Brazilian Journal of Biology®

ISSN 1519-6984 (Print)
ISSN 1678-4375 (Online)

Original Article

Genetic population structure of the Vietnamese ginseng (*Panax vietnamensis* Ha et Grushv.) detected by microsatellite analysis

Estrutura genética populacional do ginseng vietnamita (*Panax vietnamensis* Ha et Grushv.) detectada por análise de microssatélites

D. D. Vu[a*], M. T. Nguyen[b], M. D. Nguyen[c], P. L. H. Nguyen[d], T. T. X. Bui[d], K. L. Phan[b], D. G. Vu[e], Q. T. Pham[f] and T. P. T. Nguyen[d*]

[a]Join Vietnam - Russia Tropical Science and Technology Research Centre, Institute of Tropical Ecology, Hanoi, Vietnam

[b]Vietnam Academy of Science and Technology – VAST, Vietnam National Museum of Nature, Department of Experimental Taxonomy and Genetic Diversity, Hanoi, Vietnam

[c]Vietnam Academy of Science and Technology – VAST, Institute of Genome Research, Hanoi, Vietnam

[d]Vietnam Academy of Science and Technology – VAST, Institute of Ecology and Biological Resources, Department of Molecular Systematics and Conservation Genetics, Hanoi, Vietnam

[e]Hanoi University of Industry – HaUI, HaUI Institute of Technology, Hanoi, Vietnam

[f]Vietnamese Academy of Forest Sciences, Hanoi, Vietnam

Abstract

Panax vietnamensis Ha et Grushv. is a precious medicinal species native to the tropical forests of Vietnam. Due to habitat loss and over-harvesting, this species is endangered in Vietnam. To conserve the species, we investigated genetic variability and population structure using nine microsatellites for 148 individuals from seven populations across the current distribution range of *P. vietnamensis* in Vietnam. We determined a moderate genetic diversity within populations (H_O = 0.367, H_E = 0.437) and relatively low population differentiation (the Weir and Cockerham index of 0.172 and the Hedrick index of 0.254) and showed significant differentiation (P < 0.05), which suggested fragmented habitats, over-utilization and over-harvesting of *P. vietnamensis*. Different clustering methods revealed that individuals were grouped into two major clusters, which were associated with gene flow across the geographical range of *P. vietnamensis*. This study also detected that ginseng populations can have undergone a recent bottleneck. We recommend measures in future *P. vietnamensis* conservation and breeding programs.

Keywords: admixture, bottlenecks, conservation genetics, ginseng, fragmentation.

Resumo

Panax vietnamensis Ha et Grushv. é uma espécie medicinal preciosa nativa das florestas tropicais do Vietnã. Por causa da perda de hábitat e da colheita excessiva, essa espécie está ameaçada de extinção no Vietnã. Para conservá-la, investigamos a variabilidade genética e a estrutura populacional usando nove microssatélites para 148 indivíduos de sete populações em toda a distribuição atual de *P. vietnamensis* no Vietnã. Determinamos uma diversidade genética moderada dentro das populações (HO = 0,367 e HE = 0,437) e diferenciação populacional relativamente baixa (índice de Weir e Cockerham de 0,172 e índice de Hedrick de 0,254), com diferenciação significativa (P < 0,05), o que sugeriu fragmentação de hábitats, sobreutilização e sobre-exploração de *P. vietnamensis*. Diferentes métodos de agrupamento revelaram que os indivíduos foram agrupados em dois agrupamentos principais, que foram associados ao fluxo gênico em toda a área geográfica de *P. vietnamensis*. Este estudo também detectou que as populações de ginseng podem ter sofrido um gargalo recente. Recomendamos medidas em futuros programas de conservação e melhoramento de *P. vietnamensis*.

Palavras-chave: mistura, gargalos, genética da conservação, ginseng, fragmentação.

1. Introduction

Vietnamese ginseng (*Panax vietnamensis* Ha et Grushv.), one of the most important medicinal plants of the family Araliaceae, was found for the first time in the Ngoc Linh mountain range in Vietnam (Ha and Grushvitzky, 1985). Ginseng, a perennial herb, grows in small groups scattered on the slopes and ravines in the herbaceous story of tropical forests. Ginseng grows slowly, taking about 8 to 9 years to reach maturity. Its pharmacological property

*e-mail: duydinhvu87@gmail.com; ntptrang@yahoo.com
Received: May 27, 2022 – Accepted: September 16, 2022

is that ginsenosides with rich saponin compounds are accumulated, such as high content of ocotillo -type saponins (majonoside R_2) in the rhizome and exhibit the functions of immune-enhancing, anti-fatigue, and anti-cancer and improve cardiovascular function (Konoshima et al., 1998, 1999; Tran et al., 2002). In the 1970s and 1980s, the ginseng plants could be found widely in the tropical forests of Kon Tum and Quang Nam provinces, but currently, its original distribution area has been reduced to only a few habitats in Kon Tum and Quang Nam. In addition, due to the increasing demand for traditional herbal medicines, locals have over-harvested this plant. Consequently, it is being developed as a horticultural crop in these two provinces (Chien et al., 2011). *P. vietnamesis* is assessed as critically endangered on the basis of IUCN criteria (Hammer and Khoshbakht, 2005) and the national categories (Ministry of Science and Technology, 2007).

Genetic diversity is evaluated by gene flow, genetic drift, mutation and natural selection (Hamrick et al., 1992), and genetic variation and genetic structure are related to combinate factors including distribution range, evolutionary history, life cycle and mating systems (Frankham et al., 2002. Thus, genetic diversity has a critical role and reflects the adaptability of a species to changes of its environments (Reed and Frankham, 2003; Flower et al., 2018). Habitat loss and fragmentation affect distribution area and population size, and increase isolation among populations (Templeton et al., 1990). Furthermore, habitat fragmentation can influence gene flow between populations and resistance to environmental stochasticity (Bijlsma et al., 2000 ; Laurance, 2004; Sebbenn et al., 2012). Genetic variability in small populations may be decreased through genetic drifts and increase of homozygosity for common alleles (Gijbels et al., 2015), and reduce the viability and the evolutionary potential in the future (Bijlsma et al., 1997). Therefore, it is important to acknowledge that understanding the genetic structure of populations is essential to determining evolutionary characteristics, which can lead to establishing a species conservation program (Hamrick and Godt, 1996a).

