



# Target Capture Reveals the Complex Origin of Vietnamese Ginseng

Hien Thi Thu Le<sup>1,2\*</sup>, Linh Nhat Nguyen<sup>1</sup>, Hang Le Bich Pham<sup>1</sup>, Hao Thi My Le<sup>3</sup>, Toan Duc Luong<sup>3</sup>, Hue Thi Thu Huynh<sup>1,2</sup>, Van Tuong Nguyen<sup>4</sup>, Hai Van Nong<sup>1,2</sup>, Irene Teixidor-Toneu<sup>5</sup>, Hugo J. De Boer<sup>5</sup> and Vincent Manzanilla<sup>6\*</sup>

## OPEN ACCESS

### Edited by:

Nina Rønsted,  
National Tropical Botanical Garden,  
United States

### Reviewed by:

Angela Jean McDonnell,  
Chicago Botanic Garden,  
United States  
Tobias Andermann,  
University of Gothenburg, Sweden

### \*Correspondence:

Hien Thi Thu Le  
hienlethu@igr.ac.vn  
Vincent Manzanilla  
vincent.manzanilla@gmail.com

### Specialty section:

This article was submitted to  
Plant Systematics and Evolution,  
a section of the journal  
Frontiers in Plant Science

**Received:** 12 November 2021

**Accepted:** 21 June 2022

**Published:** 13 July 2022

### Citation:

Le HTT, Nguyen LN, Pham HLB,  
Le HTM, Luong TD, Huynh HTT,  
Nguyen VT, Nong HV, Teixidor-Toneu I,  
De Boer HJ and Manzanilla V (2022)  
Target Capture Reveals the Complex  
Origin of Vietnamese Ginseng.  
Front. Plant Sci. 13:814178.  
doi: 10.3389/fpls.2022.814178

<sup>1</sup> Institute of Genome Research, Vietnam Academy of Science and Technology, Hanoi, Vietnam, <sup>2</sup> Graduate University of Science and Technology, Vietnam Academy Science and Technology, Hanoi, Vietnam, <sup>3</sup> Soils and Fertilizers Research Institute, Vietnam Academy of Agricultural Sciences, Hanoi, Vietnam, <sup>4</sup> Institute of Biotechnology, Vietnam Academy of Science and Technology, Hanoi, Vietnam, <sup>5</sup> Natural History Museum, University of Oslo, Oslo, Norway, <sup>6</sup> Baseclear BV, Leiden, Netherlands

The global market of the medicinal plant ginseng is worth billions of dollars. Many ginseng species are threatened in the wild and effective sustainable development initiatives are necessary to preserve biodiversity at species and genetic level whilst meeting the demand for medicinal produce. This is also the case of *Panax vietnamensis* Ha & Grushv., an endemic and threatened ginseng species in Vietnam that is locally cultivated at different scales and has been the object of national breeding programs. To investigate the genetic diversity within cultivated and wild populations of *P. vietnamensis* we captured 353 nuclear markers using the Angiosperm-353 probe set. Genetic diversity and population structure were evaluated for 319 individuals of Vietnamese ginseng across its area of distribution and from wild and a varying range of cultivated areas. In total, 319 individuals were sampled. After filtering, 1,181 SNPs were recovered. From the population statistics, we observe high genetic diversity and high genetic flow between populations. This is also supported by the STRUCTURE analysis. The intense gene flow between populations and very low genetic differentiation is observed regardless of the populations' wild or cultivated status. High levels of admixture from two ancestral populations exist in both wild and cultivated samples. The high gene flow between populations can be attributed to ancient and on-going practices of cultivation, which exist in a continuum from understorey, untended breeding to irrigated farm cultivation and to trade and exchange activities. These results highlight the importance of partnering with indigenous peoples and local communities and taking their knowledge into account for biodiversity conservation and sustainable development of plants of high cultural value.

**Keywords:** crop domestication, *Panax*, population genomics, Vietnam, target capture sequencing

## INTRODUCTION

Ginseng has been used in traditional medicine in China for thousands of years (Robbins, 1998). Today, ginseng is used collectively to refer to several plant species, mainly in the Araliaceae genera *Panax* L. and *Eleutherococcus* Maxim. The economic value of ginseng in the global medicinal plant trade is estimated to be in excess of US\$2.1 billion (Baeg and So, 2013). The most commonly used species in the genus are *Panax ginseng* Meyer (Korean ginseng), *P. quinquefolius* L. (American ginseng), *P. notoginseng* (Burkill) F. H. Chen (Chinese ginseng), *P. japonicus* (T. Nees) C. A. Mey. (Japanese ginseng), *P. pseudoginseng* Wall. (Himalayan ginseng) and *P. vietnamensis* Ha and Grushv. (Vietnamese ginseng) (Yang, 2021). The majority of commercialized ginseng material is from cultivation and controlled sustainable wild harvest, whereas material from uncontrolled depletive wild harvesting seems to play a minor and decreasing role (Ichim and de Boer, 2020). So far, academic enquiry has unproportionally focused on *P. ginseng*, *P. quinquefolius*, and *P. notoginseng* (Zhou et al., 2005; Pan et al., 2016; Xia et al., 2016; Fan et al., 2020). Less is known of the genetic and chemical variation of other ginseng species and whether cultivation predominates over wild harvesting (Zhou et al., 2020).

In Vietnam, a number of ginseng species occur in the wild. The most recent studies recognize five taxa including three species, *P. bipinnatifidus* Seem., *P. stipuleanatus* H. T. Tsai and K. M. Feng and *P. vietnamensis* Ha and Grushv. and two varieties, *Panax vietnamensis* var. *fuscidiscus* K. Komatsu, S. Zhu and S. Q. Cai (Zhu et al., 2003; Nguyen, 2005; Phan et al., 2013) and *P. vietnamensis* var. *langbianensis* Duy, V. T. Tran and L. N. Trieu (Nong et al., 2016). *P. pseudoginseng* Wall (Lào Cai, Hà Giang, and Cao Bằng provinces) and *P. ginseng* Meyer (Lào Cai province) are species that were imported for cultivation, but are rarely found in the wild (Nguyen, 2005). The three wild species are referenced in the Vietnam Red Data Book as threatened (Dang, 2006). The locally occurring *P. vietnamensis* has a long history of use, but it was only described as a species new to science in 1985 (Ha and Grushvitzky, 1985). In Vietnamese, this species has the vernacular names “sâm Ngọc Linh” (“sâm” = ginseng), “sâm Việt Nam,” “sâm khu Năm,” “sâm đót trúc,” “củ Ngải rơm,” or “Thuốc giầu”. The complete area of distribution of *P. vietnamensis* is unknown, but in Vietnam it is found wild in the Ngọc Linh mountains straddling the provinces of Quảng Nam and Kon Tum. Vietnamese ginseng had long been used by the local Xơ Đăng ethnic group living at the foot of Mount Ngọc Linh, who used it to treat a variety of illnesses (Dang, 1999). For the last few decades, a growing interest in development of genetic resources and self-sufficiency drove domestic bioprospecting. In 1973, pharmacists Dao Kim Long and Nguyen Chau Giang discovered a *Panax* species locally known as ‘sâm đót trúc’ growing at a height of 1,800 m in the Kon Tum province, and published it as *Panax articulatus* K. L. Dao *nom. inval.* Ha Thi Dung and Grushvitzky later formally described the species as *P. vietnamensis* Ha and Grushv. based on samples collected from this area (Ha and Grushvitzky, 1985). This new ginseng was promoted as a resource

to treat wounded soldiers, miners, and local people. In 1974, a preliminary comparative analysis of the chemical composition of this Vietnamese ginseng with Asian and American ginseng showed that it was rich in ginsenosides (Dao and Nguyen, 1991). Subsequent pharmacological studies corroborated its health benefits and resulted in a wide awareness of the species through mass media in Vietnam. The exploitation, commercialization and use of this Vietnamese ginseng boomed in the 1980s and led to a sharp decrease in availability of material sourced from the wild (Dao and Nguyen, 1991). From 1979, cultivation studies have been conducted in a more systematic manner at Trà Linh Medicinal Plant Center (Quảng Nam province) by propagating plants both asexually and sexually, increasing its cultivation area, and through encouraging ethnic minorities to plant cultivars (Dao and Nguyen, 1991).

