CHILEANJAR
CHILEAN JOURNAL of AGRICULTURAL RESEARCH

RESEARCH

Genetic diversity, population structuring and expansion history of *Coronilla scorpioides*

Lamia Hakim[1], Souhir Kabtni[1], Dorra Sdouga[1], Neila Trifi-Farah[1], and Sonia Marghali[1*]

[1]Université de Tunis El Manar, Faculté des Sciences Mathématiques Physiques et Naturelles de Tunis, LR99ES12, Campus Universitaire, 2092 Tunis, Tunisie. *Corresponding author (sonia.marghali@fst.utm.tn).

Received: 21 July 2021; Accepted: 22 November 2021; doi:10.4067/S0718-58392022000100065

ABSTRACT

The Tunisian area flora has a diverse and interesting vegetation cover. To date, many species, however, are not sufficiently explored and properly exploited. Despite their medicinal potential, *Coronilla scorpioides* (L.) W.D.J. Koch species has still not been studied. In this study, we aim to investigate the genetic diversity and the population structuring of 10 Tunisian populations. Ten DNAs pooled represented the populations of *C. scorpioides* species are used in this molecular analysis. Bulked segregant analysis (BSA) was used in order to identify a genetic polymorphism associated to a specific trait by comparing DNA pools with contrasting phenotypes by sequence-related amplified polymorphism (SRAP) markers. Moreover, 10 DNA samples for each population were examined by five retrotransposon-microsatellite amplified polymorphism (REMAP) combination in order to elucidate the genetic composition of this species. Our results suggested that *C. scorpioides* species have a high genetic diversity that can be related to their geographical distribution. This genetic richness could be the source of their agronomic variability in which two contrasted populations originated from Kairouan (P210 and P224) were genetically distinguished by the BSA-SRAP method. The occurrences of evolution were elucidated, and the ancestral populations were revealed. As the first research based on the genetic structuring and diversity of *C. scorpioides*, our finding illustrates the affluences' biodiversity of Tunisian area and the genetic richness of the studied species. This study could be the genetic guideline of *C. scorpioides* species that can be used to meet the needs of medicine and the national economy.

Key words: BARE1, biodiversity, *Coronilla scorpioides*, genetic evolution, SRAP markers.

INTRODUCTION

Several species of the Fabaceae family are known to have high nutritional value and they are consumed in the forms of pasture seeds, oil seeds or dried nuts, or as a fibre sources (Amiriyan et al., 2019). *Coronilla scorpioides* (L.) W.D.J. Koch, rarely noted as scorpion's tail, is a species that belongs to the family of Fabaceae plants, genus *Coronilla*. Since antiquity, various benefits of this genus have been mentioned, involving its use as a medicinal plant due to the occurrence of cardiac glycosides including furanocoumarins and coumarins, in several species (Zoghlami and Zouaghi, 2003). Indeed, among the Fabaceae family, furanocoumarins, especially psoralen, bergapten, xanthotoxin, are produced only by two genera, which are *Psoralea* and *Coronilla* genera (Szewczyk and Bogucka-Kocka, 2012). In particular, *C. scorpioides* is the best natural source for extracting psoralen which treats many dermatological diseases such as psoriasis and vitiligo (Innocenti et al., 1996). Despite its biochemical potential and its contribution to pasture development, this species is poorly studied and still until now unexploited (Zoghlami and Zouaghi, 2003). *Coronilla scorpioides* is an annual species preferentially autogamous with a chromosomal number of $2n = 12$, spreads in the crop field, pasture, and the undergrowth (Zoghlami and Zouaghi, 2003; Abdelguerfi and Abdelguerfi-Laouar, 2004; Cano et al., 2012). In terms of geographical distribution, the species could be found in the Mediterranean basin as well as in the south of Europe, Macaronesia, and the Middle East (Cano et al., 2012). All previous studies dealt only with the distribution, biochemical potential and morphological variability of this

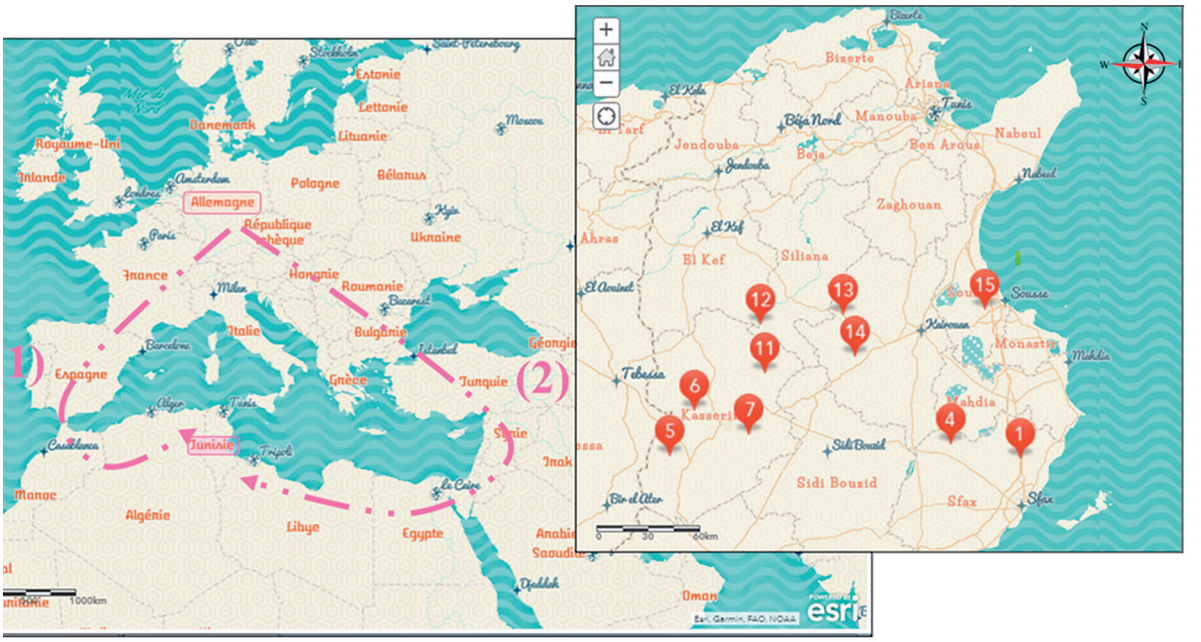
species. In the Tunisian territory, this species spreads from the higher semi-arid region to the lower arid region; although the difference between semi-arid and arid climate may be minor, the biochemical composition of plants varies between these areas (Kabtni et al., 2020). This adaptation to varied climate could be related to higher genome variability. According to Zoghlami and Zouaghi (2003), Tunisian populations of *C. scorpioides* have shown great morphological variation, which could be related to their genetic diversity (Zoghlami and Zouaghi, 2003). Indeed, the populations from Kairouan (P224 and P210) represented two contrasting populations across agronomic traits such as biomass production and DM digestibility. The population P210 was characterized by an important biomass production and a weak DM digestibility. In contrast, the population P224 which presented the lowest rate concerning the biomass production and a high level of digestibility of the DM in vitro (Zoghlami and Zouaghi, 2003). Moreover, it is important to remember that the genetic variation of *C. scorpioides* has not been investigated yet and no information is available about the genome variability of this species. Knowledge of genetic variability and diversity remains fundamental to guarantee and ensure the long-term survival of a given species. Therefore, the degree of genetic diversity could be considered as an indicator of the adaptability of a plant to its environment. To validate this hypothesis, it is important to examine the genetic variability of some populations of *C. scorpioides* (Yulita et al., 2020). This study aims to investigate the genetic diversity of 10 natural *C. scorpioides* populations selected in the Tunisian area. In addition, we attempt to select genotypes, which correspond to reproducible electrophoretic profiles with clear bands by the use of sequence-related amplified polymorphism (SRAP) markers. As well as, the geographical distribution and the evolution events of this species are closely studied using retrotransposon-microsatellite amplified polymorphism (REMAP) makers. Results from this study could provide valuable information to develop an exploitation strategy of this species and to enhance its natural potential as well as its genetics richness.