In recent years, molecular techniques have been widely used in the analysis of genetic diversity and have contributed to the conservation and management of endangered species (Bruford et al., 2017). In previous studies, Zhuravlev et al. (2008) investigated and compared the genetic diversity of *Panax ginseng* CA Mayer, using three different methods and showed low levels with allozymes and high with AFLP and SSR methods. However, they showed that it was difficult to establish effective conservation strategies using allozymes. Although, AFLP markers indicated high genetic diversity, which was impossible to calculate allele frequencies directly to estimate genetic variability within and among populations. And microsatellites (SSRs) have been shownsuccessfully to describe genetic diversity, because of their codominance and polymorphism. Similarly, markers of ISSRs using the analysis of genetic diversity of P. *stipuleanatus* Tsai (Trieu et al., 2016) in Vietnam and RAPD for *P. ginseng* in Russia (Zhuravlev et al., 2004) also indicated a restriction in the estimates of genetic diversity within populations and differentiation among populations. Studies using SSRs in *Panax* species have been used (Kim et al., 2007; Park et al., 2009; Jo et al., 2009; Van Dan et al., 2010; Liu et al., 2011; Reunova et al., 2014). The set of SSRs was developed from expressed sequence tags (EST) to analyze population genetics (Vu et al., 2020). In the present study, we used EST-SSRs developed from *P. vietnamensis* to investigate the levels of genetic diversity and population structure of *P. vietnamensis*, and to provide a platform for conservation, restoration and sustainable utilization of this endangered species in Vietnam.

2. Materials and Methods

2.1. Studied locations and sample collection

Leaf samples were randomly collected from 148 plants for seven known populations of *P. vietnamensis* in two provinces of Quang Nam and Kon Tum (Table S1), representing the natural distribution range of this species. The original habitats at all studied sites were heavily influenced by human activities in the 1980s and 1990s, as a consequence of agricultural expansion and logging for commercial purposes. The current populations are fragmented and isolated in surviving habitats. The spatial distribution and age class structure of the habitats are altered.

2.2. DNA isolation and microsatellite amplification

Samples were placed in plastic bags containing silica gel at the field; transferred to Molecular Biology Laboratory, Institute of Ecology and Biological Resources; and stored at -30°C until DNA extraction. Total genomic DNA was extracted from the samples using the modified CTAB method described by Doyle and Doyle (1990). Approximately 100 mg of the sample was ground in liquid nitrogen by Mixer mill MM 400. Total DNA amount was checked using fluorimetry and NanoDrop 2000C (Thermo Sci., USA) and then diluted to a concentration of 10 ng/µl. Polymerase chain reaction (PCR) was performed in 20 µl reaction volume containing 10 µl Dream Taq Green PCR Master Mix (Thermo Fisher Scientific, Massachusetts, MA), 2 µl pure water (ddH_2O), 10 pmol each primer, and 10 ng of the extracted DNA template. Nine EST-SSR primers which were developed by Vu et al. (2020) were used for this study. PCR protocol was performed on a GeneAmp PCR System 9700, as follows: 94°C for 2 min, followed by 35 cycles of 94°C for 30 s, suitable temperature for 30 s for each primer at 55°C and 1 min extension at 72°C, and a final extension at 72°C for 10 min. Amplified products were stored at -20°C for further analysis. Amplified products were electrophoresed using Sequi-Gen®GT DNA electrophoresis system in a 6% polyacrylamide gel in 1 x TAE buffer and then visualized by a GelRed™ Nucleic Acid Gel Stain. Alleles were sized using Gel-Analyzer software of GenoSens1850 with 25 bp DNA ladder (Invitrogen).

2.3. Molecular analysis, genetic diversity and genetic structure

We used the Micro-Checker (Van Oosterhout et al., 2004) to check for evidence of null when excess homozygosity

was detected. Genetic parameters include the number of alleles per locus (N_A), effective alleles (A_E), allelic richness (A_R), the observed (H_O) and expected (H_E) heterozygosity across loci and populations, the fixation index (F_{IS}: the inbreeding coefficient), the differentiation index between pairwise populations [the F-statistics of Weir and Cockerham (1984) F_{ST} and Hedrick (2005) G'_{ST}] and total expected heterozygosity (H_T) were calculated using GENALEX (Peakall and Smouse, 2012). In addition, allelic richness (A_R) and the inbreeding coefficient (F_{IS}) were calculated using FSTAT (Goudet, 2001) based on SSR allele frequencies. Tests for genotype linkage disequilibrium and departure from Hardy-Weinberg equilibrium were implemented using Arlequin for each population, based on 10,000 permutations. We corrected the F_{IS} values for null allele frequencies based on the individual inbreeding model (IIM) using INEst (Chybicki and Burczyk, 2009). Gene flow among populations (N_m) was calculated using the value: $N_m = [(1/F_{ST}) - 1]/4$. We used Bottleneck 1.2 (Piry et al., 1999) to test recent bottleneck events for each population via the infinite allele model (IAM), the stepwise mutation model (SMM) and the two-phase model (TPM). We evaluated the significance of these tests by the one-tailed Wilcoxon signed-rank test. The proportion of the stepwise mutation model was set to 70% under default settings. We used Arlequin 3.1 (Excoffer et al., 2005) to test the significance for variance components in the analysis of molecular variance (AMOVA). A neighbor-joining (NJ) tree was performed for genetic association among populations based on the F_{ST} values using POPTREE2 (Takezaki et al., 2010). A principal coordinate analysis (PCoA) was also performed using GenAlEx v.6.5 based on the G'_{ST} values. The Bayesian clustering approach was performed to determine population structure using STRUCTURE v.2.3.4 (Pritchard et al., 2000). Setting the admixture model with correlated allele frequencies, ten separate runs of the number of groups in the data set (K) were implemented for K between 1 and 10 at 100,000 Markov Chain Monte Carlo repetitions and at 500,000 burn–in period. We determined the optimal value of K, using Structure Harvester (Earl and von-Holdt, 2012) and StrcutureSelector (Li and Liu, 2018) to detect the number of groups that best fit the dataset based on the ΔK of Evanno et al. (2005). The estimators for the median of medians (*MedMedK)*, the median of means (MedMeanK), the maximum of medians (MaxMedK) and the maximum of mean (MaxMeanK) criteria (Puechmaille, 2016) were also conducted to detect the most likely number of clusters using StrcutureSelector. The Clumpark program (Kopelman et al., 2015) was used to generate a graphical representation of the Structure results.