Several studies have investigated the relations of *P. vietnamensis* to other ginseng species. In 2001, Komatsu et al. compared the genetic characteristics of *P. vietnamensis* based on 18S and the *matK* gene sequences and showed that it was completely homogeneous between *P. vietnamensis* and *P. quinquefolius* on the 18S gene, but there was a difference of 10 nucleotides on the *matK* gene (Komatsu et al., 2001). In 2003, Zhu et al. described a new sub-species of *P. vietnamensis* distributed in Yunnan, China, and named it *P. vietnamensis* var. *fuscidiscus* K. Komatsu, S. Zhu and S. Q. Cai. This new subspecies is different from *P. vietnamensis* at 4 nucleotides located on the *trnK* gene (1 nucleotide on the 5' extended region, 2 nucleotides on the *matK* gene and 1 nucleotide on the 3' extended region). In 2016, Nong et al. described a new variety *P. vietnamensis* var. *langbianensis* Duy, V. T. Tran and L. N. Trieu from the Lam Vien plateau, and a later ISSR study of 115 individuals from two populations showed it to have high genetic diversity (Le et al., 2019). Recent molecular studies using whole plastome data have shown that *P. vietnamensis* is sister to the widespread *P. japonicus* which ranges from India to Japan (Manzanilla et al., 2018). Other studies have investigated the differentiation between *P. vietnamensis* and *P. ginseng* (Vasyutkina et al., 2018). Studies on *P. vietnamensis* continue to shed new light on intra- and interspecific evolutionary relationships (Nguyen et al., 2020; Vu et al., 2020).

The short documented cultivation history of *P. vietnamensis* shows a patchwork of wild populations, nurseries, woodlot cultivation, and cultivation farms. Given this complex socio-ecological context, this study investigates the population genomics of wild and cultivated *P. vietnamensis*. Specifically, we enquire if wild and cultivated populations are distinct genetically. We part with two contrasting hypotheses: (1) a loss in genetic diversity in cultivars is due domestication processes, or (2) wild and cultivated plants cannot be distinguished genetically due to local plant population management practices whereby local farmers supplement their cultivated stock with wild specimens or favor reproduction between wild and cultivated specimens through woodlot cultivation. Our study has implications for biodiversity conservation best practices that involve local farmers as actors for *P. vietnamensis in-situ* and *ex-situ* conservation.

## MATERIALS AND METHODS

### Study Sites

Previous studies identified *P. vietnamensis* as a narrow endemic occurring only on Ngọc Linh mountain under the canopy of natural forests (Nguyen, 2005; Dang, 2006). It is the ginseng species with the southernmost distribution and has different chemical constituents compared to other *Panax* species (Yamasaki, 2000; Le et al., 2015). The Ngọc Linh mountain area covers 18 communities in five districts including Muồng Hoong, Ngọc Linh, Xốp (Đắk Glei district, Kon Tum province), Đắk Na, Măng Ri, Ngọc Lây, Ngọc Yêu, Tê Xăng, Văn Xuôi (Tu Mơ Rông district, Kon Tum province) and Trà Cang, Trà Don, Trà Don, Trà Leng, Trà Linh, Trà Nam, Trà Tập (Nam Trà My district, Quảng Nam province), Ch'ôm (Tây Giang district, Quảng Nam province), Phước Lộc (Phước Sơn district, Quảng Nam province). In this study, collected sample sites include: Mướng Hoong, Ngọc Linh and Xốp communes (Đắk Glei district, Kon Tum province); Trà Cang, Trà Linh, Trà Nam communes (Nam Trà My district, Quảng Nam province); Ch'ôm (Tây Giang district, Quảng Nam province); Phước Lộc (Phước Sơn district, Quảng Nam province) (Figure 1; Supplementary Tables S1, S2).

### Plant Sampling

*P. vietnamensis* has an erect posture with green or slightly purple colored stems. The tuberous root has a spindle shape with 2.4–4 cm long and diameter of 1.5–2 cm. The outer surface of the roots has a brown or yellowish gray color. The root has a solid body which is hard to break. Leaf and/or root samples of 319 individuals of *P. vietnamensis* were collected in the Ngọc Linh mountains in Quảng Nam and Kon Tum provinces of which 293 came from cultivars (both from hamlets and the Trà Linh Medicinal Plant Center, representing 265 leaf and 28 root samples) and 27 were wild (19 leaf and 8 root samples; Supplementary Tables S1, S2). These samples belonged to 19 populations, including five samples harvested from the wild and 14 from cultivated populations, including the Trà Linh Medicinal Plant Center and the Ngọc Linh Ginseng Center of Nam Trà My district (Supplementary Tables S1, S2). Leaf samples were preserved in silica-gel within 24 h after collecting.

### Library Preparation, Target Enrichment, and Sequencing

Genomic DNA from 319 *P. vietnamensis* samples was extracted using the GeneJET Plant Genomic DNA Purification Kit (ThermoFisher Scientific, USA) following the manufacturer's standard protocol for lignified, polyphenol-rich plant tissues. Extracted total DNA was quantified for integrity and quality by using gel electrophoresis and a NanoDrop 2000 spectrophotometer (ThermoFisher Scientific, USA). Colored and impure samples were cleaned with NEB Monarch Genomic DNA Purification Kit (New England Biolabs, USA). We assessed the DNA integrity on a Fragment Analyzer (Advanced Analytical Technologies, USA) with DNF-488 High Sensitivity Genomic DNA Analysis Kit (Supplementary Table S1). Dual indexed libraries were prepared using the protocol of Meyer and Kircher (2010). We targeted 353 low copy nuclear genes using the

Angiosperm 353 probe set described in Johnson et al. (2018). An advantage of this approach is that the reduced representation library approach of target capture reduces the total sequencing costs of the project, while at the same time generating re-usable data for other studies employing this kit. Moreover, this kit has been used for population level study (Van Andel et al., 2019; Wenzell et al., 2021). We prepared and pooled 319 equimolar libraries in 21 capture reactions with an average 250 ng of input DNA per pool. The RNA probes were hybridized for 24 h before target baiting, and 16 PCR cycles were carried out after enrichment following the MyBaits v.3 manual. The enriched libraries were pooled in equimolar amounts and sequenced on two Illumina HiSeq 3000 lanes (150 bp paired-end).

