



## OPEN The impact of environmental and climatic variables on genetic diversity and plant functional traits of the endangered tuberous orchid (*Orchis mascula* L.)

Mohammad Mafakheri<sup>1</sup>✉, Mehdi Bakhshipour<sup>2,7</sup>, Mina Omrani<sup>3,7</sup>, Hamid Gholizadeh<sup>4,7</sup>, Najmeh Rahimi<sup>5,7</sup>, Ali Mobaraki<sup>2</sup> & Mehdi Rahimi<sup>6</sup>

Understanding how environmental factors shape patterns of genetic and phenotypic variations in a species is necessary for conservation and plant breeding. However, these factors have not yet been completely understood in tuberous orchid species used to make 'Salep', an important ingredient in traditional medicine and beverages in middle eastern countries and India. In many areas, increasing demand has pushed species to the brink of extinction. In this study, 198 genotypes from 18 populations of the endangered species *Orchis mascula* L. spanning a large-scale climatic gradient in northern Iran were used to investigate patterns of genetic diversity and plant functional traits. Populations were sampled from three land cover types (woodland, shrubland, and pastureland/grassland). Plant height, stem length, number of flowers, bulb fresh and dry weight, glucomannan, and starch concentrations showed high variation among populations and were significantly related to land cover type. In general, genetic diversity was high, particularly in those from eastern Hyrcanian; additionally, populations showed a high level of genetic differentiation ( $G'_{st} = 0.35$ ) with low gene flow ( $Nm = 0.46$ ). The majority of genetic differentiation occurred within populations (49%) and land cover types (20%). The population structural analysis using the AFLP marker data in  $K = 4$  showed a high geographical affinity for 198 *O. mascula* genotypes, with some genotypes having mixed ancestry. Temperature and precipitation were found to shape genetic and phenotypic variation profoundly. Significant isolation by the environment was observed, confirming the strong effect of environmental variables on phenotypic and genetic variation. Marker-trait association studies based on MLM1 and MLM2 models revealed significant associations of P-TGG + M-CTT-33 and E-AGG + M-CGT-22 markers with plant height and glucomannan content. Overall, a combination of large-scale climatic variables and land cover types significantly shaped genetic diversity and functional trait variation in *O. mascula* populations.

Environmental factors play a key role in shaping patterns of genetic diversity and phenotypic variation in plant species<sup>1</sup>. They often display plasticity and genetic variation along environmental gradients, as these gradients can affect gene flow among populations or create extreme abiotic conditions. Interactions between the environment, phenotypic and genetic diversity have been studied in different plant species at the population level, emphasizing the effect of climate gradients (e.g., temperature and precipitation)<sup>2,3</sup>. Genetic diversity is crucial since it can act as a buffer against abiotic or biotic stress, anthropological disturbance, inbreeding depression, and interaction with

<sup>1</sup>Department of Plant Sciences, University of California - Davis, Davis, CA 95616, USA. <sup>2</sup>Department of Horticultural Sciences, Faculty of Agricultural Sciences, University of Guilan, P.O. Box, Rasht 41635-1314, Iran. <sup>3</sup>School of Science and Technology, Faculty of Science, Agriculture, Business and Law, University of New England, Armidale, NSW 2351, Australia. <sup>4</sup>Department of Plant Biology, Faculty of Basic Sciences, University of Mazandaran, Babolsar, Mazandaran, Iran. <sup>5</sup>Department of Chemistry and Biochemistry, New Mexico State University, Las Cruces, NM, USA. <sup>6</sup>Department of Biotechnology, Institute of Science and High Technology and Environmental Sciences, Graduate University of Advanced Technology, Kerman, Iran. <sup>7</sup>These authors contributed equally: Mehdi Bakhshipour, Mina Omrani, Hamid Gholizadeh and Najmeh Rahimi. ✉email: mmfakheri@ucdavis.edu

invasive species<sup>2,4</sup>. Populations with a history of overexploitation and fragmentation can face genetic depletion that can severely impact the survival and evolution of the species by negatively affecting fitness characteristics<sup>5–7</sup>. Under extreme conditions such as prolonged drought or high temperature, high genetic diversity is a significant advantage given the higher possibility of having genotypes/alleles that vary in their response to extremes<sup>8</sup>, enabling adaptation to environmental changes<sup>9</sup>. A species' capability to occupy new niches can also be affected by the loss of genetic diversity as well as damaging population growth and breeding<sup>10–12</sup>. A population's response to the environment (land cover, climate, etc.) can additionally be examined using functional traits, traits that are thought to have a profound influence on plant fitness<sup>13–15</sup>. The plasticity of these traits under various environmental conditions is a fundamental mechanism that plants rely on to react to environmental changes. The convoluted relationship between genetic diversity, functional traits, and environmental heterogeneity of the locations at the population level has been the subject of quite a few studies<sup>16,17</sup> and deserves more attention.

Tuberous orchids have a rich history of traditional use and have suffered from overcollection. The dried tubers of specific wild terrestrial orchids are milled to make a flour known as 'salep'<sup>18</sup>, which traditionally has been an essential ingredient of confectioneries<sup>19</sup>. In addition to conferring aromatic and flavor to formulations, it also is a thickening and stabilizing chemical agent<sup>20</sup>. Salep is integral to several medicines, beverages, and ice-creams<sup>21</sup>. Its critical rheological properties are due to the presence of polysaccharides, especially glucomannan (GLN; 16–55%)<sup>18</sup>, that is, natural fibers with high water solubility. These also have specific medicinal benefits such as normalizing blood sugar, inhibiting liver irregularities, and alleviating pancreas stress<sup>22</sup>. In recent years the harvest of tuberous orchids has increased due to high prices fueled by international demand, mainly from Iran's neighboring countries such as India, Turkey, and Pakistan<sup>23</sup>. Increasing public awareness of salep's benefits and growing demand is a dangerous combination that may jeopardize the long-term viability of natural populations of these orchids.

Besides the ongoing over-exploitation of the wild population of tuberous orchids, climate change may also hasten the extinction of orchid species<sup>24,25</sup>. Agricultural expansion, urbanization, and silviculture have exacerbated the loss of genetic diversity in orchid populations<sup>26,27</sup>. Such problematic circumstances have already taken their toll on orchids in Iran, and some orchid species, including *Orchis mascula* L., have been classified as endangered<sup>26</sup>. Therefore, conservation strategies are needed to maintain the genetic diversity in the remaining populations and assure the species' long-term viability<sup>28,29</sup>. However, the absence of reliable information on the orchid's population status or the ecological consequences of severe over-collection prevents the implementation of robust conservation plans.

Although in Iran, as in other countries, legislators have enacted detailed regulations in favor of germplasm protection. Unfortunately, these regulations have not been enforced properly, and current measures are not limiting the collection or marketing of natural resources, particularly orchid tubers<sup>23,27</sup>. Recently, numerous small-scale regional studies on the population genetic status of tuberous orchids in the west of Iran have been reported<sup>30–32</sup>. However, the data required to develop conservation programs to protect genetic diversity or to develop the germplasm stocks in breeding programs can only be gained through sufficiently large-scale analyses, preferably at the landscape level<sup>33,34</sup>. *O. mascula* is a relatively small perennial orchid widely distributed across Europe, western Asia, and northern Africa. In Iran, this species is distributed primarily in the north (Alborz Mountains) and the west (the Zagros Mountains)<sup>35,36</sup>. It is one of several widely collected terrestrial orchids with numerous traditional medicinal applications. Overcollection of tubers has reduced *O. mascula* population sizes to the point that this species is now endangered; thus, fundamental large-scale studies of the functional traits and genetic diversity responses are needed. This study aimed to sample a high number of individuals from 18 populations of the terrestrial orchid *O. mascula* in northern Iran to address the following questions: (1) does phenotypic plasticity in functional traits exist in response to environmental heterogeneity? (2) how is genetic diversity and population structure affected by the land cover pattern and climatic variables? (3) do any population(s) have superior biochemical and phenotypic traits to be utilized in breeding programs?

## Results

**Phenotypic and chemotypic plasticity in response to environmental conditions.** Land cover significantly affected most of the traits ( $p \leq 0.05$  and  $p \leq 0.01$ ); the same applies to the population effect (Table 1). The variance component of PC1 and PC2 under land cover effect was considerably high (54.88 and 34.71, respectively). As agriculturally important characters, starch content and bulb fresh weight (BFW; 154.75 and 32.34, respectively) indicated the highest variance component, while floral traits generally had the lowest values (Table 1). Plant height (PLH) with 80.97% and PC1 with 74.52% showed the highest percent of variation (%var). In the other source of variation, populations as 'random effect', variance component was generally low, and PC1(48.77) and glucomannan (GLN, 22.71) had the highest values. The two essential traits, BFW and dry bulb weight (BDW) observed to have the highest %var (83.10 and 64.49) among the traits. Interestingly, %var among individuals within populations (error) for floral traits was often higher than other traits.

Phenotypic traits varied substantially between populations. For example, in traits associated with stem and leaves, the populations from the east side of the Hyrcanian ecoregion, Klaleh (KAH), Ziarat (ZIR), Azarshahr (AZS), and Maraveh Tapeh (MVT) populations overall had the highest values in general with the last population MVT, holds the most significant values. When it comes to floral traits, the variation across the populations is almost intangible (Table 2). The MVT and Talesh (TAS) populations with 3.79 and 3.27 g were found to have the highest BFW, noticeably more than the majority of the populations. The former, with 2.22 g, had the highest value for BDW. Another population from that region, KAH with 1.23 g, also showed a high value for this economically important trait. The difference between the populations became more apparent for starch and GLN contents, in which populations from Golestan, the eastern part, had higher levels, and those from Golestan were superior regarding GLN content in general. This pattern was further analyzed by assessing the relation of these

Traits	Source of variation							
	Land cover			Populations			Individuals	
	F-ratio	Var comp	%var	F-ratio	Var comp	%var	Var comp	%var
PCA1	7491.01**	54.88	74.52	526.89**	48.77	23.48	4.11	1.98
PCA2	341.63 ns	34.71	10.26	161.97**	14.83	74.88	2.94	14.85
Plant height (cm)	2743.99**	0.29	80.97	111.39**	10.13	14.95	2.76	4.07
Stem length (cm)	1535.42**	1.66	65.06	173.95**	15.98	29.95	2.65	4.98
Number of leaves/plant	11.59 ns	1.30	28.01	6.145**	0.55	52.79	0.20	19.18
Leaf length (cm)	70.96 ns	7.05	26.54	41.46**	3.79	60.54	0.80	12.90
Leaf width (cm)	74.39**	20.88	51.65	7.40**	0.63	25.25	0.58	23.08
Inflorescence length (cm)	372.11**	0.05	68.28	18.38**	1.55	15.05	1.72	16.66
Number of flowers	1724.93**	35.01	62.07	171.95**	15.62	29.49	4.46	8.42
Length of external tepals (cm)	3.09**	0.02	46.34	0.21**	2.01	12.81	0.04	40.84
Width of external tepals (cm)	0.0084**	0.00	10.31	0.001 ns	5.00	0.00	0.00	89.68
Length of internal tepals (cm)	1.17**	0.03	33.00	0.13**	3.00	14.90	0.03	52.08
Width of internal tepals (cm)	0.0055 ns	0.02	3.94	0.002**	6.00	9.70	0.00	86.35
Length lip (cm)	1.80**	0.02	16.23	0.13**	1.01	19.46	0.01	24.30
Width lip (cm)	1.15**	0.00	14.34	0.16**	0.01	18.79	0.01	24.85
Length of median lobe of the lip (cm)	1.17**	0.01	15.88	0.16**	0.01	10.78	0.01	23.33
Median lobe width of the lip (cm)	0.023*	0.00	17.19	0.005**	0.00	11.34	0.00	71.45
Length of lateral lobe (cm)	0.56*	-0.01	17.34	0.10**	0.00	12.49	0.00	30.16
Sidelobe width (cm)	0.13**	0.07	8.63	0.009**	0.00	17.56	0.00	33.79
Bulb F weight (g)	5.12 ns	32.34	0.00	9.70**	0.88	83.10	0.18	16.89
Bulb D weight (g)	2.46 ns	6.03	29.32	1.67**	0.15	64.49	0.01	6.18
Glucomannan (g/100 g)	1202.94*	0.19	43.14	247.57**	22.71	48.16	4.09	8.68
Starch (g/100 g)	302.68**	154.75	64.10	29.38**	2.67	28.43	0.70	7.46
Total Anthocyanins (mg/g)	15.97 ns	2.03	20.85	6.11**	0.55	59.66	0.18	19.48

**Table 1.** Results of unbalanced REML mixed-model ANOVA of first two principal components and 19 phenotypic and 3 biochemical traits for 198 individual plants from 18 natural populations. Estimates for the percent of variation (%var) distributed between land cover, among populations within the land cover, and among individuals within populations (error) were estimated by treating all hierarchical levels as random effects. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

two compounds against land cover, where a reverse relationship emerged (Fig. 1); the content of GLN increased in pastures/grasslands while decremented in woodland. Starch, however, seemed more affected by land cover types, but in an opposite manner, in pastureland/grassland, its content was down to 2.5 g/100 and maximized in the woodland.

**Hierarchical clustering on principal components (HCPC) and cluster analysis.** The HCPC provides information on the factor map resulting from principal component analysis based on phenotypic and biochemical traits of the populations with a three-dimensional hierarchical tree, thereby providing a clearer view of the clustering pattern (Fig. 2). Dimension 1 explains 66.59% of the phenotypic variation, while the second dimension stands for only 23.94%. Results indicate three clusters that are projecting distinct positioning of the populations.

