Ciência e Agrotecnologia, 47:e011922, 2023
http://dx.doi.org/10.1590/1413-7054202347011922

Ciência e Agrotecnologia
SCIENCE AND AGROTECHNOLOGY
Agricultural Sciences
eISSN 1981-1829

Genetic variability in genotypes of safflower via SSR molecular marker

Variabilidade genética em genótipos de Cártamo via marcador molecular SSR

Altacis Junior de Oliveira[1*], Marco Antonio Aparecido Barelli[1], Taniele Carvalho de Oliveira[1], Nilo Leal Sander[1], Rafhael Felipin Azevedo[1], Claudete Rosa da Silva[2]

[1]Universidade do Estado de Mato Grosso, Departamento de Agronomia, Cáceres, MT, Brasil
[2]Universidade Federal Rural da Amazônia, Departamento de Agronomia, Parauapebas, PA, Brasil
*Corresponding author: altacismarquesfig@hotmail.com
Received in September 6, 2022 and approved in January 25, 2023

ABSTRACT

The safflower is an oleaginous plant belonging to the Asteraceae family. It is used as a raw material for various purposes. These plants are popular for the quality and quantity of oil produced and, and thus, studying their genetic variability using markers is necessary for determining genetic resources to conduct breeding programs. Therefore, we evaluated the genetic variability of safflower genotypes using Simple Sequence Repeat (SSR) molecular markers. The study was conducted at the State University of Mato Grosso "Carlos Alberto Reyes Maldonado", in the Campus of Cáceres-MT. In total, 121 safflower genotypes from the Germplasm collection were evaluated using 21 SSR markers. The programs GenAlEx 6.5, GENES, and Structure were used to analyze the data. We identified 158 alleles at 21 loci among the genotypes. The expected heterozygosity (He) was high (0.551 – 0.804), but the observed heterozygosity (Ho) was low (0.000 – 0.502), and the indices of the endogamy coefficient (F) were positive in all loci and all populations, with an overall average of 0.958. The genetic differentiation (FST) values among populations were low, with an average of 0.010, which suggested a low population structure. The modified Tocher clustering and the UPGMA hierarchical clustering yielded 19 and 15 distinct groups, respectively. The genetic structure showed two populations, with few intermixes in the genome. The evaluated safflower genotypes showed genetic variability, and these genetically different variants might be used in breeding programs to obtain cultivars adapted to Brazil.

Index terms: *Carthamus tinctorius* L.; genetic improvement; genetic variability.

RESUMO

O cártamo é uma oleaginosa da família Asteraceae, cuja matéria-prima serve para diversos fins. A cultura se destaca pela qualidade e quantidade de óleo produzido, neste sentido o estudo da variabilidade genética com o uso de marcadores é uma etapa inicial na exploração dos recursos genéticos em um programa de melhoramento. Diante disso, estimou-se a variabilidade genética de genótipos de cártamo via marcadores moleculares Simple Sequence Repetition (SSR). O estudo foi conduzido na Universidade Estadual de Mato Grosso "Carlos Alberto Reyes Maldonado", Campus de Cáceres-MT, onde foram avaliados 121 genótipos de cártamo da coleção de Germoplasma, utilizando 21 marcadores SSR. Foram utilizados os programas GenAlEx 6.5, GENES e software Structure para os respectivos resultados. O número de alelos detectados entre os genótipos considerando os 21 loci foi de 158, a He pode ser considerada alta, variando de 0,551 a 0,804, já a heterozigosidade observada (Ho) foi baixa, variando de 0,000 a 0,502, e índices de coeficiente de endogamia (F) foram positivos em todos os locus e em todas as populações, possuindo uma média geral de 0,958. Os valores de medida de diferenciação genética (FST) entre as populações foram encontrados baixos em média de 0,010, sugerindo baixa estrutura populacional. O agrupamento de Tocher modificado obteve 19 grupos, e o agrupamento hierárquico de UPGMA 15 grupos distintos. A estruturação genética demostrou duas populações, com poucas intermixagem no genoma. Os genótipos de cártamo avaliados possuem variabilidade genética, sendo possível explorar esta variabilidade em um programa de melhoramento genético visando obter cultivares adaptados ao Brasil.

Termos para indexação: *Carthamus tinctorius* L; melhoramento genético; variabilidade genética.

INTRODUCTION

Carthamus tinctorius L. has been cultivated and used for more than 4,000 years (Moura et al., 2015). It is an oilseed from the Asteraceae family that is mainly grown for extracting oil, which is used for human consumption (Queiroga; Girão; Albuquerque, 2021) and to make lubricants, biofuels, soaps, varnishes, and animal feed (Golkar, 2014; Kumar et al., 2016).

Safflower is an important oilseed used around the world (Kim et al., 2016; Sharifi; Namvar, 2017). It is cultivated in more than 60 countries, and the global production of safflower in 2017 was around 734,000 tons, cultivated in an area of approximately 725,000 hectares; Turkey, Mexico, and China were the largest producers of safflower with yields of 1,826, 1,565, and 1,429 kg ha^{-1}, respectively (Food and Agriculture Organization of the United Nations - FAO, 2019).

In Brazil, culture has attracted the attention of researchers and industries due to the quantity and quality of oil produced (Silveira et al., 2017; FAO, 2019). Studies on the culture, mainly on genetic improvement, are limited. Thus, further research can help in the selection of genotypes adapted to specific regions, which can increase crop yield (Singh; Nimbkar, 2016).

Evaluating the genetic variability using markers is necessary for using genetic resources in plant breeding programs (Saadaoui et al., 2017). Determining the genetic variability of a breeding collection via SSR (Simple Sequence Repeat) molecular markers helps in identifying genotypes with desirable characteristics for developing new cultivars (Kiran et al., 2017). Golkar and Mokhtari (2018) used SSRs in the safflower genotype to evaluate genetic variability and structure. Ambreen et al. (2018) evaluated association mapping for important agronomic traits in the main collection of safflower (*Carthamus tinctorius* L.) using microsatellite markers and found associations between molecular markers and traits, which can facilitate marker-assisted breeding and the identification of genetic determinants of trait variability.