3. Results

3.1. Genetic diversity

We detected null allele frequencies at seven loci ($P < 0.05$) (Table S2). Genotypic linkage disequilibrium was examined for *P. vietnamensis*. Fifty-nine out of 252 tests were significant at the 5% level. In total, 38 different alleles were identified via nine SSR loci among 148 *Panax vietnamensis* individuals in seven populations, and the genetic parameters of this species are presented in Table S2. Alleles per locus (N_A) ranged from two at one locus to six at two loci. Locus panv2 and panv4 both had the highest A value (6), while locus panv6 had the lowest N_A (2). High allelic richness (A_R) was detected at the two loci panv2 (4.2) and panv4 (4.0) and the lowest A_R was determined at locus panv6 (2.0). Private alleles were detected at one locus for Phuoc Loc and Ngok Lay, two loci for Tra Leng and Tra Cang and three loci for Ngok Linh. No private allele was found in two populations Tra Linh and Mang Ri. The observed heterozygosity (H_O) ranged from 0.269 (panv9) to 0.455 (panv7), and the expected heterozygosity (H_E) ranged from 0.267 (panv6) to 0.463 (panv2). Similarly, the total expected heterozygosity (H_T) was the highest (0.71) at the locus panv2 and the lowest (0.31) at the locus panv6. Eight of the nine loci had a positive fixation index (F_{IS}), indicating an excess of homozygotes and inbreeding. All studied loci had no signs ($P > 0.05$). The fixation index overall populations for each locus ranged from -0.051 (panv6) to 0.309 (panv9).

At the population level, the proportion of polymorphism (PPL) ranged from 88.89% in Mang Ri to 100% in the remaining six populations, an average of 98.41% (Table 1). The Mang Ri population also had the lowest values of alleles (20), alleles for each locus (2.2), effective alleles (1.2) and allelic richness (2.16), whereas these values were the highest in the Phuoc Loc population (30 alleles, $N_A = 2.4$, $A_E = 2.4$ and $A_R = 3.22$). The observed (H_O) and expected (H_E) heterozygosity ranged from 0.17 (Mang Ri) to 0.441 (Tra Leng) and 0.187 (Mang Ri) to 0.573 (Phuoc Loc), with an average of 0.367 and 0.437, respectively (Table 1). High genetic diversities were detected in the four populations of Tra Leng, Tra Cang, Tra Linh and Phuoc Loc, while these values were low in the remaining three populations (Ngok Linh, Ngok Lay and Mang Ri. The fixation index (F_{IS}) ranged from -0.003 (Ngok Linh) to 0.276 (Tra Linh), an average of 0.163, showing a significant deficiency in heterozygosity ($P < 0.001$). Significant positive F_{IS} values were detected in five populations of Tra Leng, Tra Cang, $P < 0.01$; Tra Linh, Phuoc Loc $P < 0.001$; and Ngok Lay, $P < 0.05$. Based on the individual inbreeding model (F_{IS}IIM), the inbreeding corrected for null alleles ranged from 0.046 (Ngok Linh) to 0.201 (Tra Linh), an average of 0.125, and also showed homozygosity excess. Of course, the F_{IS}IIM value was lower compared to F_{IS}. The inbreeding coefficient calculated for the total populations (F_{IT}) ranged from 0.094 (panv6) to 0.465 (panv2 and panv9), an average 0.296, suggesting homozygote excess in the *P. vietnamensis* populations. Significant heterozygous deficits were detected in the three populations of Tra Cang, Tra Linh and Phuoc Loc ($P < 0.05$), based on the Bottleneck analysis (Table 1). These results suggest that evidence of a recent bottleneck has appeared in some of the studied populations.

3.2. Genetic structure

The genetic differentiation for each locus varied from 0.045 to 0.348, an average of 0.172; and 0.05 to 0.681, an average of 0.254 for F_{ST} and G'_{ST}, respectively (Table S3) and showing moderate genetic differentiation. Gene flow (N_m)

Table 1. Genetic diversity values and results of bottleneck tests for seven *P. vietnamensis* populations.

Populations	N	PPL (%)	Alleles	N_A	A_E	A_R	H_O (SE)	H_E (SE)	F_{IS} (SE)	F_{IS} IIM	P value of bottleneck		
											A	B	C
Tra Leng	22	100	28	3.1	2.1	2.96	0.441	0.523	0.179**	0.09	0.001	0.002	ns
Tra Cang	18	100	27	3	2.2	2.98	0.434	0.513	0.183**	0.139	0.003	0.014	0.024
Tra Linh	20	100	28	3.1	2.3	3.07	0.419	0.56	0.276***	0.201	0.001	0.001	0.003
Phuoc Loc	25	100	30	3.3	2.4	3.22	0.437	0.573	0.256***	0.101	0.001	0.001	0.005
Ngok Linh	19	100	25	2.8	1.5	2.68	0.322	0.312	-0.003	0.046	ns	ns	ns
Ngok Lay	25	100	25	2.8	1.8	2.3	0.347	0.392	0.135*	0.106	ns	ns	ns
Mang Ri	19	88.89	20	2.2	1.2	2.16	0.170	0.187	0.119	0.192	ns	ns	ns
Mean		98.41			1.9 (0.098)		0.367 (0.066)	0.437 (0.023)	0.163	0.125			

N: sample size; PPL: proportion of polymorphic loci; N_A: alleles per locus; A_E: effective alleles; A_R: allelic richness; H_O and H_E: observed and expected heterozygosities; F_{IS}: fixation index; F_{IS}IIM: corrected inbreeding coefficient for null alleles; A: heterozygosity deficit, one-tailed test; B: heterozygosity excess, one-tailed test; C: heterozygosity excess or deficit, two tailed test; SE: standard error. *P < 0.05. **P < 0.01. ***P < 0.001.

