Genet Resour Crop Evol (2024) 71:2341–2353
https://doi.org/10.1007/s10722-023-01783-6

RESEARCH ARTICLE

Genetic diversity of *Citrullus colocynthis* populations using phytochemical analysis and SCoT marker variations

Abdelfattah Badr · Hoida Zaki

Received: 4 August 2023 / Accepted: 16 October 2023 / Published online: 7 November 2023

Abstract *Citrullus colocynthis* L. Schard (bitter melon) is a drought-resistant medicinal plant growing in Egypt and many other countries in the arid environments of the World. In Egypt, it is abundant in several locations in Egypt's Eastern Desert, extending from the Nile River eastward to the Red Sea, particularly in the Kosseir region on the Red Sea coast. It has a wide range of applications in traditional medicine due to its anti-inflammatory, antibacterial, antioxidant, and anesthetic features. In this study, the genetic diversity was explored using chemical analysis of the secondary metabolites in seed extract in 15 populations from different sites in the Eastern Desert of Egypt to correlate the chemical variation with genetic differences among populations as revealed by DNA fingerprinting using the Start Codon Targeted (SCoT) markers. A total of 81 chemical compounds were identified from the 15 populations. Retention time, peak area percentage, molecular weight, and chemical formula were determined for each compound. Ten SCoT primers produced 137 bands in the 15 populations of these; 85 bands were polymorphic, 50 were monomorphic, and two bands were unique to a single population. Populations located in the southern part of the Eastern Desert have similar levels of phytochemical and genetic diversity and are differentiated from populations in the northern part. On the other hand, populations in the middle part show a small resemblance to other populations in the north and south, indicating an impact of the ecology on the genetic differentiation and the chemical composition of the secondary metabolites in *C. colocynthis*.

Keywords *Citrullus colocynthis* · Genetic diversity · SCoT marker · GC–MS analysis

Introduction

Colocynth, bitter apple, bitter gourd, and bitter melon are all names given to the fruit of the *Citrullus colocynthis* (L.) Schrader, a wild perennial, plant of the family Cucurbitaceae. It is a perennial (rarely annual) prostrate plant that has a bitter taste and potential for poisoning (Heuzé et al. 2021). *Citrullus colocynthis* is predominantly a drought-resistant vine that has spread widely across the Sahara-Arabian Peninsula of Africa and the Mediterranean Basin (Dane et al. 2007; Li et al. 2022). It is abundant in Egypt, particularly in the Kosseir region on the Red Sea coast

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10722-023-01783-6.

A. Badr
Botany and Microbiology Department, Faculty of Science, Helwan University, Cairo 11790, Egypt
e-mail: abadr@science.helwan.edu.eg

H. Zaki (✉)
Botany and Microbiology Department, Faculty of Science, South Valley University, Qena 83523, Egypt
e-mail: hoaida.zaki@sci.svu.edu.eg

(Boulos 2002), but it is also plentiful in several locations in Egypt's Eastern Desert (Sheded et al. 2006; Badr et al. 2018). The fruit of *C. colocynthis* has a wide range of applications in traditional medicine across several nations (Heydari et al. 2016; Torkey et al. 2009) due to its anti-inflammatory, antibacterial, antioxidant, and anesthetic qualities (Kumar et al. 2008; Rizvi et al. 2018; Hameed et al. 2020). In Egypt, notable variations exist in the morphological quantitative characteristics of *C. colocynthis* populations in the Eastern Desert of Egypt. These variants include a tall vine length (up to 195 cm), the presence of compound hairs, and a high hair density. The investigation of fruit rind color variation, including rind color pattern, as well as the examination of seed color and size, has garnered significant attention in the study conducted by Badr et al. (2018). However, little variations in qualitative characteristics between populations, specifically in terms of leaf hair density, leaf hair type, rind color pattern, main rind color, and rind stripe color, were observed. The observed variations in fruit and seed features among populations of *C. colocynthis* highlight the significance of these qualities in contributing to the diversity of this species (Badr et al. 2018).

The phytochemical analysis of the *C. colocynthis* root, leaf, flower, and fruit independently revealed the presence of alkaloids, carbohydrates, and flavonoids and the absence of tannins, gums, and mucilage. The seeds contain large levels of various fatty acids, they include myristic, palmitic, stearic, oleic, linoleic, and linolenic (Uma and Sekar 2014). Gurudeeban et al. (2010) provided a review of the chemical makeup and the therapeutic possibilities of *C. colocynthis*. More detailed results were published by Kamalakar et al. (2015) and Bourhia et al. (2020, 2021) that the seed oil of *C. colocynthis* is edible and has a similar composition to soybean oil and the seeds have a high proportion of unsaturated fatty acids (79.80%), primarily linoleic acid and oleic acid, a low percentage of saturated fatty acids (20.20%), and a very low n-3 poly-unsaturated fatty acids level (0.5%). The seed oil potential as a biodiesel feedstock was investigated by Giwa et al. (2010), and physicochemical characterization of the seed oil showed that seed fat is made up of 10.4% palmitic acid, 6.5% stearic acid, 1.70% arachidic acid, 11.7% oleic acid, 58.5% linoleic acid, and 1.6% linolenic acid (Giwa et al. 2010; Riaz et al. 2015). Fereshtian et al. (2017) studied phytochemical traits to evaluate the genetic diversity of some *C. colocynthis* accessions from Iran.

Genetic diversity in plant species provides vital information on their capacity to cope with different environmental stresses (Badr 2008) because domestication and the evolution of crops reduce the diversity of crop species (Begna 2021). Increased genetic variation improves the chances of effective plant selection, making it an essential component in utilization by evaluating its scope and range (Begna 2021; Kardos et al. 2021). Different approaches have been used to determine genetic polymorphism in many plant species. Recently, estimating genetic diversity and deciphering genetic composition have both benefited greatly from the use of molecular markers, identifying the genes implicated in critical growth mechanisms and conserving genetic variation (Badr and El-Shazly 2012; El-Shazly et al. 2020). To better estimate genetic diversity, molecular techniques have been used in tandem with morphological variations. In medicinal and aromatic plants, the chemical composition assessment of secondary metabolites provide an additional source of important information (Badr et al. 2021; Talebi et al. 2021).

Both the distribution of genes between populations and the range of possible variations within a species in addition life history criteria, population size, breeding system, pollinators, and phylogenetic position must be considered in order to comprehend genetic variation (Hamrick and Godt 1996; Schierenbeck 2016). The genetic diversity of both wild and cultivated species of the Cucurbitaceae was assessed using morphological and agronomical traits and isozymes polymorphism (Ahmed and Alamer 2018). Guo et al. (2020) reported that multiple whole-genome duplications and important morphological and molecular variations enhanced genetic diversity preceding polyploidizations in the Cucurbitaceae. These findings established a phylogenetic framework for the family Cucurbitaceae and revealed novel information about the morphological and genetic changes underlying the family's adaptive evolution. Wax gourds (*Benincasa hispida*) are a cucurbit plant with fruits that can reach lengths of 80 cm and weigh more than 20 kg. Xie et al. (2019) addressed the genetic diversity in Cucurbitaceae by identifying many genome parts and genes that could have been chosen during domestication and likely led to the large fruit size.