Hassani et al. (2020a) evaluated the morphological description, genetic diversity, and population structure of safflower (*Carthamus tinctorius* L.) mini-crop using SRAP and SSR markers and found high genetic diversity in the safflower germplasm examined by performing agromorphological and molecular analysis. The same group (Hassani et al., 2020b), conducted a Deep Analysis of the genomic diversity, population structure, and linkage disequilibrium of safflower (*Carthamus tinctorius* L.) found across Africa and Europe. They used the NGS data generated by the DArTseq technology and found that their results matched their hypothesis that safflower domestication started somewhere west of the Fertile Crescent and then expanded across Africa and Europe.

The use of SSR markers in studies on safflower can provide information on the genetic improvement of the culture. These markers can be used to determine genetic variability and population structure. Information on both these aspects is important for using the genetic diversity of safflower populations effectively.

Therefore, in this study, we estimated the genetic variability of 121 safflower genotypes via SSR molecular markers from the germplasm collection of the Laboratory of Genetic Resources & Biotechnology (LRG&B) of the State University of Mato Grosso "Carlos Alberto Reyes Maldonado" (UNEMAT), Campus of Cáceres, Mato Grosso, Brazil.

MATERIAL AND METHODS

The study was conducted under controlled temperature and humidity conditions at the Laboratory of Genetic Resources & Biotechnology (LRG&B) and in the greenhouse belonging to the LRG&B, both associated with the Department of Agronomy of the University of the State of Mato Grosso "Carlos Alberto Reyes Maldonado" (UNEMAT), University City of the Campus of Cáceres - Mato Grosso, located at "16°07'66" latitude and "57°65'29" longitude.

For collecting DNA samples, 121 safflower genotypes were sown in 500 mL plastic cups containing commercial substrate. Two seeds were sown in the greenhouse of LRG&B, with three replicates for each genotype. The seeds were irrigated daily, twice a day, until the leaf tissue was collected. The samples were collected between eight and ten days after sowing when the second pair of true leaves emerged.

We evaluated 121 genotypes from 10 populations, which included varieties from Bangladesh, Canada, Kazakhstan, China, Ethiopia, the USA, India, Iran, Pakistan, and Turkey. These populations were grouped into six regions: South Asia (India and Pakistan), Middle East (Iran and Turkey), North America (Canada and USA), East Asia (China and Bangladesh), Central Asia (Kazakhstan), and East Africa (Ethiopia) (Table 1).

While collecting the samples, tweezers were used to pluck the leaves from the plants. Care was taken to prevent contamination, and later, the samples were stored in zip lock bags and refrigerated in an ultra-freezer at –80 °C until DNA extraction was performed.

The leaf tissue was macerated in the TissueLyser for 10 min. The DNA was extracted using the *Wizard*® Genomic DNA Purification Promega kit (USA), following the manufacturer's instructions. To amplify the DNA, 21 primers were used for the SSR loci (Table 2), which represented the genetic variability of the safflowers, as described by Mokhtari et al. (2018) and Kiran et al. (2017).

Table 1: Information on the 121 genotypes of *Carthamus tinctorius* L.

Order	PI	Origin	Order	PI	Origin	Order	PI	Origin	Order	PI	Origin
1	193473	Etiópia	33	306832	Índia	66	451956	Índia	97	572431	EUA
2	195895	Marrocos	34	306833	Índia	67	506426	China	98	572439	EUA
3	237539	Turquia	35	306838	Índia	68	508068	EUA	99	572450	EUA
4	248385	Índia	37	306866	Índia	69	514625	China	100	572464	EUA
5	248620	Paquistão	38	343783	Irã	70	525457	EUA	101	572544	Canadá
6	248808	Índia	39	343930	Etiópia	71	537658	EUA	102	576981	China
7	248828	Índia	40	367833	Argentina	72	537673	EUA	103	576985	França
8	248839	Índia	42	369845	Tajiquistão	73	537680	EUA	104	613357	EUA
9	248852	Índia	43	369849	Rússia	74	537682	EUA	105	613361	EUA
10	250083	Egito	44	369854	Uzbequistão	75	537684	EUA	106	613366	EUA
11	250188	Paquistão	45	392029	Turquia	76	537697	EUA	107	613373	EUA
12	250190	Paquistão	46	392030	Turquia	77	537712	EUA	108	613380	EUA
13	250203	Paquistão	47	392031	Turquia	78	543980	China	109	613382	EUA
14	250204	Paquistão	48	393500	Irã	79	544002	China	110	613384	EUA
15	250840	Irã	49	401474	Bangladesh	80	544013	China	111	613394	EUA
16	250922	Irã	50	401475	Bangladesh	81	544028	China	112	613404	EUA
18	253540	Hungria	51	401477	Bangladesh	82	544030	China	113	613409	EUA
19	253899	Síria	52	401480	Bangladesh	83	544031	China	114	613415	EUA
20	259996	Paquistão	53	401578	Índia	84	544036	China	115	613419	EUA
21	259997	Paquistão	54	401589	Índia	85	544038	China	116	613422	EUA
22	262443	Espanha	55	405955	Irã	86	544043	China	117	613456	EUA
23	262447	Cazaquistão	56	405961	Irã	87	560178	EUA	118	613503	EUA
24	262450	Índia	57	405965	Irã	88	532639	Índia	119	613519	Irã
25	279344	Japão	58	405970	Irã	89	568787	China	120	638543	Canadá
26	283757	Índia	59	405975	Irã	90	568792	China	121	653143	EUA
27	304438	Irã	60	406006	Irã	91	568795	China	122	653149	China
28	305161	Índia	61	406007	Irã	92	568798	China	123	653162	China
29	305198	Índia	62	406015	Irã	93	568836	China	124	251978	Turquia
30	305207	Índia	63	407606	Turquia	94	568866	China			
31	305209	Índia	64	407613	Turquia	95	568870	China			
32	305540	Cazaquistão	65	451954	Índia	96	568876	China			

PI - Identification.

