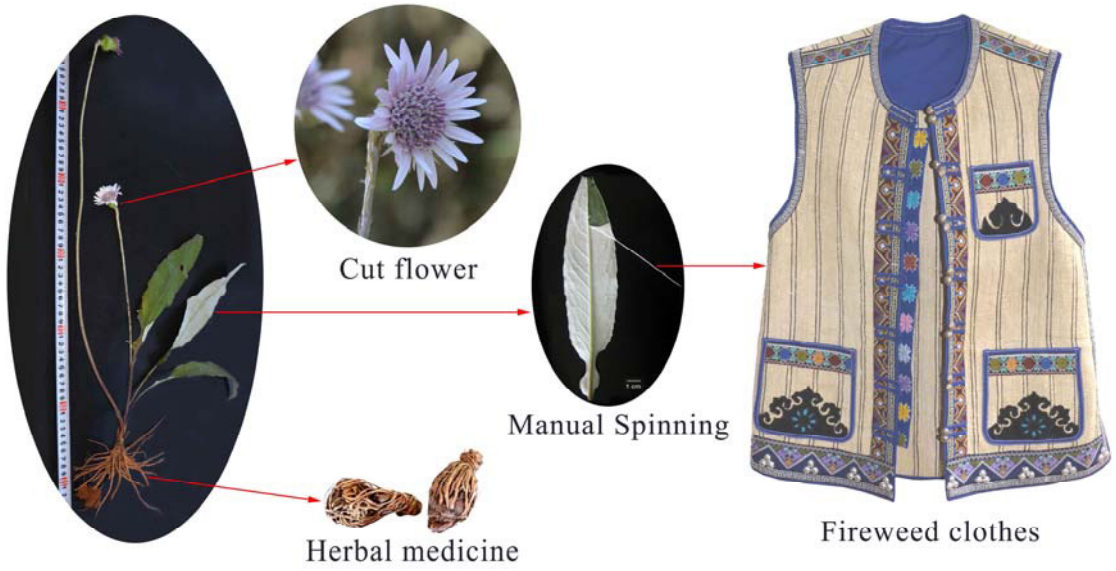

Fig. 1 *Gerbera delavayi.* Similar to many species of the Compositae family, fireweed flowers can be used as cut flowers. In addition, in Southwest China fireweeds (*Gerbera delavayi*) can be used in two applications: textile and herbal medicine. The leaf-pasted fibers can be harvested to make fireweed fiber-made clothes. The roots of this plant are collected and used as Chinese herbal medicine.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