MATERIAL AND METHODS

Plant material

Ten spontaneous populations of *Coronilla scorpioides* (L.) W.D.J. Koch were selected from the Tunisian area from the semiarid to the arid regions (Figure 1). All the details about the collection sites of these populations are summarized in Table 1. Ten plants of each population were randomly selected for DNA extraction.

Figure 1. Geographical distribution of *Coronilla scorpioides* in Tunisia and their sites of collection.



Numbers in the maps of Tunisia correspond to the sites of collection. Purple arrow with numbers in parentheses indicates the first (1) and the second (2) hypotheses of the geographical expansion event of *C. scorpioides* from Germany to Tunisia.

Table 1. Geographical distribution of *Coronilla scorpioides* populations and their sites of collection in Tunisia.

Sites of collection	Code	Populations	Localisation	GPS coordinates	Bioclimate	Habitat
1	P5	Sfax	Hencha	35°06'16.5" N, 10°44'09.8" E	Upper arid	Field of barley
4	P39	Sfax	Menzel Hached	35°10'50.5" N, 10°17'13.2" E	Upper arid	Field of wheat
5	P81	Kasserine	Lac lafiel	35°07'14.1" N, 8°27'58.1" E	Upper arid	Hill
6	P95	Kasserine	Foussena	35°21'22.0" N, 8°37'19.8" E	Upper arid	Field of wheat
7	P102	Kasserine	Sbitla	35°13'51.1" N, 8°58'23.8" E	Upper arid	Watercourses
11	P149	Kasserine	Sbiba	35°33'06.4" N, 9°04'45.6" E	Semi-arid lower	Orchard
12	P159	Siliana	Rouhia	35°48'25.0" N, 9°03'00.5" E	Upper semi-arid	Field of wheat
13	P210	Kairouan	Ouesletia	35°51'29.1" N, 9°35'19.7" E	Semi-arid lower	Ravine
14	P224	Kairouan	Haffouz	35°38'25.6" N, 9°39'45.9" E	Upper arid	Wooden orchard
15	P285	Sousse	Kala kebira	35°53'01.7" N, 10°30'28.5" E	Semi-arid lower	Pasture

The populations codes have been assigned by Zoghlami and Zouaghi (2003).

DNA extraction

The DNA extraction was carried out from seeds according to the modified CTAB method (Murray and Thompson, 1980). Briefly, 500 µL extraction buffer (0.35 M sorbitol; 0.1 M Tris-pH (8.0); 5 mM EDTA; 2 M NaCl and 2 g L^{-1} CTAB) were added to seeds already reduced on powder. After incubation in ice for 5 min, 300 µL CTAB lysis buffer, 1 µL RNAase solution, and 1 µL Protease K solution were added to the mixture. After incubation in 65 °C for 1 h, 300 µL chloroform were added and the mixture was gently shaken. Then, the mixture was centrifuged at 10 000 rpm for 20 min at 20 °C. The supernatant was pipetted and transferred in a new Eppendorf tube to which 500 µL isopropanol were added. The solution was mixed by inversion until the appearance of DNA filaments. Following this step, samples were left overnight at -20 ºC. Subsequently, samples were centrifuged at 10 000 rpm for 20 min. The pellet was then well dried and suspended in 30 µL Tris-EDTA 10:1 (Murray and Thompson, 1980). The concentration of DNA as well as the DNA purity were carried out using a spectrophotometer (NanoDrop ND-1000; Thermo Fisher Scientific, Waltham, Massachusetts, USA).

Sequence-related amplified polymorphism (SRAP) polymerase chain reaction (PCR)

The SRAP primers combinations used in this study were randomly selected from those published by Castonguay et al. (2010) (Table 2). The PCRs were performed in a 25 µL reaction volume containing 3 µL genomic DNA (35 ng), 25 mM dNTP, 0.08 mM each primer, 3 µL DreamTaq Green buffer (10X) containing $MgCl_2$, and 2.5 U DreamTaqTM DNA polymerase (Thermo Fisher Scientific, Glen Burnie, Maryland, USA). The PCRs were achieved in a thermal cycler (TProfessional TRIO Thermocycler; Biometra, Göttingen, Germany) referring to the following program: An initial denaturation was done at 94 °C for 5 min. The first five cycles consisted on denaturation, annealing and extension steps, which were applied respectively at 94 °C for 1 min, 35 °C for 1 min, and 72 °C for 1 min. Then, another 35 cycles were run with a new annealing temperature 46 °C. The final extension was induced at 72 °C for 10 min after of which the temperature reaction was maintained at 4 °C. The PCR products were separated on 2.5% agarose gels and visualized under UV light using ethidium bromide (BET) and GelDoc 2000 image analysis system (Bio-Rad, Hercules, California, USA). Compared to a DNA ladder, all bands were scored and scripted under (1) presence or (0) absence of the particular band.

Table 2. The used sequence-related amplified polymorphism (SRAP) primers sequences.