The DNA samples used in the PCR assay were diluted to 10 ng-μL^{-1} using autoclaved ultrapure water. The PCR mix was prepared as follows: 2 μL of DNA (10 ng-μL^{-1}), 0.5 μL of the deoxyribonucleotide mix (dATP, dCTP, dGTP, and dTTP) (10 mM), 1.25 μL of each primer (forward and reverse) (10 μM), 5 μL of buffer (5 X) containing magnesium (7.5 mM), 0.2 μL Taq polymerase (5 U), and 14.8 μL of autoclaved ultrapure water; the final reaction volume was 25 μL.

Following the protocol of Williams et al. (1990), the PCR assays were conducted using a Perkin Elmer model 9600 thermocycler with the following temperature program: initial denaturation phase at 94 °C for 5 min, followed by 35 cycles of denaturation at 94 °C for 30 s, annealing for 30 s (according to the temperature requirements of the specific primer) (Table 2), extension at 72 °C for 30 s, and a final extension phase at 72 °C for 5 min.

Table 2: Details of the 21 molecular SSR markers that were used to identify the molecular variability of the 121 genotypes of *Carthamus tinctorius* L.

Order	Name of sequence	Repeat type	Sequence 5′ to 3′ of the *primer*	Annealing temperature (°C)
1	CT3	$(CT)_5$...$(CT)_{10}$	AGCGAGACAACTCTGGGTGT GTGCACCAACTCTGCCTTTA	55.2
2	CT4	$(CT)_9$	GCGTTCGAGAAACATTCGAG AGCCCCCATTTGTAGGGATA	54.1
3	CT6	$(CT)_4$...$(CT)_{10}$	AAAGCCCGCCTAAAACACTT AATACGAGTTCGATTGTGTGC	58
4	CT7	$(AG)_6$...$(AGT)_5$	AGTTTTTCAAGGCAGGTGACA CCGTTTTAACTCCCAAACCA	54
5	CT10	$(AG)_4$	GTCCGATGCGGTCAAAAT GTTTCAACCACCCGGTTTC	55.2
6	CT11	$(TCCT)_3$...$(TTC)_2$	CGCCTTTCAATCCTTGTCTC CAGCACCAAGAACCATAGCA	54.1
7	CT12	(CTTT)GA$(CTTT)_2$	GGCGACGAATTTGAGTTCAC ACCAATCGAAAACGGAGATG	54.1
8	CT13	$(TCAC)_2(TTC)_2$	CTGGAAGTGACCAAACAACCT GGTTGAGTCTTATTTTGGGAGATG	54.1
9	CT14	$(TCC)_4$TCA	CTTCCTTGTTGCTCGTCACA ACACATCCCTCGTTTTCCAG	54.1
10	CT19	$(CT)_{40}$	GCTTCCGAATTGAACGGAGA ACATCATCGCCACGTCAGTT	54
11	CT26	$(GGAAA)_2$	CATCGCTACCAACCATACCC GCGGAATCGAACAACCTATC	54.3
12	CT31	$(AG)_4$	CCGACGAAGTGGTGAATAGG TATCCAAACATGCGGTTTCA	55.2
13	CT316	$(TC)_8$	CTGAACGAAAACGCAGTCAA TGGGTTTTGGATGTGAGGAT	56.5
14	CT657	$(CAT)_6$...$(CAT)_{12}$	GGCTCAACTCGACTCATCATC GCTTGTTGGGAGGGATCATA	56
15	CT309	$(TA)_8$	TTGCAAGATTGTTCGTCGAT TCCCTTCCTATTCTGGACCC	55.2
16	CT599	$(TC)_{11}$...$(TA)_6$	CATTAACCCATCCATCACCC TGTCATCTCTGAAAGCACCG	56
17	CT006	$(GT)_{12}$	CAATTCGCTTCCACCAAGAT TACTCCTACCCGCCACAAAC	55.2
18	CT820	$(AG)_7$	TTGCCCATTGTTCGTTAGGT AAAAGAAAACGCGCAGAAGA	55.2
19	CT440	$(AG)_9$	AGCAGTAGAGCATAACGCC CAGCCAGCCAGAATGATCG	59.8
20	CT518	$(AG)_{10}$	CCCTTTTTGCTCTCCAACCT CTTGGGCCTTCCTCTCTTCT	55.2
21	CT558	$(TCC)_5$...$(GA)_{21}$	ACCTTCTCCTCACCTCCGAT GTCTTAAGCTGTTCCGCCTG	55.2

The PCR product was stored at 4 °C until further experiments were conducted. The amplified products (*amplicons*) were stained with Gel Red and Blue Juice 6 X and visualized on a 3% agarose gel, using Tris borate EDTA (1%) as a buffer solution. The gel was photographed using the Locus Biotecnologia/photo documentation system *LPix Image* version 2.7 after running the gel at 60 V for 4 h.

The genetic diversity of 110 safflower genotypes was evaluated. The data on the 11 remaining genotypes were eliminated as there were less than two samples per region, which was below the minimum requirement for analysis by the GenAlEx 6.5 program (Excoffier; Laval; Schneider, 2005). The allele frequency, number of alleles, average observed heterozygosity (Ho), average expected heterozygosity (He), and inbreeding coefficient (F) were evaluated. The genetic structure of the Fst populations was measured using the same program (Wright, 1949).

The analysis of molecular variance (AMOVA) was performed to determine the distribution of genetic diversity among and within the population and between individuals, following the method described by Excoffier, Smouse, and Quattro (1992), and the significance was tested using 1,000 permutations with a 95% confidence interval.

The dissimilarity matrix resulting from the Jaccard index was analyzed by Tocher's optimization method and the UPGMA hierarchical method using the computational resource GENES (Cruz, 2013). The Bayesian cluster analysis was performed using the Structure software (Pritchard et al., 2000) to define the number of groups (K).

RESULTS AND DISCUSSION

The 21 SSR markers used in this study showed 100% polymorphism. The results for the number of alleles, observed (Ho) and expected (He) heterozygosity, and inbreeding coefficient (F) were determined (Table 3), which in turn was used to obtain information on the genetic diversity of 110 *Carthamus tinctorius* L.

Table 3: Estimation of the genetic diversity of the 110 genotypes of *Carthamus tinctorius* L. obtained from 21 SSR markers *.