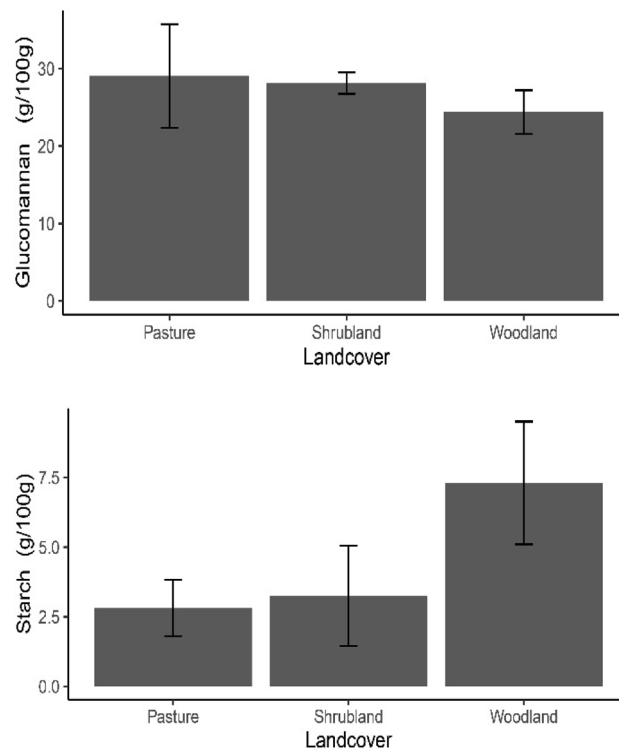
Two-way cluster analysis (heatmap) identified two main groups for phenotypic and biochemical traits (Fig. 3, column). Among biochemical traits, starch content and TAC formed a separated cluster (B) while the rest were placed in A. Traits such as GLN, number of leaves/plant (NLP), BFW, and BDW were grouped in a subcluster, and floral and leaf traits placed in AII. Two main clusters appeared at the population level (Fig. 3, row); cluster A included two subclusters; cluster AI encompassed populations from Mazandaran, Golestan, and Guilan provinces. However, populations in the subclusters were grouped based on similarity in their phenotypic or biochemical similarities, for instance, Talesh (TAS), Khalkhal (KHAL), Masouleh (MSL), Darestan (DRS), Chapouh (CHP), Khasib-Dasht (KHD) and Garmab-Dasht (GRD) all from Guilan province, clustered together. The three populations from the eastern side (MVT, KAH, and ZIR) stand out by forming a separate subcluster (AII). Four populations of Palamjan (PAJ), Mirkhamand (MKH), Kalech (KCH), and Kjour (KJR) with the lowest values of traits placed in the second cluster (B). Although discrepancies were observed in grouping patterns, clustering mainly succeeded in grouping the population according to their location or traits based on their identity (i.e., biochemical, tuber characteristics, leaves, floral features, etc.).

**Relationship between phenotypic and chemotypic traits and climatic variables.** Correlation analyses revealed that the phenotypic traits negatively correlated with the annual precipitation and geographi-

Pops	Stem and leaf traits						Floral traits				
	PLH	STL	NLP	LL	LW	IL	NF	LET	WET	LIT	WIT
<b>Guilan</b>											
TAS	29.86 ± 4.24	17.39 ± 4.15	3.87 ± 0.53	7.06 ± 2.02	4.28 ± 0.93	11.24 ± 1.72	16.46 ± 4.96	1.33 ± 0.48	0.28 ± 0.09	1.08 ± 0.52	0.25 ± 0.11
KHAL	25.42 ± 4.13	12.92 ± 2.21	3.27 ± 0.22	5.42 ± 2.16	4.17 ± 1.07	9.6 ± 1.79	17.24 ± 5.23	1.33 ± 0.19	0.26 ± 0.12	1.1 ± 0.58	0.25 ± 0.13
MSL	28.59 ± 4	17.23 ± 5.15	3.84 ± 0.38	6.15 ± 1.95	6.54 ± 0.86	11.51 ± 1.59	26.22 ± 5.37	1.39 ± 0.37	0.26 ± 0.13	1.06 ± 0.56	0.24 ± 0.14
DRS	27.17 ± 3.64	15.53 ± 4.44	3.49 ± 0.33	6.22 ± 2.11	4.37 ± 1.01	11.27 ± 1.57	22.57 ± 5.43	1.22 ± 0.38	0.28 ± 0.21	1.08 ± 0.63	0.25 ± 0.14
CHP	27.64 ± 5	17.62 ± 4.22	3.36 ± 0.6	5.79 ± 2.19	4.25 ± 1.1	10.64 ± 1.66	23.84 ± 4.97	1.49 ± 0.3	0.26 ± 0.12	1.24 ± 0.72	0.26 ± 0.12
KHD	30.63 ± 4.32	20.11 ± 5.58	3.54 ± 0.34	6.85 ± 1.98	5.49 ± 0.89	13.4 ± 1.71	25.33 ± 5.6	1.28 ± 0.25	0.27 ± 0.13	1.08 ± 0.74	0.23 ± 0.15
GRD	34.25 ± 4.56	23.8 ± 4.39	4.37 ± 0.36	8.63 ± 2.39	5.51 ± 1.29	12.56 ± 1.48	25.37 ± 5.21	1.42 ± 0.28	0.29 ± 0.11	1.27 ± 0.61	0.28 ± 0.13
<b>Mazandaran</b>											
JAD	24.67 ± 3.57	14.69 ± 2.08	3.18 ± 0.35	6.37 ± 1.91	2.54 ± 0.82	7.96 ± 1.45	18.35 ± 5.33	1.18 ± 0.13	0.27 ± 0.23	1.08 ± 0.51	0.26 ± 0.12
PAJ	21.63 ± 3.3	9.79 ± 5.25	3.27 ± 0.39	6.35 ± 2.07	2.34 ± 0.97	8.39 ± 1.68	13.97 ± 5.29	0.8 ± 0.14	0.27 ± 0.11	0.95 ± 0.59	0.23 ± 0.16
KJR	20.34 ± 4.02	10.87 ± 7.14	3.2 ± 0.31	6.35 ± 2.3	3.46 ± 1.2	7.92 ± 1.3	12.33 ± 5.6	0.82 ± 0.18	0.26 ± 0.1	0.82 ± 0.51	0.23 ± 0.13
KCH	19.37 ± 3.7	12.53 ± 4.51	3.23 ± 0.25	6.4 ± 1.94	3.29 ± 0.84	7.37 ± 1.33	12.02 ± 4.63	0.84 ± 0.11	0.26 ± 0.11	0.72 ± 0.51	0.23 ± 0.12
MKH	20.35 ± 3.76	9.9 ± 5.05	3.26 ± 0.19	6.32 ± 2.42	3.34 ± 1.32	8.31 ± 1.25	<b>10.18 ± 5.19</b>	0.78 ± 0.12	0.26 ± 0.1	0.73 ± 0.48	0.21 ± 0.13
<b>Golestan</b>											
PBD	30.42 ± 3.92	19.98 ± 3.54	3.78 ± 0.38	7.36 ± 2.48	5.34 ± 1.38	13.59 ± 2.08	23.67 ± 5.32	1.23 ± 0.31	0.28 ± 0.1	1.06 ± 0.54	0.25 ± 0.15
BAB	32.2 ± 4.57	19.55 ± 5.25	3.76 ± 0.37	7.35 ± 2.49	5.62 ± 1.39	13.78 ± 1.81	23.33 ± 5.33	1.38 ± 0.39	0.28 ± 0.11	1.13 ± 0.57	0.26 ± 0.13
KAH	35.4 ± 4.55	25.91 ± 4.89	4.53 ± 0.59	8.8 ± 2.37	5.64 ± 1.27	13.51 ± 1.82	27.01 ± 5.12	1.47 ± 0.16	0.3 ± 0.19	1.27 ± 0.55	0.23 ± 0.14
ZIR	<b>39.02 ± 4.79</b>	19.28 ± 3.56	3.13 ± 0.26	7.38 ± 2.5	3.75 ± 1.41	10.31 ± 1.23	22.74 ± 4.85	1.17 ± 0.25	0.27 ± 0.12	1.11 ± 0.56	0.25 ± 0.14
AZS	<b>39.08 ± 4.74</b>	21.19 ± 4.99	3.85 ± 0.33	6.22 ± 2.27	5.59 ± 1.18	14.87 ± 1.5	22.27 ± 5.61	1.18 ± 0.59	0.3 ± 0.16	1.04 ± 0.6	0.25 ± 0.12
MVT	<b>40.63 ± 4.31</b>	30.39 ± 7.71	6.05 ± 0.7	13.12 ± 2.36	6.4 ± 1.26	14.97 ± 1.6	<b>31.1 ± 5.13</b>	1.53 ± 0.37	0.3 ± 0.18	1.28 ± 0.53	0.27 ± 0.13
Pops	Floral traits						Bulbs		Biochemical traits		
	LLI	WL	LMLL	LMLW	LLL	SW	BFW	BDW	GLN	Starch	TAC
<b>Guilan</b>											
TAS	1.18 ± 0.2	0.93 ± 0.65	1.11 ± 0.61	0.36 ± 0.12	0.81 ± 0.48	0.32 ± 0.13	<b>3.27 ± 1.15</b>	0.81 ± 0.47	23.7 ± 7.65	2.9 ± 4.07	2.52 ± 1.5
KHAL	1.18 ± 0.29	0.92 ± 0.43	1.14 ± 0.49	0.37 ± 0.16	0.81 ± 0.49	0.31 ± 0.16	2.61 ± 1.14	0.47 ± 0.48	27.62 ± 6.65	2.23 ± 3.72	2.38 ± 1.17
MSL	1.16 ± 0.28	0.89 ± 0.69	1.05 ± 0.74	0.34 ± 0.14	0.82 ± 0.5	0.33 ± 0.13	1.13 ± 0.69	0.86 ± 0.58	29.29 ± 7.38	5.55 ± 3.8	2.59 ± 1.17
DRS	1.2 ± 0.29	0.85 ± 0.55	0.97 ± 0.88	0.36 ± 0.13	0.8 ± 0.54	0.33 ± 0.14	1.13 ± 0.69	0.97 ± 0.47	30.14 ± 6.49	7.53 ± 3.24	2.87 ± 1.2
CHP	1.02 ± 0.24	0.86 ± 0.63	1.05 ± 0.76	0.37 ± 0.13	0.84 ± 0.47	0.31 ± 0.11	1.1 ± 0.7	0.94 ± 0.48	30.25 ± 6.11	7.55 ± 3.49	2.85 ± 1.15
KHD	1.23 ± 0.21	0.86 ± 0.65	1.04 ± 0.74	0.38 ± 0.12	0.83 ± 0.47	0.31 ± 0.14	2.38 ± 0.94	1.1 ± 0.57	29.47 ± 6.45	7.32 ± 3.57	2.59 ± 1.08
GRD	1.37 ± 0.25	1.13 ± 0.61	1.33 ± 0.69	0.41 ± 0.12	1.04 ± 0.49	0.37 ± 0.14	2.63 ± 0.75	1.11 ± 0.52	34.49 ± 7.01	5.41 ± 3.55	1.49 ± 0.93
<b>Mazandaran</b>											
JAD	1.14 ± 0.21	0.93 ± 0.84	1.17 ± 0.32	0.38 ± 0.11	0.86 ± 0.49	0.32 ± 0.16	1.09 ± 0.64	0.94 ± 0.44	24.43 ± 6.43	8.52 ± 3.69	2.84 ± 1.21
PAJ	0.83 ± 0.24	0.61 ± 0.62	0.83 ± 0.71	0.34 ± 0.11	0.63 ± 0.44	0.22 ± 0.13	1.46 ± 0.75	0.65 ± 0.5	21.67 ± 5.55	<b>10.57 ± 4.07</b>	3.73 ± 1.03
KJR	0.78 ± 0.23	0.66 ± 0.22	0.78 ± 0.7	0.34 ± 0.12	0.63 ± 0.45	0.22 ± 0.13	1.42 ± 0.69	1.16 ± 0.48	21.21 ± 6.08	10.51 ± 3.82	3.67 ± 1.07
KCH	0.81 ± 0.18	0.63 ± 0.57	0.84 ± 0.68	0.33 ± 0.09	0.67 ± 0.44	0.22 ± 0.12	<b>1.46 ± 0.75</b>	1.2 ± 0.52	19.8 ± 6.32	<b>10.78 ± 4.35</b>	3.79 ± 1.08
MKH	0.74 ± 0.2	0.63 ± 0.39	0.83 ± 0.69	0.34 ± 0.11	0.67 ± 0.51	0.22 ± 0.12	1.51 ± 0.79	1.24 ± 0.56	19.08 ± 6.15	<b>11.18 ± 4.22</b>	3.88 ± 1.18
<b>Golestan</b>											
PBD	1.19 ± 0.19	0.94 ± 0.63	1.15 ± 0.67	0.38 ± 0.12	0.86 ± 0.45	0.33 ± 0.13	1.17 ± 0.68	1.01 ± 0.47	<b>31.99 ± 6.15</b>	6.48 ± 3.29	2.79 ± 1.01
BAB	1.16 ± 0.21	0.88 ± 0.52	1.15 ± 0.66	0.38 ± 0.1	0.82 ± 0.44	0.32 ± 0.12	1.18 ± 0.66	1.02 ± 0.45	31.29 ± 7.78	6.33 ± 3.9	2.66 ± 1
KAH	1.32 ± 0.23	1.15 ± 0.32	1.33 ± 0.7	0.42 ± 0.11	1.02 ± 0.5	0.35 ± 0.14	<b>1.5 ± 0.71</b>	1.23 ± 0.49	<b>34.77 ± 6.92</b>	5.59 ± 3.66	1.44 ± 0.89
ZIR	1.14 ± 0.22	0.89 ± 0.64	1.11 ± 0.7	0.37 ± 0.12	0.86 ± 0.45	0.31 ± 0.14	0.74 ± 0.66	1.07 ± 0.45	31.33 ± 6.95	7.21 ± 3.52	2.54 ± 0.95
AZS	1.18 ± 0.21	0.94 ± 0.62	1.12 ± 0.68	0.37 ± 0.13	0.82 ± 0.46	0.33 ± 0.14	<b>2.2 ± 1.74</b>	1.12 ± 0.63	29.23 ± 7.23	7.51 ± 3.9	2.59 ± 1.02
MVT	1.32 ± 0.21	1.17 ± 0.61	1.34 ± 0.83	0.42 ± 0.11	1.04 ± 0.52	0.37 ± 0.13	<b>3.79 ± 1.32</b>	<b>2.22 ± 0.47</b>	<b>38.15 ± 7.47</b>	4.86 ± 3.68	4.04 ± 0.96

**Table 2.** Means ± standard errors of 19 phenotypic and 3 biochemical traits in 198 individuals from 18 natural populations of *O. mascula*. Significant values are in bold.

cal region altitude and positively correlated with longitude. The important traits, PLH, number of flower (NF) and NLP responded positively to annual mean temperature ( $r=0.7$  and  $0.6$ ,  $P<0.01$ ), the same applies to GLN ( $r=0.7$ ,  $P<0.01$ ) and weakly with longitude. BFW and BDW weakly and negatively correlated to land cover ( $r=-0.4$ ), but there was no correlation between land cover and other phenotypic traits (Fig. 4). The starch content significantly correlated with land cover and geographical altitude ( $r=0.7$ ,  $r=0.6$ ;  $P<0.01$ ) and weakly with annual precipitation ( $r=0.5$ ). Also, it was negatively correlated with longitude ( $r=-0.4$ ) and showed no correlation with annual mean temperature. However, the GLN content significantly decreased with the increase in geographical altitude, annual precipitation, and land cover ( $r=-0.7$ ,  $-0.6$ ,  $-0.6$ ,  $P<0.01$ ), respectively. A seesaw correlation was unveiled between GLN and starch ( $r=-0.9$ ,  $P<0.01$ ), as observed earlier in Table 2 and Fig. 1.