Meanwhile, several molecular markers have found widespread use in Cucurbitaceae plants. The RAPDs was used for the analysis of genetic diversity among watermelon accessions (Ghanavati et al. 2018; Pandey et al. 2019; Ebadi et al. 2022). Genetic diversity in the species of the genus *Citrullus* in Iran was addressed by assessing the Polymorphism Information Content (PIC) using the SSR polymorphism (Avval 2017). Michael et al. (2019) also used SSR markers to address the genetic variation of *Cucurbita pepo* accessions that show resistance to phytophthora crown rot. *C. colocynthis* populations were compared for genetic diversity using variation in morphological traits, RAPD and ISSR markers, and genome sequencing by Ghanavati et al. (2018), Pandey et al. (2019) and Badr et al. (2018).

The Start Codon Targeted (SCoT) marker is a recent molecular fingerprinting method that was first developed by Collard and Mackill (2009). These authors found that the repeatability of SCoT markers depends on more than just primer length and annealing temperature; hence, these markers are highly reproducible. The ATG start codon in plant genes is flanked by a short, conserved sequence that serves as the basis for the SCoT marker. The current status and recent development of the SCoT marker were recently reviewed (Rai 2023). Xanthopoulou et al. (2015) examined the genetic diversity across landraces of squash (*Cucurbita pepo*) stored in the Greek Gene bank. The dioecious crop *Trichosanthes dioica* was analyzed using SCoT, ISSR, and SRAP markers to determine its genetic variation, population structure, and sex identity (Kumar and Agrawal 2019). The SCoT fingerprinting has proven useful in analyzing gene flow and structure of population for many Cucurbitaceae species, and they were applied to compare two distinct bottle gourd populations (Bhawna et al. 2017), assessing genetic diversity between Pumpkin accessions growing in the North of Saudi Arabia (Abdein 2018) and estimating of genetic diversity and population structure in bitter gourd accessions collected from several regions of India (Tyagi et al. 2020).

The objective of this study is to conduct a comparative analysis of the genetic diversity using SCoT markers, as well as the chemical composition, in populations of the bitter melon in Egypt. Furthermore, the study aims at utilizing the diversity in volatile compounds across 15 *C. colocynthis* populations, which span five distinct ecosystems, to assess the variation among populations and its value for conservation and utilization of the bitter melon in Egypt.

Materials and methods

Plant sampling

Citrullus colocynthis accessions were obtained from various natural locations in the Egyptian Eastern Desert as mature fruiting plants. Accessions from the 15 populations of *Citrullus colocynthis* were collected from the locations, GPS coordinates, and elevations listed in Table 1 and plotted on a map of the Eastern Desert of Egypt (which extends from the Nile River eastward to the Red Sea coast (Fig. 1). The 15 sites extend from Wadi Abo Had in Ras Garib (1) in the north to Wadi Hodein in the Shalatein province (14) and Wadi Allaqi, South of Aswan (15).

DNA isolation and SCoT marker analysis

For DNA isolations, seeds of all populations were germinated on wet blotting in Petri dishes for two weeks, and fresh, actively growing young leaves were used for DNA extraction. Using the technique provided by the manufacturer, we were able to successfully isolate complete genomic DNA using the DNeasy Plant Mini Kit (Qiagen, Germany). A NanoDrop 2000 Spectrophotometer from Thermo Scientific in Germany was used to determine the quantity of DNA present in the samples as well as their level of purity.

SCoT reactions and PCR thermocycling profile

For the SCoT DNA fingerprinting, Ten SCoT primers were employed, and the primer codes and DNA base sequences are given in Table 2. A 25 µl reaction volume of the DNA TE solution containing 12.5 µl Master Mix (sigma), 2.5 µl primer (10 pcmol), 3 µl template DNA (10 ng), and 7 µl dH_2O was used to perform the amplification reaction. The PCR thermocycling profile was set at an initial denaturation cycle of 5 min at 94 °C; then, the PCR amplification was carried out in a Perkin-Elmer/GeneAmp® PCR System 9700 (PE Applied Biosystems). Denaturation occurred at 94 °C for one minute, annealing occurred

Table 1 The sites, GPS location data, and elevation of localities from which accessions of the 15 examined populations of *Citrullus colocynthis* were collected from sites ranging from Wadi Abo Had in Ras Garib (1) in the north to Wadi Hodein in the Shalatein province (14) and Wadi Allaqi, South of Aswan (15)

Population	Site name	GPS Locations data		Elevation (m)
		Longitude	Latitude	
1	Ras Garib (Wadi Abo Had)	32°37′41.16"E	28°11′40.56"N	537
2	Wadi Qena (Naq El-Teir)	32°55′12.36"E	26°44′15.36"N	257
3	Qena-Safaga Road (Wadi Umm Taghir)	33°35′2.04"E	26°40′53.76"N	618
4	Qena-Safaga Road (Wadi El-Markh)	33°22′9.84"E	26°32′45.60"N	531
5	Qena-Safaga Road (Wadi El-Markh)	33°3′14.76"E	26°21′28.44"N	240
6	South Valley University Botanic Garden	32°44′38.76"E	26°11′35.52"N	86
7	Qeft-Quseir Road (Wadi El Nakheil)	34°11′38.04"E	26°6′23.40"N	70
8	Qeft-Quseir Road (Wadi Abu Zidan)	33°49′14.16"E	26°0′3.60"N	420
9	Qeft-Quseir Road (Wadi El-Laqeita)	33°15′36.36"E	25°55′1.92"N	187
10	Wadi Abbad (Bir Abbad)	33°1′21.36"E	25°0′55.80"N	110
11	Wadi Gimal	35°0′57.18"E	24°37′30.10"N	187
12	Wadi Shait	33°5′7.80"E	24°32′49.92"N	111
13	Aswan University Campus	32°51′44.28"E	23°59′54.60"N	163
14	Shalatein (Wadi Hodein)	35°34′14.60"E	23°2′18.70"N	53
15	Wadi Allaqi	33°13′2.28"E	22°46′15.24"N	250

at 50 °C for one minute, and elongation occurred at 72 °C for one minute in each cycle. In the final cycle, at a temperature of 72 °C, the primer extension time increased to a total of 7 min (Jiang et al. 2014).