SSR *locus*	N° of Alleles	He	Ho	F
CT3***	8	0.747	0.000	1.000
CT6***	6	0.661	0.000	1.000
CT4***	7	0.746	0.000	1.000
CT006**	7	0.718	0.000	1.000
CT10***	9	0.748	0.053	0.940
CT7***	8	0.714	0.012	0.987
CT11***	7	0.713	0.000	1.000
CT12***	6	0.551	0.000	1.000
CT13***	6	0.661	0.000	1.000
CT14**	7	0.729	0.000	1.000
CT19***	6	0.653	0.000	1.000
CT26***	11	0.804	0.502	0.383
CT31***	7	0.727	0.000	1.000
CT309***	7	0.709	0.000	1.000
CT316***	8	0.718	0.159	0.818
CT440***	8	0.733	0.000	1.000
CT518***	7	0.739	0.000	1.000
CT558***	8	0.757	0.000	1.000
CT559**	8	0.754	0.000	1.000
CT657***	9	0.755	0.000	1.000
CT820***	8	0.746	0.015	0.983

He = expected heterozygosity; Ho = observed heterozygosity; F = inbreeding coefficient. Significance level of the Hardy-Weinberg balance: *P<0.05; **P<0.01; ***P<0.001.

In total, 158 alleles were detected among the genotypes at the 21 loci. The number of alleles ranged from six (CT6, CT12, CT13, and CT19) to 11 (CT26), with eight alleles per locus on average (Table 3). Kiran et al. (2017) evaluated the genetic divergence of 148 safflower genotypes using 48 molecular SSR markers and found that the number of alleles was 2–15, which was higher than that recorded in this study. However, the average number of alleles per locus in their study was four, which was lower than that of our study. Mokhtari et al. (2018) studied the genetic divergence of 103 safflower genotypes using 32 SSR molecular markers and found a lesser number of alleles than that in our study, ranging from two to four and an average of three alleles per locus.

The number of alleles is an important parameter to determine the genetic diversity among populations before they are used in breeding programs. As the number of alleles in a population increases, its diversity also increases, which in turn increases the chance of identifying favorable genotypic combinations. Therefore, this parameter is greatly influenced by the number of genotypes evaluated, and it increases with the sample size (Petit; Mousadik; Pons, 1998). These findings explained the results obtained in this study. The number of alleles found in this study was lower than that reported by Kiran et al. (2017) and higher than that reported by Mokhtari et al. (2018), probably because the number of genotypes evaluated in those studies was different.

The expected heterozygosity (He) in our study was high (0.551 to 0.804), with an average of 0.718. This average value was higher than the values reported by Lee et al. (2014) and Bahmankar, Nabati, and Dehdari, (2017), which were 0.386 and 0.537, respectively, indicating that the genetic diversity in those studies was lower.

The observed heterozygosity (Ho) was low (0.000 to 0.502), with an average of 0.035. Ambreen et al. (2018) evaluated 124 safflower genotypes using 93 SSR primers and also found a low Ho of 0.112. The low polymorphism might be due to the predominance of sexual reproduction and self-pollination, which increases the rate of homozygosity.

The results of the inbreeding coefficient index (F) were positive for all loci and in all populations, with an average of 0.958, which is expected for autogamous breeding plants.

The F values were high and positive, probably due to the level of He relative to that of Ho for each locus and in each population. The F values indicated that inbreeding was prevalent. The F analysis can be used to measure the deficiency or excess of heterozygous genotypes present in a population. This analysis estimates the probability of two alleles being identical by descent, with a coefficient that can range from –1 to 1. Negative values indicate the presence of more heterozygotes than expected, while positive values indicate more homozygotes, and zero indicates that the process is random.

The results of AMOVA (Molecular Variance Analysis) showed that 91% of the total genetic variability occurred within populations, 5% of variability occurred between populations, and 4% of variability occurred between individuals. The variability between regions was 0% (Table 4).

In general, the average genetic differentiation (FST) between populations was low (0.010), suggesting a low population structure. Weir (1996) stated low FST indicates a similar frequency of alleles within each population, and high FST indicates different allele frequencies in the populations.

Kiran et al. (2017) found similar results after evaluating 148 safflower genotypes, in which the results of AMOVA showed that 85% of the genetic variation was explained by individuals within populations and 15% of the variation was explained between populations. Their findings also indicated a low population structure.

Table 4: Analysis of molecular variance (AMOVA) of *Carthamus tinctorius* L. genotypes from 10 populations belonging to six distinct geographic regions.

Source of Variation	Degree of Freedom	Sum of Squares	Components of Variation	Variation (%)
Between regions	5	135.28	27.05	0%
Between populations	4	110.78	27.69	5%
Within populations	100	1960.21	19.60	91%
Among individuals	110	46.00	0.41	4%
Total	219	2252.29	-	100%

The grouping performed by the modified Tocher method, with a Bayesian Jaccard similarity index, allowed the clustering of genotypes into 19 groups based on 21 SSR markers (Table 5). Group I included the highest number of genotypes (14) of the total evaluated (12.74%). These genotypes were a part of the populations from India and Pakistan (South Asia), Iran and Turkey (Middle East), China (East Asia), and Kazakhstan (Central Asia).

Groups II and III consisted of 11 genotypes, corresponding to 10% of the total genotypes evaluated. These genotypes were a part of the populations from Pakistan and India (South Asia), Iran (Middle East), Ethiopia (East Africa), Bangladesh and China (East Asia), and the USA (North America), respectively.

Groups VI, VII, XIII, XIV, and XVI included 33 (29.99%) genotypes. These groups were a part of the populations from India (South Asia), the USA and Canada (North America), China (East Asia), Turkey and Iran (Middle East), and Ethiopia (East Africa).

Groups IV, V, VIII, and IX consisted of six genotypes each (5.45%), which were a part of the populations from China (East Asia), the USA (North America), India (South Asia), Turkey and Iran (Middle East), and Bangladesh (East Asia). We also found four groups (X, XII, XV, and XIII) with three genotypes each (2.73%), which were a part of the populations from Kazakhstan (Central Asia), China (East Asia), the USA (North America), India and Pakistan (South Asia), Turkey (Middle East), and Ethiopia (East Africa).

