Phylogeography and genetic structure of Papever bracteatum populations in Iran based on genotyping-by-sequencing (GBS)

Razieh Rahmati
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK)
Zahra Nemati
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK)
Mohammad Reza Naghavi
University of Tehran
Simon Pfanzelt
Botanical Garden München-Nymphenburg
Amir Rahimi
Urmia University
Ali Ghaderi Kanzagh
Islamic Azad University
Frank R. Blattner (blattner@ipk-gatersleben.de)
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK)

Article

Keywords:

Posted Date: August 1st, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3155533/v1

Abstract

Papaver bracteatum, known for its high thebaine content and absence of morphine, has emerged as a promising alternative to opium poppy for codeine production. In this study, our objective was to create a diverse panel representing the natural variation of this species in Iran. To achieve this, we employed genotyping-by-sequencing (GBS) to obtain genome-wide distributed single-nucleotide polymorphisms (SNPs) for phylogeographic analysis, population structure assessment, and evaluation of genetic diversity within *P. bracteatum* populations. A total of 244 *P. bracteatum* individuals from 13 distinct populations formed seven genetic groups, along with one highly admixed population. We observed a split between the populations inhabiting the Alborz and Zagros Mountain systems. In between these mountain ranges the population of Kachal Mangan exhibited a high degree of genetic admixture. Our results suggest that habitat fragmentation, climate change, limited seed dispersal, and human pressure on the species' habitats are potential factors contributing to the genetic isolation of *P. bracteatum* populations. Our findings underscore the urgency of implementing conservation measures to safeguard the wild populations as important genetic resources for future breeding approaches in this medicinally important species.

Introduction

The genus *Papaver* has about 70–100 species, which are mainly found in temperate regions of the Northern Hemisphere[1,2]. It contains numerous annual and biennial species as well as a number of perennials. This genus is of commercial interest because of the medicinally important alkaloids present in several *Papaver* groups[3]. The economically most important alkaloids are morphine, codeine and thebaine. These alkaloids are known to occur in large quantity in the opium poppy, *Papaver somniferum* L.[4–7]. Among these alkaloids, thebaine cannot be converted into illicit drugs except through a series of complicated and inefficient chemical processes, which are assumed to be too complicated to be used to produce illegal drugs[8–10]. In contrast to opium poppy, where morphine is the main alkaloid accumulated in capsules, thebaine is the dominant alkaloid that is synthetized in different tissues and occurs in higher level in roots and capsules in Persian poppy, *Papaver bracteatum* Lindl.[11–14]. Therefore, *P. bracteatum* is a good substitute for farmers to replace opium poppy by the legally much less problematic Persian poppy, as an increasing worldwide demand of its alkaloids arose over the last decades to be used by the pharmaceutical industry as an alternative to traditional opium poppy[15].

Papaver bracteatum ($2n = 2x = 14$) is a perennial species that along with the polyploid species *P. orientale* L. ($2n = 4x = 28$) and *P. pseudo-orientale* (Fedde) Medw. ($2n = 6x = 42$) constitute *Papaver* section *Oxytona* Bernh.[16,17]. *Papaver bracteatum* is distributed at elevations between 1500 m and 2500 m with its main occurrence areas in the Zagros Mountains in west and northwestern Iran, the Alborz Mountains in the north of Iran, and the southern parts of the Caucasus in northwestern Iran, northeastern Turkey, Georgia and Azerbaijan[8,14,17]. In the north its distribution area is overlapping with the two polyploid species of the section and hybrids seem to occur[16].

The knowledge of genetic variation and population structure of the species can contribute to germplasm conservation and breeding of the plant. Despite medicinal importance of *P. bracteatum*, there is no comprehensive study inferring genetic diversity and genetic structure of *P. bracteatum* populations, mainly due to lack of enough DNA polymorphisms found in the up to now conducted studies[18,19]. Genotyping-by-sequencing (GBS) is one of the most efficient approaches for the study of populations and closely related species complexes, as it detects genome-wide single-nucleotide polymorphisms (SNPs) and allows their use for genotyping[20–22]. Here, we employed genome-wide distributed SNP markers obtained using GBS for individuals from 13 natural populations of *P. bracteatum* in Iran (Fig. 1) to arrive at a diversity panel that represent the natural diversity of the species in Iran. The specific goals of this study are: (i) to infer phylogeographic relationships, (ii) to resolve population structure into different gene pools, (iii) to analyze gene flow and admixture between intraspecific gene pools, and (iv) to test the impact of environment on gene flow for *P. bracteatum* populations. From this we want to arrive at a better understanding of the genetic characteristics of *P. bracteatum*, which could underpin future utilization and protection of Iranian germplasm resources of Persian poppy.

Results

Ploidy level estimations

Genome size estimations against a tomato reference resulted in a 2C genome size of 6.12 pg (± 0.14) for *P. bracteatum* and 12.97 pg (± 0.04) for the measured *P. orientale* accession. Thus, all of our analyzed populations of Persian poppy were assumed to be diploid and *P. orientale* from Ahar tetraploid.

Genotyping-by-sequencing

After removing the barcodes and quality filtering, our GBS data consisting of *P. bracteatum* together with *P. orientale* resulted in on average 5764 loci (min: 3400, max: 5923). For the dataset consisting of *P. bracteatum* individuals only, we obtained on average 11,690 loci (min:9152, max: 11,978) per individual.

Population structure and genetic diversity analyses

The Bayesian population assignment analyses using 6027 unlinked high-quality SNPs in *P. bracteatum* provide information to elucidate the allelic patterns of the populations and admixture patterns between individuals (Fig. 2). At K = 2 populations are divided into two subgroups representing Alborz Mts. vs. Zagros Mts. populations with admixture signals for the populations in between the two mountain systems (Kachal Mangan, Somagh, Tarom Sofla). At K = 3 the northern (Pasve, Qale Samur) and southern (Belch Sur, Hajmne, Harzakhani) Zagros Mts. populations split. At K = 4 the geographically intermediate populations with an admixture signal in the analysis before, now together with Kaftarchak, form a group of their own. All Kachal Mangan individuals show strong admixture, a signal that is maintained through all steps up to K = 7. With further increasing K the Alborz Mts. populations split, first the easternmost group, then the populations in the west of the Alborz Mts. system. Thus, at K = 7 the 13 analyzed populations are grouped in seven allelic groups plus one with a strong admixture signal. The geographic position of the latter (Kachal Mangan) is the easternmost in the Zagros Mts. and it occupies a kind of intermediate position between both mountain systems (Fig. 1), which correlates with its mixed allelic constitution in this analysis: Alleles are mainly shared with the Zagros Mts. populations but also with Tarom Sofla and, to a small extend, Somagh.

In our PCA the first two principal components explained 26.34% of the total molecular variation. The results (Fig. 3) are consistent with the admixture plots at K = 5 by separating the populations along their geographical origins and placing Kachal Mangan individuals in a central position.

F_{ST} values were used to quantify the level of genetic structure among populations. The pairwise F_{ST} value-derived differentiation between 13 populations of *P. bracteatum* (Table 1) ranges from very little between populations within the Zagros Mts., i.e. Hajmne vs. Belch Sur and Hazarkhani (0.026), to very strong between the geographically most distant populations, i.e. Pasve vs. Plour and Larijan (0.26). The average F_{ST} value is equal to 0.16, indicating high differentiation among populations.