The amplification results were visualized as bright bands under UV light after electrophoresis of the reaction mixture on a 1.5% agarose gel stained with ethidium bromide (0.5ug/ml) in 1X TBE buffer at 100 V. The PCR gels were photographed using a UV light and the Gel Documentation System (BIO-RAD 2000).

Preparation of plant extracts

Ten grams of powdered plant seeds were soaked in the organic solvent of 85% ethanol and then shaken for 12 h to extract the active components and after being filtered through Whatman filter paper No. 1, the extract was then chilled to 4°C for its next use.

GC–MS analysis

The GC–MS analysis of the chemical makeup of *Citrullus colocynthis* seed extract samples was done using Trace GC1310-ISQ mass spectrometer (Thermo Scientific, Austin, TX, USA) equipped with a direct capillary column TG-5MS (30 m × 0.25 mm × 0.25 m film thickness). The temperature in the column oven was increased from its initial setting of 50 °C to its final destination of 230 °C at a pace of 5 °C per minute, where it remained for 2 min. The heat was increased to 290 °C at a rate of 30 degrees per minute and then held for 2 min. The carrier gas used was helium, and the flow rate was kept constant at one milliliter per minute. The injector line temperature was maintained at 250 °C, while the MS transfer line temperature was maintained at 260 °C. Autosampler AS1300 paired with GC in split mode was used to automatically inject 1: 1 diluted samples with a solvent delay of 3 min. At ionization voltages of 70 eV, EI mass spectra were acquired in full scan mode for the m/z range of 40–1000. An ion source temperature of 200 °C was selected (Adams 2017). To determine the identities of the constituents, the retention times and mass spectra of each component were compared to the information included in the WILEY 09 and NIST 11 databases.

Data analysis The genetic distance between the populations was determined by the presence or absence of markers using SCoT fingerprinting. Coefficients of Euclidean distance were determined using the

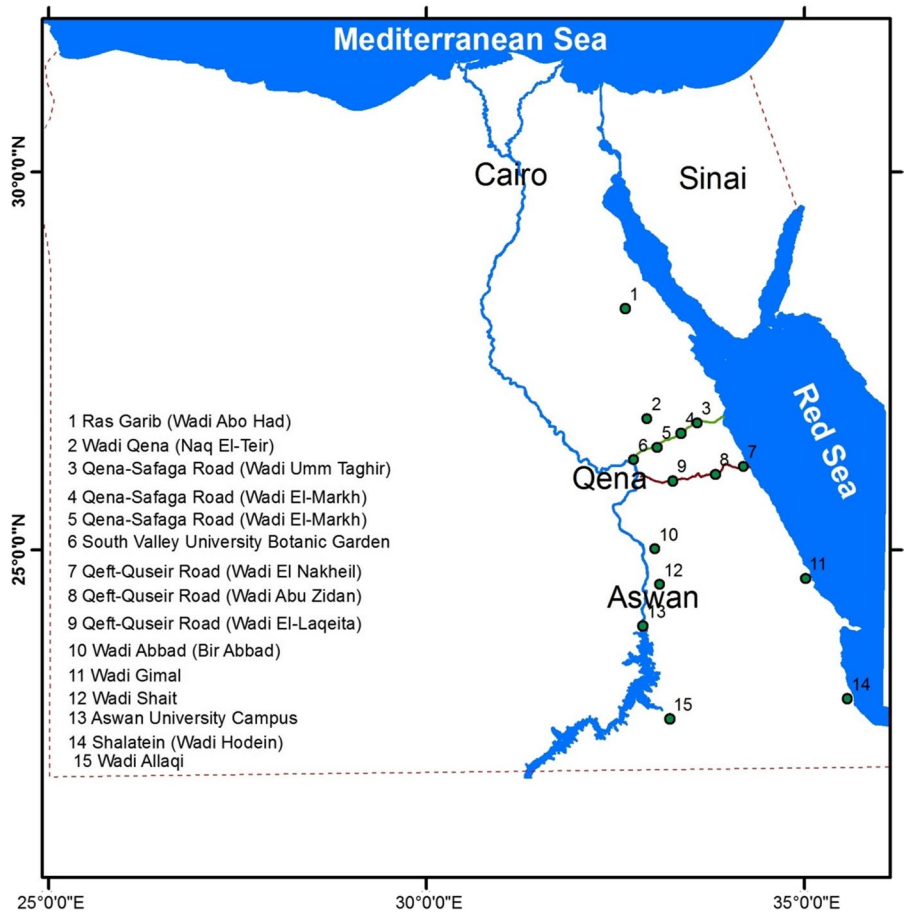
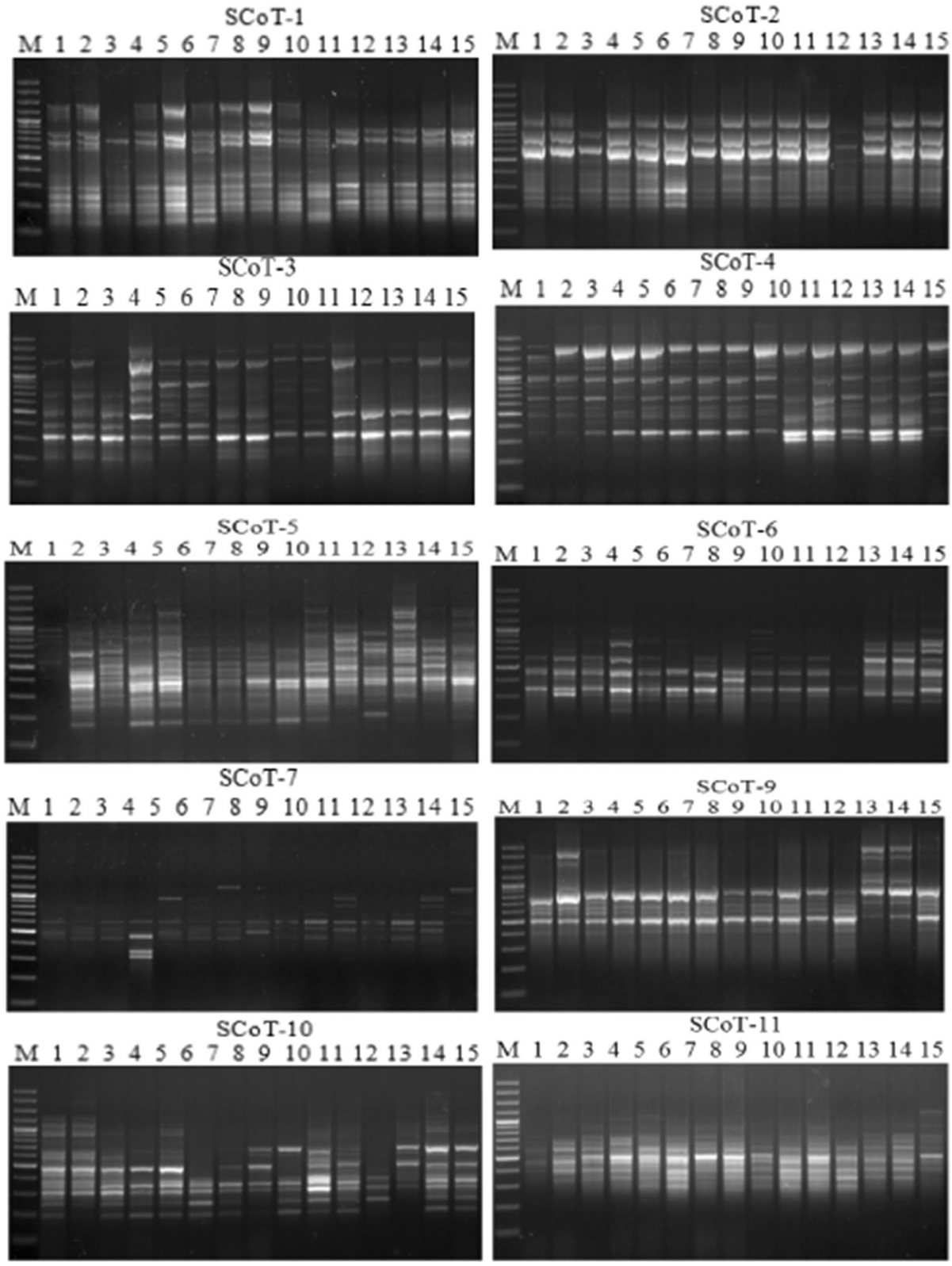