Most groups formed by the modified Tocher method included individuals collected from different populations and regions of the world, which indicated low variability between the evaluated genotypes. Reis et al. (2015), Araújo et al. (2019), and Hassani et al. (2020b) also found a low association between genetic diversity and the collection sites when evaluating characteristics of interest. Thus, the geographical origin is a poor indicator of genetic diversity, and it might not reflect greater genetic distance, which was the case in this study.

Table 5: Representation of the cluster generated by the modified Tocher optimization method based on the dissimilarity between the 110 genotypes of *Carthamus tinctorius* L.

Groups	Genotypes	Percentage of genotypes
I	20, 21, 22, 23, 24, 25, 26, 27, 42, 43, 45, 58, 61 and 82	12.74
II	11, 12, 13, 14, 15, 16, 17, 40, 41, 52 and 54	10.00
III	6, 59, 60, 77, 78, 79, 88, 89, 90, 93 and 96	10.00
IV	91, 92, 94, 95, 108 and 109	5.45
V	10, 33, 34, 35, 36 and 37	5.45
VI	8, 70, 71, 72, 73, 74, 76 and 85	7.27
VII	2, 3, 4, 87, 105, 106 and 107	6.36
VIII	29, 30, 44, 46, 62 and 110	5.45
IX	48, 49, 50, 51, 53 and 55	5.45
X	32, 75 and 86	2.73
XI	9 and 81	1.82
XII	7, 68 and 69	2.73
XIII	57, 66, 67, 83, 84, 98, 99, 100 and 101	8.18
XIV	63, 64, 65, 97 and 104	4.54
XV	5, 28 and 47	2.73
XVI	38, 39, 102 and 103	3.64
XVII	18 and 19	1.82
XVIII	1, 31 and 80	2.73
XIX	56	0.91

Groups XI and XVII consisted of only two genotypes each (1.82%), which were a part of the populations from India/Pakistan (South Asia) and China (Group XI) (East Asia), and Iran (Middle East) (Group XVII). Group XIX consisted of only one genotype (0.91%), suggesting that it was the most divergent of all the evaluated genotypes, and it came from the population of Iran (Middle East). This result indicated that genotype 56 was the most divergent relative to the other genotypes and should be considered for plant breeding programs. Similar results were obtained by Cordeiro et al. (2020), who reported the formation of groups with only one genotype using the modified Tocher method.

Based on the dendrogram obtained by the UPGMA hierarchical method with a significant cut at 90%, the genotypes were divided into 15 groups (Figure 1). The highest number of genotypes were found in Group I (23 genotypes), Group IV (11 genotypes), and Group IX (10 genotypes). All genotypes in the three groups belonged to populations from India and Pakistan (South Asia), Iran and Turkey (Middle East), Kazakhstan (Central Asia), China (East Asia), the USA (North America), and Ethiopia (East Africa).

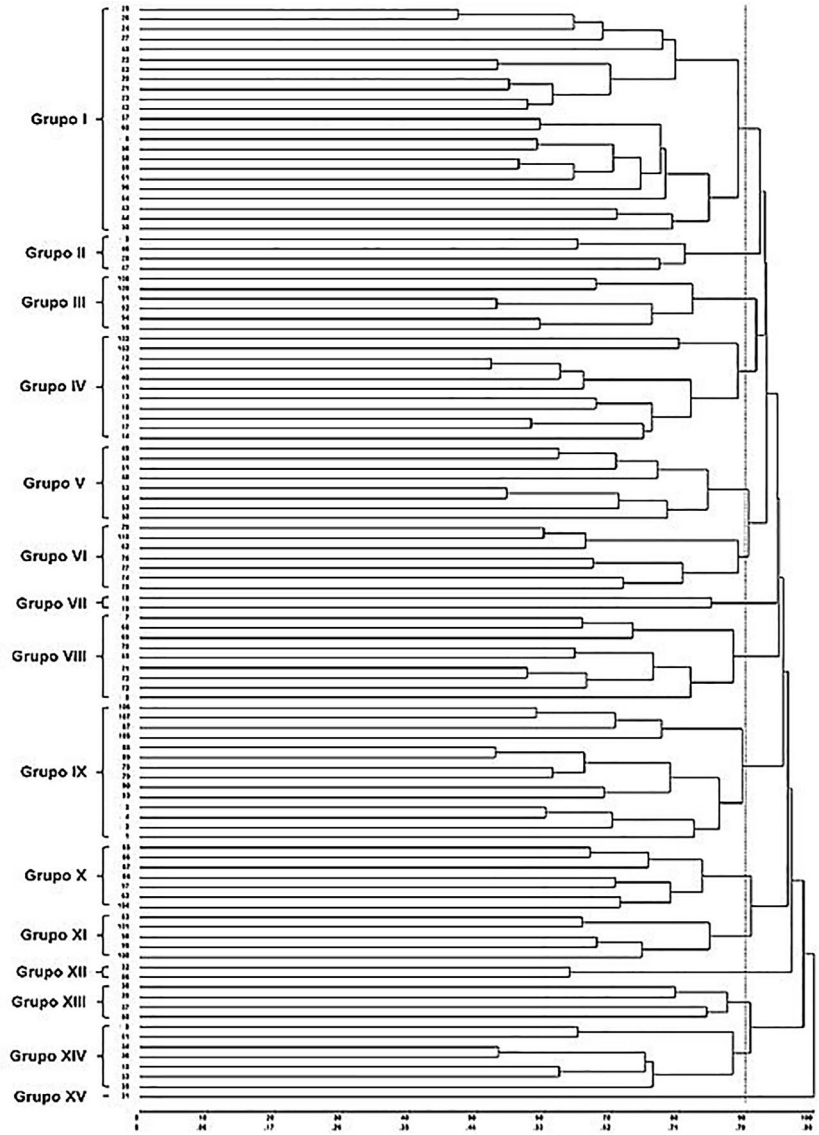

Figure 1: A dendrogram of the grouping of 110 safflower genotypes constructed by the UPGMA method and based on the dissimilarity estimated from molecular characteristics.