Table 1
Weighted pairwise comparison of genetic differentiation among 13 populations of *P. bracteatum*

	Bal	Bsr	Hjn	Hkn	KMn	Kch	Ljn	Psv	Plr	QSr	Sgh	TSl	Ysh
Bal	*0*												
BSr	*0.176*	*0*											
Hjn	*0.161*	*0.026*	*0*										
Hkn	*0.155*	*0.047*	*0.026*	*0*									
KMn	*0.139*	*0.071*	*0.065*	*0.061*	*0*								
Kch	*0.154*	*0.196*	*0.175*	*0.166*	*0.155*	*0*							
Ljn	*0.154*	*0.223*	*0.200*	*0.195*	*0.183*	*0.200*	*0*						
Psv	*0.218*	*0.170*	*0.156*	*0.151*	*0.118*	*0.251*	*0.260*	*0*					
Plr	*0.154*	*0.218*	*0.195*	*0.190*	*0.183*	*0.202*	*0.038*	*0.260*	*0*				
QSr	*0.195*	*0.138*	*0.132*	*0.125*	*0.096*	*0.223*	*0.239*	*0.072*	*0.236*	*0*			
Sgh	*0.181*	*0.198*	*0.180*	*0.169*	*0.153*	*0.158*	*0.238*	*0.253*	*0.234*	*0.217*	*0*		
TSl	*0.168*	*0.169*	*0.154*	*0.149*	*0.118*	*0.162*	*0.218*	*0.212*	*0.216*	*0.188*	*0.148*	*0*	
Ysh	*.0.034*	*0.168*	*0.154*	*0.149*	*0.133*	*0.147*	*0.147*	*0.212*	*0.148*	*0.189*	*0.170*	*0.161*	*0*

Since our analyses show individuals from Kachal Mangan as admixed (Fig. 2), another pairwise F_{ST} comparison was conducted to measure differentiation between these individuals and populations from other subgroups. The pairwise F_{ST} reveals lower differentiation of Kachal Mangan towards populations Tarom Sofla (0.118), Belch Sur/Hajmne/Hazarkhani (0.056) and Pasve/Qale Samur (0.089) in comparison to the remaining combinations (> 0.15; Table 2).

Table 2
Pairwise F_{ST} values between Kachal Mangan and other population groups according to the clustering in the admixture analysis

KMn						
Plr/Ljn	**Ysh/Bal**	**Kch**	**Sgh**	**TSl**	**BSr/Hjn/Hkn**	**Psv/QSr**
0.185	0.151	0.155	0.153	0.118	0.056	0.089

The range of estimated heterozygosity within and between populations provides insights into the genetic diversity and population structure of different populations (Fig. 4). Heterozygosity refers to the presence of different alleles at a specific genetic locus, indicating genetic variation within a population. Here, the boxplot analysis reveals that the populations of Somaq from western parts of the Alborz Mts, as well as Youshand Larijan in the central Alborz Mts, exhibit higher levels of heterozygosity compared to other populations, including even the admixed population of Kachal Mangan from Kurdistan province in the Zagros Mts.

Phylogenetic inferences

The maximum parsimony (MP) analysis of the dataset including *P. orientale* as outgroup resulted in 80 equally parsimonious trees of 28,095 steps length with a consistency index (CI) of 0.30 and a retention index (RI) of 0.67. The MP analysis excluding *P. orientale* resulted in 120 equally parsimonious trees with a length of 21,890 steps (CI = 0.20, RI = 0.56).

Schematic representations of the strict consensus trees of both analyses are provided in Fig. 4, the detailed consensus trees as Supplementary Figures S2 and S3.

The initial analysis, where we included *P. orientale* individuals from Ahar, showed that these tetraploids have the highest similarity within *P. bracteatum* in the geographically closest populations in the north of the Zagros Mts. This results in the Pasve/Qale Samur populations occupying the position as sister group to all other *P. bracteatum* individuals in the MP tree (Fig. 5a). However, with our sampling design including *P. orientale* from just one area and not separating homoeologous alleles in the tetraploid, we cannot conclude that this is correct. This means that a biogeographical interpretation of the trees, which would imply that *P. bracteatum* originated in the Zagros Mts. and afterwards colonized the Alborz Mts. from west to east, is not possible based on this dataset. Thus, we analyzed the dataset in addition without the *P. orientale* individuals. This MP analysis was essentially an unrooted analysis where the tree was afterwards drawn to make it easily comparable with the result of the analysis including *P. orientale*. With midpoint rooting or putting the root according to the result of the population assignment analyses, the Zagros and Alborz Mts. groups would be sisters with the Kachal Mangan population at an intermediate position. This topology is indicated by an asterisk at the root position in Fig. 5b.

Within *P. bracteatum* we can detect in both analyses groups, which are compatible with the population assignment analyses (Fig. 2). Geographically close populations group also together in the phylogenetic trees, and individuals collected from different populations but grouped into single units at K = 7 in the population assignment analysis, like Pasve/Qale Samur or Balade/Youssh, result also in the phylogenetic trees intermingled and not separated according to their collection sites (Fig. 4, Supplementary Fig. S3). This indicates close relationships of populations from within the two mountain systems, and the Kachal Mangan population in the east of the Zagros Mts. combines alleles from both main groups and accordingly occupies an intermediate position. The hybrid nature of this population (Fig. 2) results in very low bootstrap support values for this group and also for the branches around their individual positions in the trees. Within the Alborz Mts. populations a geographic progression from west to east can be derived from this dataset (Figs. 1 and 4).

The main differences of a maximum-likelihood analysis (ML) of *P. bracteatum* towards the results of the MP analysis concerns (i) the position of the Kachal Mangan individuals, which in ML (Supplementary Fig. S4) form a grade at the base of the Alborz Mts. populations, while in MP Kachal Mangan individuals form grades at the base of the Alborz and Zagros Mts. populations (Fig. 5). In both cases, bootstrap support values for the Kachal Mangan groups and the clades around are rather low, which is indicative of the hybrid nature of this population. (ii) The second difference can be found within the group of central Alborz Mts. populations where Balade/Youssh and Larijan/Plour are sister groups in ML instead of the latter grouping within a grade of Balade/Youssh individuals in MP. Still, relationships among the populations reflect also in ML clearly the results of the population structure analyses.

Ecoclimatic niche modeling

The best potential distribution-envelop model was obtained using the following settings: regularization value 1 and set of tested feature classes LQH, respectively. Figure 6 shows the resulting map of our ecoclimatic niche modeling approach. Suitable ecoclimatic conditions were found in the mountain ranges from the eastern part of the Taurus in Turkey, the Armenian Highlands, the northern part of Zagros Mts., Alborz to Zagros Mts. and extended eastwards to the mountain systems of Central Asia (Hindukush and Himalayas). Thus, our prediction of areas with possible suitable climate conditions for *P. bracteatum* are larger than the realized distribution area of the species, with absence or very rare occurrences of the species in Central Asia and southwestern Turkey.

Discussion

Papaver bracteatum is known as a potential source of raw material for the production of codeine, without the involvement of morphine. Despite pharmaceutical importance of *P. bracteatum*, genetic variation and genetic structure of natural populations are still unclear. Hence, to investigate phylogeography and population structure of the species in Iran, we here used genome-

wide distributed SNP markers obtained with high-throughput genotyping-by-sequencing (GBS) on accessions belonging to 13 natural populations of *P. bracteatum* and one population of its closely related species *P. orientale*.

Based on our result, the 13 populations were clustered into seven subgroups: (i) Larijan and Plour, (ii) Balade and Yoush, both subgroups occurring in the central Alborz Mts., (iii) Kaftarchack, (iv) Somagh and (v) Tarom Sofla, all from western parts of the Alborz Mts., (vi) Belch Sur, Hajmne and Hazarkhani from Kurdistan province in the Zagros Mts., and (vii) Pasve and Qale Samur from Azerbaijan province in the northern Zagros Mts. The successive subdivisions of these populations in population-assignment groups from K = 2 to K = 7 (Fig. 2) is reflected in the clades that were obtained by phylogenetic analyses (Fig. 4) when midpoint rooting of the tree is applied. We found a clear geographic pattern within Iranian *P. bracteatum* populations with close relationships for the Zagros Mts. and the Alborz Mts. populations, respectively. Within the Alborz Mts. we can deduce also a colonization pattern from west to east, as the easternmost populations are nested within a grade of populations from further west. For the materials from within the Zagros Mts. a colonization pattern is not obvious. All these populations show at K = 7 in our population assignment analysis very low signals of admixture.