Fig. 1 A map of Egypt showing the sites of *Citrullus colocynthis* 15 studied populations in the Eastern Desert of Egypt extending from the River Nile eastwards the Red Sea coast

Table 2 Primers codes, DNA base sequences, number of polymorphic, monomorphic, unique, and total bands produced by each primer, and polymorphism percentage calculated by the analysis of Start Codon Targeted fingerprinting in the examined *Citrullus colocynthis* populations

Primer	Sequence	Monomorphic bands	Polymorphic bands	Unique bands	Total bands	Polymorphism (%)	Mean band frequency
SCoT-1	5'-ACGACATGGCGACCACGC-3'	6	9	0	15	60%	0.6
SCoT-2	5'-ACCATGGCTACCACCGGC-3'	9	4	0	13	31%	0.8
SCoT-3	5'-ACGACATGGCGACCCACA-3'	5	10	0	15	67%	0.7
SCoT-4	5'-ACCATGGCTACCACCGCA-3'	7	6	0	13	46%	0.8
SCoT-5	5'-CAATGGCTACCACTAGCG-3'	5	12	0	17	71%	0.6
SCoT-6	5'-CAATGGCTACCACTACAG-3'	4	10	0	14	71%	0.6
SCoT-7	5'-ACAATGGCTACCACTGAC-3'	1	7	2	10	90%	0.4
SCoT-9	5'-ACAATGGCTACCACTGCC-3'	5	7	0	12	58%	0.6
SCoT-10	5'-ACAATGGCTACCACCAGC-3'	3	13	0	16	81%	0.6
SCoT-11	5'-ACAATGGCTACCACTACC-3'	5	7	0	12	58%	0.7
total		50	85	2	137		

PAST Software version 3.22 using the formula given by (Legendre and Legendre 2012). Using the PAST Software and the un-weighted pair group with arithmetic average (UPGMA) (Sokal and Michener 1958),


SCoT-1
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
SCoT-2
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
SCoT-3
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
SCoT-4
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
SCoT-5
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
SCoT-6
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
SCoT-7
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
SCoT-9
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
SCoT-10
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
SCoT-11
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

◂**Fig. 2** Photographs illustrating the Start Codon Targeted fingerprinting profile produced by the ten primers (given in Table 2) in the examined populations of *Citrullus colocynthis*

distance cluster trees were created to estimate genetic diversity among the studied populations. Evaluation of the SCoT markers' potential to estimate genetic variability was performed using the heterozygosity index (H), polymorphic information content (PIC), marker index (MI), effective multiplex ratio (E), discriminating power (D), arithmetic mean of H (H.av), and resolving power (RP) (Amiryousefi et al. 2018). Principle component analysis (PCA) was performed based on SCoT & GC-mass data using MVSP ver.3.1 (Multivariate Statistical Program).

Compounds isolated by GC–MS analysis were assigned numerical values as 0 for absence and 1 for presence in binary matrices for the construction of cluster trees and Principal Component scatter diagram using the PCA to illustrate the diversity between populations based on variation in the types of the secondary metabolites in the seeds of the examined populations. For all sampled *C. colocynthis* populations, the identified compounds' data was analyzed using the PAST Software ver. 3.22 (Hammer et al. 2001). Details of the GC–MS data used for the analysis are available in Supplementary Table 1.

Results

After optimizing the amount of template DNA, Taq DNA polymerase, and $MgCl_2$ concentrations, and the thermal profile of the amplification procedure, genomic DNA extracted from bulk samples of the 15 populations of *C. colocynthis* was amplified using 10 SCoT primers. Photographs of SCoT fingerprinting are shown in Fig. 2, and Table 2 lists the number of polymorphic, monomorphic, unique, and total bands produced by each primer, as well as the polymorphism percentage obtained from the study of SCoT fingerprinting in the sampled *C. colocynthis* populations.

As shown in Table 2, the ten primers produced a total of 137 bands, including 85 polymorphic ones, while the 15 populations of *C. colocynthis* shared 50 monomorphic bands. and only two unique bands were scored by primer 7. The number of bands generated by primer SCoT 5 was the largest (17), while the number generated by primer SCoT 7 was the lowest (10). All but one of the bands produced by the primers SCoT 7 are polymorphic; on the other hand, the SCoT 10 profile showed the highest number of polymorphic bands (13 bands from a total of 16), and the SCoT 2 profile showed the lowest number of bands (four polymorphic bands from a total of 13). Polymorphism was found in varying amounts, from 90% in SCoT 7 to only 31% in SCoT 2 primer. Details of the scoring of the numbers of monomorphic, polymorphic, unique, and total bands recorded in the SCoT fingerprinting of the 15 *C. colocynthis* populations are given in Supplementary Table 2. The maximum number of polymorphic bands (52) was found in population 5. Populations 12 and 9 had much fewer polymorphic bands (30 and 32, respectively). Only two unique bands appeared in population 4 produced by SCoT 7 primer. The percentages of polymorphic bands ranged between 51% in population 5 to 37.5% in population 12.