Nine groups had four to nine genotypes each. These groups included Group VIII with nine genotypes, Group V with eight genotypes, Groups VI, X, and XIV with seven genotypes each, Group III with six genotypes, Group XI with five genotypes, and Groups II and XIII with four genotypes. These genotypes belonged to populations from Turkey (Middle East), India and Pakistan (South Asia), the USA and Canada (North America), China (Central Asia), Iran (Middle East), Bangladesh (East Asia), and Ethiopia (East Africa), respectively.

Groups VII and XII consisted of two genotypes each, and group XV had one genotype. Thus, it had the most divergent genotype relative to the other groups. Groups with only one genotype are more divergent than the others. These genotypes can be used in breeding programs (Rotili et al., 2012). We found no geographic structure based on the similarity between the genotypes of the same population or region since some genotypes of the same population or region were allocated to different groups.

The consistency of the obtained dendrogram was evaluated by the co-phenetic correlation coefficient (CCC), which measured the correlation between the distances recovered from the dendrogram with the original distance matrix proposed by Sokal and Rohlf (1962). Based on the CCC, the results of the t-test conducted for the grouping method showed a significant value ($P \leq 0.01$) between groups for the mean grouping method (UPGMA). The correlation coefficient ($r \geq 0.62$) suggested variability in the consistency of the grouping pattern between the genotypes. Lira et al. (2021) reported similar results after evaluating 124 safflower genotypes, with a CCC of 0.70. Correa et al. (2020) evaluated the phenotypic dissimilarity in nine genotypes of sunflower, which also belongs to the Asteraceae family. They found a CCC of 0.65, and the results of their t-test were significant ($P < 0.01$).

Both grouping methods (modified Tocher and UPGMA) showed similarities in the grouping of genotypes. Groups X and XVII formed by the modified Tocher grouping, were similar to Groups XII and VII formed by the UPGMA method. The genotypes allocated in Groups I, XIX, and part of the modified Tocher group III were all allocated in Group I obtained by the UPGMA method. The other groups formed did not show similarity in the safflower genotype groupings. More groups were formed when the genotypes were grouped by the Tocher method than by the hierarchical UPGMA method. Oliveira et al. (2019) found similar results by applying these methodologies, where the UPGMA and Tocher methods could efficiently categorize the genotypes, although the number of groups formed was different.

Using the Bayesian method, proposed by Evanno, Regnaut, and Goudet (2005) and the Structure software, we identified the structure of this set of evaluated genotypes. The data from Delta K showed only one peak ($K = 2$). The data had the highest peak and the greatest adequacy between the suggested groups, assuming that was the real K value.

By analyzing the population structure (Figure 2), the safflower genotypes were placed into two groups, which matched the results obtained from the Delta K variation graph.

Group I consisted of 58 genotypes belonging to populations from Ethiopia (East Africa), Turkey and Iran (Middle East), India and Pakistan (South Asia), Kazakhstan (Central Asia), China (East Asia), and the USA (North America). Group II consisted of 63 genotypes belonging to populations from Turkey and Iran (Middle East), Bangladesh (East Asia), India (South Asia), China (East Asia), and the USA and Canada (North America).

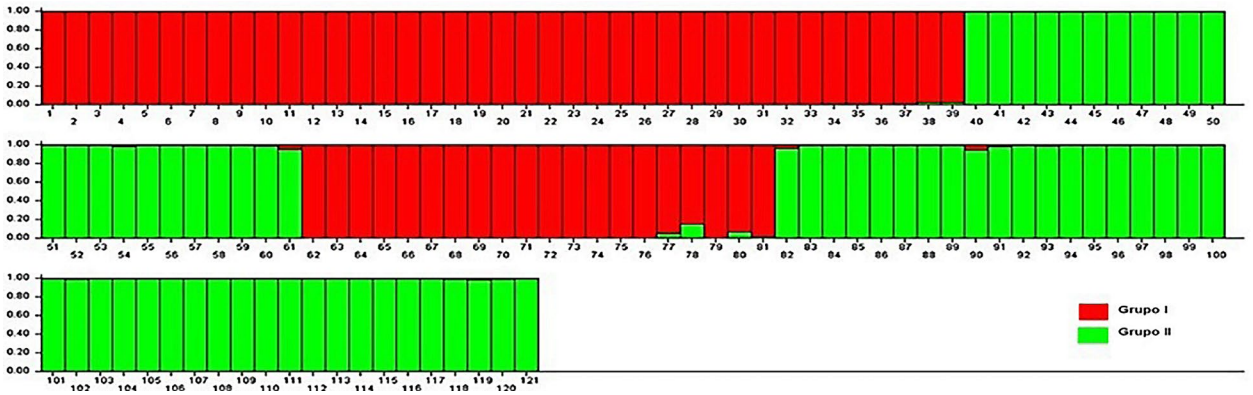

Figure 2: The population structure of 10 populations and six regions of *Carthamus tinctorius* L. included 121 genotypes based on 21 molecular SSR markers; K = 2. Each vertical bar represents a genotype and the percentage of adherence to each group.

Similar results were reported by Mokhtari et al. (2018), who evaluated the genetic diversity and genetic structure of the population of *Carthamus tinctorius* L. using SSR markers. In that study, the genotypes could be divided into two groups.

The presence of introgressions was related to the genotypes that contained different colors in the same bar of the bar plot (Figure 2). The average proportion of introgressed fragments was approximately 13%, and the genotypes were placed into two or more groups. The two groups evaluated showed introgression but with greater intensity in genotype 78 of Group I. The occurrence of introgression in the safflower genotypes was low, considering that only 11 genotypes showed introgression, coming from individuals of other populations.

A low rate of gene introgression was reported by Asfaw, Blair, and Almekinders (2009), who conducted genetic studies on landrace beans from Ethiopia and Kenya. Delfini et al. (2021) found an introgression of approximately 55% by analyzing the genetic diversity, population structure, and linkage disequilibrium (LD) in common bean accessions. Fisseha et al. (2016) studied accessions of beans from Ethiopia and reported high introgression (58%).

By performing principal coordinates analysis (PCoA), we determined the spatial distribution of the ten populations and how the 110 safflower genotypes presented themselves within the populations. The first two coordinates explained 5.74% of the total variation among the accessions, with dimensions 1 and 2 explaining 2.99% and 2.75%, respectively (Figure 3).