In contrast to the populations mentioned above, the Kachal Mangan population from the east of Zagros Mts. that geographically is in an intermediate position between the western and eastern populations (Fig. 1), was obtained as admixed in all analyses from K = 4 to K = 7. Also in phylogenetic analyses, this population showed all signs of pronounced admixture, with (i) an intermediate position between western and eastern groups in the trees, (ii) low bootstrap support values in all analyses for its position but also for the branches that connect its subgroups to other populations, and (iii) non-monophyly for the Kachal Mangan individuals forming subclades that are sister to either Zagros or Alborz populations, depending on the allelic composition of the individuals. Analysis of F_{ST} values (Tables 1 and 2) and PCA (Fig. 3) support this central position of the Kachal Mangan individuals with alleles shared with its neighboring populations. This pattern could either be interpreted as the Kachal Mangan area being the center of origin of the species and still maintaining a diverse set of ancient alleles, which got partly lost when the species spread east- and westwards out of this area. Alternatively, Kachal Mangan might be a meeting point where eastern and western populations came into contact, resulting in introgression and accordingly admixed allelic signals for its individuals. Genetic distances and heterozygosity within the Kachal Mangan population are not larger than in the other populations around, which might indicate that it is not an ancient population but a Zagros population that was introgressed by Alborz Mts.-derived gene flow.

In contrast to our result, Qaderi et al.[18] using inter simple-sequence repeats (ISSR) and start-codon targeted (SCoT) markers clustered six populations of *P. bracteatum* into three distinct subgroups without any signal of admixture. Furthermore, we here observed a relatively high fixation index (F_{ST} = 0.16) as a measure of genetic differentiation among *P. bracteatum* populations, which is still low in comparison with previous studies of Persian poppy presented so far (Hadipour et al.[19]: G_{ST} = 0.52; Qaderi et al.[18]: G_{ST} = 0.53). This suggests a relatively low level of gene flow and clear genetic differentiation among populations from different geographical regions. The resolving power of markers in the inference of phylogenetic relatedness and evolutionary events is mainly determined by the level of polymorphism detectable. Therefore, the higher differentiation observed among populations in the previous studies can be attributed to the markers used, where fast-evolving marker regions like microsatellite-targeting ISSRs show probably less shared alleles in comparison to more slowly evolving ones like GBS, where SNPs are mostly detected in the under-methylated (i.e. functional) genomic parts. However, GBS arrives at a large number of genome-wide distributed SNP markers within *P. bracteatum*, which might not be comparable to the limited number of informative genomic areas probed by diverse methods like ISSR, amplified fragment length polymorphisms (AFLP) and SCoT in previous studies[18,19].

In *P. bracteatum* different mechanisms influence gene flow and population connectivity. On the one hand, these are the abilities of the plants to disperse pollen and seeds, which is influenced not only by the kind of pollinators and morphological seed structures that support seed transportation but also by geographic settings in the species' occurrence areas. The widely open *Papaver* flowers are unspecialized regarding their pollinators[23], which often are beetles, flies and wild bees. This means that pollen is mostly transported in a range of a few meters up to a few hundred meters[24]. Also, seed dispersal in *Papaver* is

mostly restricted to distances of centimeters to few meters[25], but might increase when seeds are transported by water after shedding from the capsule. However, obligate outcrossing of *P. bracteatum*[17] might restrict rare water- or wind-mediated long-distance dispersals, as single plants that are established away from the nearest population will not set seeds due to their inability of selfing. In addition, the adaptation of *P. bracteatum* to mountain habitats might have resulted in population contacts in lowlands during glacial cold cycles in the Pleistocene but prevents the colonization of intervening lowlands during inter-glacial periods and under current climate conditions. Thus, the differentiation we see today might be influenced by population contacts during glacial maxima, resulting in gene flow and admixture where the different genotypes had met. Under extant climate conditions *P. bracteatum* populations might be much clearer isolated, a tendency that will increase with climate warming. This notion is supported by the predicted potential distribution area of the species (Fig. 6) that is larger than the currently realized distribution area. This could indicate relatively low colonization abilities of the species, resulting in slow migration into suitable areas if further away from extant populations. In addition, human impact on the habitats of the already fragmented populations of the species will increase pressure on the species[26–28]. The natural habitat of *P. bracteatum* are rocky slopes and grassy meadows at elevations mainly between 2,000–2,500 m[13]. These areas are under increasing pressure from livestock grazing in Iran. Overgrazing not only affects the *P. bracteatum* populations directly but entire natural habitats, as it significantly reduced the vegetation cover, which leads to the decline in pollinator activity in the areas. Consequently, the chances of inter-population pollination get more limited, and isolated, genetically more homogeneous populations are expected.

Due to low gene flow and high genetic variation between *P. bracteatum* populations, future conservation programs for this species should include representative populations of the diverse genotype groups we describe here to arrive at the highest genetic variation for both seed and germplasm conservation. We show that our GBS-based approach is able to infer the phylogeography and the population structure of *P. bracteatum*. We studied populations from the main distribution area of the species only. Thus, it would be of interest to include now also populations from the more peripheral occurrence areas of the species in Turkey, Armenia and also from highlands east of the Alborz Mts. This would add to the general understanding of the population dynamics and genotypic diversity of *P. bracteatum* and improve knowledge of colonization patterns in the Irano-Turanian floristic region[29]. Our study based on a large number of genome-wide distributed SNP markers proved to reliably distinguish the genetic variations in highly homogenous plant populations of *P. bracteatum* and accessions could be utilized for future investigations and improvement programs.

Conclusions

This study provides comprehensive insights into the genetic diversity and unique population dynamics of *P. bracteatum* in Iran. A high-resolution phylogeny obtained by using a large number of GBS-derived SNP markers showed that Iranian *P. bracteatum* populations consist of seven genetic groups with a major split between the populations of the Alborz and the Zagros Mt. systems. Within these areas, five and two subgroups exist, respectively. One population from the eastern Zagros Mts. showed clear signals of introgression from its neighboring Zagros and Alborz populations, while otherwise, the populations seem clearly separated. The low dispersal abilities of the species together with increasing habitat fragmentation due to climate change and human impact on the populations might warrant in and ex-situ conservation measures for this valuable medicinal plant.

Materials and methods

Plant materials, DNA extraction and genome size measurement

We included 264 individuals consisting of 244 *P. bracteatum* and 20 *P. orientale* samples in the phylogenetic analyses (Supplementary Table S1). The materials were collected in accord with the Iranian regulations regarding plant genetic resources and determined by A.G.K. All materials were included in the ex-situ collection and herbarium of the Azad University of Tehran (Tehran, Iran) under the collection numbers given in Supplementary Table S1. *Papaver bracteatum* accessions were

assigned to thirteen natural populations. The sampling sites (Fig. 1) were in the central Alborz Mts. of Balade (Bal), Youssh (Ysh), Larijan (Ljn) and Plour (Plr), west of the Alborz Mts. with the regions Kaftarchak (Kck), Somagh (Sgh), Tarom Sofla (TSl), and the Zagros Mts. regions of Kachal Mangan (KMn), Pasve (Psv), Qale Samur (QSr), Belch Sur (BSr), Hajmne (Hjn) and Hazarkhani (Hkn). The *P. orientale* population was collected close to Ahar in northwestern Iran. Leaves were harvested from the plants and immediately dried and stored in silica gel. Genomic DNA was extracted from the dried leaves using the DNeasy Plant mini kit (QIAGEN) according to the instructions of the manufacturer.

To confirm the uniformity of ploidy level of the analyzed accessions, genome sizes of one individual of each population of *P. bracteatum* and three individuals of *P. orientale* were measured based on fresh leaves and using propidium iodide (PI) as a stain in flow cytometry with a Cyflow Space (Sysmex Partec) flow cytometer, following the procedure described by Jakob et al.[30]. The tomato cultivar 'Stupicke' was used as the size standard (2C DNA content = 1.96 pg).