Table 2. Analysis of molecular variance from natural populations of *P. vietnamensis* produced from Arlequin.

	df	Sum of squares	Variance components	Total variation (%)	Fixation indices
Among populations	6	81.978	0.277	14.47	F_{IS} = 0.196***
					F_{ST} = 0.145***
					F_{IT} = 0.313***
Among individuals within populations	141	276.654	0.322	16.81	
Within inviduals	148	195	1.317	68.72	
Total	295	553.632	1.917		

df: degree of freedom. ***P < 0.001.

for each locus varied from 0.468 (panv1) to 5.29 (panv8), an average of 1.653. Both The analysis of molecular variance (AMOVA), F_{ST} and G'_{ST} analyses were conduced to detect genetic variation between the studied populations and groups. Hierachical AMOVA showed that 14.47% of the total molecular variation found among seven populations, and 68.72% of the total variation was distributed within individuals (Table 2). The remaining variation occurred among individuals within populations (16.81%). This result was also confirmed by the overall F_{ST}, G'_{ST} and N_m values (Table S2). Moreover, the genetic differentiation among populations (F_{ST} and G'_{ST}) varied from 0.033 to 0.232, and 0.042 to 0.515, with F_{ST} and G'_{ST}, respectively. For both F_{ST} and G'_{ST} values, the highest levels detected between populations Mang Ri and Tra Leng, whereas the lowest levels detected between Ngok Linh and Mang Ri. The genetic variation between populations was significant (P < 0.01), based on 999 permutations. The genetic groups of *P. vietnamensis* populations were detected by different clustering methods. Two groups were generated using the Neighbor-joining (NJ) analysis (Figure S1). The four populations of Phuoc Loc, Tra Cang, Tra Leng and Tra Linh from Quang Nam were clustered together with a bootstrap value of 99%. The remaining three populations of Ngok Linh, Ngok Lay and Mang Ri from Kon Tum were clustered into the second group. The G'_{ST} values were used for principal coordinate analysis (PCoA). The first and second principal coordinate explained 65.42% and 13.88% of the variation within the genetic data, respectively (Figure S1). The PCoA analysis showed that the Ngok Lay population was separated from the populations in Kon Tum, and Tra Leng was also separated from the populations in Quang Nam. The Bayesian analysis, performed by STRUCTURE detected that the most likely number of genetic clusters was 2 (ΔK = 372.5) and showed that all studied individuals exhibited admixture from two groups (Figure S2). The percentage of the ancestry of each population and individuals in the two groups was presented in each color. One group (orange) was predominant in three Kon Tum populations Ngok Linh, Ngok Lay and Mang Ri with strong ancestral values 78.9%, 70.1% and 96.4%, respectively (Figure 1; Table S4). The second group (blue) was predominant in the remaining four Quang Nam populations Tra Leng, Tra Cang, Tra Linh and Phuoc Loc, with ancestral values 95.7%, 90.5%, 84.7% and 82%, respectively. However, at K = 3, the Quang Nam populations were divided into two subgroups. One group included only one population of Tra Leng with the highest ancestral value (68%). The other group included three populations Tra Cang, Tra Linh and Phuoc Loc with the ancestral values 48.4%, 51.8% and 70.7%, respectively (Figure 1; Table S4). At K = 4, the Kon Tum populations were differentiated into two subgroups. One subgroup included only one population Ngok Lay with an ancestral value of 45.8%. The second subgroup included the remaining two populations Ngok

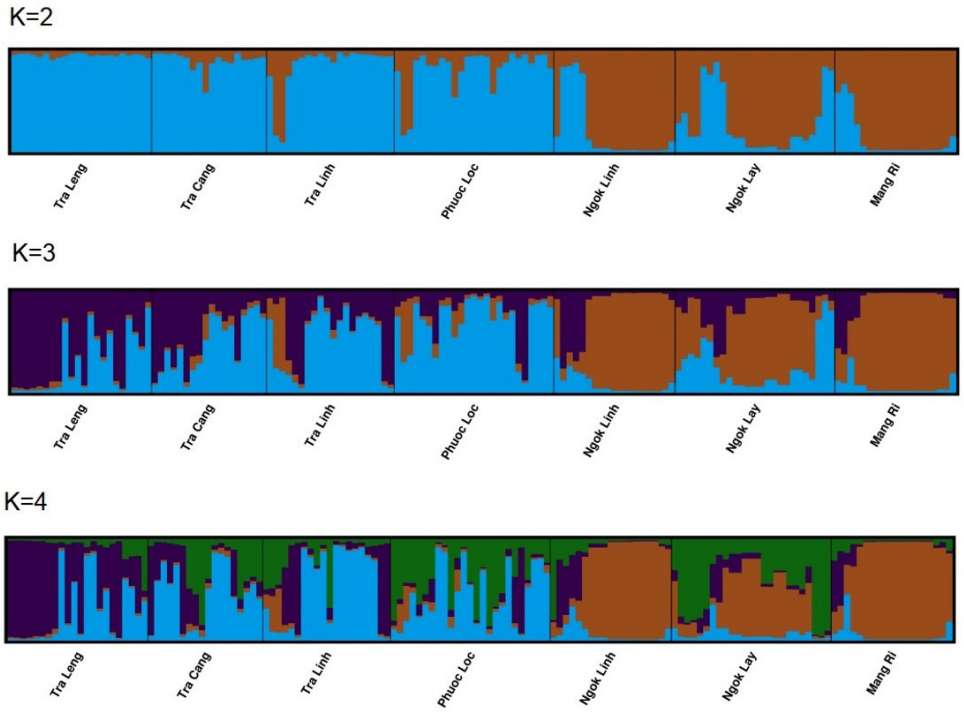

Figure 1. Bar plots for seven *P. vietnamensis* populations at K = 2, 3 and 4.