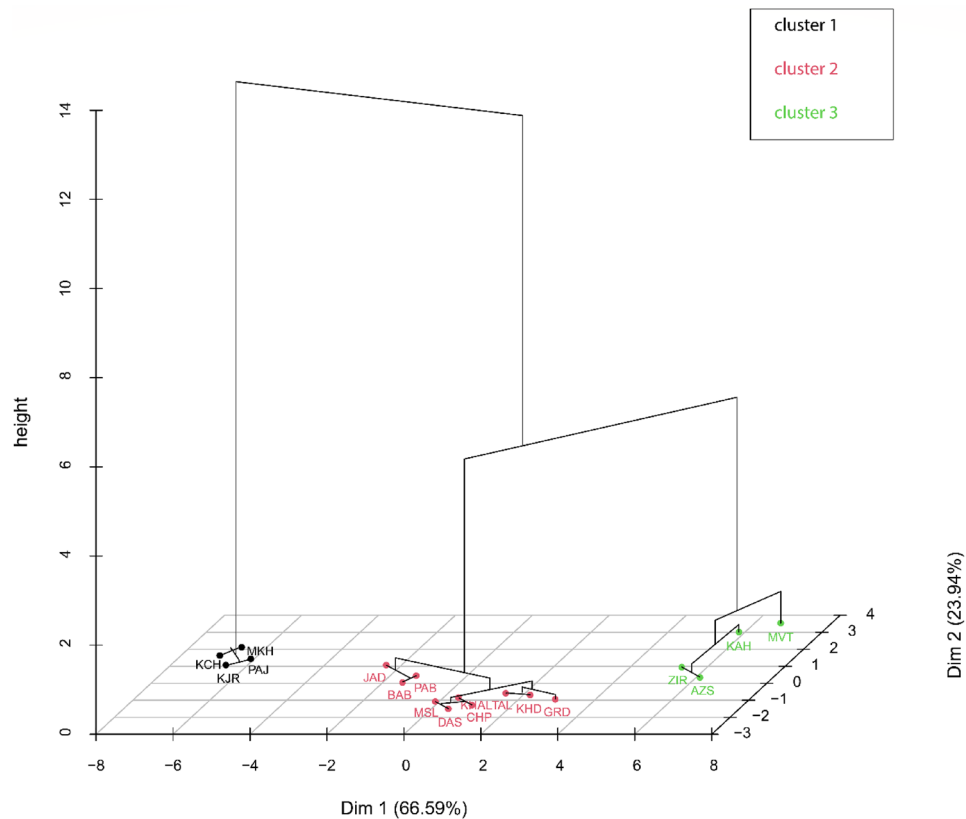
**Figure 1.** The graphs display the influence of land cover types on glucomannan (GLN; g/100 g) and starch (g/100 g) content. The difference in GLN content between pasture and woodland was significant at  $*P < 0.05$  and nonsignificant between pasture and shrubland. Starch content under woodland land cover type versus pasture and shrubland was significant at  $*P < 0.05$ .

The former compounds correlated weakly with the increase of TAC, and the latter indicated a relatively positive relationship.

**Population genetics.** *Genetic diversity and differentiation of *O. mascula* populations.* At the population level, TAS, MSL, and CHP had the highest observed alleles ( $N_a = 2$ ), while AZS had the lowest value (1.65) among populations with an average of 1.89. The highest effective alleles ( $N_e = 1.64$ ) were related to the TAS population, and the lowest  $N_e$  (1.21) belonged to the ZIR population with an average of 1.46 (Table 3). Among the eighteen studied populations, the average polymorphism percentage ( $P\%$ ) was 31.14%, while MVT from Golestan had the highest (65.63%), and PAJ was the lowest (18.75%). With an average of 0.35, the highest value of Shannon's information index ( $I$ ) was observed in the JAD population (0.49), while MSL and PBD populations (0.25) had the lowest level. An average of 0.42 expected heterozygosity ( $H_e$ ) and 0.30 for Nei's gene diversity index ( $H$ ) were observed. The parameters of  $H_e$  and  $H$  were the highest in the KAH population (0.55, 0.39, respectively), and the MSL population (0.31, 0.18, respectively) showed the lowest. The average of the populations' total genetic diversity ( $H_t$ ) was 0.37, and the genetic diversity within populations ( $H_s$ ) was 0.24.

The coefficient of genetic differentiation among populations ( $G'_{st}$ ) was 0.35, indicating that 35% of the genetic variation was among the populations, and 65% was within the populations. The  $G'_{st}$  was higher than average, considering that  $G'_{st}$  between 0.05 and 0.15 is defined as moderate and values over 0.30 as high<sup>37</sup>. Thus,  $G'_{st}$  reflected the significant genetic differentiation among the populations. These findings were consistent with the relatively low gene flow value ( $Nm = 0.46$ ) observed among populations (Table 3). The principal coordinate analysis (PCoA) for 198 individuals of *O. mascula* revealed the presence of two groups (Fig. 5). Almost all 118 individuals of populations from Guilan (7 populations) and Mazandaran (5 populations) gathered into group I, and group II contained 80 individuals from 6 populations of Golestan. The pattern observed in PCoA confirms the clustering pattern revealed in heatmap visualization of two-way cluster analyses. Analysis of molecular variance (AMOVA) showed that value for genetic diversity occurred among populations ( $\Phi_{iPR} = 0.388$ ,  $p = 0.001$ ) with a significant share of total genetic diversity (48%), whereas a larger proportion of variation (52%) existed within populations ( $\Phi_{iPT} = 0.509$ ,  $p = 0.001$ ; Table 4). Differentiation among the three regions accounted for 20% of the genetic variance that existed among the three regions ( $\Phi_{iRT} = 0.374$ ,  $p = 0.001$ ).

*Population structure.* Structure analysis divided the populations into two clusters at  $K = 2$  (Figs. 6 and 7a), Cluster I, and Cluster II (Fig. 7b). Cluster I contained 117 individuals, among which 75 were from Guilan province and 42 individuals were from Mazandaran province (Fig. 7a, red). Cluster II included 63 individuals from Golestan province (Fig. 7a, green). However, 18 Golestan and Mazandaran individuals were identified as genetic



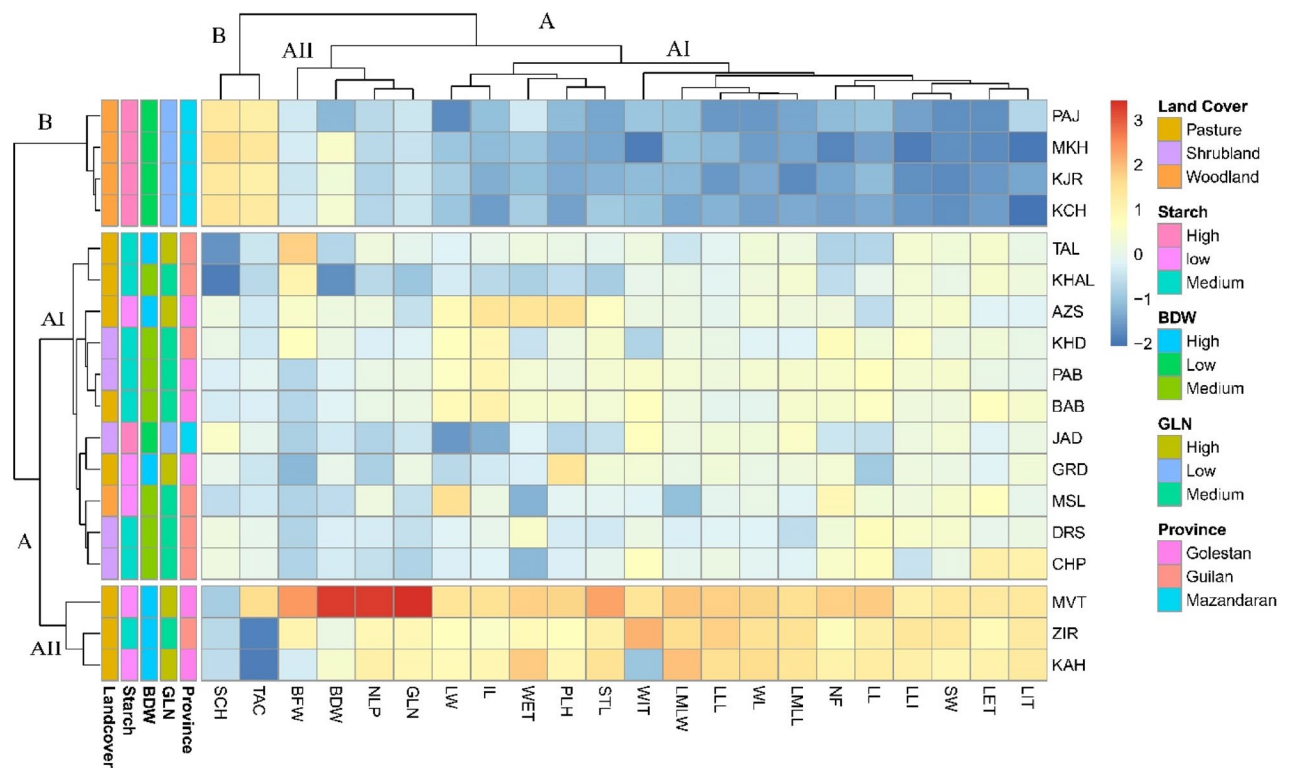
**Figure 2.** Hierarchical Clustering on Principal Components (HCPC) based on phenotypic and biochemical traits of 18 *O. mascula* natural populations. Dimension 1 explained 66.59% of the phenotypic variation, while the second dimension was only 23.94%. The populations clustered into three groups reflecting the influence of land cover types. For an explanation of abbreviations, see Tables 6 and 7.

admixtures (Fig. 7b, cyan). Structure analysis was further evaluated at  $K=4$  to see the consistency of the populations' membership, which changed to some level (Fig. 7c); Cluster I included 33 individuals, and 100% of their members came from the Guilan province. Cluster II had 24 individuals from Guilan province. Cluster III comprised 57 individuals, of which 15 were from Guilan and 42 from Mazandaran. However, this cluster had 24 individuals who shared some of them with Cluster IV (57 individuals), all belonging to Golestan populations. The admixture individuals were at the east of Mazandaran, more toward Golestan province (Fig. 7c, cyan color).

**Influence of environmental variables on genetic diversity.** Correlation analyses revealed that population polymorphism significantly decreased with the increase of annual precipitation and geographical altitude ( $r = -0.9, -0.7, p < 0.01$ , respectively) (Fig. 8), which suggested a pattern of low genetic diversity at high precipitation and elevation. The population polymorphism weakly and negatively correlated to land cover ( $r = -0.4$ ), it also weakly and positively correlated to longitude ( $r = 0.4$ ) and strongly to annual mean temperature ( $r = 0.6, p < 0.01$ ). The population Nei's genetic diversity demonstrated a negative correlation to annual precipitation ( $r = -0.6, p < 0.01$ ), and positively to longitude ( $r = 0.6, p < 0.01$ ). This parameter also potently and negatively correlated to land cover ( $r = -0.6, p < 0.01$ ), precipitation in June ( $r = -0.4$ ) and geographical altitude ( $r = -0.5$ ). The population expected heterozygosity also significantly correlated to annual precipitation ( $r = -0.7, p < 0.01$ ), negatively correlated to geographical altitude ( $r = -0.5$ ), precipitation in June ( $r = -0.3$ ), land cover ( $r = -0.4$ ), positively with longitude ( $r = 0.5$ ), annual mean temperature ( $r = 0.4$ ), mean temperature in May ( $r = 0.4$ ) and mean temperature in June ( $r = 0.4$ ) (Fig. 8).