Forward primers		Reverse primers	
Names	Primers sequences	Names	Primers sequences
F8	5' GTA GCA CAA GCC GGA AT 3'	R9	5' GAC ACC GTA CGA ATT TGA 3'
F14	5' TGA GTC CAA ACC GGA CC 3'	R15	5' CGC ACG TCC GTA ATT CCA 3'
F13	5' CGA ATC TTA GCC GGC AC 3'	R8	5' GAC ACC GTA CGA ATT TGC 3'
F11	5' CGA ATC TTA GCC GGA TA 3'	R7	5' GAC ACC GTA CGA ATT GAC 3'
F12	5' TGA GTC CAA ACC GGA GC 3'	R14	5' CGC ACG TCC GTA ATT AAC 3'

Retrotransposon-microsatellite amplified polymorphism (REMAP) PCR reactions

The REMAP PCR reactions were performed using 50 ng genomic DNA, 25 mM dNTPs, 0.08 mM each primer, 3 µL 10X DreamTaq Green buffer containing $MgCl_2$ and 2.5 U DreamTaqTM DNA polymerase (Thermo Fisher Scientific). The final reaction volume was adjusted to 25 µL with ultrapure water (R UltraPure, pH 7, DNase-free, protease-free, RNase-free, Invitrogen, Waltham, Massachusetts, USA). One fixed reverse primer (LTR reverse 7286: 5'-GGAATTCATAGCATGGATAATAAACGATTATC-3') with seven forward arbitrary SSR primers generate seven REMAP combinations (Table 3). All REMAP combinations were tested with two PCR programs (Table 4). The PCR products were separated by electrophoresis on 1.8% agarose gel. Staining and visualization of the gel, and band scoring were the same as well as described for SRAP markers.

Bulked segregant analysis

Bulked segregant analysis (BSA) was used in conjunction with SRAP analysis to screen the genetic diversity of 10 *C. scorpioides* populations. This method consists of the bulks of equivalent DNA amounts of 10 plants randomly selected from a single population. Ten-pooled DNA representatives of each population were used in the bulked segregation analysis to track the inter populations' diversity.

Data analysis

The amplified fragments scored as binary matrices were used for analysis. The number of multilocus genotypes (MLG), Shannon-Wiener index of MLG diversity (Shannon and Weaver, 1949), Simpson's index (Simpson, 1949), Evenness, and Nei's unbiased gene diversity index (Nei, 1978) were estimated. By evaluating their resolving power, the ability of the most insightful primers to discriminate between the genotypes was assessed as follows: $Rp = \Sigma\ Ib$ ($Ib = 1-(2x\ |0.5-p|)$; p is the proportion of individuals containing the allele). In order to estimate the informativeness of the marker, the polymorphism information content (PIC) was calculated for each primer combination by the formula as follows: $PIC = 1 - \Sigma\ (i = 1)\ Pi^2$ (Pi: the frequency of i alleles). The genetic diversity indexes, the population's differentiation, and the populations structuring were computed by R program (R Foundation for Statistical Computing, Vienna, Austria) using adegenet, ade4, and ape packages. The principal component analysis (PCA), phylogenic and bayesian cluster analysis were computed by adegenet packages (Jombart et al., 2020) implemented in R program (version 3.5.1) (Fox, 2005).

Table 3. Polymorphism and incidence of used retrotransposon-microsatellite amplified polymorphism (REMAP) primers combinations (long terminal repeats [LTR] reverse 7286 + simple sequence repeat [SSR] primers).

Primers combination	Arbitrary SSR marker code	SSR motifs	Successful PCR program	Na	Np	%P	Rp	PIC
AB	B 8386	$(GTG)_7C$	Program II	9.00	9.00	100	10.694	0.887
AD	D 8081	$(GA)_9C$	Program II	5.00	5.00	100	3.854	0.790
AG	G 8387	$(CA)_{10}G$	Program I	6.00	6.00	100	2.320	0.790
AH	H 8565	$GT(CAC)_7$	Program I	7.00	7.00	100	6.306	0.832
AE	E 8385	$(CAC)_7G$	Program I	9.00	9.00	100	7.395	0.857
AC	C 8082	$(CT)_9G$	No one	-	-	-	-	-
AF	F 8564	$(CAC)_7T$	No one	-	-	-	-	-

A-H: Long terminal repeats (LTR) reverse primers 7286; PCR: polymerase chain reaction; Na: total number of amplified bands; Np: number of polymorphic bands; %P: percentage of polymorphic bands; Rp: resolving power; PIC: polymorphism information contents.

Table 4. Polymerase chain reaction (PCR) programs for retrotransposon microsatellite amplified polymorphism (REMAP) markers.

Program I
94 °C for 1 min 50 s 35 cycles: 94 °C for 30 s, 58 °C for 45 s, 72 °C for 1 min 72 °C for 5 min
Program II
92 °C for 5 min 40 cycles: 92 °C for 45 s, 56 °C for 45 s, 72 °C for 1 min 72 °C for 10 min

RESULTS

BSA analysis SRAP primers diversity

The 15 SRAP primers combinations generated 111 bands, of which 107 have been classified as polymorphic. In fact, the total number of bands (Na) are ranged from five for F12R14, F11R15, F13R8 primers to 11 for F13R9 and F12R9 primers combinations (Table 5). Moreover, the lowest number of generated polymorphic bands (Np = 2) are scored for the primer combination F11R15. In total, all other SRAP markers have generated a high number of polymorphic bands with a maximum of 11 loci for F13R9 and F12R9 primers. Most of these markers produce an important PIC. However, the resolving power are varied with an average of 8.186.

Population's diversity and clustering of *C. scorpioides*

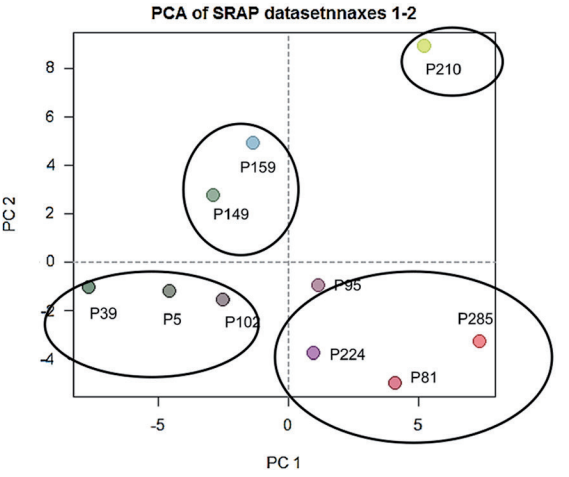
According to the PC1 of the PCA analysis (Figure 2), the populations from Kasserine (P81 and P95), from Kairouan (P210 and P224), and from Sousse (P285) are differentiated from the other populations by the presence of some specific markers. In fact, the involvement of several primers' combinations F12R7, F13R14, and F12R15, is actually associated

Table 5. Genetic diversity index using the bulked segregant analysis (BSA)-sequence-related amplified polymorphism (SRAP) method.

SRAP combinations	Na	Np	%P	Rp	PIC
F13R9	11.00	11.00	100	12.20	0.898
F12R9	11.00	11.00	100	10.40	0.898
F11R8	9.00	8.00	88.88	13.60	0.884
F12R7	6.00	6.00	100	6.80	0.826
F13R14	7.00	7.00	100	7.40	0.823
F11R7	6.00	6.00	100	6.00	0.791
F12R14	5.00	5.00	100	4.20	0.730
F13R7	10.00	10.00	100	11.40	0.880
F11R9	6.00	6.00	100	6.00	0.768
F11R14	8.00	8.00	100	9.00	0.853
F12R8	7.00	7.00	100	6.60	0.830
F11R15	5.00	2.00	40	8.40	0.789
F13R15	9.00	9.00	100	8.80	0.868
F12R15	6.00	6.00	100	6.20	0.797
F13R8	5.00	5.00	100	5.80	0.789

Na: Number of bands; Np: number of polymorphic bands; %P: percentage of polymorphic bands; Rp: resolving power; PIC: polymorphism information content.