Heterozygosity index, PIC, marker index, effective multiplex ratio, discriminating power, arithmetic mean of H, and resolving power are shown in Table 3. As marker parameters. Each primer's PIC value is calculated by averaging the PIC values across all loci. The PIC values for the different primers varied from 0.415 (SCoT 2) to 0.336 (SCoT 6), with an average of 0.36. The ratio of polymorphic fragments to the total is called the effective multiplex ratio (E). Our results showed that the primer SCoT 5 produced the highest effective multiplex ratio (11). Effective multiplex ratio (E) values ranged from 3.667 for the SCoT 7 primer to 8.967 on average across all primers. To evaluate the overall value of the marker system, each SCoT primer's marker index was calculated. The MI ranged from 0.020 for primers SCoT 1, SCoT 3, SCoT 5, and SCoT 9 to 0.011 for primer SCoT 7. On average, the MI was 0.0179 for each primer. The primers' ability to discriminate among samples is measured by a metric called "resolving power" (R). The average RP was 4.8267 for all primers, with RP = 7.60 for the SCOT 10 primer and RP = 2.40 for the SCOT 2 primer.

The index of heterozygosity (H) is the likelihood that a given population member its heterozygous for the locus in question. The average RP was 0.4227, with the SCoT 6 primer having the highest H index (0.484) and the SCoT 2 primer the lowest (0.274). The discriminating power (D) of a test measures

Table 3 Marker parameters calculated for each Start Codon Targeted primer for the 15 *Citrullus colocynthis* populations

	H	PIC	E	H. av	MI	D	RP
SCoT-1	0.468	0.343	9.400	0.002	0.020	0.608	5.867
SCoT-2	0.274	0.415	10.867	0.001	0.015	0.302	2.400
SCoT-3	0.447	0.352	9.933	0.002	0.020	0.562	5.200
SCoT-4	0.301	0.407	10.600	0.002	0.016	0.336	4.400
SCoT-5	0.457	0.348	11.000	0.002	0.020	0.582	7.200
SCoT-6	0.484	0.335	8.267	0.002	0.019	0.652	6.267
SCoT-7	0.464	0.344	3.667	0.003	0.011	0.867	2.933
SCoT-9	0.470	0.342	7.467	0.003	0.020	0.614	3.067
SCoT-10	0.482	0.336	9.533	0.002	0.019	0.646	7.600
SCoT-11	0.380	0.380	8.933	0.002	0.019	0.447	3.333
Avg./primer	0.4227	0.3602	8.9667	0.0021	0.0179	0.5616	4.8267

H Heterozygosity index, *PIC* Polymorphic Information Content, *E* Effective multiplex ratio, *H. av* Arithmetic mean of heterozygosity, *MI* Marker Index, *D*: Discriminating power, *RP* Resolving power

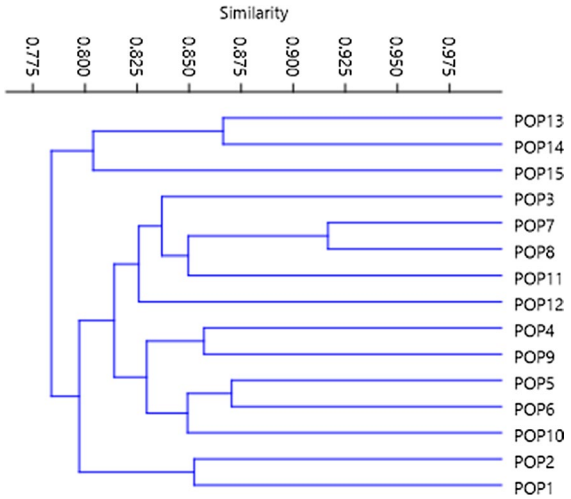


Fig. 3 A distance un-weighted pair group with arithmetic average tree illustrating the relationships among the *Citrullus colocynthis* populations based on the analysis of Start Codon Targeted analysis

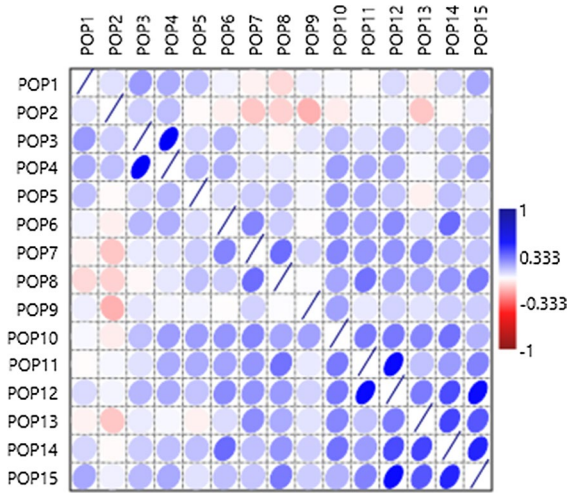


Fig. 4 Correlation map based on Start Codon Targeted & GC analysis by PAST using Pearson coefficient

how likely it is that any two randomly selected test individuals will have different test results. The average discriminating power (D) across all primers was 0.562, with the SCoT 7 primer having the highest D (0.867) and the SCoT 2 primer having the lowest D (0.302).

The Euclidean similarity index among the 15 populations of *C. colocynthis* was calculated (Supplementary Table 3). The highest degree of similarity was found between populations 7 and 8, while the lowest degree was found between populations 13 and 5. A UPMGA tree constructed using the PAST software (Fig. 3) depicted the genetic diversity present in the sampled populations based on SCoT polymorphism. In this dendrogram, three populations, 13 to 15, and populations 1 and 2, were separated into two small clusters. The main cluster was divided into two subclusters: one comprising populations 6, 7, 11, and 12, and the other composed of the remaining populations (4, 9, 5, 6, and 10).

Pearson correlations between the *C. colocynthis* populations based on SCoT analysis using PAST are shown in Fig. 4. A highly positive correlation was found between populations 3 and 4 and populations 11 and 12. In contrast, a negative correlation was found between populations 1 and 6 and between populations 2 and 7, 8, 9, and 13. In general, more

positive correlations were recorded in populations 13, 14, and 15.

Many phytochemical substances were isolated from the *C. colocynthis* seed ethanol extracts using GC–MS analysis. A total of 81 combinations were identified from the 15 populations. Retention time, peak area percentage, molecular weight, and chemical formula were all determined for each compound (Supplementary Table 1). The highest molecular weight was 624 for cholest-5-en-3-yl palmitate, and the lowest molecular weight was 184 for 3-heptyne-2,5-diol,6-methyl-5-(1-methylethyl, scored in population 1. The compound with the lowest retention time (11.7 min) was 6-tetradecene, whereas 1-Hexacosanol had the highest retention time (32.14 min) (Supplementary Table 1).