In testing for isolation by distance (IBD), the relationship between geographical and genetic distance and geographic and phenotypic distance were insignificant (Table 5). These results indicate the lack of effect of IBD. On the contrary, in isolation by environment (IBE), relationships between genetic and phenotypic distances with environmental variables were revealed to be significant, showing the influence of essential variables such as mean temperature or annual precipitation on controlling the variation among the populations. Other tests also exhibited significant correlations between phenotypic and genetic distance. For example, in other partial Mantel tests, the correlation between residuals of genetic  $\times$  geographic with phenotypic distance or phenotypic  $\times$  genetic with geographic was nonsignificant. The dominant effect of the environment was observed in all the comparisons (Table 5).





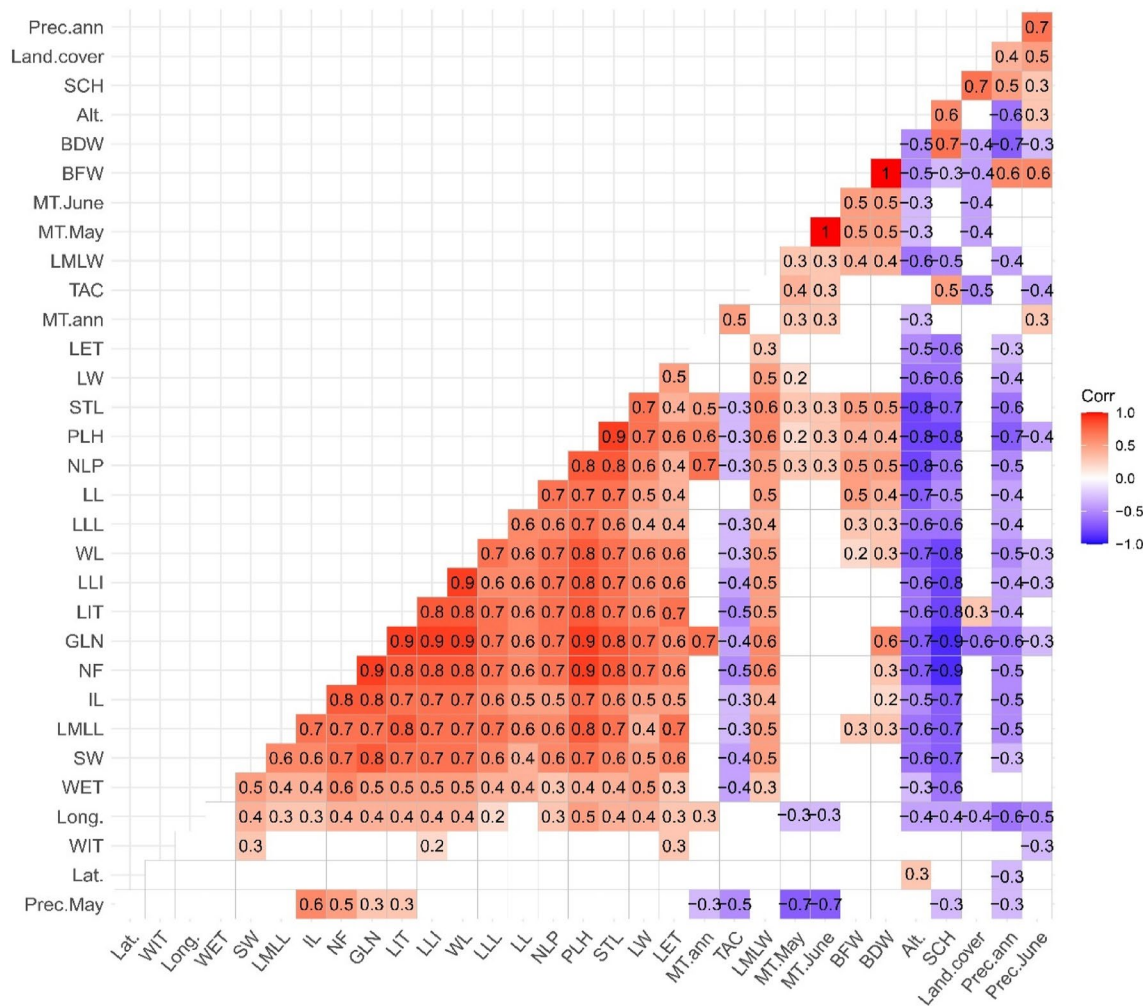
**Figure 3.** Heatmap visualization of two-way cluster analyses for 19 phenotypic and 3 biochemical traits (row tree) and 18 natural populations (column tree) evaluated in the growing season of 2020 in northern Iran, mainly Hyrcanian ecoregion. Column annotations indicate grouping based on 18 populations, 3 provinces, bulb dry weight, GLN, starch, and spike length. For an explanation of abbreviations, see Tables 6 and 7.

**Marker-trait association.** Applying two models of MLM1 and MLM2 to find a possible marker-trait association (Table S1) indicated strong comparisons of traits vs. AFLP markers with  $r^2 > 0.40$  at  $P < 0.01$ , where three of the markers revealed a simultaneous association with several important traits. For example, marker P-CCA + M-AGA-49 showed significant simultaneous associations with quantitative traits PLH, STL, NLP, LL, IL, NF, LLI, LMLL, BFW, and BDW on both MLM1 and MLM2 models (Table S1). Marker P-TGG + M-CTT-33 had concurrently significant associations with quantitative traits LW, IL, NF, LET, LIT, LLI, LMLL, LMLW, and biochemical traits GLN and TAC based on MLM1; and showed significant associations with quantitative traits of STL, LL, LW, NF, LET, LIT and biochemical trait GLN on MLM2 model. In addition, primer E-AGG + M-CGT-22 showed significant associations with quantitative traits PLH, IL, NF, WL, LMLL, and biochemical traits GLN and TAC based on MLM1. It showed significant associations with quantitative traits of PLH, SL, NLP, IL, NF, LLI, WL, and biochemical traits GLN and TAC based on MLM2 model (Table S1).

## Discussion

**Phenotypic plasticity in the context of environmental heterogeneity.** It's often asserted that relying on phenotypic markers for genetic diversity assessment is an inferior approach given the significant influence of environmental variables<sup>38–40</sup>. While this is to some extent true, phenotypic traits are still valuable tools to unmask the responsiveness of the species across climatic gradients. Such information can reflect the fundamental pattern for species biodiversity conservation and find agronomically important traits with possible application in crop improvement programs. In this study, we examined the effects of environmental factors on genetic diversity and functional traits in *O. mascula*. To uncover the impact of sources of variation at the population level on phenotypic plasticity, an unbalanced REML mixed-model ANOVA was employed, showing the notable influence of two factors, land cover and the population (Table 1).

In this study, we observed a strong pattern of phenotypic response to environmental variables, revealing the importance of land cover type and precipitation. *O. mascula* is relatively tolerant to shade; however, deep shade can significantly jeopardize its survival. Several comprehensive studies have used this species to understand the influence of light availability on reproduction performance<sup>41,42</sup>. The dominant pattern witnessed in this study was the better performance of *O. mascula* in pasturelands/grasslands of east Hyrcanian, which could be linked to the availability of light and soil nutrients (i.e., not content) as opposed to the woodlands of Mazandaran and low altitudes of Guilan. The determinative effect of light was studied by<sup>41</sup>, who found that the overall growth rate was higher in coppiced than undisturbed woodland. Although plant species respond differently to shade, many usually do not succeed in flowering under deep shade<sup>42</sup>. In addition to light, higher nutrient availability in clear-cuts and pastures<sup>43</sup> may have facilitated better growth performance.



**Figure 4.** Correlogram of 19 phenotypic and 3 biochemical traits with 9 environmental and climatic variables evaluated in 18 *O. mascula* populations in northern Iran, Hyrcanian ecoregion. For an explanation of abbreviations, see Tables 6 and 7.

High variation in floral traits, particularly in labellum length and width, has been reported recently in *Orchis purpurea* Huds.<sup>44</sup>, and in *O. mascula* and *O. pauciflora* Ten.<sup>45</sup>. The results of the current study were, to a large extent, consistent with these studies. In the comprehensive phenotypic study of Ebrahimi, et al.<sup>26</sup>, quantitative floral traits showed considerably higher variation than the results reported here; however, the number of flowers had the highest coefficient of variation (41.34%), similar to our results (62.07%, Table 1). Their study area was Abr Forest, where *O. mascula* manifested considerably higher variation among almost all qualitative and quantitative floral traits indicating high ornamental values in these populations.

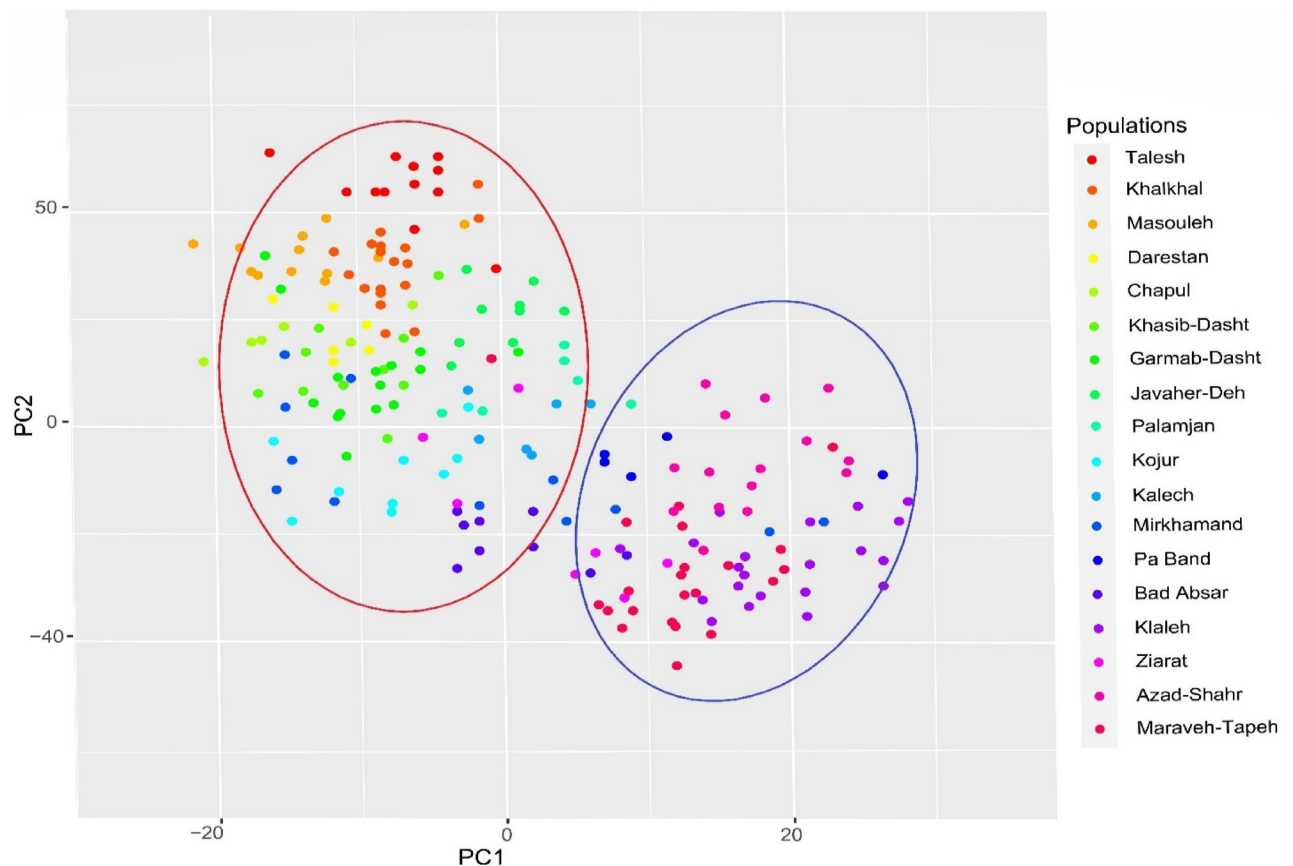
Growth and the probability of flowering are in most tuberous orchids primarily associated with access to essential nutrients and the carbohydrates stored in the belowground organs<sup>46,47</sup>. Interestingly, two primary compounds in underground organs of *O. mascula* (GLN and starch) showed an inverse relationship (Fig. 1), similar to what has been previously reported by Tekinşen and Güner<sup>48</sup> on several salep species, including *Orchis italica* Poir. and *Orchis morio* L.. The starch and GLN content reported on round-type tubers like *O. mascula* in Iran (6.15% and 22.1%, respectively)<sup>22</sup> were notably lower than our observation (Table 2; 11.13 g/100 starch and 38.15 g/100 GLN). The dry matter produced in bulbs was considerably higher in populations in the east or at high elevations in the west. High nutrient and light availability together with a higher amount of assimilated carbohydrates, can significantly contribute to better growth and reproduction. Nonetheless, in terrestrial orchids relying on stored material belowground may not be sufficient at the beginning of growth or reproduction when demand reaches its peak<sup>49</sup>. Improving local light conditions in woodland orchids by coppicing may directly or indirectly positively affect orchids<sup>50</sup> and lead to larger plants with a higher flower number and larger inflorescence size. Similarly, the dry matter under the ground can result from the increased light conditions and explain why the difference in land cover significantly influenced plant performance in this study (Fig. 4).