Linh and Mang Ri with ancestral values 70.5% and 80.7%, respectively (Figure 1; Table S4). The Puechmaille approach showed three clusters based on the estimator MedMeanK, MedMedK, MaxMedK and MaxMeanK (Figure S3). Although the K = 3 model indicated a lower Δ*K* value than the delta*K* values at K = 2, this model was supported from the results of the Puechmaille estimators.

4. Discussion

Previously, Vu et al. (2020) developed nine SSR primers from *Panax vietnamensis* and evaluated the genetic diversity of this species in Ngoc Linh Nature Reserve in Vietnam, and showed that *P. vietnamensis* has a moderate genetic diversity level with the observed and expected heterozygosity, 0.422 and 0.479, respectively. In the present study, similar results was also recorded with values of 0.367 and 0.437, respectively. Similar values of genetic diversity were observed for *P. ginseng* (Reunova et al., 2014), *P. notoginseng* (Liu et al., 2011) and *P. vietnamensis* (Reunova et al., 2011). Another study showed that low genetic diversity was detected for some *Panax* species, such as *P. stipuleanatus* using ISSR markers (Trieu et al., 2016), and *P. ginseng* using allozymes, RAPD and ISSRs (Koren et al., 2003; Zhuravlev et al., 2008; Reunova et al., 2010). In the present study, high genetic diversity was detected in four Quang Nam populations Tra Leng, Tra Cang, Tra Linh and Phuoc Loc, compared with that in the three Kon Tum populations Ngok Linh, Ngok Lay and Mang Ri. These may indicate that the *P. vietnamensis* habitat is restricted to small areas. The current populations were degraded and fragmented into subpopulations with few individuals. This suggests that habitat degradation and excessively harvesting of *P. vietnamensis* led to low genetic heterozygosity. Low genetic diversity can be related to high degree of habitat disturbance and small population sizes in Kon Tum. Our study showed that high allelic richness was observed in four Quang Nam populations, such as Phuoc Loc (A_R = 3.22, Tra Linh (A_R = 3.07, Tra Cang (A_R = 2.98) and Tra Leng (A_R = 2.96). These populations should be a priority for genetic conservation (Petit et al., 1998). The results showed a heterozygosity deficit in all studied populations, except for the Ngok Linh population. The mean observed heterozygosity was 0.367, whereas the mean expected heterozygosity was 0. 437. This suggests that *P. vietnamensis* has been affected by its restricted and fragmented range. This species is found only in two provinces Quang Nam and Kon Tum in Viet Nam. The average of significant heterozygosity deficit (F_{IS} = 0.163, F_{IS}IIM = 0.125) was estimated from148 plants across seven populations showing the existence of biparental inbreeding within the populations. Low plant density could be a major factor leading high levels of inbreeding. The deficiency of heterozygosity was detected using Bottleneck analysis and also showed a decrease in the population size of *P. vietnamensis*.

To study the genetic structure of *P. vietnamensis*, genetic differentiation among populations were analyzed. Our results showed that a moderate population differentiation (the F-statistics of Weir and Cockerham F_{ST} = 0.172 and AMOVA F_{ST} = 0.145 and Hedrick G'_{ST} = 0.254) was revealed. Population differentiation was low within the same province, an average of F_{ST} = 0.056 and G'_{ST} = 0.15 in

Quang Nam; F_{ST} = 0.048 and G'_{ST} = 0.091 in Kon Tum, and high in different provinces with F_{ST} = 0.145 and G'_{ST} = 0.345. Similarly, AMOVA analysis indicated that most of genetic variation was distributed within the populations (85%). This can be a consequence of the decrease in genetic divergence among populations. *P. vietnamensis* is a long-living and outcrossing species. The species is insect-pollinated by insects and predominantly outcrossed. Pollen dispersal could contribute mainly to the gene flow and population structure of *P. vietnamensis*. Moreover, its seeds can be dispersed by animals (rodents and birds). Low genetic differentiation between populations reflects high gene flow (Hamrick and Godt, 1996b; Brütting et al., 2012). Previous studies showed low population differentiation and strong gene flow in *P. ginseng* (Zhuravlev et al., 2008; Reunova et al., 2014) and *P. stipuleanatus* (Le et al., 2016). Our results determined high gene flow between *P. vietnamensis* populations (N_m = 1.653) and indicated that the number of migrants per generation inferred from the nine studied loci.

Different clustering approaches presented the genetic structure within *P. vietnamensis* and visualized the genetic relationships of its populations. NJ tree based on the F_{ST} values and PCoA based on the G'_{ST} values identified two different clusters from seven populations (Figure S1). These results were consistent with the geographical distribution of the *P. vietnamensis* populations. The admixture model-based method performed in the Structure program also confirmed that the two clusters were optimal for the 148 sampled individuals. One cluster included four populations in Quang Nam (Tra Leng, Tra Cang, Tra Linh and Phuoc Loc) and exhibited strong similarity of the bar plot pattern, whereas the second cluster included the remaining three populations in Kon Tum (Ngok Linh, Ngok Lay and Mang Ri). This suggests that the existence of the genetic structure for this species is the consequence of gene flow, regarding populations into the identified genetic clusters. However, three or four clusters can be formed from the Structure analysis. The Puechmaille method also revealed three genetic clusters based on the estimators MedMeanK, MedMedK, MaxMedK and MaxMeanK.

In conclusion, in the present study, we determined the moderate genetic variability within populations and low population differentiation of *P. vietnamensis* in the tropical forests of Vietnam. This study also indicated that gene flow is relatively high in the same province and low in different provinces. Therefore, we recommend that all studied populations might be considered for species conservation. Populations with high allelic richness and genetic variability could be prioritized for species consideration.