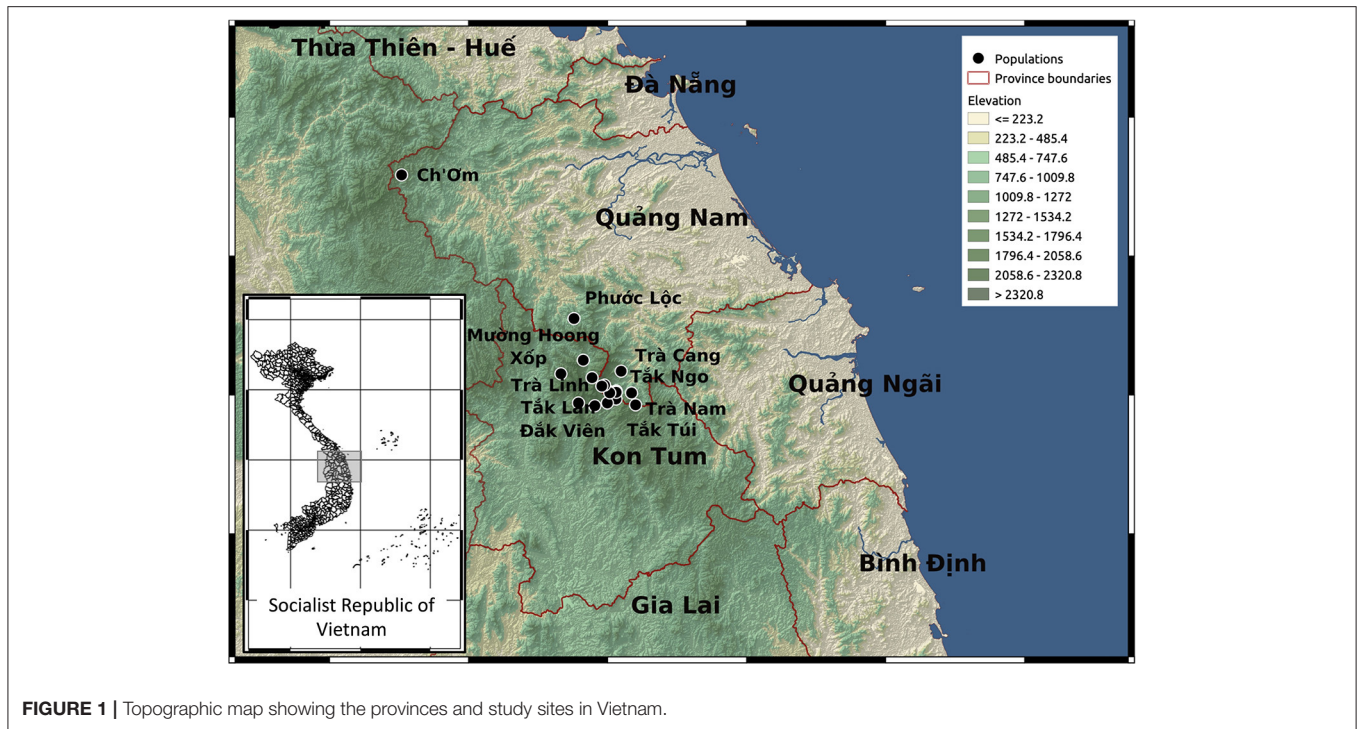
### Bioinformatics Analysis

Raw sequencing reads were checked for quality using FastQC v0.11.9 (Andrews, 2010) and MultiQC v1.9 (Ewels et al., 2016). Trimmomatic v0.36 (Bolger et al., 2014) was used to remove adapter sequences and to filter low quality bases with a sliding window of 10 and mean quality equals 20 and a minimum read length of 20 bp. Trimmed reads were mapped by using the BWA\_MEM algorithm in BWA v0.7.17 (Li and Durbin, 2010). The 353 supercontigs from *Aralia cordata* were used as reference (Johnson et al., 2018). Coverage of each marker for each sample was determined by BEDtools v2.29.2 (Quinlan, 2014). Following the recommendations from Manzanilla et al. (2021), markers with coverage lower than 100X were excluded from the dataset. Duplicate reads were marked and removed from BAM files with the MarkDuplicates program in Picard v2.23.1 (Wysoker et al., 2015). Single nucleotide polymorphism (SNP) calling was performed using DeepVariant v1.1.0 (Ip et al., 2020). We performed additional filtering steps using vcftools v0.1.16 (Danecek et al., 2011) and retained SNPs with mapping quality above 50, with mean depth values ranging from 30 to 500, with minor allele frequency  $\geq 0.10$ , and with missing data  $< 20\%$ . The filtering resulted in a final dataset of 317 markers with total length equal to 570,100 bp for 282 samples. For reproducibility purposes, all the scripts used during the data processing are available on Open Science Framework (<https://osf.io/w9mgc/>) and on GitHub (<https://github.com/vincentmanz/Ginseng>), all the sequencing data have been deposited under the NCBI BioProject accession (BioProject PRJNA788747).

### Population Structure

Population structure was inferred based on 1,181 SNPs from the 317 target capture markers resulting from the filtering described above using STRUCTURE v2.3.4 (Pritchard et al., 2000). Prior analyses, VCF files format was converted with PGDSpider v2.1.1.5 (Lischer and Excoffier, 2012). STRUCTURE analyses were done using the correlated allele frequency method by defining prior population structure or location. Population structure was inferred by estimating the optimum number of clusters ( $K$ ) values ranging from one to 20.

Many of the 317 markers have multiple SNPs and extracting multiple SNPs per marker yields blocks of closely linked SNPs. The STRUCTURE program permits inclusion of weakly linked SNPs with some degree of non-independence. To overcome



**FIGURE 1** | Topographic map showing the provinces and study sites in Vietnam.

the possible effects of linked SNPs, we sub-sampled our data set. From the original data set, we created 10 sets where we randomly choose 1 SNP per marker. To analyze these 10 data sets, we used the same pipeline as for the original data set, using STRUCTURE with 3 replicates and for  $k = 1$  to  $k = 8$  for 1,000,000 generations. Then we compiled results for the 10 data sets to estimate the best number of ancestral populations. We performed a paired  $t$ -test to evaluate if the populations' admixture values from the randomization set were significantly different from those obtained for the full SNPs data set. Runs were set with 1,000,000 iterations and a burn-in of 100,000. Selection of the  $K$  value based on the calculation of delta  $K$  was performed with STRUCTURE HARVESTER v0.6.94 (Earl and VonHoldt, 2012; **Supplementary Figure S1**). CLUMPP v1.1.2 (Jakobsson and Rosenberg, 2007) was used to obtain the average permuted individual and population Q-matrices throughout the three replicates for each  $K$ -value. Those matrices were used as input for distruct v1.1 (Rosenberg, 2004) which was used to obtain bar plots where each individual is represented as a segment divided into  $K$  colors that represent the estimated membership coefficients from each cluster.