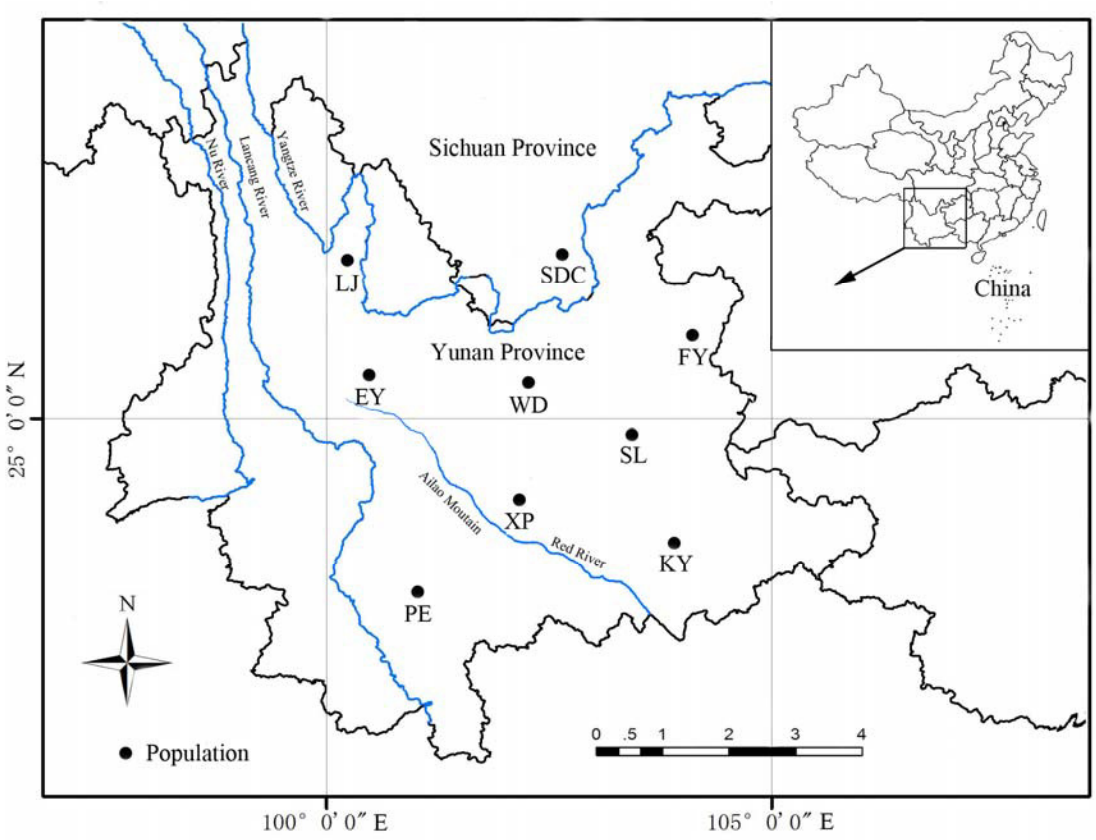

Fig. 2 Geographic distribution of the studied *Gerbera delavayi* populations. Nine wild fireweed populations were mainly distributed in five different areas. LJ and EY were located in northwest Yunnan. WD, FY, SL, and XP were distributed in central Yunnan. KY and PE lived in South and Southeast Yunnan, respectively. SDC was from South Sichuan.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

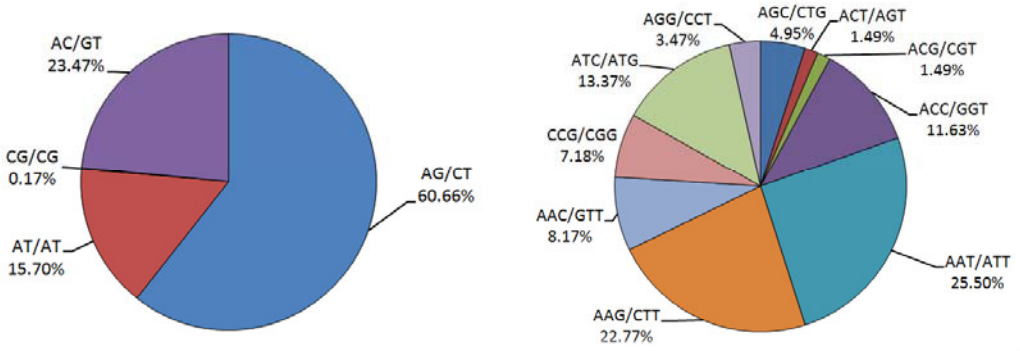

Fig. 3 SSR frequencies and distribution of different dinucleotides and trinucleotides in the *Gerbera delavayi* transcriptome.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

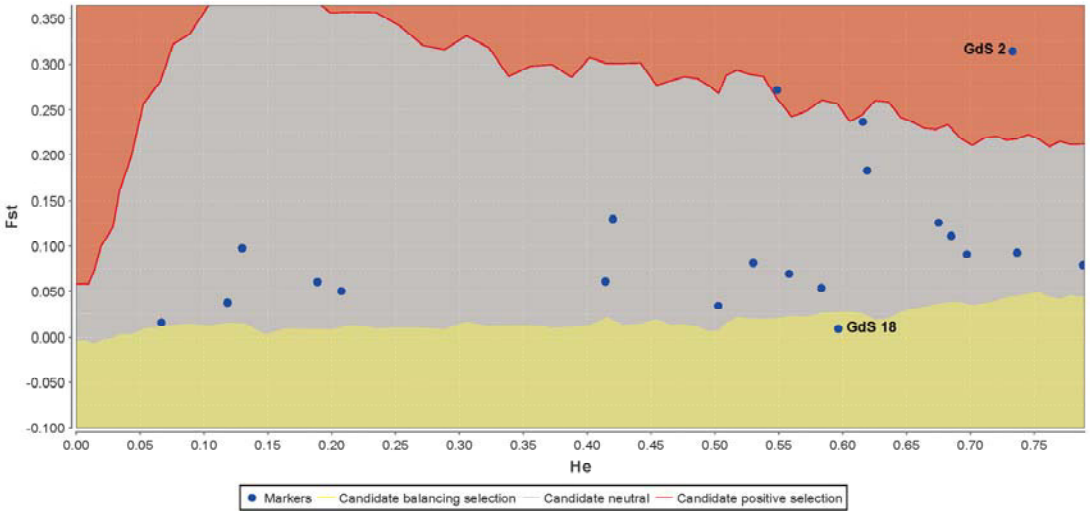

Fig. 4 F_{ST} outlier test for 21 genic-SSR loci in nine *Gerbera delavayi* populations. Gds 2 is located in the red area, indicating this locus is under positive selection. Gds 18 is distributed in the yellow area, indicating this locus is subject to balancing selection.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