Another major player shaping the observed patterns of phenotypic variation is precipitation, which climatically decreased from 800 mm in the west of Hyrcanian to 250 mm in the east. The low rainfall may impose abiotic stress on *O. mascula* populations in Golestan and lead to increased phenotypic variation. It's possible that increased dry matter of tubers in areas with lower precipitation encourages the accumulation of carbohydrate



Population	$N_a$	$N_e$	P%	$I$	$H_e$	$H$	$H_t$	$H_s$	$G'_{st}$	$N_m$
TAS	2	1.64	27.84	0.37	0.39	0.25				
KHAL	1.94	1.48	38.07	0.31	0.41	0.23				
MSL	2	1.55	23.58	0.25	0.31	0.18				
DRS	1.93	1.63	25.57	0.41	0.42	0.26				
CHP	2	1.61	20.17	0.35	0.46	0.31				
KHD	1.99	1.51	25.85	0.29	0.41	0.27				
GRD	1.88	1.45	31.82	0.36	0.44	0.30				
JAD	1.89	1.55	22.44	0.49	0.48	0.3				
PAJ	1.99	1.49	18.75	0.39	0.39	0.25				
KJR	1.97	1.51	20.74	0.39	0.43	0.27				
KCH	1.95	1.43	22.44	0.42	0.46	0.29				
MKH	1.98	1.52	38.64	0.28	0.45	0.32				
PBD	1.88	1.32	28.98	0.25	0.46	0.31				
BAB	1.76	1.34	30.68	0.37	0.51	0.33				
KAH	1.73	1.36	<b>44.03</b>	0.29	<b>0.55</b>	<b>0.39</b>				
ZIR	1.67	1.21	31.25	0.35	0.52	<b>0.37</b>				
AZS	1.65	1.33	<b>44.03</b>	0.42	0.49	<b>0.35</b>				
MVT	1.69	1.38	<b>65.63</b>	<b>0.46</b>	0.52	<b>0.36</b>				
<b>Average</b>	<b>1.89</b>	<b>1.46</b>	<b>31.14</b>	<b>0.35</b>	<b>0.42</b>	<b>0.3</b>	<b>0.37</b>	<b>0.20</b>	<b>0.35</b>	<b>0.46</b>

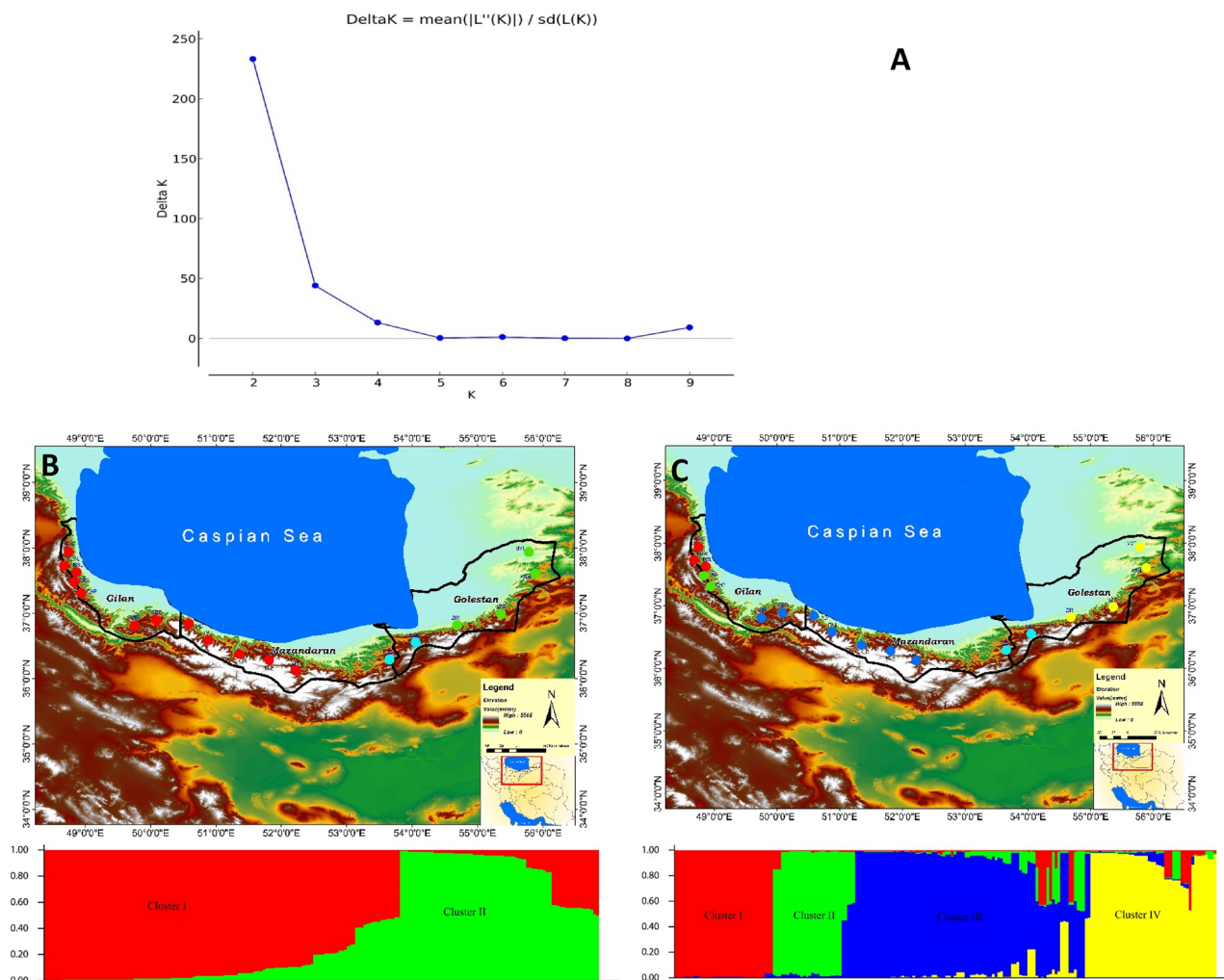
**Table 3.** Genetic diversity statistics of 18 populations of *Orchis mascula* based on AFLP. Observed number of alleles ( $N_a$ ), the effective number of alleles ( $N_e$ ), Shannon's diversity index ( $I$ ), expected heterozygosity ( $H_e$ ), Nei's gene diversity index ( $H$ ), total gene diversity ( $H_t$ ), gene diversity within populations ( $H_s$ ), coefficient of genetic differentiation ( $G'_{st} = (H_t - H_s)/H_t$ ), gene flow ( $N_m$ ). Significant values are in bold.



**Figure 5.** Principal coordinates analysis of 198 individuals of 18 populations of *Orchis mascula* based on the genetic similarity matrix derived from AFLP markers divided the populations into two groups. For an explanation of abbreviations, see Table 6.

	Source	d.f	Sum of squares	Variance components	% of variation	Phi statistics	P value
Three types of Land cover	<b>Among groups</b>	2	1735.907	10.171	20	<b>PhiRT: 0.347</b>	0.001
	<b>Among Pops</b>	15	2964.190	16.076	31	<b>PhiPR: 0.388</b>	0.001
	<b>Within Pops</b>	180	4555.086	25.306	49	<b>PhiPT: 0.509</b>	0.001
	<b>Total</b>	197	9255.182	51.553			

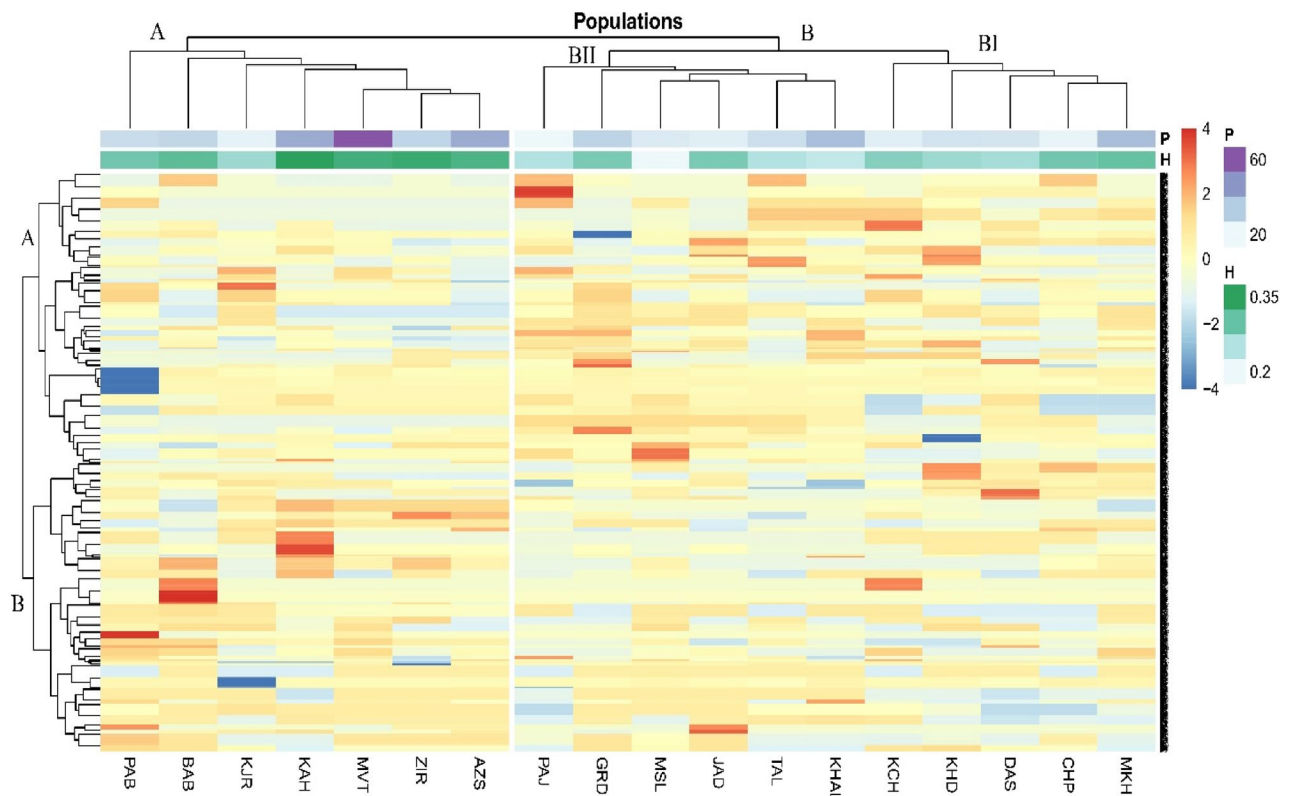
**Table 4.** Result of analysis of molecular variance (AMOVA) for 18 natural populations of *Orchis mascula*. Significant values are in bold.



**Figure 6.** Map of 18 populations of and Bayesian admixture proportions identified by STRUCTURE of 198 individual plants of *Orchis mascula*. (A) The optimum number of subpopulations was determined using  $\Delta K$  in the Bayesian clustering method,  $K=2$ ; (B)  $K=2$  and (C)  $K=4$ . The colors used in the maps are approximately associated with STRUCTURE clusters. Cyan-colored circles represent the 'admixed' populations. For an explanation of abbreviations, see Table 6.

compounds in *O. mascula*. Aggregation of carbohydrate compounds is a common drought tolerance strategy<sup>51</sup>. Low predictability of precipitation may contribute to a relative increase in floral trait variation on the eastern side. March-Salas, et al.<sup>52</sup> reported enhanced within-individual and within-population variability in flowering phenology by less foreseeable precipitation patterns in *Onobrychis viciifolia* Scop. populations in a multigeneration experiment.

**Population genetics and structure.** Preserving genetic diversity is crucial since it can buffer against upcoming short-term and long-term climate-change-driven stress, disturbance, inbreeding depression, and interactions with hostile species<sup>2,53,54</sup>. In situations where high genetic diversity can be a significant advantage, the population faces unpredictable levels of abiotic stress in the environment, such as high temperature or water

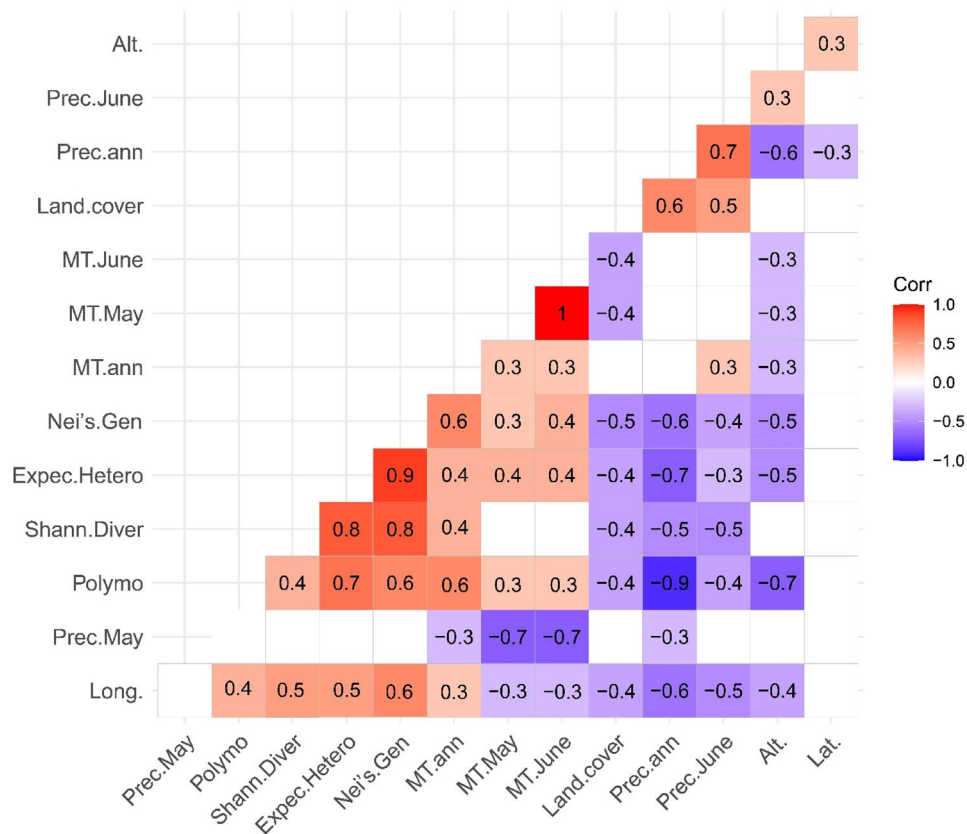


**Figure 7.** Heatmap visualization of two-way cluster analyses for 18 natural populations of *Orchis mascula* (row tree) and alleles generated by AFLP markers (column tree) evaluated in the growing season of 2020 in north of Iran, Hyrcanian ecoregion. For an explanation of abbreviations, see Tables 3 and 6.

stress<sup>8</sup>. Population genetic diversity is the reflection of mating system, habitat fragmentation, and climatic elements, which survival of plant species depends on as the greater genetic diversity is, the higher chance of coping with various climatic changes.