Acknowledgements

We thank Tay Bac Pharmaceutical Trading One Member Co., Ltd., Than Uyen, Lai Chau province, Vietnam for support of this study. This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 106.06 – 2018.310.

References

BIJLSMA, P., VAN ARENDONK, J.A.M. and WOOLLIAM, J.A., 2000. A general procedure for predict rates of inbreeding in populations undergoing mass selection. *Genetics*, vol. 154, no. 4, pp. 1865-1877. https://doi.org/10.1093/genetics/154.4.1865.

BIJLSMA, R., BUNDGAARD, J., BOEREMA, A.C. and VAN PUTTEN, W.F., 1997. Genetics and environmental stress, and the persistence of populations. In: R. BIJLSMA and V. LOESCHCKE, eds. *Environmental stress, adaptation and evolution*. Basel: Birkhauser Verlag, pp. 193-207. http://dx.doi.org/10.1007/978-3-0348-8882-0_11.

BRUFORD, M.W., DAVIES, N., DULLOO, M.E., FAITH, D.P. and WALTERS, M., 2017. Monitoring changes in genetic diversity. In: M. WALTERS and R.J. SCHOLES, eds. *The GEO handbook on biodiversity observation networks*. New York: Springer, pp. 107-128. http://dx.doi.org/10.1007/978-3-319-27288-7_5.

BRÜTTING, C., WESCHE, K., MEYER, S. and HENSEN, I., 2012. Genetic diversity of six arable plants in relation to their Red List status. *Biodiversity and Conservation*, vol. 21, no. 3, pp. 745-761. http://dx.doi.org/10.1007/s10531-011-0212-z.

CHIEN, H.X., TAI, N.T., TRUC, N.B., TINH, T.X., THAO, L.B., LUAN, T.C. and NHUT, D.T., 2011. Factors effecting in vitro microrhizome formation of *Panax vietnamensis* Ha et Grushv. and quantitation of saponin content of there from generated plantlets grown in Ngoc Linh Mountain. *Journal of Biotechnology*, vol. 9, pp. 317-331.

CHYBICKI, I.J. and BURCZYK, J., 2009. Simultaneous estimation of null alleles and inbreeding coefficient. *The Journal of Heredity*, vol. 100, no. 1, pp. 106-113. http://dx.doi.org/10.1093/jhered/esn088. PMid:18936113.

DOYLE, J.J. and DOYLE, L.J., 1990. Isolation of plant DNA from fresh tissue. *Focus*, vol. 12, pp. 13-15.

EARL, D.A. and VON-HOLDT, B.M., 2012. Structure harvester: a website and program for visualizing structure output and implementing the Evanno method. *Conservation Genetics Resources*, vol. 4, no. 2, pp. 359-361. http://dx.doi.org/10.1007/s12686-011-9548-7.

EVANNO, G., REGNAUT, S. and GOUDET, J., 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, vol. 14, no. 8, pp. 2611-2620. http://dx.doi.org/10.1111/j.1365-294X.2005.02553.x. PMid:15969739.

EXCOFFER, L., LAVAL, G. and SCHNEIDER, S., 2005. Arlequin v. 3.5. an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, vol. 1, pp. 47-50.

FLOWER, C.E., FANT, J.B., HOBAN, S., KNIGHT, K.S., STEGER, L., AUBIHL, E., GONZALEZ-MELER, M.A., FORRY, S., HILLE, A. and ROYO, A.A., 2018. Optimizing conservation strategies for a threatened tree species: in situ conservation of white ash (*Fraximus americana* L.) genetic diversity through insecticide treatment. *Forests*, vol. 9, no. 4, p. 202. http://dx.doi.org/10.3390/f9040202.

FRANKHAM, R., BAILOU, J.D. and BRISCOE, D.A., 2002. *Introduction to conservation genetics*. Cambridge: Cambridge Univeristy Press.

GIJBELS, P., HERT, K.D., JACQUEMYN, H. and HONNAY, O., 2015. Reduced fecundity and genetic diversity in small populations of rewarding versus deceptive orchid species: a meta-analysis. *Ecology and Evolution*, vol. 148, pp. 153-159.

GOUDET, J., 2001 [viewed 16 September 2022]. *FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3)* [online]. Available from: http://www.unil.ch/izea/softwares/fstat.html

HA, T.D. and GRUSHVITZKY, I.V., 1985. New species in *Panax* (Araliaceae) in Vietnam. *Le Journal de Botanique*, vol. 70, pp. 518-522.

HAMMER, K. and KHOSHBAKHT, K., 2005. Towards a 'red list' for crop plant species. *Genetic Resources and Crop Evolution*, vol. 52, no. 3, pp. 249-265. http://dx.doi.org/10.1007/s10722-004-7550-6.

HAMRICK, J.L. and GODT, M.J.W., 1996a. Conservation genetics of endemic plant species. In: J.S. AVISE and J.L. HAMRICK, eds. *Conservation genetics: case histories from nature*. New York: Chapman and Hall, pp. 281-304. http://dx.doi.org/10.1007/978-1-4757-2504-9_9.

HAMRICK, J.L. and GODT, M.J.W., 1996b. Effects of history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 351, pp. 1683-1685.

HAMRICK, J.L., GODT, M.J.W. and SHERMAN-BROYLES, S.L., 1992. Factors influencing levels of genetic diversity in woody plant species. *New Forests*, vol. 6, no. 1-4, pp. 95-124. http://dx.doi.org/10.1007/BF00120641.

HEDRICK, P.W., 2005. A standardized genetic differentiation measure. *Evolution*, vol. 59, no. 8, pp. 1633-1638. http://dx.doi.org/10.1111/j.0014-3820.2005.tb01814.x. PMid:16329237.