Cluster dendrograms constructed using the Euclidean distance based on a variation of the chemical compounds found in ethanol seeds extracts of *C. colocynthis* growing in different habitats in the Eastern desert of Egypt are shown in Fig. 5. The 15 populations are delimited as two major clusters, one comprising five populations (5, 1, 2, 3, and 4), and the other comprises the other ten populations, close similarity was evident between subclusters of populations including 13, 14, and 15 as one sub-cluster and populations 6, 7, 10, 11 and 12 as another subcluster whereas the two populations 8 and 9 were clearly differentiated as single branches in the tree indicating small resemblance between themselves and the other populations.

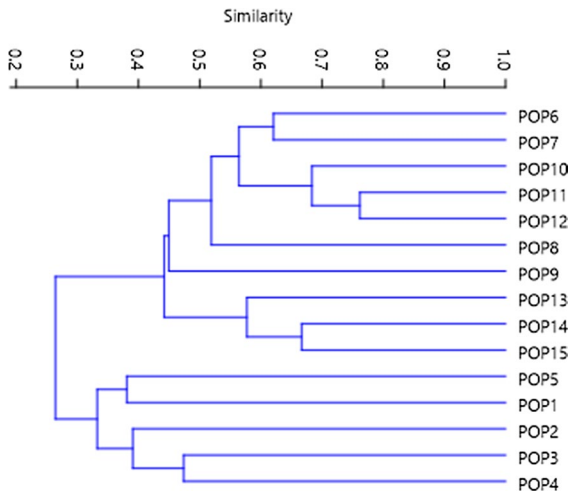


Fig. 5 Un-weighted pair group with arithmetic average distance tree illustrating the relationships among the *Citrullus colocynthis* 15 populations based on the analysis of chemical constituents revealed by GC–MS analysis

Based on the results of the SCoT and GC–MS analyses, the populations were grouped together using principal component analysis (PCA; Fig. 6). The first two axes were used to organize the 15 different populations. Populations 1, 3, and 5 were clustered together to form a group, while populations 8 through 15 were clustered together in the other. Population 2 and population 9 were each separated from all other populations. Thus, population 9 was distinguished from all other populations in the cluster dendrogram based on variation of the chemical compounds and in PCA analysis of the SCoT and GC–MS data.

Discussion

In this study, phytochemical composition and molecular data are utilized to build a precise and more detailed description of the genetic resources of *C. colocynthis* natural population in Egypt. To relate variation in the phytochemical composition with the genetic differentiation of the examined populations, the polymorphism in SCoT markers determined the genetic distance between populations. The resolving power of the SCoT markers was effective in differentiating the populations of *C. colocynthis* by a Euclidean distance dendrogram. The analysis of the SCoT data depicted the genetic diversity in the sampled populations, differentiating the three populations, 13 to 15, and populations 1 and 2 as two small distant clusters. These results generally indicate the differentiation of northern populations from the southern populations of *C. colocynthis* in the Eastern Desert of Egypt. This finding is supported by a highly positive correlation between populations 3 and 4 and populations 13, 14, and 15. The genetic diversity was assessed in 38 accessions of *C. colocynthis* plants from the semi-arid region of Rajasthan in India using RAPD and ISSR markers showed high level of genetic differences in relation to their geographical distributions (Verma et al. 2017). Also, in line with a study conducted by Al-Nablsi et al. (2022), there exists a significant variation in the genetic diversity, as determined by RAPD markers, across various accessions of a *C. colocynthis* population inhabiting the hot arid deserts of the United Arab Emirates.

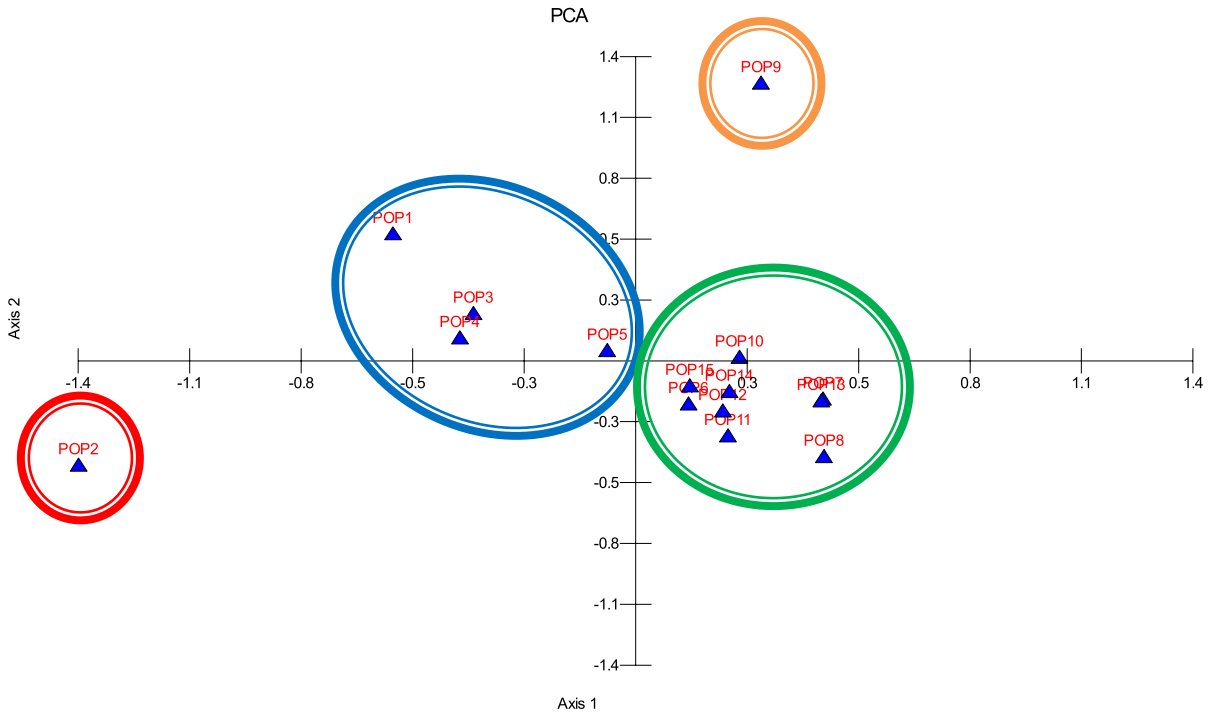


Fig. 6 Principal Component Analysis (PCA) of 15 populations of *Citrullus colocynthis* based on Start Codon Targeted & GC-mass analysis by using MVSP

The phytochemical composition of the sampled populations varied widely between the populations. A total of 81 different chemical compounds were isolated from the 15 *C. colocynthis* populations. Many of these compounds appeared in previous research on *C. colocynthis* (Gurudeeban et al. 2010; Hussain et al. 2014; Ahmed et al. 2019). About 150 compounds were identified as potentially intriguing bioactive chemicals in a study of the phytochemical composition of *C. colocynthis* seed extract conducted by Bourhia et al. (2020). These findings have strong supporting evidence from GC–MS analysis of the plant extract, which identified polyphenols and flavonoids as the bioactive chemicals responsible for the plant's antioxidant action (Skrovankova et al. 2015). Many of the compounds reported by these two authors were also recorded in the current study.