Principal component analysis (PCA) of the multilocus genotypes was conducted to visualize potential groupings of the 282 individuals using R package "genstat" (Gogarten et al., 2019). Population statistics were obtained using the R package "hierfstat" (Goudet, 2005). Based on the filtered SNPs, we computed basic statistics, the observed heterozygosity ( $H_o$ ) and mean diversity ( $H_s$ ) within populations. Among populations, we estimated the total genetic diversity ( $D_{ST}$ ), and the corrected  $D_{ST}$  ( $D_{STP}$ ).  $F_{ST}$  and corrected  $F_{ST}$  ( $F_{STP}$ ) were assessed as well as  $F_{IS}$  following Nei per overall loci.

We measured the population differentiation as defined by Jost ( $D_{EST}$ ). Finally, we estimated the overall gene diversity ( $H_T$ ) and the corrected  $H_T$  ( $H_{TP}$ ). We visualized metadata related to the population geographic situation (populations, commune, district, province) and to the cultivation status (wild or cultivated) on the PCAs, and assessed the correlation between the metadata and the principal components using the R package *PCAtools*.

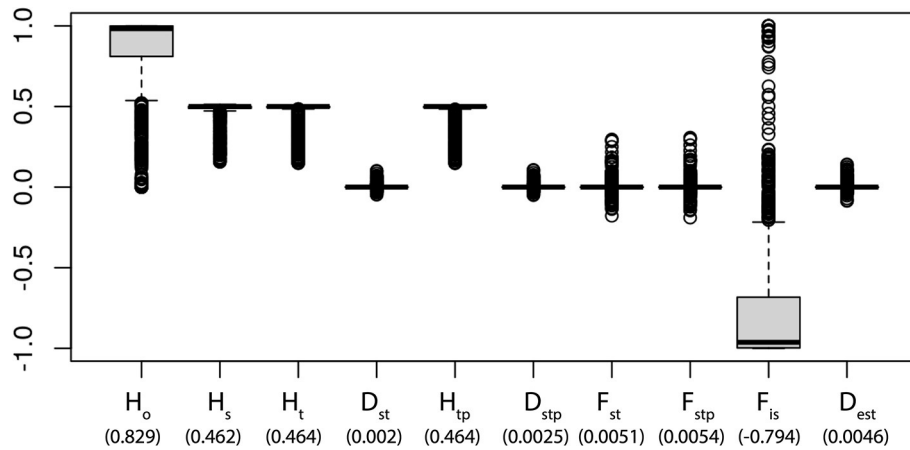
## RESULTS

### Sequencing and SNP Calling

Sequencing of the 353 targeted regions in 319 *P. vietnamensis* samples yielded a total of 1.359 billion reads from the two sequencing lanes combined for all the samples. After trimming, on average, 3.8 millions of reads per sample were retained (**Supplementary Material**) with a duplication level of 83%. Thirty-seven individuals and 36 nuclear markers presented a coverage under the threshold and were discarded (**Supplementary Figure S2**). The final dataset consists of 317 markers with total length equal to 570,100 bp for 282 samples. The SNPs pipeline analysis provided 1,181 SNPs with 8.39% of missing data on average.

### Genetic Diversity

The overall gene diversity ( $H_T = 0.4644$ ) and at the sub-population level ( $H_s = 0.462$ ) are identical, which indicates that all populations mix freely and all the samples are essentially part of a single panmictic population (**Figure 2**). This is corroborated by a pairwise comparison between populations based on Nei's genetic distances (**Supplementary Figure S3**).  $F_{ST}$  measures the



**FIGURE 2** | Summary statistics of genetic variation existing in *Panax vietnamensis* by 1,181 SNPs. The vertical line = median; boxes show quantiles; and points show outliers.  $H_o$ , heterozygosity within population;  $H_s$ , genetic diversity within population;  $H_t$ , overall gene diversity;  $H_{tp}$ , corrected  $H_t$ ;  $D_{st}$ , gene diversity among samples;  $D_{stp}$ , corrected  $D_{st}$ ;  $F_{st}$ , fixation index;  $F_{stp}$ , corrected  $F_{st}$ ;  $F_{is}$ , inbreeding coefficient per overall loci;  $D_{est}$ , measure of population differentiation. Numbers in parentheses below each category indicate the average value.

amount of genetic variance that can be explained by population structure based on Wright's F-statistics. A  $F_{ST}$  value of 0 indicates no differentiation between the subpopulations while a value of 1 indicates complete differentiation. The low genetic structure found shows that there is a relatively high gene flow ( $F_{ST} = 0.0051$ ), which is consistent with the on-going exchange of seeds and plants in the region. The low fixation index within populations ( $F_{IS} = -0.7944$ ) and a lack of extensive differentiation among populations ( $F_{ST}$ ), is concordant with the low population differentiation ( $D_{ST} = 0.0047$ ).

## Population Structure

The STRUCTURE analyses indicate that both wild and cultivated samples derive from the same two ancestral populations (Supplementary Figure S4). The STRUCTURE analysis showed a stabilization in the log likelihood value after  $K = 2$  (Supplementary Figure S1), suggesting  $K = 2$  has the optimal number of subpopulations identified within the genetic data also supported by the  $\Delta K$  method (Supplementary Figures S1, S5). The admixture plot shows a wide range of admixture patterns among the individuals (Figure 3). Chung Tam, Mãng Lùng, Mãng Rương, Mướng Hoang, Tắc Lan, Tắc Ngo, Tắc Ràng, and Trà Linh farm include admixed individuals and individuals from one ancestral population. Eleven populations present only admixed individuals with various proportions. This admixture pattern over the different populations suggests that the 19 populations did not show any distinct population genetic structure based on the dataset (Figure 4).

The principal components account for a large proportion of the multidimensional variability, 59% for the PCA1 and PCA2 (Supplementary Figure S6). We plotted the principal components with the metadata (population, commune, district, province) and to the cultivation status (wild or cultivated) (Figure 5; Supplementary Figures S7–S9). We tested if there was a correlation between the metadata and the PCs but no PCs has a strong correlation with any of the metadata