Genotyping-by-sequencing

To obtain genome-wide SNPs, GBS analyses[20] were used for 244 *P. bracteatum* and 20 *P. orientale* individuals. For assessment of reproducibility, four replicates for one sample were included in GBS. To prepare the library for each sample, 200 ng of genomic DNA was digested using two restriction enzymes, *Pst*I-HF (NEB) and *Msp*I (NEB). Library preparation, individual barcoding and single-end sequencing on the Illumina NovaSeq were performed following Wendler et al.[31]. Library preparation and sequencing were conducted at the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany.

The quality of raw fastq data was assessed using the fastqc toolkit[32]. Adapter trimming of GBS sequence reads was conducted with Cutadap[33] within ipyrad v.0.9.56[34]. GBS reads were clustered in ipyrad. We conducted two runs of ipyrad resulting in two GBS alignments concatenating all filtered loci: (i) one with all individuals, including also *P. orientale*, and (ii) excluding the latter and restricting the dataset to *P. bracteatum* individuals. We set the minimal number of samples to possess a certain locus to 90% of the individuals and the clustering threshold of reads within and between individuals to 0.85 and 0.90 in the first and second datasets, respectively. The default settings were used for the other parameters. For GBS-based phylogenetic analyses, both datasets were used whereas for population assignment analyses and principal component analysis (PCA), we only used the second dataset.

Population structure and genetic diversity analyses

Single-nucleotide polymorphisms obtained by the ipyrad pipeline were subjected to quality control using plink 1.9[35] with the following QC parameters: minor allele frequency < 0.01, Hardy-Weinberg p-value < 0.000001, individuals with missing genotypes > 0.1 and missing rate per SNP > 0.1[35]. Population assignment analysis was conducted using the model-based Bayesian clustering software in Admixture 1.3[36]. The number of ancestral populations (K) was tested for a range from 1 to 13. The optimal K was determined based on cross validation error curve and the results were graphically represented using the R package ggplot2[37] for K = 2 to K = 13, with an optimal K = 7 for this dataset.

Principle component analysis (PCA) was conducted in plink 1.9[35] and the results were visualized using a python script (https://github.com/Siavash-cloud/3D-PCA-plot).

We used seven genetic subgroups identified in the population assignment analysis to test for genetic differentiation. The program Arlequin v 3.5[38] was used to calculate pairwise genetic distances (F_{ST}) between populations according to Weir and Cockerham[39]. F_{ST} values were interpreted according to Del Carpio et al.[40], where an F_{ST} of 0 shows no differentiation between populations and a value of 1 shows complete differentiation. Populations were assumed to have little differentiation when F_{ST} values were below 0.05, moderate differentiation with values between 0.05 and 0.15, strong differentiation for F_{ST} values between 0.15 and 0.25, and very strong differentiation for values higher than 0.25. The heterozygosity within each thirteen populations was estimated based on counts of site patterns across clustered reads within ipyrad. To estimate the heterozygosity within each population, ipyrad compares the number of heterozygous sites (sites where individuals have

different alleles) to the total number of sites analyzed. The distribution of heterozygosity was shown as boxplot using ggplot2 in R package[37].

Phylogenetic inference

To infer the phylogenetic relationships based on the GBS data, maximum parsimony (MP) analyses were conducted in Paup* 4.0a169[41], with initially including all 264 individuals and defining *P. orientale* as outgroup, followed by an analysis of the dataset including only *P. bracteatum* individuals. For both analyses the concatenated alignment length was 659,080 bp. We conducted two-step MP analyses[42] with an initial heuristic search with 500 random-addition sequences (RAS) restricting the number of stored trees to 25 per repetition, TBR branch swapping and steepest descent not enforced. The resulting trees were then used as starting trees in a heuristic analysis with maxtree set to 10,000. Clade support was evaluated by 500 bootstrap re-samplings with settings as before but excluding the initial RAS step.

In addition, phylogenetic relationships of *P. bracteatum* accessions were inferred using maximum likelihood (ML) based on the concatenated alignment applying the GTRGAMMA model and 100 parsimony starting trees in RAxML v.8.2.12[43]. Statistical support was assessed via rapid bootstrapping with 500 replicates.

Ecoclimatic niche modelling

To obtain a map of habitat suitability and model the ecoclimatic niche of *P. bracteatum*, a maximum entropy approach was employed, as implemented in Maxent 3.3.3k[44]. Occurrence records were based on sampling locations of plant material used for phylogeographic analyses. Additional records were drawn from GBIF (www.gbif.org) and checked for plausibility of geographic coordinates. Current bioclimatic data were downloaded from the WorldClim database (version 2.1, climate data for 1970–2000)[45]. To avoid collinearity among bioclimatic layers, five layers were selected using hierarchical clustering. The used bioclimatic layers were annual mean temperature, mean diurnal range (the mean of the difference of maximum and minimum monthly temperatures), isothermality, temperature annual range and precipitation of coldest quarter. Pairwise Pearson's correlation coefficients between the selected bioclimatic layers were < 0.7. The 'ENMeval' R package[46] was used to build and compare several ecoclimatic niche models across a range of different settings. Ten thousand background points were drawn randomly from the study region, which was set to 25°–70°E and 27°–48°N. Regularization values were 0.5, 1, 2, 3, 4; the sets of tested feature classes were L, LQ, H, LQH, LQHP, LQHPT; test data were partitioned using the "checkerboard1" setting; and the number of kfolds for evaluation was set to 4. The best model was selected based on ΔAICc values.

Declarations

Acknowledgements

We would like to thank C. Koch for help with plant cultivation and A. Himmelbach and S. König for performing Illumina sequencing. We would also like to thank Dr. Siavash Salek Ardestani for his help with data visualization.

Author contributions

M.R.N., F.R.B., Z.N., R.R. conceived the experiments, A.R. collected Zagros Mts. samples, A.Gh.K. collected Alborz Mts. samples. R.R., Z.N., F.R.B., M.R.N, S.P. coordinated study, R.R. and Z.N. performed experiments, Z.N., R.R., F.R.B., S.P. analyzed data, F.R.B., R.R., Z.N. wrote the initial manuscript. All authors contributed to and approved the final version of the manuscript.

Data availability statement

All sequence data are available through the NCBI nucleotide database under BioProject IDs PRJNA995928 (*P. bracteatum*) and PRJNA998183 (*P. bracteatum* and *P. orientale*).

Additional information

Competing interest

The authors declare no competing interests.