Figure 2. Principal component analysis (PCA) of 10 populations using bulked segregant analysis (BSA)-sequence-related amplified polymorphism (SRAP) method.



Each population has been represented by point with a specific colour and the circles indicate the groups.

with PC1 construction. However, the second PCA axis (PC2), was defined by the presence of loci generated from primers' combinations F12R9, F11R7, F12R8, F13R7, and F12R15. The PCA analysis differentiated *C. scorpioides* populations into four groups. The population from Kairouan (P210) presented the first group. This group is genetically contrasted with the third group composed by populations from Sfax (P39, P5), and from Kasserine (P102). The second group was formed by populations from Kasserine (P95, P81), Kairouan (P224), and Sousse (P285). The fourth group is constituted by populations from Kasserine (P149) and from Siliana (P159). The PCA analysis did not differentiate *C. scorpioides* populations according to their geographical distributions.

Based on Euclidean distances, *C. scorpioides* populations were differentiated into four clusters by the Neighbour-joining method (Figure 3). The populations P159 (Siliana) and P210 (Kairouan) were genetically closest to each other, and formed cluster I. The population P149 (Kasserine) constituted the cluster II, which is nearest to cluster I. In addition, the cluster III is composed by three sub-clusters (III.a, III.b, and III.c): The cluster III.a (P95: Kasserine) and the cluster III.b (P224: Kairouan) are the closest sub-clusters. Nevertheless, the cluster III.c, the most distant, was formed by two populations P81 (Kasserine) and P285 (Sousse). Likewise, cluster IV was divided into two sub-clusters: The cluster IV.a and cluster IV.b, which are respectively formed by population from Kasserine (P102) and populations from Sfax (P5 and P39). It is important to mention that clusters III and IV are deviated from cluster I by the same genetic distance.

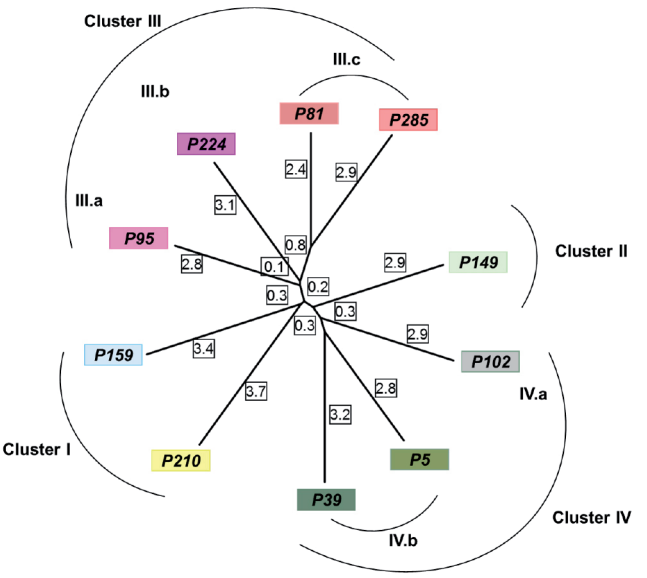
Evolution analysis and population structure, REMAP polymorphism

Seven REMAP primers combinations were tested with two PCR programs (Table 4) and only two combinations (7286+8082 and 7286+8564) did not show any amplification. Five REMAP primers combinations generated 36 distinguished, clear and polymorphic banding patterns. The total number of bands as well as the number of polymorphic bands varied according to the primer combinations with an average of 7.2 per primer (Table 3). Indeed, the total number of bands (Na) changed from 6 for the primer combination AG to 9 for AB and AE primer pairs. Besides, the results showed a considerable resolving power (Rp) for the AB combinations (Table 4). However, the AG and the AD combinations showed the lowest resolving power. The use of REMAP markers revealed 36 loci in which there were 22 qualified as uninformative loci generated by the following primers combinations AB AD AH and AE (Table 3).

Population genotyping and genetic variability

To evaluate the genetic diversity of each population, genetics indexes were calculated using only the informative locus (Table 6). Our results revealed 65 genotypes dispersed in 100 individual and constructed 10 *C. scorpioides* populations.

Figure 3. Neighbour joining cladogram using Euclidian distance between studied populations.



The colours correspond to ten populations represented in the principal component analysis (PCA) analysis.

Table 6. Genetic diversity index using the retrotransposon-microsatellite amplified polymorphism (REMAP) markers.

Populations	N	MLG	H	Lambda	E.5	Hexp
P285	10	10	2.30	0.900	1.000	0.360
P224	10	10	2.30	0.900	1.000	0.433
P210	10	8	2.03	0.860	0.934	0.268
P159	10	10	2.30	0.900	1.000	0.314
P149	10	9	2.16	0.880	0.952	0.463
P102	10	9	2.16	0.880	0.952	0.314
P95	10	6	1.64	0.780	0.850	0.198
P81	10	4	1.31	0.720	0.945	0.276
P39	10	8	2.03	0.860	0.934	0.248
P5	10	9	2.16	0.880	0.952	0.254
Total	100	65	3.86	0.967	0.628	0.406

N: Number of individuals; MLG: number of multilocus genotypes; H: Shannon-Wiener index of MLG diversity (Shannon and Weaver, 1949); Lambda: Simpson's index (Simpson, 1949); E.5: Evenness; Hexp: Nei's unbiased gene diversity (Nei, 1978).

The genotypic richness measured as the number of observed genotypes, was high in populations P285 (Sousse), P224 (Kairouan), and P159 (Siliana), while the lowest value was recorded in populations from Kasserine (P81 and P95). The Shannon index and the Simpson index were variable (Table 6). Indeed, the populations of Kasserine (P81 and P95) represented the most homogeneous populations, and this are mentioned by the lowest values of number of multilocus genotypes (MLG). However, the genotypes structuring populations (P285, P224, and P159) are regularly abounded. The Nei's unbiased genetic diversity measured among populations indicated that lesser and greater genetic diversity was seen respectively among the populations from Kasserine (P95 and P149) (Table 6).