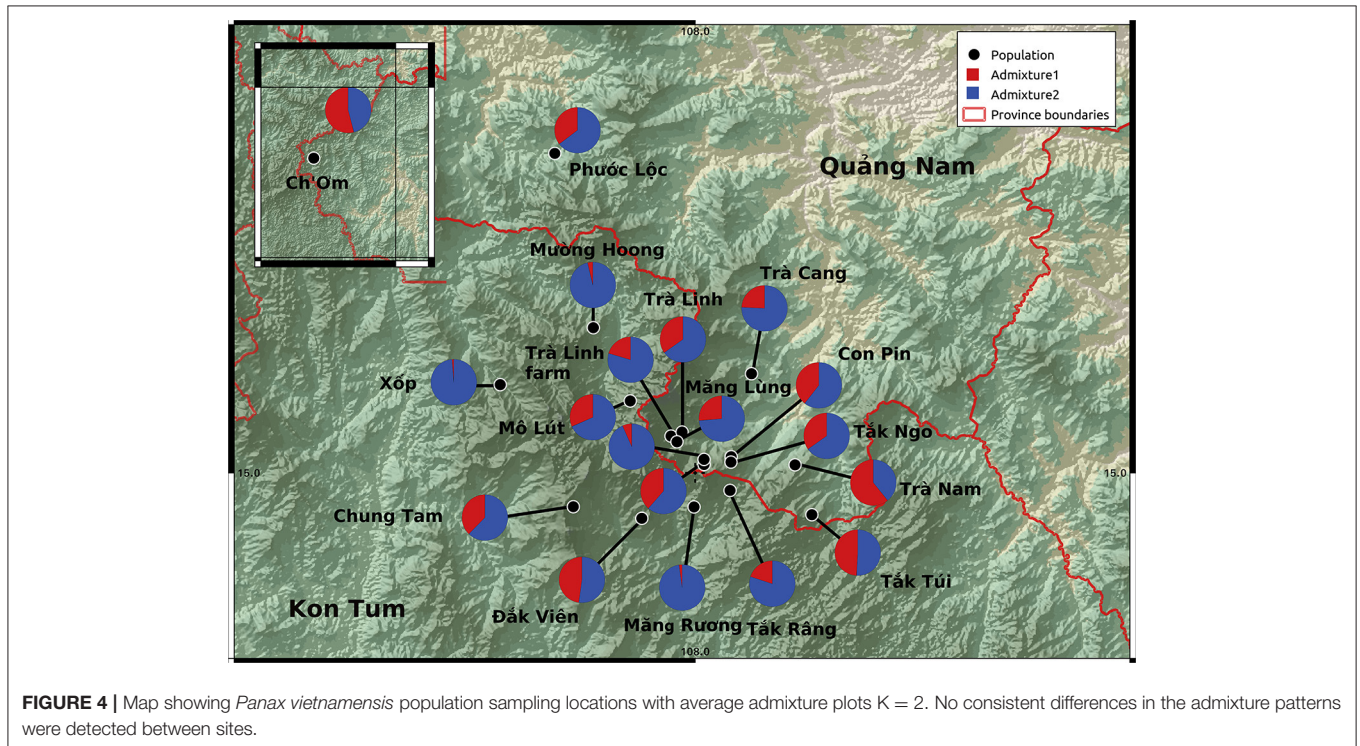
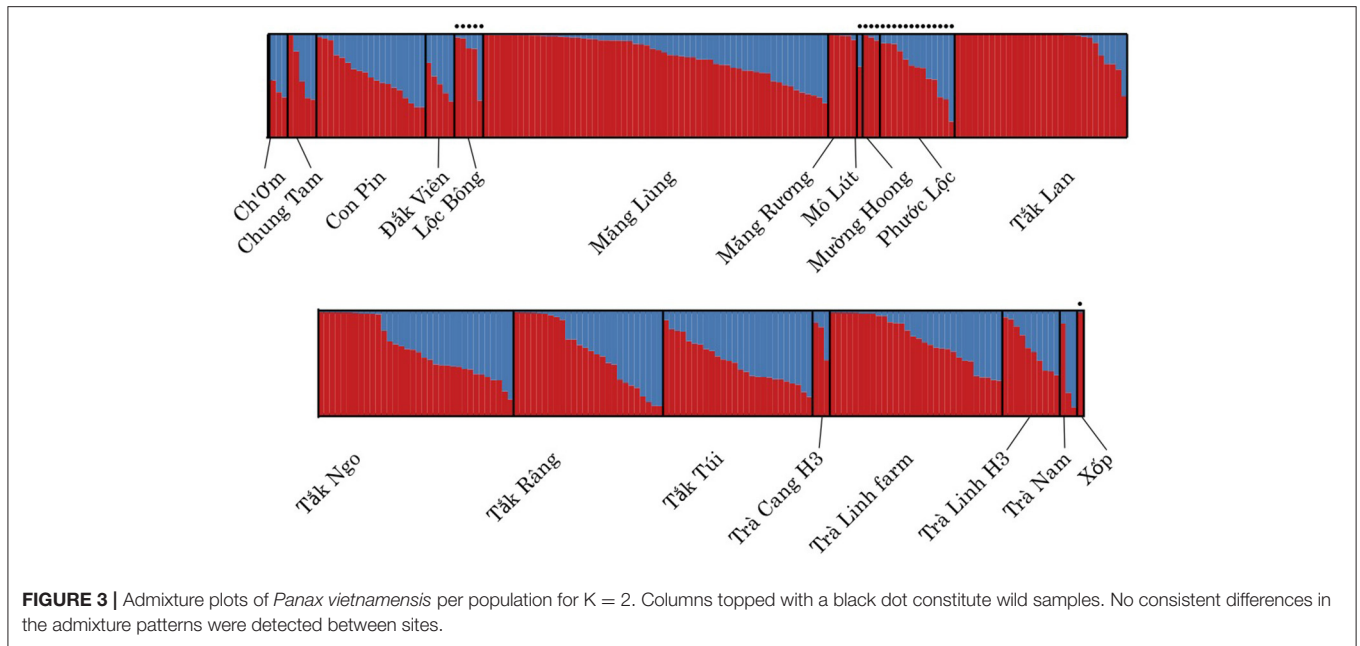
(Supplementary Figure S10), except for the Province and the PCA2.

The sub-sampled data sets show a high standard deviation on the likelihood estimates (Supplementary Material—SNP). The sub-sampling has reduced the number of SNPs and this introduces some stochasticity in the analysis. Despite these variations, all the sets converged to an optimal number of population of two (Supplementary Material and <https://rpubs.com/vincentmanz/A>). We compiled the average admixture value for each population. The results of the sub-sampling sets are highly consistent. The comparison of the averaged admixture proportions in the 19 populations between the sub-sampled and the full data sets, show that the admixture values of the randomized SNPs are not significantly different from the full data set ( $p$ -value = 0.6321). The presence of linked markers in our data set does not affect the STRUCTURE estimates.