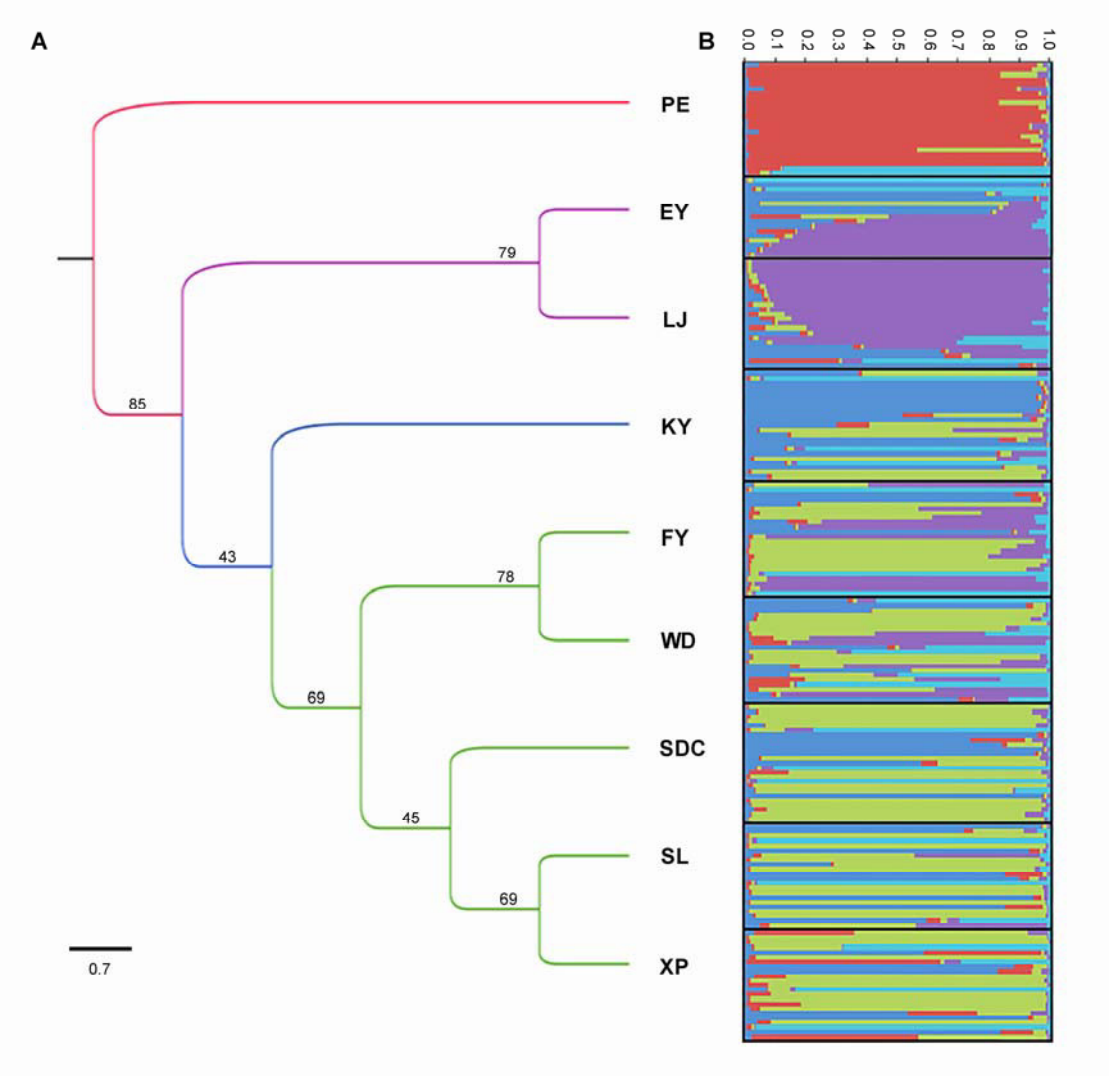

Fig. 5 Clustering of nine *Gerbera delavayi* populations. A: the unrooted UPGMA tree of nine *Gerbera delavayi* populations. The bootstrap values are indicated at major branches. Cluster Ⅰ (PE), Ⅱ (EJ and LJ), Ⅲ (KY), and Ⅳ (FY, WD, SDC, SL and XP) are indicated by red, purple, blue, and green colors, respectively. B: population structure of nine *Gerbera delavayi* populations. Each individual is shown with a vertical line, and the membership proportion is shown on the y axis.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