The analysis of the GC–MS compounds variation by the UPGMA distance method and the PCA scatter diagram generally indicated differentiation of the populations sampled in the northern parts of the Eastern desert of Egypt (1, 2, 3, 4, and 5) from populations sampled from the southern parts of the (13, 15 and 15). In Iran, 17 accessions of *C. colocynthis* from different parts of Iran, variation in morphological and phytochemical traits differentiated one accession from Kerman, which had highest amount of total phenolic compound. Based on fruit traits and phytochemical compound variation, colocynth plants in Iran had high genetic diversity that may be useful for breeding cultivated forms of this plant (Fereshtian et al. 2017). Ebadi et al. (2022) reported significant morphological and molecular diversity in watermelon accessions, which they attributed to an assumption that most useful genes, such as resistance to pests, diseases, and environmental stresses and genes for product quality, are typically distributed evenly across the genome. Since genetic variability was found to be greater within populations, they reasoned that collecting large samples from these populations would be the most effective conservation strategy.

In Egypt's Eastern Desert, populations of *C. colocynthis* are geographically distinct along a north–south axis of the Qena- Kosseir Road, as evidenced by morphological variation and polymorphism in the ISSR markers (Badr et al. 2018). Ahmed

and Alamer (2018) also reported that *C. colocynthis* populations are genetically differentiated based on allozyme markers diversity. However, Ahmed and Alamer (2018) didn't find a correlation between genetic variation and physical distance and concluded that any population that has a high genetic diversity can be studied for genetic analysis using isozymes to reveal its genetic differentiation. The results of the current study indicated that populations in the southern part of the Eastern Desert representing three sites shared higher phytochemical and genetic diversity compared to other populations. On the other hand, populations in the middle part show a small resemblance to each other and populations in the North and South of the Eastern Desert. These results demonstrate an impact of the ecology in the area on the genetic differentiation and the chemical composition of the secondary metabolites in *C. colocynthis*.

Acknowledgements We'd like to express our special appreciation to Prof. Dr. Mohamed Ismail Abdelkareem, a professor of Applied Geology and Remote Sensing at South Valley University's Faculty of Science, for advice and help in constructing the map of the sites of the examined colocynth populations.

Author contributions AB and HZ both made substantial contributions to the study's design, implementation, analysis, and writing.

Funding Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).

Declarations

Conflict of interest The authors declare no conflict of interest.

References

Abdein MAE-H (2018) Genetic diversity between pumpkin accessions growing in the Northern Border Region in Saudi Arabia based on biochemical and molecular parameters. Egypt J Bot 58(3):463–476. https://doi.org/10.21608/ejbo.2018.3612.1171

Adams RP (2017) Identification of essential oil components by gas chromatography/mass spectrometry. 5, online. Texensis Publishing, Gruver, TX USA

Ahmed M, Ji M, Qin P, Gu Z, Liu Y, Sikandar A, Iqbal MF, Javeed A (2019) Phytochemical screening, total phenolic and flavonoids contents and antioxidant activities of *Citrullus colocynthis* L. and *Cannabis Sativa* L. Appl Ecol Environ Res 17(3):6961–6979. https://doi.org/10.15666/aeer/1703_69616979

Ahmed SM, Alamer KH (2018) Genetic differentiation of *Citrullus colocynthis* (L.) populations depending on allozyme diversity. Egypt Jo Bot 58(3):581–590. https://doi.org/10.21608/ejbo.2018.3595.1170

Al-Nablsi S, El-Keblawy A, Mosa KA, Soliman S (2022) Variation among individuals of *Citrullus colocynthis* from a desert population in morphological, genetic, and germination attributes. Trop Ecolol 63:171–182. https://doi.org/10.1007/s42965-021-00190-1

Amiryousefi A, Hyvonen J, Poczai P (2018) iMEC: online marker efficiency calculator. Appl Plant Sci 6(6):e1159. https://doi.org/10.1002/aps3.1159

Avval SE (2017) Assessing polymorphism information content (PIC) using SSR molecular markers on local species of *Citrullus colocynthis*. Case Study: Iran, Sistan-Balouchestan Province

Badr A (2008) Molecular approaches to plant systematic and evolution. Taeckholmia 28:127–167

Badr A, El-Shazly HH (2012) Molecular approaches to origin, ancestry and domestication history of crop plants: Barley and clover as examples. J Genet Engi Biotechnol 10(1):1–12. https://doi.org/10.1016/j.jgeb.2011.08.002

Badr A, Kamel M, Zaki H (2018) Genetic diversity of Colocynth (*Citrullus colocynthis* Schrader) populations in the Eastern Desert of Egypt as revealed by morphological variation and ISSR polymorphism. Feddes Repert 129(3):173–184. https://doi.org/10.1002/fedr.201700011

Badr A, El-Shazly HH, Sakr M, Farid MM, Hamouda M, Elkhateeb E, Ahmad HS (2021) Genetic diversity and volatile oil components variation in *Achillea fragrantissima* wild accessions and their regenerated genotypes. J Genet Eng Biotechnol 19:166. https://doi.org/10.1186/s43141-021-00267-3

Begna T (2021) Effects of crop evolution under domestication and narrowing genetic bases of crop species. Open J Plant Sci. https://doi.org/10.17352/ojps.000032

Bhawna G, Abdin MZ, Arya L, Verma M (2017) Use of SCoT markers to assess the gene flow and population structure among two different populations of bottle gourd. Plant Gene 9:80–86. https://doi.org/10.1016/j.plgene.2016.09.001

Boulos L (2002) Flora of Egypt. Vol. 3. (Verbenaceae—Compositae), Al Hadara Publ., Cairo, Egypt. https://doi.org/10.1111/j.1756-1051.2002.tb01389.x

Bourhia M, Messaoudi M, Bakrim H, Mothana RA, Sddiqui NA, Almarfadi OM, El Mzibri M, Gmouh S, Laglaoui A, Benbacer L (2020) *Citrullus colocynthis* (L.) Schrad: chemical characterization, scavenging and cytotoxic activities. Open Chem 18(1):986–994. https://doi.org/10.1515/chem-2020-0124

Bourhia M, Bouothmany K, Bakrim H, Hadrach S, Salamatullah AM, Alzahrani A, Alyahya HK, Albadr NA, Gmouh S, Laglaoui A, El Mzibri M, Benbacer L (2021) Chemical profiling, antioxidant, antiproliferative, and antibacterial potentials of chemically characterized extract of *Citrullus colocynthis* L. seeds. Separations 8(8):114. https://doi.org/10.3390/separations8080114