JO, B.H., SUH, D.S., CHO, E.M., KIM, J., RYU, G.H. and CHUNG, K.W., 2009. Characterization of polymorphic microsatellite loci in cultivated and wild *Panax ginseng*. *Genes & Genomics*, vol. 31, no. 2, pp. 119-127. http://dx.doi.org/10.1007/BF03191145.

KIM, J., JO, B.H., LEE, K.L., YOON, E.S., RYU, G.H. and CHUNG, K.W., 2007. Identification of new microsatellite markers in *Panax ginseng*. *Molecules and Cells*, vol. 24, no. 1, pp. 60-68. PMid:17846499.

KONOSHIMA, T., TAKASAKI, M., ICHIISHI, E., MURAKAMI, T., TOKUDA, H., NISHINO, H., DUC, N.M., KASAI, R. and YAMASAKI, K., 1999. Cancer chemopreventive activity of majonoside-R2 from Vietnamese ginseng, *Panax vietnamensis*. *Cancer Letters*, vol. 147, no. 1-2, pp. 11-16. http://dx.doi.org/10.1016/S0304-3835(99)00257-8. PMid:10660083.

KONOSHIMA, T., TAKASAKI, M., TOKUDA, H., NISHINO, H., DUC, N.M., KASAI, R. and YAMASAKI, K., 1998. Anti-tumor-promoting activity of majonoside-R2 from Vietnamese ginseng, *Panax vietnamensis* Ha et Grushv. *Biological & Pharmaceutical Bulletin*, vol. 21, no. 8, pp. 834-838. http://dx.doi.org/10.1248/bpb.21.834. PMid:9743252.

KOPELMAN, N.M., MAYZEL, J., JAKOBSSON, M., ROSENBERG, N.A. and MAYROSE, I., 2015. CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, vol. 15, no. 5, pp. 1179-1191. http://dx.doi.org/10.1111/1755-0998.12387. PMid:25684545.

KOREN, O.G., POTENKO, V.V. and ZHURAVLEV, Y.N., 2003. Inheritance and variation of allozymes in *Panax ginseng* C.A. Meyer (Araliaceae). *International Journal of Plant Sciences*, vol. 164, no. 1, pp. 189-195. http://dx.doi.org/10.1086/344758.

LAURANCE, W.F., 2004. Forest-climate interactions in fragmented tropical landscapes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, vol. 359, no. 1443, pp. 345-352. http://dx.doi.org/10.1098/rstb.2003.1430. PMid:15212089.

LI, Y.L. and LIU, J.X., 2018. Structure selector: a web-based software to select and visualize the optimal number of clusters using multiple methods. *Molecular Ecology Resources*, vol. 18, no. 1, pp. 176-177. http://dx.doi.org/10.1111/1755-0998.12719. PMid:28921901.

LIU, H., XIA, T., ZUO, Y.J., CHEN, Z.J. and ZHOU, S.L., 2011. Development and characterization of microsatellite markers for *Panax notoginseng* (Araliaceae), a Chinese traditional herb. *American Journal of Botany*, vol. 98, no. 8, pp. e218-e220. http://dx.doi.org/10.3732/ajb.1100043. PMid:21821584.

MINISTRY OF SCIENCE AND TECHNOLOGY and VIETNAM ACADEMY OF SCIENCE AND TECHNOLOGY, 2007. *Red book of Vietnam. Part II - plants*. Hanoi, Vietnam: House for Science and Technology.

PARK, S.-W., HYUN, Y.-S. and CHUNG, K.-W., 2009. Genetic polymorphism of microsatellite markers in *Panax ginseng* C.A. Meyer. *Journal of Ginseng Research*, vol. 33, no. 3, pp. 199-205. http://dx.doi.org/10.5142/JGR.2009.33.3.199.

PEAKALL, R. and SMOUSE, P.E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research an update. *Bioinformatics*, vol. 28, no. 19, pp. 2537-2539. http://dx.doi.org/10.1093/bioinformatics/bts460. PMid:22820204.

PETIT, R.J., MOUSADIK, E.I. and PONS, O., 1998. Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, vol. 12, no. 4, pp. 844-855. http://dx.doi.org/10.1046/j.1523-1739.1998.96489.x.

PIRY, S., LUIKART, G. and CORNUET, J.-M., 1999. Computer note. BOTTLENECK: a computer program for detecting recent reductions in the effective size using allele frequency data. *Journal of Heredity*, vol. 90, no. 4, pp. 502-503. http://dx.doi.org/10.1093/jhered/90.4.502.

PRITCHARD, J.K., STEPHENS, M. and DONNELLY, P., 2000. Inference of population structure using multilocus genotype data. *Genetics*, vol. 155, no. 2, pp. 945-959. http://dx.doi.org/10.1093/genetics/155.2.945. PMid:10835412.

PUECHMAILLE, S.J., 2016. The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, vol. 16, no. 3, pp. 608-627. http://dx.doi.org/10.1111/1755-0998.12512. PMid:26856252.

REED, D.H. and FRANKHAM, R., 2003. Correlation between fitness and genetic diversity. *Conservation Biology*, vol. 17, no. 1, pp. 230-237. http://dx.doi.org/10.1046/j.1523-1739.2003.01236.x.

REUNOVA, G.D., KATS, I.L., MUZAROK, T.I. and ZHURAVLEV, Y.N., 2010. Polymorphism of RAPD, ISSR and AFLP markers of the *Panax ginseng* C.A. Meyer (Araliaceae) genome. *Russian Journal of Genetics*, vol. 46, no. 8, pp. 938-947. http://dx.doi.org/10.1134/S1022795410080053. PMid:20873202.

REUNOVA, G.D., KOREN, O.G., MUZAROK, T.I. and ZHURAVLEV, Y.N., 2014. Microsatellite analysis of *Panax ginseng* natural population in Russia. *Chinese Medicine*, vol. 5, no. 4, pp. 231-243. http://dx.doi.org/10.4236/cm.2014.54028.