Principal component analysis (PCA) and phylogenetic analysis

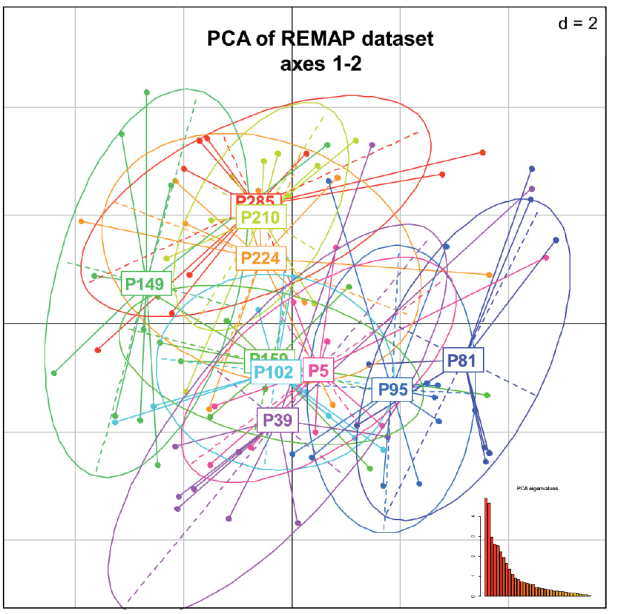
In order to differentiate *C. scorpioides* populations, a PCA was carried out using REMAP markers (Figure 4). Based on their genotypes, the populations were not scattered on either side of the plane of the PCA, with individuals within populations overlapping each other. However, three groups were clearly distinguished despite the overlap of some populations (P102: Kasserine and P159: Siliana) in the centre of the PCA. Indeed, the first group was constituted by the centroids of populations from Sousse (P285), Kairouan (P224 and P210), and Kasserine (P149). Populations from Sfax (P5 and P39) constituted the second group. The third group is represented by two populations from Kasserine (P81 and P95), which are geographically close (Figure 4). Because populations of the same group came from different locations and many populations overlap, it is virtually difficult to correlate marker-based population distributions with geographical location.

The bootstrapped Neighbour-joining trees computed with Nei's distance, showed the presence of four groups (Figure 5). In fact, the population P95 from Kasserine constituted the cluster I. Populations from Kasserine (P81, P149, and P102) were grouped together in the cluster II. The cluster III was composed by populations from Sfax (P5 and P39). The cluster IV was subdivided into two sub-clusters: IV.a and IV.b. Siliana (P159) population was clustered with population of Sousse (P285) highlighting the sub-cluster IV.a, while Kairouan populations (P224 and P210) were grouped together in the sub-cluster IV.b. Based on their genetic distance, the clustering of *C. scorpioides* populations, collected from various regions in Tunisian area, was correlated to their geographical provenance. According to the phylogenetic tree using the Neighbour-joining method, population from Kasserine (P95) had the least evolutionary process, which can be interpreted as the ancestral population. Even so, two populations (P102 and P159) had the most evolutionary process.

Bayesian model-based clustering analysis. In order to analyse the genetics composition of *C. scorpioides* populations, the Bayesian model-based clustering approach was used. The number of K-means clustering is determined, and the number of clusters was ranged between 3 and 8 (Figure 6) and the discriminant analysis of principal components (DAPC) was computed. According to the model at K = 3, *C. scorpioides* genotypes were classified in three clusters. The membership probability of each population element indicated its assignment in the cluster. As shown in Figure 7, the cluster 1 consisted essentially of populations from Sfax (P5 and P39), populations from Kasserine (P149) and Siliana

(P159). Some individuals of this cluster were found in other populations with low frequency. The second cluster included two populations from Kasserine (P81 and P95); however, three major populations from Kairouan (P210 and P224) and Sousse (P285) formed the third cluster (Figure 7).

Figure 4. Principal component analysis of 10 populations using retrotransposon microsatellite amplified polymorphism (REMAP) markers.



Each population has been represented by a specific colour. The colours of the groups are independent to the colours chosen in previous Figures 2 and 3.

Figure 5. Bootstrapped Neighbour joining trees computed by Nei's genetic distance.

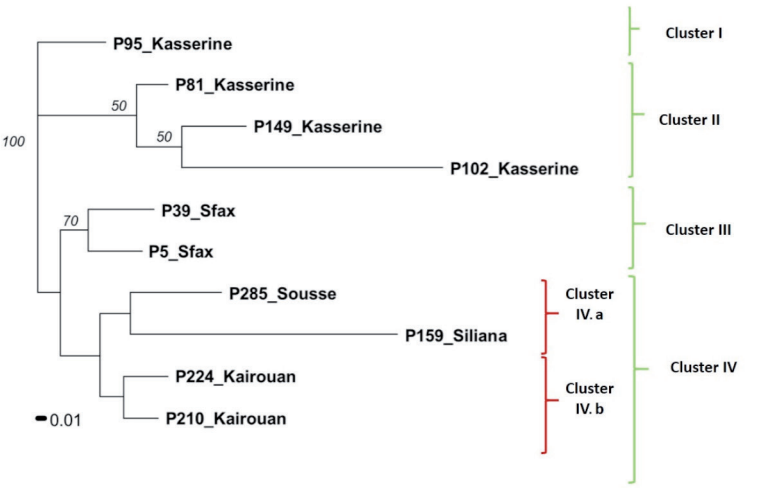


Figure 6. Bayesian information criterion (BIC) value and K ranging from 1 to 7 based on genotypes of *Coronilla scorpioides* using retrotransposon-microsatellite amplified polymorphism (REMAP) markers.