## DISCUSSION

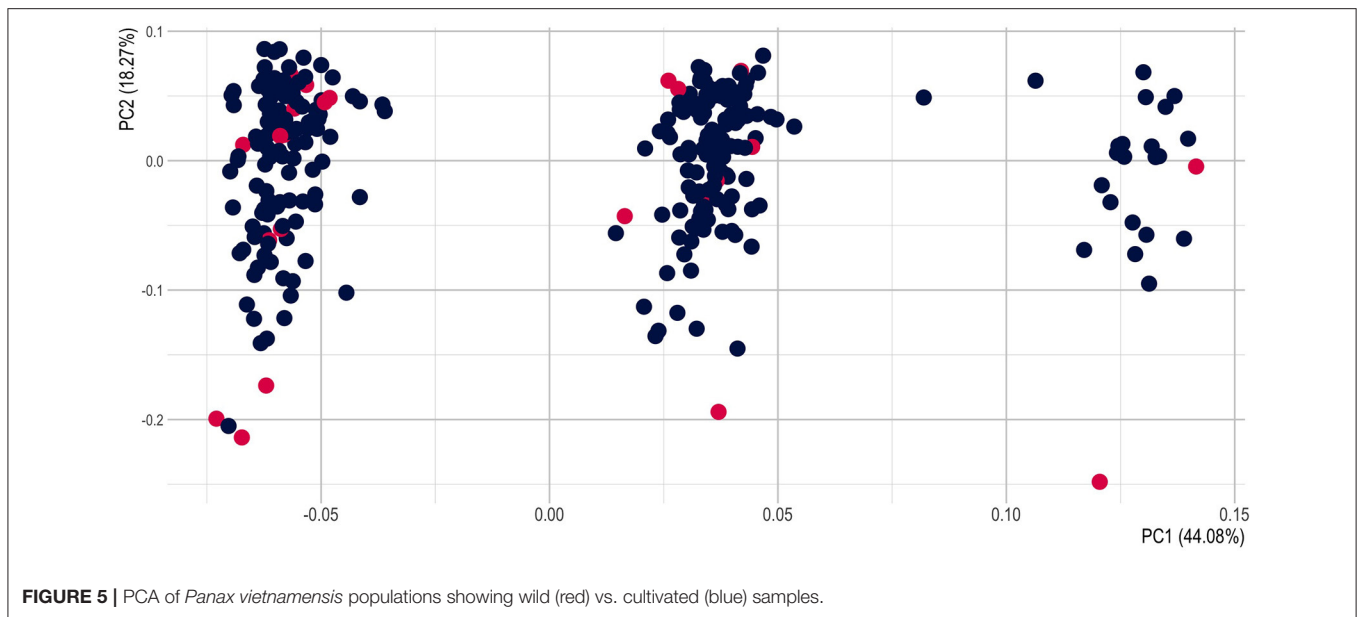
The genetic diversity in the 1,181 SNP dataset is very high, supporting outcrossing of *P. vietnamensis*. The structure and population genomic analyses do not show any segregation between the 19 populations. Furthermore, there is no difference between cultivated individuals and those found in the wild. High genetic diversity and gene flow between *P. vietnamensis* populations have recently been attributed to insect dispersal (Vu et al., 2020). Strong gene flow and insufficient splitting time between cultivated populations has also been observed for *P. notoginseng* (Pan et al., 2016), and was attributed to non-random mating among the individuals of the population. Here we argue that high genetic diversity and the lack of differentiation between populations is due to extensive ancestral woodlot cultivation, management and exchange and trade practices.

Cultivated plants usually present profound morphological differences when compared to their parental populations, along with a reduced genetic diversity due to the bottleneck processes



that occur during domestication (Doebly et al., 2006), yet this is not the case for *P. vietnamensis*. This could be the result of a short domestication history of *P. vietnamensis* combined with panmixia, or very weak genetic differentiation, or a continued cultivation understorey to meet consumer preference for wild materials (Liu et al., 2021). The high heterozygosity can be explained by the artificial gene flow among the population due to the trade between the farmers in the region. It is likely that the extensive pattern of admixture that we observe in both the wild

and cultivated samples is due to anthropic influences, as observed for other plant species of high cultural value (Stefenon et al., 2008; Martínez-Castillo et al., 2014; Wiehle et al., 2014). The little time depth of *P. vietnamensis* domestication history seems to result from a lack of deeper historical documentation rather than being indicative of a recent domestication process. While Vietnamese ginseng was only formally described in 1985, ginseng has been known for centuries in Vietnam. We can assume that ginseng was at least known from the fifteenth



century when ambassadors from Vietnam in Beijing would have learnt about its value (Ock, 2019). The medicinal use of ginseng in neighboring China and more broadly in Southeast Asia predates this, and we can assume that this was also known in Vietnam. Ock proposes that Vietnamese-Korean trade as early as the nineteenth century would have brought Korean ginseng to Vietnam (Ock, 2019). Later, ginseng would have reached Vietnam as imperial gifts or through private trade from China, and from Ryukyu (islands between Japan and Taiwan), which were hubs of international trade (Ock, 2019). It is not possible to assert if local knowledge about the medicinal properties of *P. vietnamensis* predated the trade of ginseng from abroad. If local populations were aware of the properties of ginseng, trade would have facilitated the exploitation of *P. vietnamensis*. Knowledge of the medicinal use of ginseng would have traveled from the coast to the mountains and local populations there, and awareness that similar species occurred naturally in that area, people would have started harvesting and trading these species. Eventually, transplanting young individuals to villages or local forests provided better control of the resource as it matured (Koh, 2019). Koh (2019) concludes that as cultivation of ginseng expanded, wild ginseng harvesting decreased. Yet, it might be that the expansion of cultivation included woodlot or understorey cultivation, and the high degrees of management of wild populations present today could result in the observed lack of differentiation among these. Alternatively, the occurrence of Vietnamese ginseng might not have been known to either local ethnic minorities or people living along the coast. In that case either the coastal trade in ginseng did not extend to the mountain areas or no link was made between traded ginseng material and the local wild populations of Vietnamese ginseng. Nevertheless, local ethnic minorities living in the Annamite mountains might have been using Vietnamese ginseng themselves, and transplanting and mixing populations in the process. For instance, Vietnamese ginseng had long been used

by the local Xơ Đăng ethnic group living at the foot of Mount Ngọc Linh.

Given the migration history of the region, with many ethnic minorities in Vietnam having migrated from or through China, it is possible that ginseng cultivation and use traditions had been picked up and brought by migrants. Despite the fact that *P. vietnamensis* is mostly found in the high mountain area of central Vietnam, it is possible that migrants brought *P. vietnamensis* with them from China and that *P. vietnamensis* is in fact a selected variety of *P. japonica*. For instance, *P. vietnamensis* var. *fuscidiscus* is found in northern Vietnam. Further research is needed to fully understand if people's migratory history shaped the evolution and distribution of *P. vietnamensis*. In the absence of written historical records, a cross-cultural, comparative review of names and uses from all these ethnic minorities (Teixidor-Toneu et al., 2018) could provide further evidence of the history of ginseng presence and cultural use in the region.

The blurred distinction we observe here between wild and cultivated *P. vietnamensis* samples highlights the importance of participatory conservation approaches that fully integrate local knowledge and practice. The continued development of ginseng forest farming is desirable as a sustainable strategy to achieve demands for wild produce while preserving the species wild populations (Stefenon et al., 2008). Strategies that are co-designed between scientists and indigenous peoples and local communities can successfully and ethically conserve culturally important species (Sterling et al., 2017) and are better adapted to local conditions (Maldonado et al., 2016).

## CONCLUSIONS

The population structure of a total of 319 individuals of *Panax vietnamensis* was analyzed based on 1,181 SNP markers. Our study shows that all known *P. vietnamensis* populations, wild and cultivated, are part of one single panmictic population

and cannot be differentiated genetically. Both wild and cultivated samples derive from two ancestral populations and show varying degrees of admixture. Given the continuum of cultivation practices from fertilized and irrigated fields in specialized cultivation centers and small-scale farms to understorey untended transplants that are later harvested, and the exchange and trade activities throughout the region, we attribute the observed high genetic diversity and gene flow to anthropic influence. These results showcase the importance of integrating and understanding local ecological knowledge and practice for the sustainable management of culturally valuable plants.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: National Center for Biotechnology Information (NCBI) BioProject database under Accession Number PRJNA788747.

## AUTHOR CONTRIBUTIONS

HTTL, HVN, HJDB, and VM conceived and initiated the project and designed the experimental study. LNN, HLPB, HTML, TDL, and VTN collected data. LNN and VM analyzed data. VM, LNN, HTTL, and HJDB interpreted results. HD, ITT, HTTL, LNN, and VM wrote the article. All authors read, commented, and approved the final manuscript version.