In this study, we found that genetic variation among *O. mascula* populations follows the pattern of environmental heterogeneity and is significantly affected by climatic variables such as temperature, precipitation, and land cover. Genetic diversity ( $H$ ) at the population level with an average of 0.30 was observed to be higher than the previous studies on *O. mascula*; for example, Jacquemyn, et al.<sup>41</sup> reported an average of 0.16 or Gholami, et al.<sup>31</sup> 0.10. However, the latter used ISSR markers with a small sample size (5). Strong geographical affinity seemed to be followed by the pattern of genetic diversity among populations, analogous to phenotypic variation, the level of genetic diversity was notably higher in the populations of Golestan, where KLH had the highest  $H$  (0.39) and other populations from this province similarly showed high  $H$  (Table 3). As an orchid species with a mixed breeding system, *O. mascula* in this study was possibly affected by several factors. Land cover type appears to have been a significant player because the genetic variation among populations was considerably higher in pastureland/grassland populations. This response could be supported by the light availability of the eastern populations against woodland of the west and Mazandaran province; brighter conditions mean the prevalence of pollinators in addition to a higher number of flowers (as was observed in phenotypic diversity). Thus, the possibility of pollinators visiting more flowers available before leaving due to the food-deceiving nature of this species is higher; however, the gene flow ( $Nm$ ) mean was low.

The genetic structuring of populations was strong in  $K=2$  and  $K=4$  (Figs. 6 and 7a), further supporting the high genetic differentiation ( $G'_{st}$ : 0.35, Table 3). It could be argued that the large geographic distance from the west to the east of the Hyrcanian ecoregion ( $\approx 500$  km) strongly contributes to the limited gene flow. However, this does not seem to be the case, as we found no evidence for IBD. On the other hand, the genetic structure mainly supports the prominent effect of the environment on shaping this pattern among populations. Low gene flow is often caused by habitat loss and fragmentation, leading to reductions in population size and increased spatial isolation<sup>55,56</sup>. Low within-population genetic diversity (52%) and high differentiation ( $G'_{st}$ : 0.35) in this study are contrary to previous results reported on *O. mascula*, showing a high genetic diversity within population (92%) and low  $G'_{st}$  (0.125)<sup>41</sup>. Moreover, the woodland type did not affect genetic partitioning. This is in sharp contrast with the results of this study, where 20% percent of the total genetic diversity was partitioned among three groups of land covers (woodland, shrubland, and pastureland). Differences in scale can explain these contrasting results. Whereas Jacquemyn, et al.<sup>57</sup> studied 15 populations in the local landscape with slight variation in climatic conditions, in our study, populations were studied across a much larger scale leading to strong differences in climatic conditions between eastern and western populations and significant IBE. This suggests



**Figure 8.** Correlogram of genetic diversity indices and geocology variable. Abbreviations: precipitation in May and June (Prec. June, Prec.May); Polymorphism (%; Polymo.); Shannon Diversity Index (Shann.Diver); Expected heterozygosity (Expec.Hetero); Nei's genetic diversity (Nei's.Gen); Mean Temperature annual (MT.ann), in May and June (MT.June and MT.May); Annual precipitation(Prec.ann); Altitude-(Alt.), Longitude(Long.). For an explanation of other abbreviations, see Table 3.

Simple Mantel tests	<i>r</i>	<i>P</i> -value
Genetic × Geographic	0.112	n.s
Genetic × Environment	<b>0.556</b>	0.001
Phenotypic × Geographic	0.159	n.s
Phenotypic × Environment	<b>0.531</b>	0.001
Phenotypic × Genetic	<b>0.492</b>	0.013
Partial Mantel tests		
Genetic × Geographic   Phenotypic	0.098	n.s
Genetic × Environment   Phenotypic	<b>0.506</b>	0.001
Phenotypic × Genetic   Geographic	0.269	n.s
Phenotypic × Environment   Genetic	<b>0.518</b>	0.001

**Table 5.** Correlations between Nei's genetic distance, phenotypic distance, climatic (19 environmental variables), and geographic (km) differences among 18 populations of *Orchis mascula* tested with simple and partial Mantel tests for IBD and IBE. Significant values are in bold.

that populations have genetically adapted to the different precipitation regimes across Hyrcania and that this genetic adaptation has contributed to the observed high genetic differentiation. These results further support the findings of Siepielski, et al.<sup>58</sup>, who showed that precipitation was a dominant driver of natural selection in plant and animal populations for over 150 species.

**Implication for breeding efforts.** Given its economic importance, *O. mascula* population sizes have been reduced beyond recovery in many areas in the west, northwest, and north of Iran. Developing a breeding program could be a sustainable option to counteract the further loss of populations. Selection of the most suitable

Populations	Sample no	Latitude (°N)	Longitude (°E)	Altitude (m)
<b>Guilan</b>				
Talesh(TAS)	12	37° 53' 48.95" N	48°44'55.37" E	1432
Khalkhal(KHAL)	18	37° 59' 9.93" N	48° 28' 2.86" E	1888
Masouleh(MSL)	12	37° 19' 22.07" N	48° 57' 23.32" E	1275
Darestan(DAR)	6	36° 58' 11.24" N	48° 55' 37.56" E	1113
Chapul(CHP)	6	37° 8' 30.17" N	49° 9' 32.94" E	752
Khasib-Dasht(KHD)	9	36° 56' 46.81" N	50° 4' 1.99" E	1976
Garmab-Dasht(GRD)	12	36° 53' 4.28" N	50° 13' 1.44" E	697
<b>Mazandaran</b>				
Javaher-deh(JAD)	9	36° 51' 46.56" N	50° 27' 20.62" E	2632
Palamjan(PAJ)	6	36° 45' 44.12" N	50° 30' 58.92" E	1852
Kojur(KJR)	9	36° 22' 50.82" N	51° 45' 57.01" E	1609
Kalech (KCH)	6	36° 16' 50.21" N	51° 39' 38.44" E	2841
Mirkhamand(MKH)	12	36° 19' 7.95" N	51° 53' 19.53" E	2513
<b>Golestan</b>				
Pa Band(PBD)	6	36° 32' 43.74" N	54° 11' 51.16" E	2208
Bad Absar(BAB)	9	36° 26' 3.46" N	53° 55' 36.74" E	1284
Klaleh(KAH)	21	37° 42' 0.83" N	55° 57' 31.31" E	830
Ziarat(ZIR)	9	36° 40' 27.36" N	54° 30' 4.76" E	1538
Azad-Shahr(AZS)	15	37° 6' 16.54" N	55° 23' 7.23" E	1845
Maraveh-Tapeh(MVT)	21	37° 51' 25.82" N	55° 33' 19.18" E	394

**Table 6.** Details of geographical locations where populations of *Orchis mascula* were sampled.

populations and cultivation of local plant material could represent the first steps in a sustainable management of *O. mascula* populations in the wild. Our results showed that populations from Golestan possessed several important desirable traits, including a high GLN content and bulb dry weight that meets the high-quality threshold<sup>48</sup>. Another approach to probe for favorable traits is the percent of variation (%var) populations display<sup>59</sup>. These traits mentioned above also had the highest variation percentage under the effects of land cover and population (Table 1). A phenotypic study on *O. mascula* populations by<sup>26</sup> also indicated high variation for tuber characteristics, BFW & BDW in particular.

Additionally, marker-assisted phenotypic trait selection may provide heritable traits with high economic value. In this context, marker-trait associations based on MLM1 and MLM2 models revealed significant associations of P-CCA + M-AGA-49 with important traits such as PLH, BFW, and BDW, P-TGG + M-CTT-33 and E-AGG + M-CGT-22 with PLH and GLN (Table S1). The associated traits observed here are important in the adaptability of *O. mascula* as they were found to be most responsive in the populations coping with stressful conditions. These linked markers with another quantitative trait might also show a pleiotropic effect or may be located very closely<sup>60</sup>. Marker-trait association in terrestrial orchids, including *O. mascula*, has been previously reported by Gholami, et al.<sup>31</sup>, who also observed a significant association between ISSR markers and tuber dry weight. Notable traits with high potential for selection have been identified in this study. However, it should be noted that while all measurements were conducted on samples collected from natural populations, a common garden experiment can provide better insights into this trait relationship and its reliability for further application.

Our study illustrates the functional trait variation in response to environmental variables in relation to genetic diversity, offers empirical evidence that supports a combination of different datasets in population-level studies, and observes the interactions from a broader perspective to gain a more reliable estimation. Readily measurable functional traits and their incorporation with population genetics provide a deeper understanding of species' ability to adapt to environmental changes and facilitate using this variation for applications such as plant breeding<sup>61</sup>. Our findings in *O. mascula* can be used as the foundation of a plant breeding program that allows for the cultivation of *O. mascula* in an agricultural setting, thereby reducing the demand for wild-collected tubers. Given the current limitations of the legal protection provided for *O. mascula*, this may be the best way to conserve natural populations of this species, as well as other species of tuberous orchids.

## Material and methods

**Study sites and sampling.** In the spring of 2020, eighteen populations of *O. mascula* were sampled in the Hyrcanian ecoregion (Table 6), following Flora Iranica and Flora of Iran<sup>62,63</sup>. This is an exceptionally rich biodiversity area containing relict species in Tertiary period forests covering extensive regions between Iran and Azerbaijan. Anthropogenic threats to wild species in these areas include exploitation of natural populations and habitat destruction through the construction of housing and industrial activity. It should be noted that in 2020 accessing the populations of *O. mascula* for this study was possible because graduate researchers are automatically allowed to pursue research within their field of study, exempt from common regulations to collect plant material unless the plant of interest is in the protected, disputed areas near borders or the collection is for com-



Quantitative traits	Biochemical traits	Primer sequences 5' → 3'
1. Plant height (PLH; cm)	Glucomannan content (GLN; g/100 g)	P-TGG+M-CTT
2. Stem length (SL; cm)	Starch content (g/100 g)	P-TGG+M-TTG
3. Number of leaves/plant (NLP)	Total content of anthocyanin (mg/gr) (TAC)	P-CCA+M-AGA
4. Leaf length (LL; cm)		P-GCA+M-CTC
5. Leaf width (LW; cm)		P-GCA+M-AGT
6. Inflorescence length (IL; cm)		P-ACC+M-AGA
7. Number of flowers (NF; cm)		P-GTT+M-CTT
8. Length of external tepals (LET; cm)		P-CCA+M-CTT
9. Width of external tepals (WET; cm)		P-GTT+M-ATC
10. Length of internal tepals (LIT; cm)		P-ACC+M-AGT
11. Width of internal tepals (WIT; cm)		E-AAC+M-CTT
12. Length of lip (LL; cm)		E-AAC+M-TGT
13. Width of lip (WL; cm)		E-AAC+M-CCA
14. Length of median the lobe of the lip (LMLL; cm)		E-AAG+M-TGT
15. Length of Median lobe width of the lip (LMLW; cm)		E-ACA+M-AAG
16. Length of the lateral lobe (LLL; cm)		E-ACC+M-CAG
17. Sidelobe width (SW; cm)		E-AGA+M-CAG
18. Bulb Fresh weight (BFW; g)		E-AGC+M-CGA
19. Bulb Dry weight (BDW; g)		E-AGC+M-CTT
		E-AGG+M-CGT

**Table 7.** Characteristics used in the phenotypic and biochemical analyses of *Orchis mascula* populations. List of primers' combinations used in this study is also presented in the last column.

mercial purpose. In this case, our study was a conservation/breeding attempt entirely in line with the current rules to access endangered plant species; thus, official permission or license was not needed. However, in Jul. 2021, new regulations were announced (the complete list can be found at <https://rc.majlis.ir/fa/law/show/1667090>) in which access to any wild population requires official permission. Dr. Gholizadeh identified the species identity of the samples, and voucher specimens from the populations were deposited in the University of Guilan Herbarium (GUM; no. 5631 to 5649), available for botanical studies with an official request.

A total of 198 individuals from 18 populations, with at least 20 m distance between individuals in each population, were collected from these populations. Nineteen quantitative phenotypic traits related to the size of the plants and flowers (Table 7) were recorded for each individual. A digital caliper was used to measure the quantitative traits with a precision of 0.01 mm. Using a digital scale, the fresh weight of bulbs was recorded, and subsequently, bulbs were dried in the oven at 72 °C and measured for dry weight afterward.

Young and fresh leaves were collected from each individual, immediately dried with silica gel, and transported to the laboratory for DNA extraction. For each population, the land cover was characterized by assessing the percentage of vegetation (e.g., woodland, shrubland, and pastureland/grassland). Using Google Earth Pro®, high-resolution aerial images of the locations were acquired. The interpretation of the images was based on visual classification using the ArcGIS 10.4® software (URL: <https://www.esri.com/en-us/arcgis/products/arcgis-desktop/resources>), a program used to calculate the areas and percentages of land use and land cover by each category.

**Chemotyping.** *Glucomannan and starch contents.* To produce salep powder from bulbs, a traditional method previously described by Tekinşen and Güner<sup>48</sup> was used. For this purpose, the bulbs were washed with cold water, cleaned from dirt and mud, and boiled for 10–15 min in milk; then, the samples were dried at 21 ± 2 °C until they hardened (7–10 days). The dried specimens were first cut into small pieces and then powdered into flour by milling. Further analysis was carried out to measure the components of interest in the flour.