References

1. Kadereit, J. W., Schwarzbach, A. E. & Jork, K. B. The phylogeny of *Papaver* s. l. (Papaveraceae): Polyphyly or monophyly? Plant Syst. Evol. 204, 75–98 (1997).
2. Zhou, J. *et al.* Complete chloroplast genomes of *Papaver rhoeas* and *Papaver orientale*: Molecular structures, comparative analysis, and phylogenetic analysis. Molecules 23, 437 (2018).
3. McNicholas, L. F. & Martin, W. R. New and experimental therapeutic roles for naloxone and related opioid antagonists. Drugs 27, 81–93 (1984).
4. Rezaei, M., Naghavi, M. R., Hosseinzadeh, A., Abasi, A. & Nasiri, J. Spatiotemporal oscillations of morphinan alkaloids in opium poppy. J. Biosci. 43, 391–405 (2018).
5. Singh, A., Menéndez-Perdomo, I. M. & Facchini, P. J. Benzylisoquinoline alkaloid biosynthesis in opium poppy: An update. Phytochem. Rev. 18, 1457–1482 (2019).
6. Li, Y., Winzer, T., He, Z. & Graham, I. A. Over 100 million years of enzyme evolution underpinning the production of morphine in the Papaveraceae family of flowering plants. Plant Commun. 1, 100029 (2020).
7. Ozber, N. & Facchini, P. J. Phloem-specific localization of benzylisoquinoline alkaloid metabolism in opium poppy. J. Plant Physiol. 271, 153641 (2022).
8. Seddigh, M., Jolliff, G. D., Calhoun, W. & Crane, J. M. *Papaver bracteatum*, potential commercial source of codeine. Econ. Bot. 36, 433–441 (1982).
9. Ataee, N., Fooladi, J., Namaei, M. H., Rezadoost, H. & Mirzajani, F. Biocatalysts screening of *Papaver bracteatum* flora for thebaine transformation to codeine and morphine. Biocatal. Agric. Biotechnol. 9, 127–133 (2017).
10. Tisserat, B. & Berhow, M. Production of pharmaceuticals from *Papaver* cultivars in vitro. Eng. Life Sci. 9, 190–196 (2009).
11. Tarkesh Esfahani, S., Karimzadeh, G., Naghavi, M. R. & Vrieling, K. Altered gene expression and root thebaine production in polyploidized and methyl jasmonate-elicited *Papaver bracteatum* Lindl. Plant Physiol. Biochem. PPB 158, 334–341 (2021).
12. Lee, E.-J. & Facchini, P. Norcoclaurine synthase is a member of the pathogenesis-related 10/Bet v1 protein family. Plant Cell 22, 3489–3503 (2010).
13. Sharghi, N. & Lalezari, I. *Papaver bracteatum* Lindl., a highly rich source of thebaine. Nature 213, 1244–1244 (1967).
14. Nyman, U. & Bruhn, J. G. *Papaver bracteatum* – a summary of current knowledge. Planta Med. 35, 97–117 (1979).
15. Chaterjee, A., Shukla, S., Mishra, P., Rastogi, A. & Singh, S. P. Prospects of in vitro production of thebaine in opium poppy (*Papaver somniferum* L.). Ind. Crops Prod. 32, 668–670 (2010).
16. Carolan, J. C., Hook, I. L. I., Walsh, J. J. & Hodkinson, T. R. Using AFLP markers for species differentiation and assessment of genetic variability of in vitro-cultured *Papaver bracteatum* (section *Oxytona*). Vitro Cell. Dev. Biol. - Plant 38, 300–307 (2002).
17. Goldblatt, P. Biosystematic studies in *Papaver* section *Oxytona*. Ann. Mo. Bot. Gard. 61, 264 (1974).
18. Qaderi, A. *et al.* Molecular diversity and phytochemical variability in the Iranian poppy (*Papaver bracteatum* Lindl.): A baseline for conservation and utilization in future breeding programmes. Ind. Crops Prod. 130, 237–247 (2019).
19. Hadipour, M., Kazemitabar, S. K., Yaghini, H. & Dayani, S. Genetic diversity and species differentiation of medicinal plant Persian poppy (*Papaver bracteatum* L.) using AFLP and ISSR markers. Ecol. Genet. Genomics 16, 100058 (2020).
20. Elshire, R. J. *et al.* A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6, e19379 (2011).

21. He, J. *et al.* Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front. Plant Sci. 5, (2014).
22. Nemati, Z., Harpke, D., Gemicioglu, A., Kerndorff, H. & Blattner, F. R. Saffron (*Crocus sativus*) is an autotriploid that evolved in Attica (Greece) from wild *Crocus cartwrightianus*. Mol. Phylogenet. Evol. 136, 14–20 (2019).
23. McNaughton, I. H. & Harper, J. L. The comparative biology of closely related species living in the same area. I. External breeding barriers between *Papaver* species. New Phytol. 59, 15–26 (1960).
24. Miller, J. a. C. *et al.* Pollination biology of oilseed poppy, *Papaver somniferum* L. Aust. J. Agric. Res. 56, 483–490 (2005).
25. Blattner, F. & Kadereit, J. W. Patterns of seed dispersal in two species of *Papaver* L. under near-natural conditions. Flora 185, 55–64 (1991).
26. Frankham, R. Genetics and extinction. Biol. Conserv. 126, 131–140 (2005).
27. Aguilar, R., Quesada, M., Ashworth, L., Herrerias-Diego, Y. & Lobo, J. Genetic consequences of habitat fragmentation in plant populations: Susceptible signals in plant traits and methodological approaches. Mol. Ecol. 17, 5177–88 (2009).
28. Leidner, A. K. & Haddad, N. M. Combining measures of dispersal to identify conservation strategies in fragmented landscapes. Conserv. Biol. 25, 1022–1031 (2011).
29. Manafzadeh, S., Salvo, G. & Conti, E. A tale of migrations from east to west: The Irano-Turanian floristic region as a source of Mediterranean xerophytes. J. Biogeogr. 41, 366–379 (2014).
30. Jakob, S. S., Meister, A. & Blattner, F. R. The considerable genome size variation of *Hordeum* species (Poaceae) is linked to phylogeny, life form, ecology, and speciation rates. Mol. Biol. Evol. 21, 860–869 (2004).
31. Wendler, N. *et al.* Unlocking the secondary gene-pool of barley with next-generation sequencing. Plant Biotechnol. J. 12, 1122–1131 (2014).
32. Babraham Bioinformatics. FastQC a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed 20.01.2023).
33. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12 (2011).
34. Eaton, D. A. R. & Overcast, I. ipyrad: Interactive assembly and analysis of RADseq datasets. Bioinformatics 36, 2592–2594 (2020).
35. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575 (2007).
36. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664 (2009).
37. Villanueva, R. A. M. & Chen, Z. J. ggplot2: Elegant graphics for data analysis (2nd ed.). Meas. Interdiscip. Res. Perspect. 17, 160–167 (2019).
38. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10, 564–567 (2010).
39. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. Evolution 38, 1358–1370 (1984).
40. Pino Del Carpio, D. *et al.* The patterns of population differentiation in a *Brassica rapa* core collection. Theor. Appl. Genet. 122, 1105–1118 (2011).
41. Swofford, D. *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods).* Version 4.0b169. Sinauer Assoc., Sunderland (2002).
42. Blattner, F. R. Phylogenetic analysis of *Hordeum* (Poaceae) as inferred by nuclear rDNA ITS sequences. Mol. Phylogenet. Evol. 33, 289–299 (2004).
43. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313 (2014).

44. Phillips, S. J., Anderson, R. P. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190, 231–259 (2006).
45. Fick, S. E. & Hijmans, R. J. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315 (2017).
46. Muscarella, R. *et al.* ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. Methods Ecol. Evol. 5, 1198–1205 (2014).