## FUNDING

This work was performed on the High Performance Computer (HPC) system in the Institute of Genome Research, Vietnam Academy of Science and Technology, Vietnam and on resources provided by UNINETT Sigma2—The National Infrastructure for High Performance Computing and Data Storage in Norway.

## REFERENCES

- Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available online at: [www.bioinformatics.babraham.ac.uk/projects/fastqc](http://www.bioinformatics.babraham.ac.uk/projects/fastqc) (accessed May 22, 2022).
- Baeg, I.-H., and So, S.-H. (2013). The world ginseng market and the ginseng (Korea). *J. Ginseng Res.* 37, 1–7. doi: 10.5142/jgr.2013.37.1
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330
- Dang, N.T. (2006). Red data book 2004 of Vietnam. *Acad. J. Biol.* 28, 1–4. doi: 10.15625/0866-7160/v28n1.823
- Dang, N. P. (1999). *Conservation and Development of Ginseng in the 5th Zone*. Hanoi: Quang Nam Vietnam Quang Nam Health Dept.
- Dao, K. L., and Nguyen, C. G. (1991). Overview of the discovery process of Vietnamese ginseng or “sâm đót trúc” in Ngoc Linh mountain (Kon Tum). *Proc. Hist. Pharm. Sect. 5th Reg. Quang Nam – Da Nang Prov. Pharm. Assoc. Quang Nam – Da Nang*, 138–146.

This project was supported by the Ministry of Science and Technology, Vietnam under the project Transcriptome Sequencing and Analysis of *Panax vietnamensis* Ha and Grushv. (Grant No. 16/2017-HĐ-NVQG) and the European Union's Seventh Framework Program for Research, Technological Development and Demonstration under the Grant Agreement No. 606895 to the FP7-MCA-ITN MedPlant, Phylogenetic exploration of medicinal plant diversity. This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 841127.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.814178/full#supplementary-material>

**Supplementary Figure S1** | The estimated ln probability of data given the K-value. Error bars are standard deviations of 3 replicate runs.

**Supplementary Figure S2** | Heat map of Nei's genetic distances showing the average number of pairwise differences between populations.

**Supplementary Figure S3** | Coverage of the markers per sample.

**Supplementary Figure S4** | Pair plot PCAs.

**Supplementary Figure S5** | PCAs with individuals colored by populations.

**Supplementary Figure S6** | PCAs with individuals colored by HiSeq lane.

**Supplementary Figure S7** | PCAs with individuals colored by province.

**Supplementary Figure S8** | Pearson  $r^2$  correlates of the principal components back to the metadata. The asterisk symbol indicates significant correlation of the metadata on the specific PCA axis.

**Supplementary Figure S9** | Admixture plots for k equal two to eight.

**Supplementary Figure S10** | Admixture plots for  $K = 2$  sorted by samples of wild vs. cultivated origin.

**Supplementary Table S1** | Voucher specimen information.

**Supplementary Table S2** | Populations' information.

- Doebley, J. F., Gaut, B. S., and Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell* 127, 1309–1321. doi: 10.1016/j.cell.2006.12.006
- Earl, D. A., and VonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7
- Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048. doi: 10.1093/bioinformatics/btw354
- Fan, G., Liu, X., Sun, S., Shi, C., Du, X., Han, K., et al. (2020). The chromosome level genome and Genome-Wide Association Study for the agronomic traits of *Panax notoginseng*. *Iscience* 23, 101538. doi: 10.1016/j.isci.2020.101538
- Gogarten, S. M., Sofer, T., Chen, H., Yu, C., Brody, J. A., Thornton, T. A., et al. (2019). Genetic association testing using the GENESIS R/Bioconductor package. *Bioinformatics* 35, 5346–5348. doi: 10.1093/bioinformatics/btz567
- Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* 5, 184–186. doi: 10.1111/j.1471-8286.2004.00828.x