By comparing the results obtained from the Bayesian analysis (determined by the Structure software) (Figure 2) with those obtained by the UPGMA method (Figure 1) and the principal coordinates analysis (PCoA) (Figure 3), we found that the methods of analysis were similar, as eight genotypes (6, 20, 21, 22, 23, 24, 26, and 30) belonging to the populations from Pakistan, Kazakhstan, and India, originating from southern Asia and central Asia, were placed in the same groups although they were determined by the different methods of analysis.

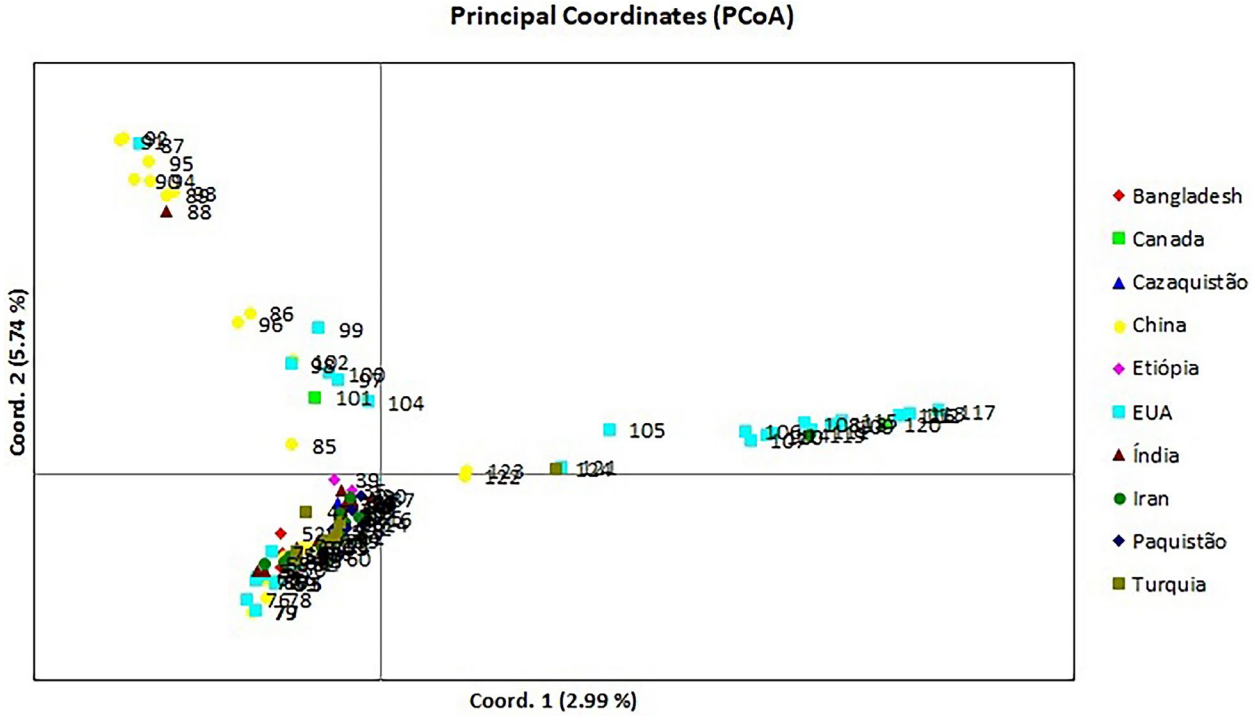

Figure 3: Principal coordinates analysis of 110 genotypes of *Carthamus tinctorius* L.

CONCLUSIONS

We found genetic variability among the genotypes of *Carthamus tinctorius* L. evaluated in this study using SSR markers. We conducted analyses based on the UPGMA and TOCHER methods, which placed the genotypes into several groups. The PCoA and Genetic Structure analyses placed the genotypes into fewer groups with a greater agglomeration of the genotypes. Our findings suggested that genotypes should be selected carefully for breeding programs, especially those involving hybridization, to maximize the genetic variability while choosing parents.

AUTHOR CONTRIBUTION

Conceptual idea: Oliveira, A. J.; Barelli, M. A. A.; Sander, N. L.; Methodology design: Oliveira, A. J.; Barelli, M. A. A.; Oliveira, T. C.; Sander, N. L. Data collection: Oliveira, A. J.; Sander, N. L.; Barelli, M. A. A. Data analysis and interpretation: Oliveira, A. J.; Barelli, M. A. A.; Oliveira, T. C.; Sander, N. L.; Azevedo, R. F.; Silva, C. R. and Writing and editing: Oliveira, A. J.; Barelli, M. A. A.; Oliveira, T. C.; Sander, N. L.; Azevedo, R. F.; Silva, C. R.

ACKNOWLEDGMENTS

To the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - Financing Code 001, for the support and scholarship granted. To the National Council for Scientific and Technological Development (CNPq) and the Research Support Foundation of the State of Mato Grosso (FAPEMAT) for financial support.

REFERENCES

AMBREEN, H. et al. Association mapping for important agronomic traits in safflower (*Carthamus tinctorius* L.) core collection using microsatellite markers. Frontiers in Plant Science, 9:402, 2018.

ARAÚJO, L. B. R. et al. Diversidade genética em pinhão manso com base em marcadores ISSR. Nativa, 7(4):363-370, 2019.

ASFAW, A.; BLAIR, M. W.; ALMEKINDERS, C. Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) landraces from the East African highlands. Theoretical and Applied Genetics, 120:1-12, 2009.

BAHMANKAR, M.; NABATI, D. A.; DEHDARI, M. Genetic relationships among Iranian and exotic safflower using microsatellite markers. Journal of Crop Science and Biotechnology, 20:159-165, 2017.

CORDEIRO, A. G. M. L. et al. Diversidade genética entre cupuízeiros nativos do Portal da Amazônia, Mato Grosso, Brasil. Scientific Electronic Archives, 13(3):51-56, 2020.

CORREA, V. R. S. et al. Dissimilaridade fenotípica em genótipos de girassol. Research, Society and Development, 9(11):e3489119814, 2020.

CRUZ, C. D. GENES: A software package for analysis in experimental statistics and quantitative genetics. Acta Scientiarum. Agronomy, 3(3):271-276, 2013.