Figures

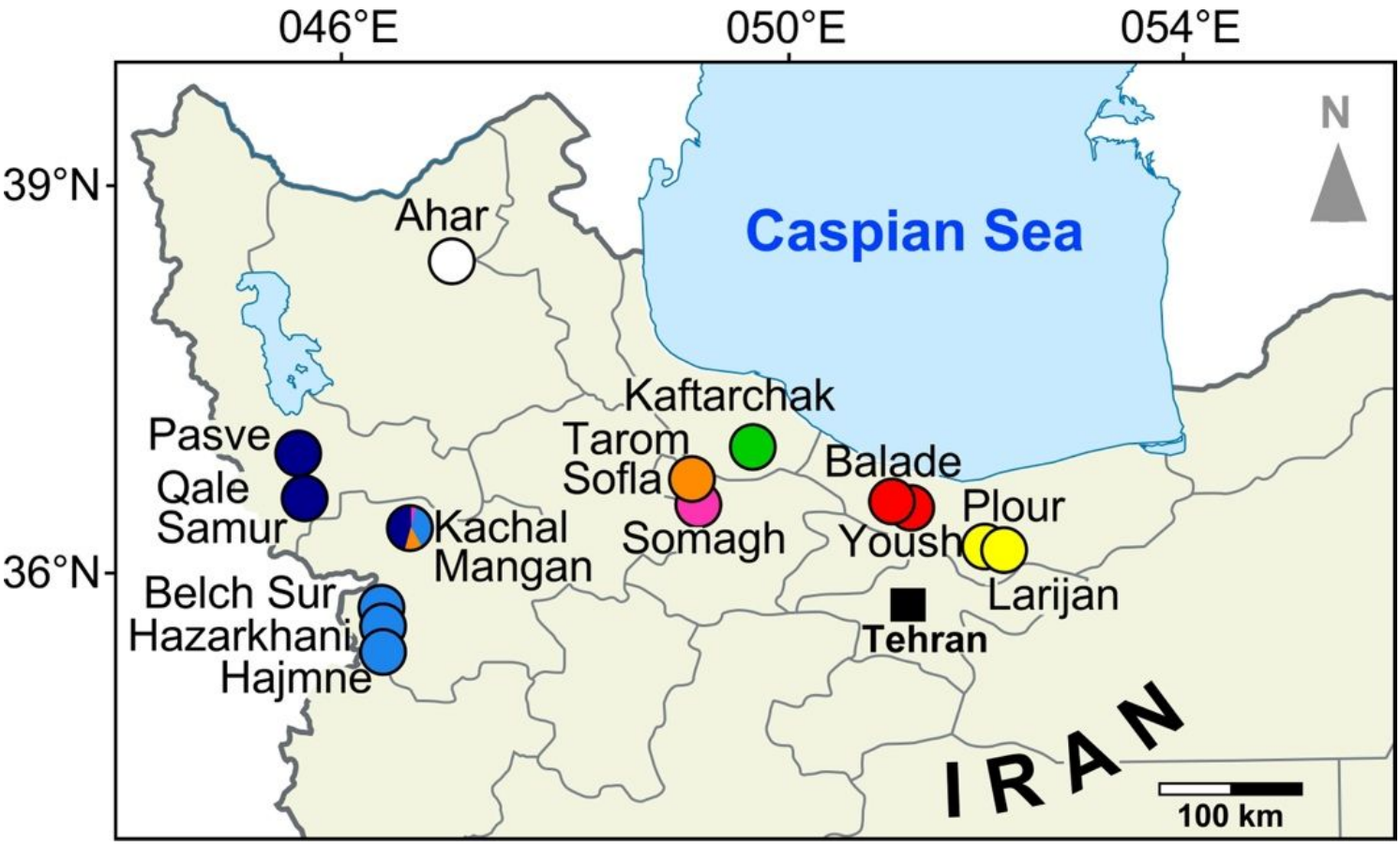

Figure 1

Map of northwestern Iran providing the geographic positions of the *P. bracteatum*(dot colors refer to population ancestry analysis) and *P. orientale* (white dot) populations included in the analyses.

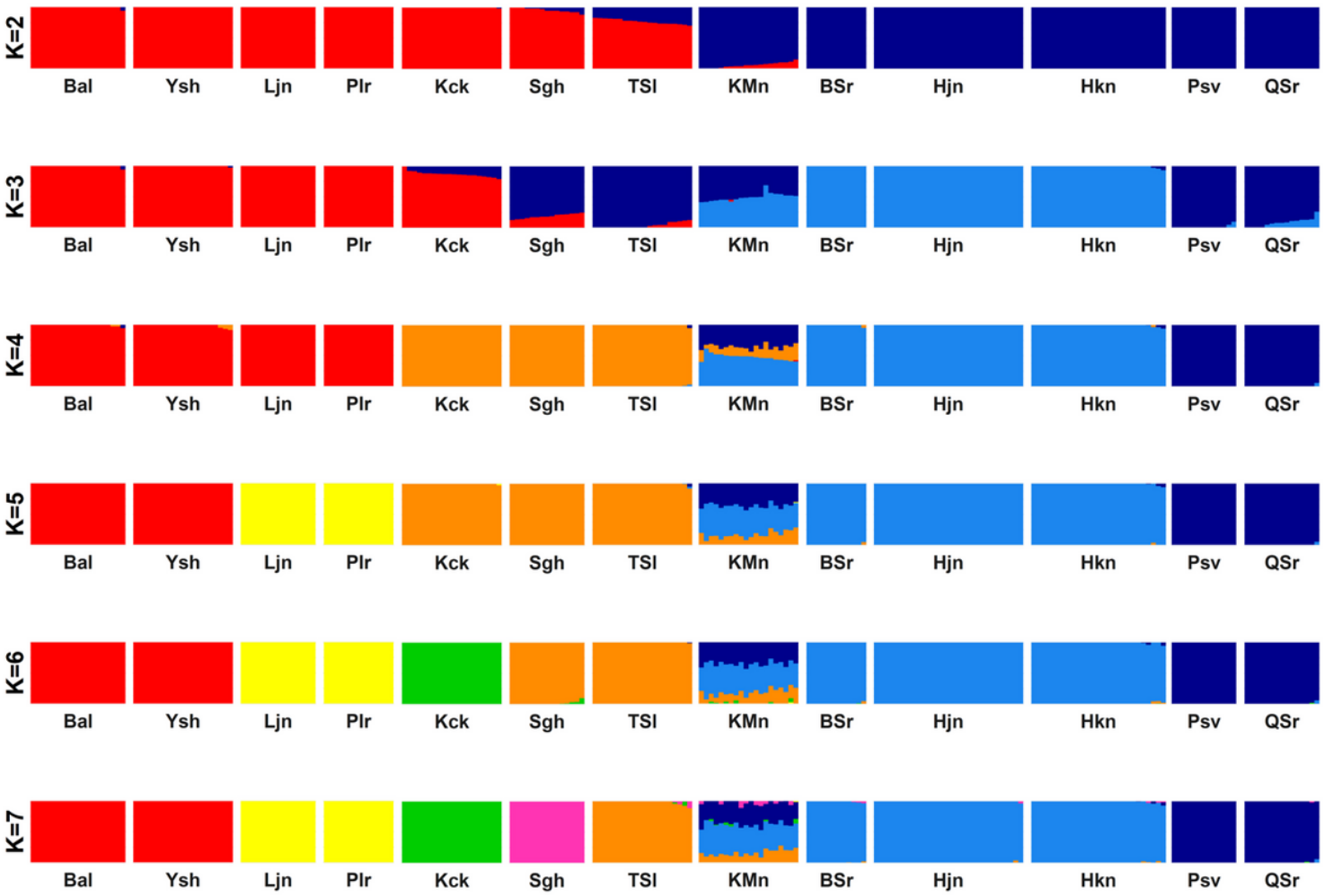

Figure 2

Population assignment analyses from K = 2 to K = 7 based on 6027 unlinked SNPs of 244 *P. bracteatum* individuals. Each box shows a population with the containing individuals represented by thin bar plots. The colors indicate the assignment of the individuals to K populations.