The samples were prepared according to procedures specified for GLN and starch using GLN (K-GLUM 10/04) and total starch (AA/AMG 11/01, AOAC Method 996.11) assay kits (Megazyme International Ireland Limited, 2004a, b). The GLN and starch contents of the samples were determined by measurement of absorbance values of prepared blind and sample solutions (for GLN A<sub>1</sub>, A<sub>2</sub>, A<sub>3</sub>; for starch ΔA, F) in a UV-Vis Spectrophotometer (Shimadzu – UV Mini 1240) at 340 nm (for GLN) and 510 nm (for starch) and using the following formula 1 and 2, respectively<sup>64,65</sup>:

$$\Delta A_{\text{glucomannan}} = (A_3 - A_1)_{\text{sample}} - (A_3 - A_1)_{\text{blank}} \times 36.8 \text{ [g/100 g]}$$

$$\text{Starch} = \Delta A \times (F/W) \times 90 \text{ [g/100 g]}$$

where ΔA = absorbance value of sample solution compared with blind solution; F = 100 (μg of glucose control)/absorbance value of glucose control (1.03); and W = the weight of the sample (100 mg). The final content was reported as g/100 g of dry matter.

**Total anthocyanin content.** To measure tepals and labellum's anthocyanin content (purple pigment), the pigment was first extracted with 0.5 mL methanol/0.1% HCl for 24 h in the dark. Then the absorbance was read at 510 nm using a UV-Vis spectrophotometer (UV1601; Shimadzu, Kyoto, Japan)<sup>66</sup>. The anthocyanin content is expressed as mg per gram of fresh weight.

**Genotyping.** Genomic DNA was extracted from silica gel dried young leaves of the individuals utilizing CTAB procedure<sup>67</sup>. The quality and quantity of extracted DNA were assayed by 1% agarose gel electrophoresis and NanoDrop<sup>®</sup> spectrophotometer and diluted with sterile distilled water to give a final 100 ng/μL. The method described by Vos, et al.<sup>68</sup> was employed with slight modification for AFLP analysis. An amount of 250 ng genomic DNA (per sample) was homogenized with EcoRI and MseI primers without a selective base in pre-amplification and with three selective bases in amplification in 20 μL master mix (2 μL Buffer, 2 μL 10 X BSA, 0.5 μL MseI, 0.5 μL EcoRI, 4.5 μL double-distilled water). The obtained combination of genomic DNA and restriction enzymes were incubated for 12 h at 30 °C. Then double-stranded EcoRI (Eco) and MseI (Mse) linkers were ligated to the restriction fragments without additional nucleotides and afterward amplified by primers with three additional nucleotides at the 3'-end (Table 1). In selective amplification, the combinations of primers were used. The PCR Thermo cycles were carried out with a program of 35 three-step cycles, including 94 °C for 1 min, 56 °C for 30 s, and 72 °C for 1 min. On 6% acrylamide gel in vertical electrophoresis, Bio-Rad, the DNA fragments were separated, and to reveal the bands, the silver staining method was applied<sup>69</sup>. The banding pattern of the amplified DNA fragments was turned into binary codes as presence (1) or absence (0), and only the consistently repeatable bands were scored. Weak or smeared bands were excluded, and the fragments of the same molecular weight were counted as the same locus.

**Data analysis. Phenotypic data.** A principal component analysis (PCA) was performed on the phenotypic data. The first two principal components were retained as they explained the highest portion of the total variation and were treated like traits and subjected to further analysis. To determine how phenotypic variation is partitioned between different types of land covers (fixed effect), among populations (random effect), and among individuals (error), an unbalanced restricted maximum likelihood (REML) mixed-model analysis of variance (ANOVA) was conducted for the 22 measured traits and PC1 and PC2 (Table 1). Means ± standard errors of the traits based on each population were also provided (Table 2). The analyses were implemented in PROC MIXED (SAS Institute, 2008, Cary, NC).

Hierarchical Clustering on Principal Components (HCPC) using "Factoextra" and "FactoMineR" packages<sup>70,71</sup> was used to assess relationships among phenotypic traits. The phenotypic diversity expressed by the sampled *O. mascula* population was further explored with a two-way clustering analysis using the "Heatmaply" package in the R software<sup>72</sup>. Given the difference in the scaling of the traits, each trait was normalized using Z-score before cluster analysis.

A correlation analysis between phenotypic traits was carried out using Pearson's correlation coefficients ( $P < 0.01$ ) using the "CorrPlot" package<sup>73</sup>. The same package was utilized to assess the relationship between genetic diversity indices and environmental factors (average temperature, annual precipitation, and during the growing seasons, altitude, longitude, and latitude).

**Molecular data. Genetic diversity.** Genetic variation indices including the observed number of alleles ( $N_a$ ), the effective number of alleles ( $N_e$ ), expected heterozygosity ( $H_e$ ), Shannon's diversity index ( $I$ ), Nei's genetic diversity ( $H$ ), total heterozygosity ( $H_t$ ), and intra-population genetic diversity ( $H_s$ ) were computed using POPGENE v3.0<sup>74</sup> (Table 3). All these calculations assumed that populations were in Hardy-Weinberg equilibrium.

**Genetic structure.** A Cluster (heatmap) analysis was performed at the population level based on the matrix of Nei's unbiased genetic distance using "ggplot2"<sup>75</sup> and "Pheatmap"<sup>76</sup> packages in the R software. A principal coordinate analysis (PCoA) was carried out using the "logisticPCA" package<sup>77</sup> in the R software to visualize the spatial structure. Bayesian model-based cluster analysis using STRUCTURE v2.3.4<sup>78</sup> was also used to infer genetic structure and define the number of clusters in the data set. The correlated allele frequencies and admixed model were applied with 50,000 burn-in and 100,000 MCMC. The assumed number of groups ( $K$ ) varied from 1 to 10, and 10 runs per  $K$  were performed. The STRUCTURE HARVESTER<sup>79</sup> determined optimum  $K$  based on  $L(K)$  and  $\Delta K$ . The structure result at  $K = 2$  and  $K = 4$  were mapped according to geographical locations using the ArcGIS 10.4<sup>®</sup> software (URL: <https://www.esri.com/en-us/arcgis/products/arcgis-desktop/resources>).

**Genetic differentiation.** To assess the magnitude of genetic differentiation, the partitioning of total genetic variation at different scales was computed with analysis of molecular variance (AMOVA) in "vegan"<sup>80</sup> in the R software. Shannon differentiation coefficient ( $G'_{st}$ ) was calculated based on the following formula,  $G'_{st} = (H_{sp} - H_{pop})/H_{sp}$  ( $H_{sp}$ , total Shannon information index;  $H_{pop}$ , average Shannon information index within the population), and the gene flow ( $Nm$ ) among populations was estimated using POPGENE v3.0<sup>74</sup>.

**Isolation by distance/isolation by environment.** To test the effects of geographic distance and environmental differences on the genetic and phenotypic structure, the genetic distances among populations were calculated using Nei's genetic distance implemented in R package "Poppr"<sup>81</sup> and phenotypic distance using the "dist" function in R software. Euclidean geographic distances were calculated using the "fossil" package<sup>82</sup>. For sampling sites, climatic variables were collected from WorldClim version 2.067. Bioclimatic variables (19) were considered as different

environmental space vectors and used the Canberra distance to calculate the distance between populations in this vector space. For isolation by distance (IBD) and isolation by environment (IBE), the relationship of genetic and phenotypic distances with geographic and environmental conditions was analyzed, respectively, using simple and partial Mantel tests implemented in the R package "vegan"<sup>80</sup>. Additionally, all other possible comparisons (geographic, phenotype, genetic, and environment) were tested.

**Trait-marker association analysis.** Two models were compared to test the marker-trait association between AFLP and phenotypic and chemotypic traits using TASSEL 4.0.1. (<http://www.maizegenetics.net/bioinformatics>)<sup>83</sup>; a Mixed Linear Model (MLM) using the kinship matrix (K) estimated by the STRUCTURE HARVESTER<sup>79</sup> as a random effect (MLM1) and an MLM using both Q and K (MLM2). Results were compared to determine a better model. Significance of associations between loci and traits was described as P-values (a probability level of 0.001).

## Data availability

The datasets used and/or analyzed during the current study available from the corresponding author on reasonable request.

Received: 1 February 2022; Accepted: 6 September 2022

Published online: 17 November 2022

## References

1. Read, Q. D., Moorhead, L. C., Swenson, N. G., Bailey, J. K. & Sanders, N. J. Convergent effects of elevation on functional leaf traits within and among species. *Funct. Ecol.* **28**, 37–45. <https://doi.org/10.1111/1365-2435.12162> (2014).
2. Isbell, F. *et al.* Biodiversity increases the resistance of ecosystem productivity to climate extremes. *Nature* **526**, 574–577. <https://doi.org/10.1038/nature15374> (2015).
3. Vellend, M. & Geber, M. A. Connections between species diversity and genetic diversity. *Ecol. Lett.* **8**, 767–781. <https://doi.org/10.1111/j.1461-0248.2005.00775.x> (2005).
4. Hart, S. P., Schreiber, S. J. & Levine, J. M. How variation between individuals affects species coexistence. *Ecol. Lett.* **19**, 825–838. <https://doi.org/10.1111/ele.12618> (2016).
5. Paschke, M. C. & Schmid, B. Relationship between population size, allozyme variation, and plant performance in the narrow endemic *Cochlearia bavarica*. *Conserv. Genet.* **3**, 131–144 (2002).
6. Soleimani, V., Baum, B. & Johnson, D. A. AFLP and pedigree-based genetic diversity estimates in modern cultivars of durum wheat [*Triticum turgidum* L. subsp. durum (Desf.) Husn.]. *Theor. Appl. Genet.* **104**, 350–357. <https://doi.org/10.1007/s001220100714> (2002).
7. Hughes, A. R., Inouye, B. D., Johnson, M. T., Underwood, N. & Vellend, M. Ecological consequences of genetic diversity. *Ecol. Lett.* **11**, 609–623. <https://doi.org/10.1111/j.1461-0248.2008.01179.x> (2008).
8. Prati, D., Peintinger, M. & Fischer, M. Genetic composition, genetic diversity and small-scale environmental variation matter for the experimental reintroduction of a rare plant. *J. Plant Ecol.* <https://doi.org/10.1093/jpe/rtv067> (2016).
9. Barrett, R. D. & Schluter, D. Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**, 38–44. <https://doi.org/10.1016/j.tree.2007.09.008> (2008).
10. Atwater, D. Z. & Callaway, R. M. Testing the mechanisms of diversity-dependent overyielding in a grass species. *Ecology* **96**, 3332–3342. <https://doi.org/10.1890/15-0889.1> (2015).
11. Cook-Patton, S. C., McArt, S. H., Parachnowitsch, A. L., Thaler, J. S. & Agrawal, A. A. A direct comparison of the consequences of plant genotypic and species diversity on communities and ecosystem function. *Ecology* **92**, 915–923. <https://doi.org/10.1890/10-0999.1> (2011).
12. Whitney, K. D. *et al.* Experimental drought reduces genetic diversity in the grassland foundation species *Bouteloua eriopoda*. *Oecologia* **189**, 1107–1120. <https://doi.org/10.1007/s00442-019-04371-7> (2019).
13. Violle, C. *et al.* Let the concept of trait be functional!. *Oikos* **116**, 882–892. <https://doi.org/10.1111/j.0030-1299.2007.15559.x> (2007).
14. Kattge, J. *et al.* TRY plant trait database-enhanced coverage and open access. *Glob. Change Biol.* **26**, 119–188. <https://doi.org/10.1111/gcb.14904> (2020).
15. König, P. *et al.* Advances in flowering phenology across the Northern Hemisphere are explained by functional traits. *Glob. Ecol. Biogeogr.* **27**, 310–321 (2018).
16. Robinson, K. M., Ingvarsson, P. K., Jansson, S. & Albrechtsen, B. R. Genetic variation in functional traits influences arthropod community composition in aspen (*Populus tremula* L.). *PLoS ONE* **7**, e37679. <https://doi.org/10.1371/journal.pone.0037679> (2012).
17. Karbstein, K., Prinz, K., Hellwig, F. & Römermann, C. Plant intraspecific functional trait variation is related to within-habitat heterogeneity and genetic diversity in *Trifolium montanum* L. *Ecol. Evol.* **10**, 5015–5033. <https://doi.org/10.1002/ece3.6255> (2020).
18. Kaya, S. & Tekin, A. R. The effect of salep content on the rheological characteristics of a typical ice-cream mix. *J. Food Eng.* **47**, 59–62. [https://doi.org/10.1016/S0260-8774\(00\)00093-5](https://doi.org/10.1016/S0260-8774(00)00093-5) (2001).
19. Ktistis, G. & Georgakopoulos, P. P. Rheology of salep mucilages. *Pharmazie* **46**, 55–56 (1991).
20. Kayacier, A. & Dogan, M. Rheological properties of some gums-salep mixed solutions. *J. Food Eng.* **72**, 261–265. <https://doi.org/10.1016/j.jfoodeng.2004.12.005> (2006).
21. Sen, M. A., Palabiyik, I. & Kurultay, S. The effect of saleps obtained from various Orchidaceae species on some physical and sensory properties of ice cream. *Food Sci. Technol.* **39**, 82–87. <https://doi.org/10.1590/fst.26017> (2019).
22. Farhoosh, R. & Riazi, A. A compositional study on two current types of salep in Iran and their rheological properties as a function of concentration and temperature. *Food Hydrocoll.* **21**, 660–666. <https://doi.org/10.1016/j.foodhyd.2006.07.021> (2007).
23. Ghorbani, A., Zarre, S., Gravendeel, B. & de Boer, H. J. Illegal wild collection and international trade of CITES-listed terrestrial orchid tubers in Iran. *Traffic Bullet.* **26**, 52–58 (2014).
24. Lenoir, J., Gégout, J.-C., Marquet, P., De Ruffray, P. & Brisse, H. A significant upward shift in plant species optimum elevation during the 20th century. *Science* **320**, 1768–1771. <https://doi.org/10.1126/science.1156831> (2008).
25. Chen, Z.-Q., Algeo, T. J. & Fraiser, M. L. Organism-environment interactions during the Permian-Triassic mass extinction and its aftermath. *Palaio* **28**, 661–663. <https://doi.org/10.2110/palo.2012.p12-102r> (2013).
26. Ebrahimi, A. *et al.* Evaluation of phenotypic diversity of the endangered orchid (*Orchis mascula*): Emphasizing on breeding, conservation and development. *S. Afr. J. Bot.* **132**, 304–315. <https://doi.org/10.1016/j.sajb.2020.05.013> (2020).
27. Ghorbani, A., Gravendeel, B., Naghibi, F. & de Boer, H. Wild orchid tuber collection in Iran: A wake-up call for conservation. *Biodivers. Conserv.* **23**, 2749–2760. <https://doi.org/10.1007/s10531-014-0746-y> (2014).