SEBBENN, A.M., LICONA, J.C., MOSTACEDO, B. and DEGEN, B., 2012. Gene flow in an overexploited population of *Swietenia macrophylla* King (Meliaceae) in the Bolivian Amazon. *Silvae Genetica*, vol. 61, no. 1-6, pp. 212-220. http://dx.doi.org/10.1515/sg-2012-0027.

TAKEZAKI, N., NEI, M. and TAMURA, K., 2010. Software for constructing population trees from allele frequency data and computing other population statistics with windows interface. *Molecular Biology and Evolution*, vol. 27, no. 4, pp. 747-752. http://dx.doi.org/10.1093/molbev/msp312. PMid:20022889.

TEMPLETON, A.R., SHAW, K., ROUTMAN, E. and DAVIS, S.K., 1990. The genetic consequences of habitat fragmentation. *Annals of the Missouri Botanical Garden*, vol. 77, no. 1, pp. 13-27. http://dx.doi.org/10.2307/2399621.

TRAN, Q.L., ADNYANA, I.K., TEZUKA, Y., HARIMAYA, Y., SAIKI, I., KURASHIGE, Y., TRAN, Q.K. and KADOTA, S., 2002. Hepatoprotective effect of majonoside R2, the major saponin from Vietnamese ginseng (*Panax vietnamensis*).

Planta Medica, vol. 68, no. 5, pp. 402-406. http://dx.doi.org/10.1055/s-2002-32069. PMid:12058314.

TRIEU, L.N., MIEN, N.T., TIEN, T.V., KET, N.V. and DUY, N.V., 2016. Genetic diversity of *Panax stipuleanatus* Tsai in north Vietnam detected by inter singple sequence repeat (ISSR) markers. *Biotechnology, Biotechnological Equipment*, vol. 30, no. 3, pp. 506-511. http://dx.doi.org/10.1080/13102818.2016.1157448.

VAN DAN, N., RAMCHIARY, N., CHOI, S.R., UHM, T.S., YANG, T.-J., AHN, I.-O. and LIM, Y.P., 2010. Development and characterization of new microsatellite markers in *Panax ginseng* (C.A. Meyer) from BAC end sequences. *Conservation Genetics*, vol. 11, no. 3, pp. 1223-1225. http://dx.doi.org/10.1007/s10592-009-9924-y.

VAN OOSTERHOUT, C., HUTCHINSON, W.F., WILLS, D.P.M. and SHIPLEY, D.F., 2004. Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, vol. 4, no. 3, pp. 535-538. http://dx.doi.org/10.1111/j.1471-8286.2004.00684.x.

VU, D.D., SHAH, S.N.M., PHAM, M.P., BUI, V.T., NGUYEN, M.T. and NGUYEN, T.P.T., 2020. De novo assembly and transcriptome characterization of an endemic species of Vietnam, *Panax vietnamensis* Ha et Grushv., including the development of EST-SSR markers for population genetics. *BMC Plant Biology*, vol. 20, no. 1, p. 358. http://dx.doi.org/10.1186/s12870-020-02571-5. PMid:32727354.

WEIR, B.S. and COCKERHAM, C.C., 1984. Estimating F-statistics for the analysis of population structure. *Evolution*, vol. 38, no. 6, pp. 1358-1370. PMid:28563791.

ZHURAVLEV, Y.N., KOREN, O.G., REUNOVA, G.D., ARTYUKOVA, E.V., KOZYRENKO, M.M., MUZAROK, T.I. and KATS, I.L., 2004. Ginseng conservation program in Russian Primorye: genetic structure of wild and cultivated populations. *Journal of Ginseng Research*, vol. 28, no. 1, pp. 60-66. http://dx.doi.org/10.5142/JGR.2004.28.1.060.

ZHURAVLEV, Y.N., KOREN, O.G., REUNOVA, G.D., MUZAROK, T.I., GORPENCHENKO, T.Y., KATS, I.K. and KHROLENKO, Y.A., 2008. *Panax ginseng* natural populations: their past, current state and perspectives. *Acta Pharmacologica Sinica*, vol. 29, no. 9, pp. 1127-1136. http://dx.doi.org/10.1111/j.1745-7254.2008.00866.x. PMid:18718182.

Supplementary Material

Supplementary material accompanies this paper.

Table S1. Collection localities of Panax vietnamensis.

Table S2. Characterization and polymorphic levels of nine microsatellite loci in Panax vietnamensis.

Table S3. Pairwise genetic differentiation between seven P. vietnamensis populations.

Table S4. Percentage of ancestry for seven P. vietnamensis populations, based on 10 runs at K = 2, K = 3 and K = 4 in STRUCTURE and compiled in STRUCTURE HARVESTER.

Figure S1. Relationships between the seven P. vietnamensis populations. A) Neighbor - joining (NJ) tree based on the FST values produced from POPTREE21 and B) Principal Coordinate analysis (PCoA) based on the G'ST values produced from GenALEX2

1Takezaki N, Nei M, Tamura K. 2010. Software for constructing population trees from allele frequency data and computing other population statistics with windows interface. Mol Evol. 27:747-752. http://dx.doi.org/10.1093/molbev/msp312.

2Peakall R, Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research an update. Bioinformatics. 28:2537-39.

Figure S2. A) The distribution of deltaK over K = 1-10, B) the mean values of distribution of probability of the data (LnP (K)) and standard deviation from 10 runs for each value of K = 1-10 in Structure1 analysis of seven P. vietnamensis populations using Structure Harvester2

1Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155: 945-959.

2Earl DA, von-Holdt BM. 2012. Structure Harvester: a website and program for visualizing structure output and implementing the Evanno method. Cons Genet Res. 4:359-361.

Figure S3. Genetic clusters inferred with estimators MedMedK, MedMeanK, MaxMedK and MaxMeanK from Structire results for seven

P. vietnamensis populations and compiled in StructureSelector1

1Li YL, Liu JX. 2018. Structure selector: a web-based software to select and visualize the optimal number of clusters using multiple methods. Mol Ecol Resour. 18:176-177. http://dx.doi.org/10.1111/1755-0998.12719.

This material is available as part of the online article from 10.1590/1519-6984.264369