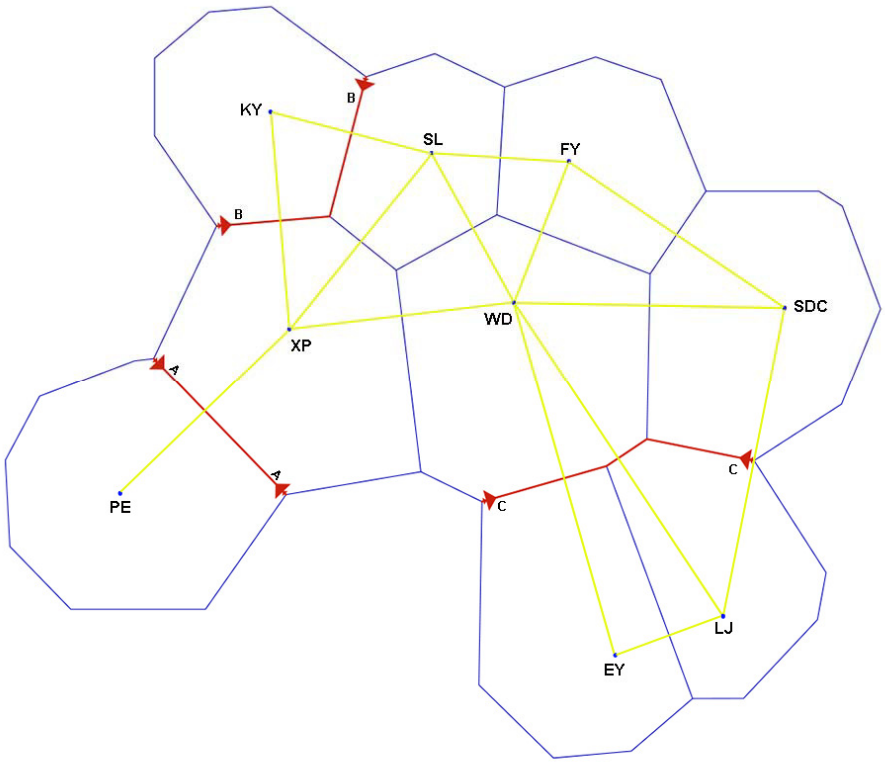

Fig. 6 Genetic barrier detected with the BARRIER Software. Genetic barriers are indicated by red bold lines with arrows. A, B, and C represent the first, second, and third genetic barrier, respectively.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

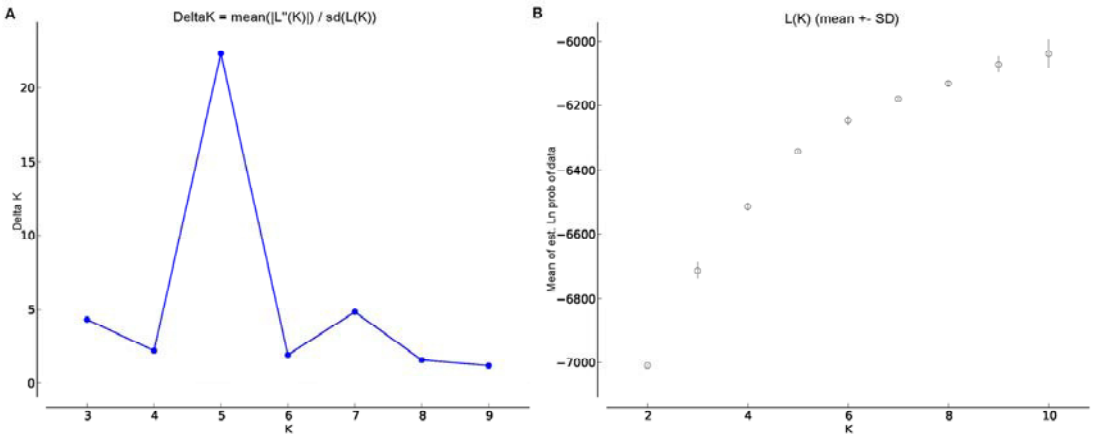

Fig. S1 Plot of the Ln P (D) ± SD and delta *k*.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