28. Barrett, S. C. & Kohn, J. R. *The application of minimum viable population theory to plants. Genetics and conservation of rare plants* 3–1 (Oxford University Press, 1991).
29. Yun, S. A., Son, H.-D., Im, H.-T. & Kim, S.-C. Genetic diversity and population structure of the endangered orchid *Pelatantheria scolopendrifolia* (Orchidaceae) in Korea. *PLoS ONE* **15**, e0237546. <https://doi.org/10.1371/journal.pone.0237546> (2020).
30. Gholami, S., Vafaee, Y., Nazari, F. & Ghorbani, A. Exploring genetic variations in threatened medicinal orchids using start codon targeted (SCoT) polymorphism and marker-association with seed morphometric traits. *Physiol. Mol. Biol. Plants* **27**, 769–785. <https://doi.org/10.1007/s12298-021-00978-4> (2021).
31. Gholami, S., Vafaee, Y., Nazari, F. & Ghorbani, A. Molecular characterization of endangered Iranian terrestrial orchids using ISSR markers and association with floral and tuber-related phenotypic traits. *Physiol. Mol. Biol. Plants* **27**, 53–68. <https://doi.org/10.1007/s12298-020-00920-0> (2021).
32. Kaki, A., Vafaee, Y. & Khadivi, A. Genetic variation of *Anacamptis coriophora*, *Dactylorhiza umbrosa*, *Himantoglossum affine*, *Orchis mascula*, and *Ophrys schulzei* in the western parts of Iran. *Ind. Crops Prod.* **156**, 112854. <https://doi.org/10.1016/j.indcrop.2020.112854> (2020).
33. Falk, D., & Holsinger, K. E. Genetic sampling guidelines for conservation collections of endangered plants (1991).
34. Dempewolf, H. *et al.* Past and future use of wild relatives in crop breeding. *Crop Sci.* **57**, 1070–1082. <https://doi.org/10.2135/cropsci2016.10.0885> (2017).
35. Renz, J. *Flora Iranica*. Part 126: Orchidaceae (1978).
36. Shahsavari, A. *Flora of Iran*. Part 57: Orchidaceae (2008).
37. Boulila, A., Béjaoui, A., Messaoud, C. & Boussaid, M. Genetic diversity and population structure of *Teucrium polium* (Lamiaceae) in Tunisia. *Biochem. Gen.* **48**, 57–70. <https://doi.org/10.1007/s10528-009-9295-6> (2010).
38. Zannou, A., Struik, P., Richards, P., Zoundjih, E. & Yam, J. (*Dioscorea* spp.) responses to the environmental variability in the Guinea Sudan zone of Benin. *Afr. J. Agric. Res.* **10**, 4913–4925. <https://doi.org/10.5897/AJAR2013.8099> (2015).
39. Sujii, P. *et al.* Morphological and molecular characteristics do not confirm popular classification of the Brazil nut tree in Acre, Brazil. *Genet. Mol. Res.* <https://doi.org/10.4238/2013.september.27.3> (2013).
40. Nadeem, M. A. *et al.* DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnol. Biotechnol. Equip.* **32**, 261–285. <https://doi.org/10.1080/13102818.2017.1400401> (2018).
41. Jacquemyn, H., Brys, R., Adriaens, D., Honnay, O. & Roldán-Ruiz, I. Effects of population size and forest management on genetic diversity and structure of the tuberous orchid *Orchis mascula*. *Conserv. Genet.* **10**, 161–168. <https://doi.org/10.1007/s10592-008-9543-z> (2009).
42. Mitchell, P. & Woodward, F. Responses of three woodland herbs to reduced photosynthetically active radiation and low red to far-red ratio in shade. *J. Ecol.* <https://doi.org/10.2307/2260575> (1988).
43. Likens, G. E., Bormann, F. H., Johnson, N. M., Fisher, D. & Pierce, R. S. Effects of forest cutting and herbicide treatment on nutrient budgets in the Hubbard Brook watershed-ecosystem. *Ecol. Monog.* **40**, 23–47. <https://doi.org/10.2307/1942440> (1970).
44. Jacquemyn, H. & Brys, R. Lack of strong selection pressures maintains wide variation in floral traits in a food-deceptive orchid. *Ann. Bot.* **126**, 445–453. <https://doi.org/10.1093/aob/mcaa080> (2020).
45. Olaya-Arenas, P., Meléndez-Ackerman, E. J., Pérez, M. E. & Tremblay, R. Demographic response by a small epiphytic orchid. *Am. J. Bot.* **98**, 2040–2048. <https://doi.org/10.3732/ajb.1100223> (2011).
46. Primack, R. B., Miao, S. & Becker, K. R. Costs of reproduction in the pink lady's slipper orchid (*Cypripedium acaule*): Defoliation, increased fruit production, and fire. *Am. J. Bot.* **81**, 1083–1090. <https://doi.org/10.2307/2446500> (1994).
47. Whigham, D. F. & O'Neill, J. P. Dynamics of flowering and fruit production in two eastern North American terrestrial orchids. In *Tipularia Discolor and Liparis Lilifolia in Population Ecology of Terrestrial Orchids* (eds Wells, T. C. E. & Willems, J. H.) 89–101 (SPB Academic Publishers, 1991).
48. Tekinşen, K. K. & Güner, A. Chemical composition and physicochemical properties of tubera salep produced from some Orchidaceae species. *Food Chem.* **121**, 468–471. <https://doi.org/10.1016/j.foodchem.2009.12.066> (2010).
49. Whigham, D. F. Biomass and nutrient allocation of *Tipularia discolor* (Orchidaceae). *Oikos* <https://doi.org/10.2307/3544398> (1984).
50. Mattila, E. & Kuitunen, M. T. Nutrient versus pollination limitation in *Platanthera bifolia* and *Dactylorhiza incarnata* (Orchidaceae). *Oikos* **89**, 360–366. <https://doi.org/10.1034/j.1600-0706.2000.890217.x> (2000).
51. Xu, W. *et al.* Drought stress condition increases root to shoot ratio via alteration of carbohydrate partitioning and enzymatic activity in rice seedlings. *Acta Physiol. Plant.* **37**, 9. <https://doi.org/10.1007/s11738-014-1760-0> (2015).
52. March-Salas, M., Fandos, G. & Fitze, P. S. Effects of intrinsic environmental predictability on intra-individual and intra-population variability of plant reproductive traits and eco-evolutionary consequences. *Ann. Bot.* **127**, 413–423. <https://doi.org/10.1093/aob/mcaa096> (2021).
53. Jump, A. S., Marchant, R. & Peñuelas, J. Environmental change and the option value of genetic diversity. *Trends Plant Sci.* **14**, 51–58. <https://doi.org/10.1016/j.tplants.2008.10.002> (2009).
54. Huang, J. *et al.* Global semi-arid climate change over last 60 years. *Clim. Dyn.* **46**, 1131–1150. <https://doi.org/10.1007/s00382-015-2636-8> (2016).
55. Crémieux, L., Bischoff, A., Müller-Schärer, H. & Steinger, T. Gene flow from foreign provenances into local plant populations: Fitness consequences and implications for biodiversity restoration. *Am. J. Bot.* **97**, 94–100. <https://doi.org/10.3732/ajb.0900103> (2010).
56. Garant, D., Forde, S. E. & Hendry, A. P. The multifarious effects of dispersal and gene flow on contemporary adaptation. *Funct. Ecol.* **21**, 434–443. <https://doi.org/10.1111/j.1365-2435.2006.01228.x> (2007).
57. Jacquemyn, H. *et al.* Multigenerational analysis of spatial structure in the terrestrial, food-deceptive orchid *Orchis mascula*. *J. Ecol.* **97**, 206–216. <https://doi.org/10.1111/j.1365-2745.2008.01464.x> (2009).
58. Siepielski, A. M. *et al.* Precipitation drives global variation in natural selection. *Science* **355**, 959–962. <https://doi.org/10.1126/science.aag2773> (2017).
59. Ene, C. O., Ogbonna, P. E., Agbo, C. U. & Chukwudi, U. P. Studies of phenotypic and genotypic variation in sixteen cucumber genotypes. *Chilean J. Agric. Res.* **76**, 307–313. <https://doi.org/10.4067/S0718-58392016000300007> (2016).
60. Pradhan, S. K. *et al.* Population structure, genetic diversity and molecular marker-trait association analysis for high temperature stress tolerance in rice. *PLoS ONE* **11**, e0160027. <https://doi.org/10.1371/journal.pone.0160027> (2016).
61. Swarup, S. *et al.* Genetic diversity is indispensable for plant breeding to improve crops. *Crop Sci.* **61**, 839–852. <https://doi.org/10.1002/csc2.20377> (2021).
62. Patzak, A. *Plantaginaceae in KH Rechinger Flora Iranica 15: 1–21* (Academische Druck und Verlagsantalt, 1965).
63. Mehrvarz Saeidi, S. *Plantaginaceae Family* Vol. 14 (Research Institute of Forests and Rangelands, 1995).
64. Limited, M. I. I. Glucomannan assay procedure KGLUM 10/04. Ireland (2004).
65. Limited, M. I. I. Total starch assay procedure (amyloglucosidase/a-Amylase Method) AA/AMG 11/01. AOAC Method 996.11. Ireland (2004).
66. Bradshaw, H., Otto, K. G., Frewen, B. E., McKay, J. K. & Schemske, D. W. Quantitative trait loci affecting differences in floral morphology between two species of monkeyflower (*Mimulus*). *Genetics* **149**, 367–382. <https://doi.org/10.1093/genetics/149.1.367> (1998).
67. Del Sal, G., Manfioletti, G. & Schneider, C. The CTAB-DNA precipitation method: A common mini-scale preparation of template DNA from phagemids, phages or plasmids suitable for sequencing. *Biotechniques* **7**, 514–520 (1989).



68. Vos, P. *et al.* AFLP: A new technique for DNA fingerprinting. *Nucl. Acid Res.* **23**, 4407–4414. <https://doi.org/10.1093/nar/23.21.4407> (1995).
69. Bassam, B. J., Caetano-Anollés, G. & Gresshoff, P. M. Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Biochem.* **196**, 80–83. [https://doi.org/10.1016/0003-2697\(91\)90120-1](https://doi.org/10.1016/0003-2697(91)90120-1) (1991).
70. Husson, F., Josse, J., Le, S., Mazet, J. & Husson, M. F. Package 'FactoMineR'. *An R package* **96**, 698 (2016).
71. Lê, S., Josse, J. & Husson, F. FactoMineR: An R package for multivariate analysis. *J. Stat. Softw.* **25**, 1–18 (2008).
72. Galili, T. in *The R User Conference, useR! 2017 July 4–7 2017 Brussels, Belgium*. 219.
73. Wei, T. *et al.* Package 'corrplot'. **56**, e24 (2017).
74. Yeh, F. POPGENE (version 1.3. 1). Microsoft Window-Bases Freeware for Population Genetic Analysis. <http://www.ualbertaca/~fyeh/> (1999).
75. Wickham, H. & Chang, W. URL: <http://CRAN.R-project.org/package=ggplot2>. An implementation of the Grammar of Graphics. **3** (2008).
76. Kolde, R. & Kolde, M. R. Package 'pheatmap'. *R package* **1**, 790 (2015).
77. Landgraf, A. J. & Lee, Y. Dimensionality reduction for binary data through the projection of natural parameters. *J. Mult. Anal.* **180**, 104668. <https://doi.org/10.1016/j.jmva.2020.104668> (2020).
78. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959. <https://doi.org/10.1093/genetics/155.2.945> (2000).
79. Earl, D. A. & VonHoldt, B. M. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Biol. J. Linn. Soc.* **4**, 359–361. <https://doi.org/10.1007/s12686-011-9548-7> (2012).
80. Oksanen, J. *et al.* Vegan: Community Ecology Package. R package version 2.4-3 (2016).
81. Kamvar, Z. N., Tabima, J. F. & Grünwald, N. J. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *Peer. J.* **2**, e281 (2014).
82. Vavrek, M. J. Fossil: Palaeoecological and palaeogeographical analysis tools. *Palaeontol. Electron.* **14**, 16. <https://doi.org/10.7717/peerj.281> (2011).
83. Bradbury, P. J. *et al.* TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635. <https://doi.org/10.1093/bioinformatics/btm308> (2007).

## Acknowledgements

The authors thank Prof. Daniel Potter and Reed Kenny for improving the English and their constructive comments.

## Author contributions

Conceptualization: M.M. and M.B.; Methodology: H.G. and M.R.; Investigation: M.M., M.B. and N.R.; Visualization: M.M., N.R., M.O. and A.M.; Writing – Original Draft Preparation: M.O., M.B. and A.M.; Writing – Review & Editing: M.M., A.M. and M.R.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-19864-4>.

**Correspondence** and requests for materials should be addressed to M.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022