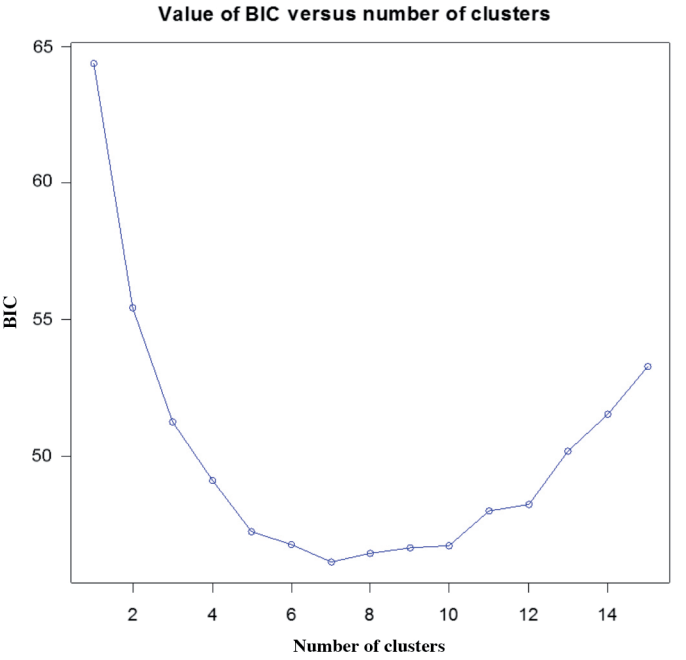
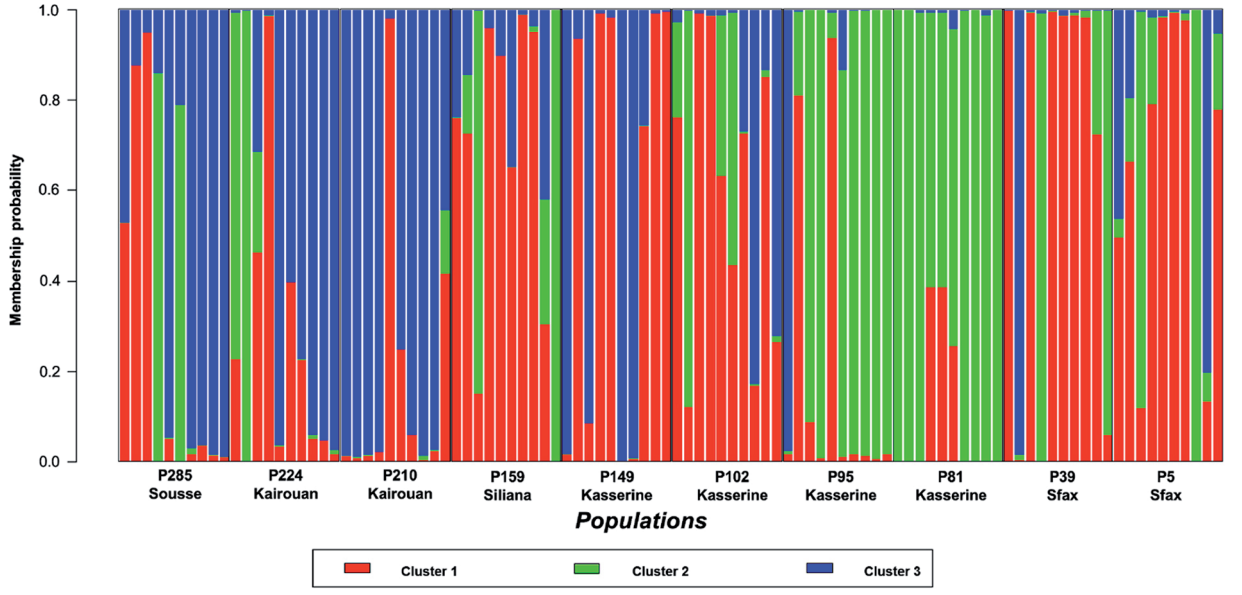


Figure 7. Genetic clustering of *Coronilla scorpioides* populations assessed with retrotransposon-microsatellite amplified polymorphism (REMAP) markers (K = 3).



The colours of the groups are independent to the colours chosen in previous figures.

DISCUSSION

Molecular analysis by SRAP markers was involved through 10 DNAs for each *C. scorpioides* populations. These DNAs were bulked by population in order to identify a genetic polymorphism associated with a specific trait of contrasting phenotypes (Castonguay et al., 2010). Indeed, to measure the extent of genetic variation in germplasm, a wide genetic base will be essential (Mandoulakani et al., 2015). The genetic diversity of *Vitis vinifera* L. accessions and other horticultural grape taxa was clarified via SRAP markers (Guo et al., 2012). The analysis of the results reveals a significant resolving power (Rp) of the SRAP primers. In the Tunisian *C. scorpioides* genome, the average revealed polymorphic bands (Np) was similar to which described in Chinese *Cynodon dactylon* (Zheng et al., 2017) and Indonesian *Dalbergia latifolia* (Yulita et al., 2020). According to our findings, the Np in *C. scorpioides* was higher than those mentioned in 152 similar studies (Robarts and Wolfe, 2014). The percentage of polymorphic loci (76.26%) obtained in this study suggests that SRAP marker could be considered reliable to detect polymorphism at the *C. scorpioides* population (Yang et al., 2015). The variation in the number of amplified bands is due to the occurrence of annealing sites in the genome (Li and Quiros, 2001; Robarts and Wolfe, 2014). Indeed, during the thermal cycles, not all primers could be annealed to their complementary sequences. Consequently, it result in the absence of amplified bands and even some non-specific products (Yulita et al., 2020). Similarly to *V. vinifera*, the SRAP markers used with *C. scorpioides* have shown a great average of the PIC (Guo et al., 2012). These results are higher than those obtained with *C. dactylon* and *Festuca-Lolium* populations (Cheng et al., 2016; Zheng et al., 2017). In general, SRAP markers have shown lower polymorphism compared to others markers such as ISSR and SSR markers (Cheng et al. 2016; Zheng et al., 2017). In fact, molecular markers have different natures of polymorphism (Yulita and Rahmat, 2019). Actually, 45% of the SRAP amplified products are known genes. Thus, the coding regions (genes) have a low mutation rate than non-coding regions, not conserved (Yulita et al., 2020).

The BSA approach coupled with SRAP markers differentiate *C. scorpioides* populations according to their quantitative traits (Zoghlami Khelil et al., 2005). Our findings confirm what was described in others studies. In fact, P81 (Kasserine) and P224 (Kairouan) are genetically distinguished from P210 (Kairouan) by a specific locus. In the same way, these two populations have been identified as having the highest value of total phenolic contents with a high digestibility compared to P210 (Zoghlami Khelil et al., 2005). However, the P210 has been reported as the richest populations in tannins contents, total mineral contents and plant biomass (Zoghlami Khelil et al., 2005). In our study, P210 is genetically distinguished from P244 and P81 by the presence of some locus. The association of this information could help to identify some functional polymorphisms and could contribute to the expression of related quantitative trait loci (QTL). The BSA-SRAP methodology could be a useful tool for this target (Robarts and Wolfe, 2014). In fact, this strategy has been successfully used to link some genomic variation in alfalfa to their freezing tolerance (Castonguay et al., 2010). Targeting the genome coding region, SRAP markers, have a potential that extends beyond compared to the most usually used multiloci markers. Indeed, SRAP markers have the ability to summarize locus with specific biological function (Guindon et al., 2019). Indeed, an important development has been shown in the application of SRAP markers since their publication in 2001 (Li and Quiros, 2001; Robarts and Wolfe, 2014). Like other dominant markers, SRAP markers have revealed their capacity to elucidate genetic variability (Li and Quiros, 2001). It is important to note that this is the first study interested in the genetic structuring of *C. scorpioides*'s populations. According to Euclidean genetic distance, the classification of *C. scorpioides* populations in cladogram showed a large divergence of the *C. scorpioides* populations. Despite the lowest geographical distance between populations from Kairouan (P210 and P224), they are considered genetically contrasted (Figure 3). In fact, the genetic differentiation of *C. scorpioides* populations is not associated to their geographical distribution (Alasaad et al., 2008). However, some studies have found a higher correlation between the clusters and origins of the accessions differentiated by SRAP markers (Alasaad et al., 2008). Therefore, theories about ecological mechanisms (niche adaptation and interspecific interaction), as well as studies on taxonomy affinity and genetics for conservation, could be explored by applying SRAP (Robarts and Wolfe, 2014). The classification in cladogram results is aligned with that shown in the PCA analysis. Therefore, the genetics distance revealed by populations clustering could relate to their QTL and environmental adaptation.