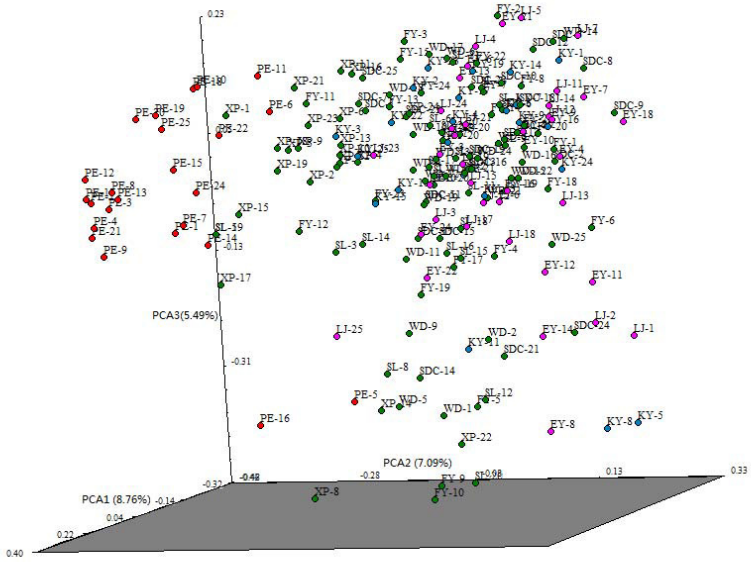

Fig. S2 Principle component analysis (PCoA) of 204 individuals. Color codes for individuals and populations are consistent with those of four genetic clusters in the unrooted UPGMA tree.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

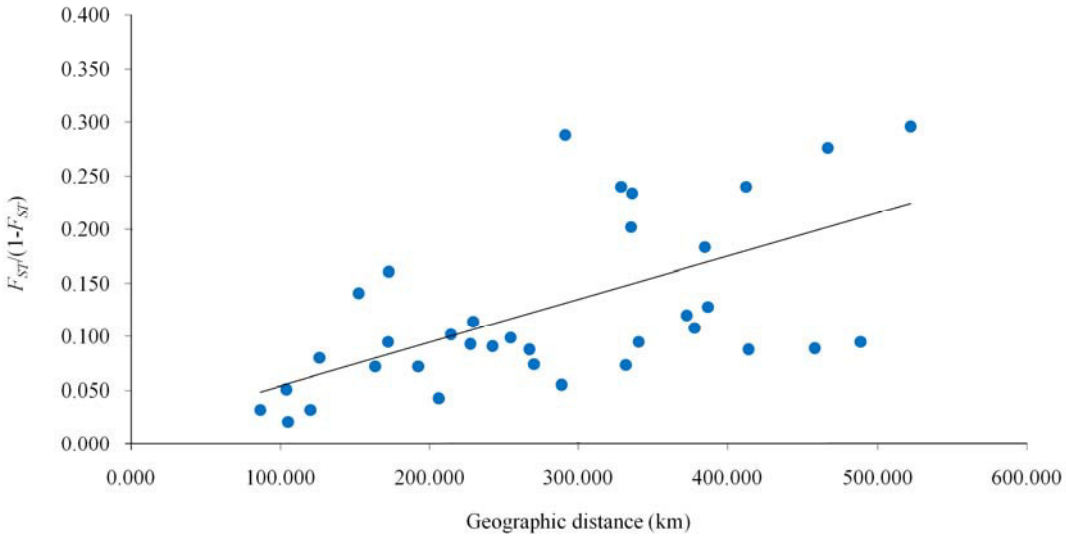

Fig. S3 Correlation between geographic and genetic distances among nine *Gerbera delavayi* populations.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

Table S1 Pairwise gene flow between nine *Gerbera delavayi* populations.

	PE	EY	LJ	KY	FY	WD	SDC	SL	XP
PE	0.000								
EY	0.708	0.000							
LJ	0.902	7.815	0.000						
KY	0.866	2.799	2.624	0.000					
FY	1.038	2.657	3.426	2.201	0.000				
WD	1.038	2.528	2.836	2.691	11.655	0.000			
SDC	0.586	3.373	2.438	2.836	3.481	5.484	0.000		
SL	1.066	2.086	1.962	2.438	4.958	7.815	4.558	0.000	
XP	1.549	1.229	1.353	1.766	2.762	2.624	2.327	3.426	0.000

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

Table S2 Pairwise F_{ST} (above diagnol) and gene flow (below diagnol) between four genetic clusters.

	Cluster Ⅰ	Cluster Ⅱ	Cluster Ⅲ	Cluster Ⅳ
Cluster Ⅰ	0.000	0.221**	0.224**	0.143**
Cluster Ⅱ	0.881	0.000	0.077**	0.072**
ClusterⅢ	0.866	2.997	0.000	0.067**
ClusterⅣ	1.498	3.222	3.481	0.000

Cluster Ⅰ: PE; Cluster Ⅱ: EJ and LJ; Cluster Ⅲ: KY; Cluster Ⅳ: FY, WD, SDC, SL, and XP. *P<0.05, **P<0.01.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2656v1 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016