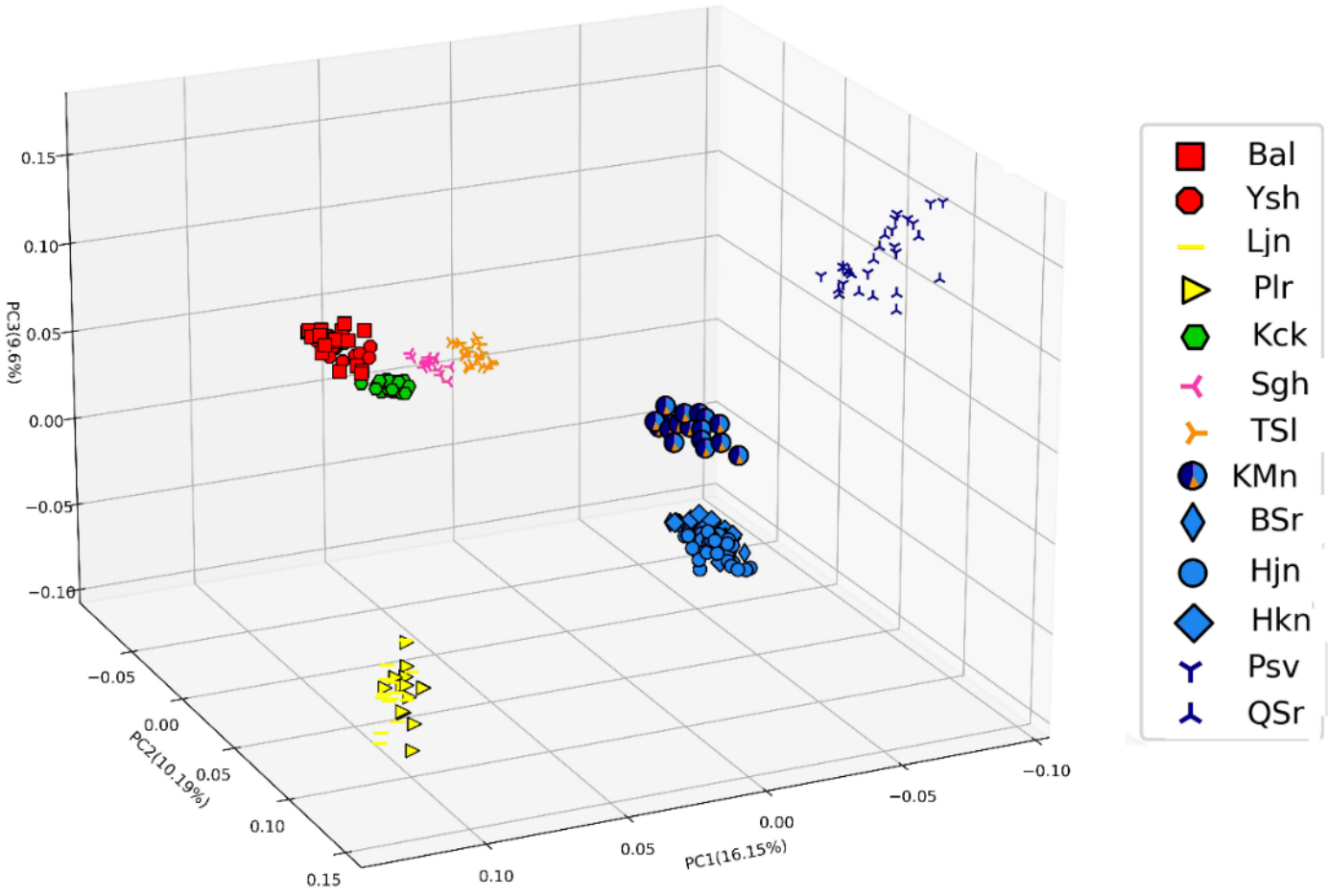

Figure 3

Plot of the first three axes of the Principle Component Analysis based on 6027 genome-wide SNPs scored in 244 individuals.

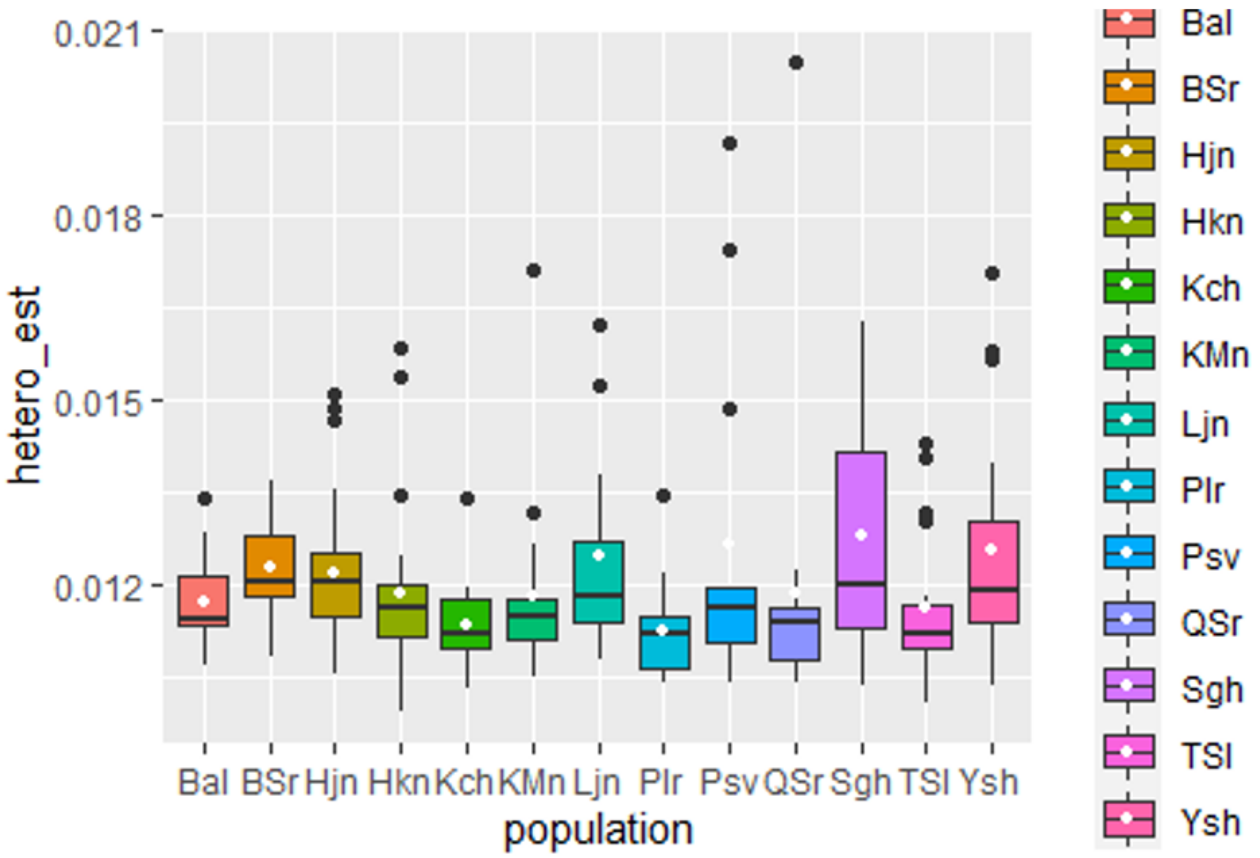

Figure 4

Boxplot showing individual estimated heterozygosity across loci within each population. Median values are represented by thick black lines, with the box indicating quartiles. Whiskers extend to 1.5× the interquartile range or the maximum value. Black circles indicate individuals with heterozygosity outside the range, while white circles represent the mean heterozygosity for each population.