The most widely used to date are therefore retroelements, mainly long terminal repeats (LTR) retrotransposons and short interspersed nuclear elements (SINE). REMAP, based on direct PCR between element and microsatellites, seems more promising than other markers based on transposable elements, as microsatellite sequences can be closely associated

with retrotransposons (Kalendar et al., 1999). This is demonstrated in barley (Ramsay et al., 1999). In this part of our study, we explored diversity, distribution and evolutionary dynamics of *C. scorpioides* populations using the REMAP markers. The analysis of *C. scorpioides*'s populations using REMAP markers, generated a highly percentage of the polymorphic band. The percentage of the polymorphic band identified in *C. scorpioides* by REMAP markers was higher than that obtained in *Medicago sativa* (Mandoulakani et al., 2012) and *Triticum aestivum* (Holasou et al., 2019). In this sense, the insertion polymorphism resulting from the element's integration could differentiate lines, groups of lines, or even species (Ourari et al., 2020). In higher eukaryotes, essentially plants and vertebrates, there is, generally, little convergence of insertion, i.e., an insertion found at the same location in two genotypes will reflect an insertion present in an ancestor common to both genotypes. For the purposes of assessing genetic diversity, the most considered elements are those that meet the criteria of a high copy number, guaranteed stability of insertions, and high potential for polymorphism (Kalendar et al., 1999). The Shannon and the Simpson index suggest that this species has a high genetic diversity in Tunisia. According to our observations, the identified *C. scorpioides* genotypes are irregularly distributed among the studied populations. The high biodiversity of *Coronilla* genus in species sense are reported in the study of Abdelguerfi et al. (2004). In fact, the Mediterranean Fabaceae species are known by a great value of genetic diversity (Zitouna et al., 2013). The diversity of each populations indicates that populations from Siliana (P159), Kairouan (P224) and Sousse (P285) are genetically heterogenic and have the highest number of genotypes. All genotypes are regularly distributed in these populations, which reflect a high genetic diversity. However, populations from Kasserine (P81 and P95) are the most homogeneous populations which is reflected especially by low multilocus genotype (MLG) values.

According to the PCA analysis based on REMAP markers, because populations of the same group came from different locations and many populations overlap, the marker-based population distributions of *C. scorpioides* was not correlated with geographical location.

Comparing to the SRAP markers, REMAP markers could be related to the geographical origin of this species. Many studies have reported that retrotransposon-related polymorphism is associated to the evolution events in species (Kalendar et al., 1999; Basirnia et al., 2014) and to their geographical origin (Sharifi-Rigi et al., 2014). Indeed, the REMAP markers used in this study have divided *C. scorpioides* by their evolutionary rate. According to the Neighbour-joining tree, population from Kasserine (P95) could be the ancestral population, which can be confirmed by their low genetic diversity. At the same time, an adjacent population originating from Kasserine (P102) has shown a great evolutionary rate, which is proven by their highly degree of genetic diversity. In fact, the geographical proximity between populations implies an upward trend in the gene flow (Bossart and Prowell, 1998). However, P95 (Kasserine) do not cluster in a major group of BSA-SRAP analysis, which may suggest that the genetic diversity revealed by SRAP markers in *C. scorpioides* is not exclusively due to geographical origin. The clustering analysis based on the Bayesian model classified *C. scorpioides* into four groups. The genotypic clustering is correlated to the geographical distribution, which could confirm the results of the phylogeny analysis. The ancestral population originating from Kasserine (P95) constituting an extra-group confirm the genotypic composition of this population revealed by the genetic index. In 1838, *C. scorpioides* were firstly identified in Germany, Switzerland, Prussia, and Istria regions and then dispersed across the Mediterranean region (Europe, North of Africa, and Asia) (Zoghlami and Zouaghi, 2003). Therefore, there are two hypotheses that explain the existence of this species in Tunisia (Figure 1). Probably, *C. scorpioides* species migrated from Germany to Tunisia through Turkey and Egypt. However, according to our results, the ancestral population is located in the western middle of Tunisia. Therefore, the migration of this species could be from the west to the east of the Northern African areas. To validate this distribution event, this assumption could be confirmed by comparing the Tunisian populations with the Moroccan and the European populations. The ecological processes as well as the adaptation of niches and the interspecific interaction of this species could be addressed via the two markers used in this study.

CONCLUSIONS

Our results highlight the richness of Tunisian area in biodiversity and the genetic diversity of the *Coronilla scorpioides* species in this area. In addition, by using the bulked segregant analysis (BSA)-sequence-related amplified polymorphism (SRAP), we successfully correlated some agronomic traits to genetic marker. The ecological processes as well as the adaptation of niches and the interspecific interaction of this species could be addressed via the two markers used. Globally,

we have demonstrated that Tunisian *C. scorpioides* species constitute important phytogenetic resources, which must retain the attention of breeders in order to introduce it into interesting fields. Thus, the unveiling of the genetic diversity of this species could help in the biodiversity conservation from the genetic erosion caused by human expansion and climatic variations.

ACKNOWLEDGEMENTS

Authors are grateful to Professor Zoghlami Aziza for providing all studied samples for this assignment.

REFERENCES

Abdelguerfi, A., et Abdelguerfi-Laouar, M. 2004. Les ressources génétiques d'intérêt fourrager et/ou pastoral: diversité, collecte et valorisation au niveau méditerranéen. Cahiers Options Méditerranéennes 62:29.

Alasaad, S., Li, Q.Y., Lin, R.Q., Martín-Atance, P., Granados, J.E., Díez-Baños, P., et al. 2008. Genetic variability among Fasciola hepatica samples from different host species and geographical localities in Spain revealed by the novel SRAP marker. Parasitology Research 103(1):181-186. doi:10.1007/s00436-008-0952-1.

Amiriyan, M., Shojaeiyan, A., Yadollahi, A., Maleki, M., and Bahari, Z. 2019. Genetic diversity analysis and population structure of some Iranian Fenugreek (*Trigonella foenum-graecum* L.) landraces using SRAP Markers. Molecular Biology Research Communications 8(4):181-190.

Basirnia, A., Darvishzadeh, R., and Abdollahi Mandoulakani, B. 2014. Retrotransposon insertional polymorphism in sunflower (*Helianthus annuus* L.) lines revealed by IRAP and REMAP markers. Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology 150(4):641-652.

Bossart, J.L., and Prowell, D.P. 1998. Genetic estimates of population structure and gene flow: limitations, lessons and new directions. Trends in Ecology and Evolution 13(5):202-206.

Cano, A.L., Gómez, P.S., and Martínez, J.F.J. 2012. A new species of *Coronilla* (Loteae, Fabaceae) from Southeastern Spain: Evidence from morphological and molecular data. Folia Geobotanica 47(3):317-335.