- Ha, T. D. and Grushvitzky, I. V. (1985). A new species of the genus *Panax* (Araliaceae) from Vietnam. *Bot. Zhurnal* 70, 519–522.
- Ichim, M. C., and de Boer, H. J. (2020). A review of authenticity and authentication of commercial ginseng herbal medicines and food supplements. *Front. Pharmacol.* 11, 2185. doi: 10.3389/fphar.2020.612071
- Ip, E. K., Hadinata, C., Ho, J. W., and Giannoulidou, E. (2020). dv-trio: a family-based variant calling pipeline using DeepVariant. *Bioinformatics* 36, 3549–3551. doi: 10.1093/bioinformatics/btaa116
- Jakobsson, M., and Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23, 1801–1806. doi: 10.1093/bioinformatics/btm233
- Johnson, M. G., Pokorny, L., Dodsworth, S., Botigué, L. R., Cowan, R. S., Devault, A., et al. (2018). A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Syst. Biol.* 68, 594–606. doi: 10.1093/sysbio/syy086
- Koh, S. (2019). The history of ginseng cultivation in Orient. *J. Ginseng Cult.* 1, 57–66
- Komatsu, K., Zhu, S., Fushimi, H., Qui, T. K., Cai, S., and Kadota, S. (2001). Phylogenetic analysis based on 18S rRNA gene and matK gene sequences of *Panax vietnamensis* and five related species. *Planta Med.* 67, 461–465. doi: 10.1055/s-2001-15821
- Le, N. T., Nong, V. D., and Tran, V. T. (2019). Genetic diversity of *Panax vietnamensis* var. *langbianensis* populations in Lam Vien plateau-Vietnam detected by inter simple sequence repeat (ISSR) markers. *Vietnam J. Biotechnol.* 17, 651–661. doi: 10.15625/1811-4989/17/4/14720
- Le, T. H. V., Lee, G. J., Vu, H. K. L., Kwon, S. W., Nguyen, N. K., Park, J. H., et al. (2015). Ginseng saponins in different parts of *Panax vietnamensis*. *Chem. Pharm. Bull.* 63, 950–954. doi: 10.1248/cpb.c15-00369
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698
- Lischer, H. E., and Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* 28, 298–299. doi: 10.1093/bioinformatics/btr642
- Liu, H., Burkhardt, E. P., Chen, V. Y. J., and Wei, X. (2021). Promotion of *in situ* forest farmed American ginseng (*Panax quinquefolius* L.) as a sustainable use strategy: opportunities and challenges. *Front. Ecol. Evol.* 9, 141. doi: 10.3389/fevo.2021.652103
- Maldonado, J., Bennett, T. B., Chief, K., Cochran, P., Cozzetto, K., Gough, B., et al. (2016). Engagement with indigenous peoples and honoring traditional knowledge systems. *Clim. Change* 135, 111–126. doi: 10.1007/s10584-015-1535-7
- Manzanilla, V., Kool, A., Nguyen N. L., Nong V. H., Le T. T. H., and de Boer, H. J. (2018). Phylogenomics and barcoding of *Panax*: toward the identification of ginseng species. *BMC Evol. Biol.* 18, 44. doi: 10.1186/s12862-018-1160-y
- Manzanilla, V., Teixidor-Toneu, I., Martin, G., Hollingsworth, P., de Boer, H. J., and Kool, A. (2021). Using target capture to address conservation challenges: population-level tracking of a globally-traded herbal medicine. *Mol. Ecol. Resour.* 22, 212–224. doi: 10.1111/1755-0998.13472
- Martínez-Castillo, J., Camacho-Pérez, L., Villanueva-Viramontes, S., Andueza-Noh, R. H., and Chacón-Sánchez, M. I. (2014). Genetic structure within the Mesoamerican gene pool of wild *Phaseolus lunatus* (Fabaceae) from Mexico as revealed by microsatellite markers: implications for conservation and the domestication of the species. *Am. J. Bot.* 101, 851–864. doi: 10.3732/ajb.1300412
- Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* 2010, pdb-prot5448. doi: 10.1101/pdb.prot5448
- Nguyen, T. H. M., Le, T. S., and Nguyen, T. P. T. (2020). Genetic characteristics of *Panax vietnamensis* Ha and Grushv. populations based on SSR. *Acad. J. Biol.* 42, 11–19. doi: 10.15625/0866-7160/v42n1.13906
- Nguyen, V. T. (2005). The species of *Panax* L. in Vietnam. *J. Med. Mater-Hanoi* 10, 71–76.
- Nong, V. D., Le, N. T., Nguyen, D. C., and Tran, V. T. (2016). A new variety of *Panax* (Araliaceae) from Lam Vien Plateau, Vietnam and its molecular evidence. *Phytotaxa* 277, 12. doi: 10.11646/phytotaxa.277.1.4
- Ock, S. (2019). The history of awareness for ginseng in Vietnam. *J. Ginseng Cult.* 1, 78–92.
- Pan, Y., Wang, X., Sun, G., Li, F., and Gong, X. (2016). Application of RAD sequencing for evaluating the genetic diversity of domesticated *Panax notoginseng* (Araliaceae). *PLoS ONE* 11, e0166419. doi: 10.1371/journal.pone.0166419
- Phan, K. L., Le, T. S., Phan, K. L., Vu, D. D., and Pham, V. T. (2013). “Lai Chau ginseng *Panax vietnamensis* var. *fuscidiscus* K. Komatsu, S. Zhu and S. Q. Cai. Morphology, distribution and conservation status,” in *Proceedings of the 2nd VAST-KAST Workshop on Biodiversity and Bio-Active Compound* (Natural Science and Technology Published House), 65–73.
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1093/genetics/155.2.945
- Quinlan, A. R. (2014). BEDTools: the Swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinforma.* 47, 11–12. doi: 10.1002/0471250953.bi112s47
- Robbins, C. S. (1998). *American Ginseng: The Root of North America's Medicinal Herb Trade*. Washington, DC: TRAFFIC North America. p. 94.
- Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Mol. Ecol. Notes* 4, 137–138. doi: 10.1046/j.1471-8286.2003.00566.x
- Stefanon, V. M., Gailing, O., and Finkeldey, R. (2008). Genetic structure of plantations and the conservation of genetic resources of Brazilian pine (*Araucaria angustifolia*). *For. Ecol. Manag.* 255, 2718–2725. doi: 10.1016/j.foreco.2008.01.036
- Sterling, E. J., Filardi, C., Toomey, A., Sigouin, A., Betley, E., Gazit, N., et al. (2017). Biocultural approaches to well-being and sustainability indicators across scales. *Nat. Ecol. Evol.* 1, 1798–1806. doi: 10.1038/s41559-017-0349-6
- Teixidor-Toneu, I., Jordan, F. M., and Hawkins, J. A. (2018). Comparative phylogenetic methods and the cultural evolution of medicinal plant use. *Nat. Plants* 4, 754–761. doi: 10.1038/s41477-018-0226-6
- Van An del, T., Veltman, M. A., Bertin, A., Maat, H., Polime, T., Lambers, D. H., et al. (2019). Hidden rice diversity in the Guianas. *Front. Plant Sci.* 10, 1161. doi: 10.3389/fpls.2019.01161
- Vasyutkina, E., Adrianova, I. Y., Reunova, G., Nguyen, T., and Zhuravlev, Y. N. (2018). A comparative analysis of genetic variability and differentiation in *Panax vietnamensis* Ha et Grushv. and *P. ginseng* C. A. Meyer using ISSR markers. *Russ. J. Genet.* 54, 262–265. doi: 10.1134/S1022795418020163
- Vu, D. D., Shah, S. N. M., Pham, M. P., 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 Biol.* 20, 358. doi: 10.1186/s12870-020-02571-5
- Wenzell, K. E., McDonnell, A. J., Wickert, N. J., Fant, J. B., and Skogen, K. A. (2021). Incomplete reproductive isolation and low genetic differentiation despite floral divergence across varying geographic scales in *Castilleja*. *Am. J. Bot.* 108, 1270–1288. doi: 10.1002/ajb2.1700
- Wiehle, M., Prinz, K., Kehlenbeck, K., Goenster, S., Mohamed, S. A., Finkeldey, R., et al. (2014). The African baobab (*Adansonia digitata*, Malvaceae): genetic resources in neglected populations of the Nuba Mountains, Sudan. *Am. J. Bot.* 101, 1498–1507. doi: 10.3732/ajb.1400198
- Wysoker, A., Tibbetts, K., and Fennell, T. (2015). *Picard Tools*. Broad Institute.
- Xia, P., Guo, H., Zhang, Y., Deyholos, M. K., Peng, L., Jia, Y., et al. (2016). Wild *Panax vietnamensis* and *Panax stipuleanatus* markedly increase the genetic diversity of *Panax notoginseng* (Araliaceae) revealed by start codon targeted (SCoT) markers and ITS DNA barcode. *Biochem. Syst. Ecol.* 66, 37–42. doi: 10.1016/j.bse.2016.03.007
- Yamasaki, K. (2000). Bioactive saponins in Vietnamese ginseng, *Panax vietnamensis*. *Pharm. Biol.* 38, 16–24. doi: 10.1076/phbi.38.6.16.5956
- Yang, M. S. (2021). “Chinese ginseng,” in *Nutraceuticals* (Elsevier), 853–864.
- Zhou, M., Yang, G., Sun, G., Guo, Z., Gong, X., and Pan, Y. (2020). Resolving complicated relationships of the *Panax bipinnatifidus* complex in southwestern China by RAD-seq data. *Mol. Phylogenet. Evol.* 149, 106851. doi: 10.1016/j.ympev.2020.106851

- Zhou, S. L., Xiong, G. M., Li, Z. Y., and Wen, J. (2005). Loss of genetic diversity of domesticated *Panax notoginseng* F. H. Chen as evidenced by ITS sequence and AFLP polymorphism: A comparative study with *P. stipuleanatus* H. Tsai et K. M. Feng. *J. Integr. Plant Biol.* 47, 107–115. doi: 10.1111/j.1744-7909.2005.00013.x
- Zhu, S., Fushimi, H., Cai, S., Chen, H., and Komatsu, K. (2003). A new variety of the genus *Panax* from southern Yunnan, China and its nucleotide sequences of 18S ribosomal RNA gene and matK gene. *J. Jpn. Bot.* 78, 86–94.

**Conflict of Interest:** VM was employed by Baseclear BV.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Le, Nguyen, Pham, Le, Luong, Huynh, Nguyen, Nong, Teixidor-Toneu, De Boer and Manzanilla. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.