DELFINI, J. et al. Population structure, genetic diversity and genomic selection signatures among a Brazilian common bean germplasm. Scientific Reports, 11:2964, 2021.

EVANNO, G.; REGNAUT, S.; GOUDET, J. Detecting the number of clusters of individuals using the software structure: A simulation study. Molecular Ecology, 14(8):2611-2620, 2005.

EXCOFFIER, L.; LAVAL, G.; SCHNEIDER, S. Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online, 1:47-50, 2005.

EXCOFFIER, L.; SMOUSE, P. E.; QUATTRO, J. M. Analysis of molecular variance in ferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics, 131(2):479-491, 1992.

FISSEHA, Z. et al. Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) germplasm of Ethiopia as revealed by microsatellite markers. African Journal of Biotechnology, 15:2824-2847, 2016.

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS - FAO. FAOSTAT Emissions database. 2019. Available in: <https://www.fao.org/faostat/en/#data/QCL>. Access in: September, 9 2020.

GOLKAR, P. Breeding improvements in safflower (*Carthamus tinctorius* L.): A review. Australian Journal of Crop Science, 8(7):1079-1085, 2014.

GOLKAR, P.; MOKHTARI, N. Molecular diversity assessment of a world collection of safflowers genotypes by SRAP and SCoT molecular markers. Physiology and Molecular Biology of Plants, 24:1261-1271, 2018.

HASSANI, S. M. R. et al. In-depth genome diversity, population structure and linkage disequilibrium analysis of worldwide diverse safflower (*Carthamus tinctorius* L.) accessions using NGS data generated by DArTseq technology. Molecular Biology Reports, 47:2123-2135, 2020b.

HASSANI, S. M. R. et al. Morphological description, genetic diversity and population structure of safflower (*Carthamus tinctorius* L.) mini core collection using SRAP and SSR markers. Biotechnology & Biotechnological Equipment, 34(1):1043-1055, 2020a.

KIM, S. G. et al. First report of fusarium wiltca used by *Fusarium proliferatum* on safflower. Research in Plant Disease, 22(2):111-115, 2016.

KIRAN, B. U. et al. Genetic diversity of safflower (*Carthamus tinctorius* L.) germplasm as revealed by SSR markers. Plant Genetic Resources, 15(1):1-11, 2017.

KUMAR, S. et al. Utilization of molecular, phenotypic, and geographical diversity to develop compact composite core collection in the oil seed crop, safflower (*Carthamus tinctorius* L.) through maximization strategy. Frontiers in Plant Science, 7:1554, 2016.

LEE, G. A. et al. Genetic assessment of safflower (*Carthamus tinctorius* L.) collection with microsatellite markers acquired via pyrosequencing method. Molecular Ecology Resources, 14(1):69-78, 2014.

LIRA, J. P. E. et al. Safflower genetic diversity based on agronomic characteristics in Mato Grosso state, Brazil, for acrop improvement program. Genetics and Molecular Research, 20(1):gmr18698, 2021.

MOKHTARI, N. et al. Assessment of genetic diversity and population genetic structure of *Carthamus* species and Iranian cultivar collection using developed SSR markers. Journal of Genetics, 97:67-78, 2018.

MOURA, P. C. S. et al. Características gerais e ecofisiologia do cártamo (*Carthamus tinctorius* L.). Journal of Agronomic Sciences, 4:136-150, 2015. Available in: <http://www.pag.uem.br/anteriores/v4ne>. Access in: January, 26 2022.

OLIVEIRA, C. S. et al. Características de plântulas: Dissimilaridade genética entre acessos de pimenta. Revista Ciência, Tecnologia & Ambiente, 9:e09114, 2019.

PETIT, R. J.; MOUSADIK, A.; PONS, A. O. Identifying populations for conservation on the basis of genetic markers. Conservation Biology, 12(4):844-855, 1998.

PRITCHARD, J. K. et al. Association mapping in structured populations. The American Journal of Human Genetics, 67(1):170-181, 2000.

QUEIROGA, V. P.; GIRÃO, E. G.; ALBUQUERQUE, E. M. B. Cártamo (*Carthamus tinctorius* L.) tecnologias de plantio e utilização. Campina Grande: AREPB, 2021. 148p.

REIS, M. V. M. et al. Variabilidade genética e associação entre caracteres em germoplasma de pinhão-manso (*Jatropha curcas* L.). Revista Ciência Agronômica, 46(2):412-420, 2015.

ROTILI, E. A. et al. Divergência genética em genótipos de milho, no Estado do Tocantins. Revista Ciência Agronômica, 43(3):516-521, 2012.

SAADAOUI, E. et al. Castor bean (*Ricinus communis* L.): Diversity, seed oil and uses. In: PARVAIZ, A. (Eds.). Oilse ed crops: Yield and adaptation sunderen vironmental stress. Nova Jersey: Wiley Online Library. p.19-33, 2017.

SHARIFI, R. S.; NAMVAR, A. Grain filling and fatty acid composition of safflower fertilized with integrated nitrogen fertilizer and biofertilizers. Pesquisa Agropecuária Brasileira, 52(4):236-243, 2017.

SILVEIRA, L. et al. Influência alelopática do extrato aquoso de folhas de citronela (*Cymbopogon*) sobre a germinação e desenvolvimento inicial de quatro genótipos conhecidos de cártamo (*Carthamus tinctorius* L.). Acta Iguazu, 6(5):197-206, 2017.

SINGH, V.; NIMBKAR, N. Safflower. In: GUPTA, S. K. (ed). Breeding oilseed crops for sustainable production. EUA: Academic Press Elsevier. p.149-167, 2016.

SOKAL, R. R.; ROHLF, F. J. The comparison f dendrograms by objective methods. Wiley, 11(2):33-40, 1962.

WEIR, B. S. Genetics data analysis II: Methods for discrete population genetic data. Suderland: Sinauer Associates. 1996. 455p.

WILLIAMS, J. G. K. et al. DNA polymorphism simplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research, 18(22):6531-6535, 1990.

WRIGHT, S. The genetical structure of populations. Annals of Eugenics, 15(1):323-354, 1949.