Castonguay, Y., Cloutier, J., Bertrand, A., Michaud, R., and Laberge, S. 2010. SRAP polymorphisms associated with superior freezing tolerance in alfalfa (*Medicago sativa* spp. *sativa*). Theoretical and Applied Genetics 120(8):1611-1619.

Cheng, Y., Ma, X., Zhou, K., Humphreys, M.W., and Zhang, X.Q. 2016. Phylogenetic analysis of *Festuca-Lolium* complex using SRAP markers. Genetic Resources and Crop Evolution 63(1):7-18.

Fox, J. 2005. Getting started with the R commander: a basic-statistics graphical user interface to R. Journal of Statistical Software 14(9):1-42.

Guindon, M.F., Martin, E., Cravero, V., Gali, K.K., Warkentin, T.D., and Cointry, E. 2019. Linkage map development by GBS, SSR, and SRAP techniques and yield-related QTLs in pea. Molecular Breeding 39(4):54.

Guo, L., Liu, X., Liu, X., Yang, Z., Kong, D., He, Y., et al. 2012. Construction of genetic map in barley using sequence-related amplified polymorphism markers, a new molecular marker technique. African Journal of Biotechnology 11(74):13858-13862.

Holasou, H.A., Rahmati, F., Rahmani, F., Imani, M., and Talebzadeh, Z. 2019. Elucidate genetic diversity and population structure of bread wheat (*Triticum aestivum* L.) cultivars using IRAP and REMAP Markers. Journal of Crop Science and Biotechnology 22(2):139-151.

Innocenti, G., Piovan, A., Caniato, R., Filippini, R., and Cappelletti, E.M. 1996. Within-plant distribution of coumarins in *Coronilla* and *Securigera* species. International Journal of Pharmacognosy 34(2):114-118.

Jombart, T., Kamvar, Z.N., Collins, C., Luštrik, R., Beugin, M., Knaus, B.J., et al. 2020. Adegenet: Exploratory analysis of genetic and genomic data. [Data Collection]. The Comprehensive R Archive Network. Available at https://cran.r-project.org/web/packages/adegenet/

Kabtni, S., Sdouga, D., Rebey, I.B., Save, M., Trifi-Farah, N., Fauconnier, M.L., et al. 2020. Influence of climate variation on phenolic composition and antioxidant capacity of *Medicago minima* populations. Scientific Reports 10(1):1-15.

Kalendar, R., Grob, T., Regina, M., Suoniemi, A., and Schulman, A. 1999. IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. Theoretical and Applied Genetics 98(5):704-711.

Li, G., and Quiros, C.F. 2001. Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in Brassica. Theoretical and Applied Genetics 103(2-3):455-461.

Mandoulakani, B.A., Piri, Y., Darvishzadeh, R., Bernoosi, I., and Jafari, M. 2012. Retroelement insertional polymorphism and genetic diversity in *Medicago sativa* populations revealed by IRAP and REMAP markers. Plant Molecular Biology Reporter 30(2):286-296.

Mandoulakani, B.A., Sadigh, P., Azizi, H., Piri, Y., Nasri, S., and Arzhangh, S. 2015. Comparative assessment of IRAP, REMAP, ISSR, and SSR markers for evaluation of genetic diversity of alfalfa (*Medicago sativa* L.) Journal of Agricultural Science and Technology 17(4):999-1010.

Murray, M.G., and Thompson, W.F. 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Research 8(19):4321-4326.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89(3):583-590.

Ourari, M., Coriton, O., Martin, G., Huteau, V., Keller, J., Ainouche, M.-L., et al. 2020. Screening diversity and distribution of *Copia* retrotransposons reveals a specific amplification of *BARE1* elements in genomes of the polyploid *Hordeum murinum* complex. Genetica 148(2):109-123.

Ramsay, L., Macaulay, M., Cardle, L., Morgante, M., degli Ivanissevich, S., Maestri, E., et al. 1999. Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. The Plant Journal 17(4):415-425.

Robarts, D.W., and Wolfe, A.D. 2014. Sequence-related amplified polymorphism (SRAP) markers: A potential resource for studies in plant molecular biology. Applications in Plant Sciences 2(7):1400017.

Shannon, C.E., and Weaver, W. 1949. The mathematical theory of communication. University of Illinois Press, Urbana, Illinois, USA.

Sharifi-Rigi, P., Saeidi, H., and Rahiminejad, M. 2014. Genetic diversity and geographic distribution of variation of *Hordeum murinum* as revealed by retroelement insertional polymorphisms in Iran. Biologia 69(4):469-477.

Simpson, E.H. 1949. Measurement of diversity. Nature 163(4148):688-688.

Szewczyk, K., and Bogucka-Kocka, A. 2012. Analytical methods for isolation, separation and identification of selected furanocoumarins in plant material. p. 57-92. In Rao, V. (ed.) Phytochemicals A global perspective of their role in nutrition and health. IntechOpen. doi:10.5772/27495. Available at https://www.intechopen.com/chapters/32938.

Yang, P., Liu, X., Liu, X., Yang, W., and Feng, Z. 2015. Diversity analysis of the developed qingke (hulless barley) cultivars representing different growing regions of the Qinghai-Tibet Plateau in China using sequence related amplified polymorphism (SRAP) markers. African Journal of Biotechnology 9(50):8530-8538.

Yulita, K.S., Atikah, T.D., Wardani, W., and Susila. 2020. Unraveling genetic variations of *Dalbergia latifolia* (Fabaceae) from Yogyakarta and Lombok Island, Indonesia. Biodiversitas 21(2):833-841.

Yulita, K.S., and Rahmat, H.H. 2019. Population genetic structure and diversity of a critically endangered ramin [*Gonystylus bancanus* Miq. (Kurz), Thymelaeaceae] from Kalimantan and Sumatra based on sequence random amplified polymorphism. IOP Conference Series: Earth and Environmental Science 308:012067. doi:10.1088/1755-1315/308/1/012067.

Zheng, Y., Xu, S., Liu, J., Zhao, Y., and Liu, J. 2017. Genetic diversity and population structure of Chinese natural bermudagrass [*Cynodon dactylon* (L.) Pers.] germplasm based on SRAP markers. PLOS ONE 12(5):e0177508.

Zitouna, N., Marghali, S., Gharbi, M., Chennaoui-Kourda, H., Haddioui, A., and Trifi-Farah, N. 2013. Mediterranean *Hedysarum* phylogeny by transferable microsatellites from *Medicago*. Biochemical Systematics and Ecology 50:129-135.

Zoghlami, A., and Zouaghi, M. 2003. Morphological variation in *Astragalus hamosus* L. and *Coronilla scorpioides* L. populations of Tunisia. Euphytica 134(2):137-147.

Zoghlami Khelil, A., Ben-Salem, H., et Zouaghi, M. 2005. Production et caractérisation nutritionnelle de deux légumineuses pastorales: *Astragalus hamosus* et *Coronilla scorpioides*. Revue de l'Institut Nationale Agronomique de Tunisie 20:209-226.