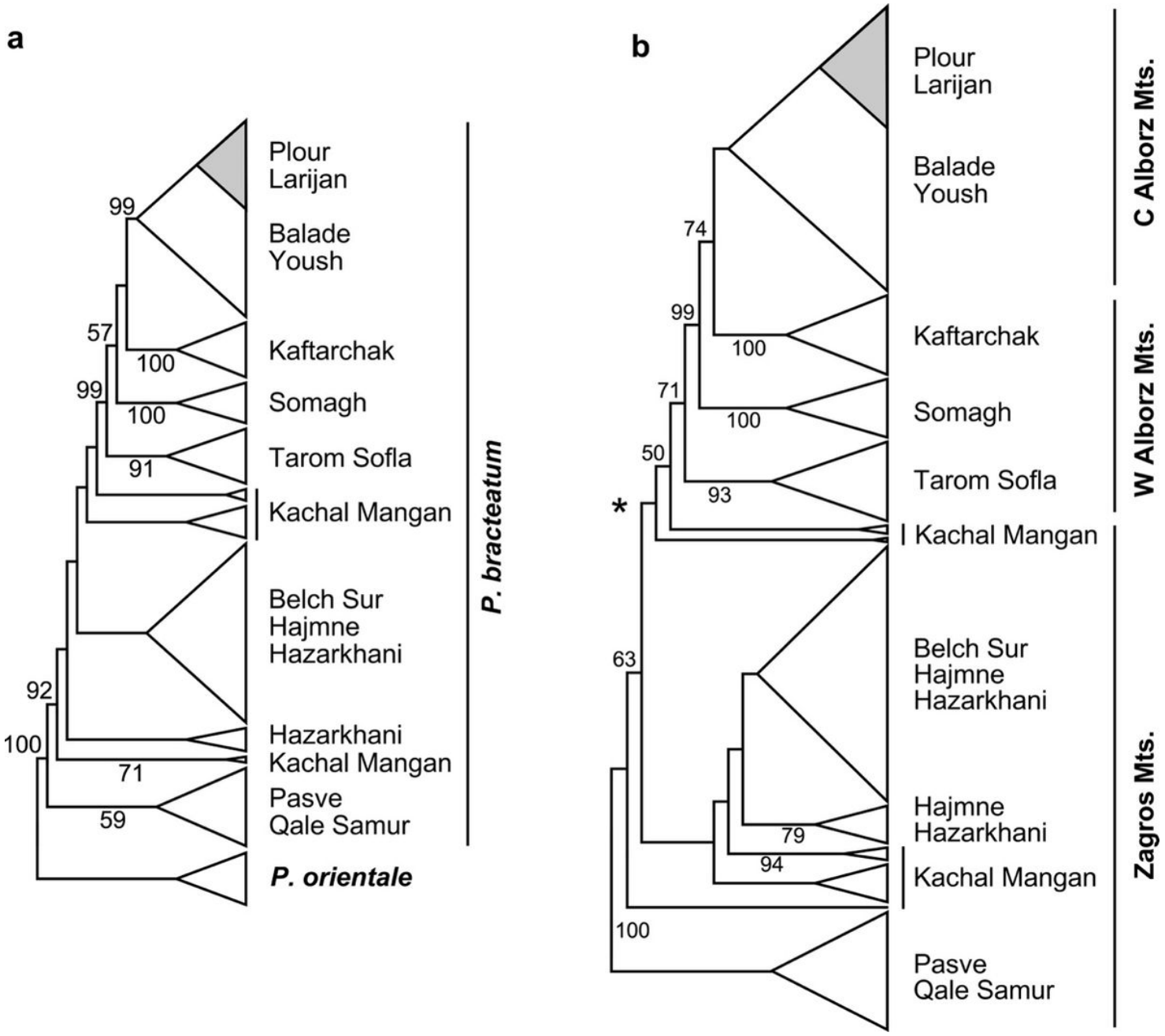

Figure 5

(**a**) Schematic representation of the strict consensus tree topology of 80 most parsimonious trees derived from an MP analysis of the GBS dataset including 264 individuals of *P. bracteatum*and *P. orientale*. Numbers along the branches provide bootstrap support values (%) for the tree backbone, names to the right the population affiliations for *P. bracteatum*. The tree was rooted with the *P. orientale* individuals. (**b**) Schematic representation of the strict consensus tree topology of 120 MP trees derived from the analysis of GBS data of 244 individuals of *P. bracteatum*. The tree was rooted with the Pasve/Qale Samur individuals to make it easier comparable with Fig. 5a. An asterisk marks the position of the root according to a population assignment analysis (Fig. 2), which separates at K = 2 the populations from Zagros vs. Alborz Mts. Numbers along branches provide bootstrap support values (%) for the tree backbone. Names to the right indicate population affiliations. The individuals from Larijan and Plour form in both trees a monophyletic unit within the polyphyletic populations from Balade and Yoush. Also the individuals from Kachal Mangan result in both analyses polyphyletic.

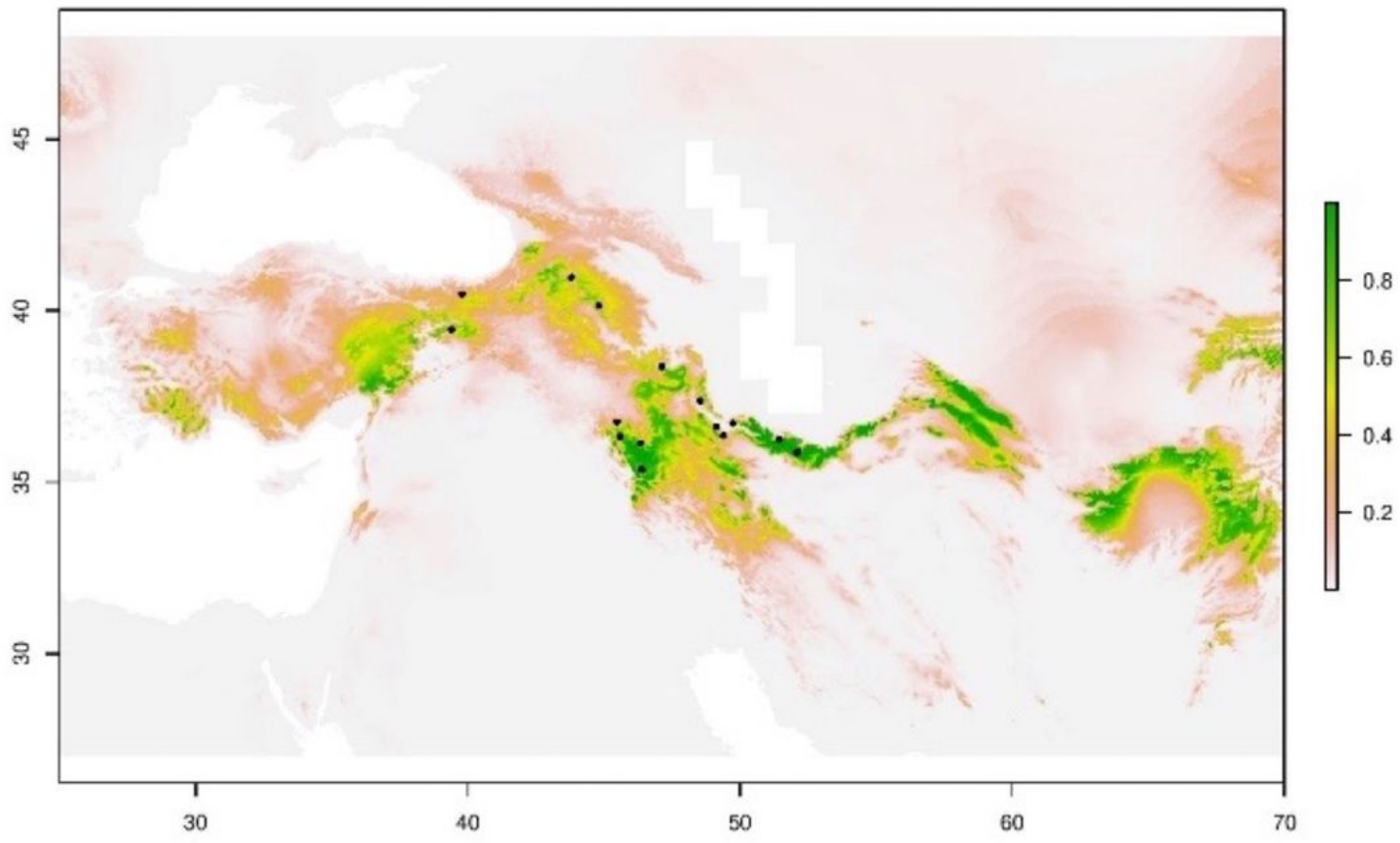

Figure 6

Ecoclimatic niche modelling for *P. bracteatum*. Black dots indicate occurrence points that were used as input for the distribution models. The scale for occurrence probability is provided to the right. The niche models concur very well with the actual distribution of the species in Turkey and Iran. For southeastern Turkey and parts of Central Asia they show however that the potential distribution area/climate-defined niche of *P. bracteatum* is much larger than the regions currently inhabited by the species.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- BlattnerSupplMaterials.pdf