Collard BCY, Mackill DJ (2009) Start codon targeted (SCoT) polymorphism: a simple, novel DNA marker technique for generating gene-targeted markers in plants. Plant Mol Biol Report 27(1):86–93. https://doi.org/10.1007/s11105-008-0060-5

Dane F, Liu JR, Zhang CK (2007) Phylogeography of the bitter apple, *Citrullus colocynthis*. Genet Resour Crop Evol 54(2):327–336. https://doi.org/10.1007/s10722-005-4897-2

Ebadi M, Soltani F, Mostofi Y, Alabboud M (2022) Analysis of genetic diversity among watermelon [*Citrullus lunatus* Thunb (Matsum.) and Nakai] accessions by phenotypic and molecular markers. J Agric Sci Technol 24(2):429–440

El-Shazly HH, Ahmed HIS, Hamouda MM, Badr A (2020) Genetic diversity and population structure of the medicinal plant *Achillea fragrantissima* (Forssk.) Sch. Bip. in the mountains of South Sinai, Egypt. Plant Gene 21:100212

Fereshtian M, Soltani F, Kashi A, Babalar M (2017) Evaluation of genetic diversity of some *Citrullus colocynthis* accessions based on morphological and phytochemical traits. Iran J Horticult Sci 48(2):2Pe293-Pe303

Ghanavati S, Masoumiasl A, Moradi F (2018) Genetic diversity investigation of bitter melon (*Citrullus colocynthis* L.) populations using morphological and molecular markers. In 2018

Giwa S, Abdullah LC, Adam NM (2010) Investigating "Egusi" (*Citrullus Colocynthis* L). seed oil as potential biodiesel feedstock. Energies 3(4):607–618. https://doi.org/10.3390/en3040607

Guo J, Xu W, Hu Y, Huang J, Zhao Y, Zhang L, Huang Ch, Ma H (2020) Phylotranscriptomics in Cucurbitaceae reveal multiple whole-genome duplications and key morphological and molecular innovations. Mol Plant 13(8):1117–1133. https://doi.org/10.1016/j.molp.2020.05.011

Gurudeeban S, Satyavani KS, Ramanathan T (2010) Bitter apple (*Citrullus colocynthis*): an overview of chemical composition and biomedical potentials. Asian J Plant Sci 9:394–401

Hameed BS, Ali Q, Hafeez M, Malik A (2020) Antibacterial and antifungal activity of fruit, seed and root extracts of *Citrullus colocynthis* plant. Biol Clin Sci Res J 2020(1):33. https://doi.org/10.54112/bcsrj.v2020i1.33

Hammer Ø, Harper DAT, Ryan PD (2001) PAST Paleontological. Statistics Software Package for Education and Data Analysis. Palaeontologia Electronica 4:9

Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. Phil Transact: Biol Sci 351(1345):1291–1298. https://doi.org/10.1098/rstb.1996.0112

Heuzé V, Tran G, Lebas F (2021) Colocynth (*Citrullus colocynthis)*. https://www.feedipedia.org/node/594

Heydari M, Homayouni K, Hashempur MH, Shams M (2016) Topical *Citrullus colocynthis* (bitter apple) extract oil in painful diabetic neuropathy: Aa double-blind randomized placebo-controlled clinical trial. J Diabetes 8(2):246–252. https://doi.org/10.1111/1753-0407.12287

Hussain AI, Rathore HA, Sattar MZA, Chatha SAS, Sarker SD, Gilani AH (2014) *Citrullus colocynthis* (L.) Schrad (bitter apple fruit): a review of its phytochemistry, pharmacology, traditional uses and nutritional potential. J Ethnopharmacol 155(1):54–66. https://doi.org/10.1016/j.jep.2014.06.011

Jiang LF, Qi XT, Zhang XQ, Huang LK, Ma X, Xie W (2014) Analysis of diversity and relationships among orchardgrass (*Dactylis glomerata* L.) accessions using start codon-targeted markers. Genet Mol Res: GMR 13(2):4406–18

Kamalakar K, Manoj GNVTS, Prasad RBN, Karuna MSL (2015) Thumba (*Citrullus colocynthis* L.) seed oil: a potential bio-lubricant base-stock. Grasas Aceites 66:055

Kardos M, Armstrong EE, Fitzpatrick SW, Hauser S, Hedrick PW, Miller JM, Tallmon DA, Funk WC (2021) The crucial role of genome-wide genetic variation in conservation. Proc Natl Acad Sci 118(48):e2104642118. https://doi.org/10.1073/pnas.2104642118

Kumar J, Agrawal V (2019) Assessment of genetic diversity, population structure and sex identification in dioecious crop, *Trichosanthes dioica* employing ISSR, SCoT and SRAP markers. Heliyon. https://doi.org/10.1016/j.heliyon.2019.e01346

Kumar S, Kumar D, Choudhary M, Saroha K, Singh N, Vashishta B (2008) Antioxidant and free radical scavenging potential of *Citrullus colocynthis* (L.) Schrad. methanolic fruit extract. Acta Pharm 58(2):215–221. https://doi.org/10.2478/v10007-008-0008-1

Legendre P, Legendre L (2012) Numerical ecology, developments in environmental modelling, 3–Elsevier, Amsterdam, The Netherlands, pp. 419

Li QY, Munawar M, Saeed M, Shen JQ, Khan MS, Noreen S, Alagawany M, Naveed M, Madni A, Li CX (2022) *Citrullus colocynthis* (L.) Schrad (Bitter Apple Fruit): promising traditional uses, pharmacological effects, aspects, and potential applications. Front Pharmacol. https://doi.org/10.3389/fphar.2021.791049

Michael VN, Moon P, Fu Y, Meru G (2019) Genetic diversity among accessions of *Cucurbita pepo* resistant to phytophthora crown rot. HortSci Horts 54(1):17–22. https://doi.org/10.21273/hortsci13506-18

Pandey A, Khan MK, Isik R, Turkmen O, Acar R, Seymen M, Hakki EE (2019) Genetic diversity and population structure of watermelon (*Citrullus sp.*) genotypes. 3 Biotech 9(6):210

Rai MK (2023) Start codon targeted (SCoT) polymorphism marker in plant genome analysis: current status and prospects. Planta 257(2):34. https://doi.org/10.1007/s00425-023-04067-6

Riaz H, Chatha SAS, Hussain AI, Bukhari SA, Hussain SM, Zafar K (2015) Physico-chemical characterization of bitter apple (*Citrullus colosynthis*) seed oil and seed residue. Int J Biosci 6:283–292

Rizvi TS, Khan AL, Ali L, Al-Mawali N, Mabood F, Hussain J, Adnan M, Al-Harrasi A (2018) In vitro oxidative stress regulatory potential of *Citrullus colocynthis* and *Tephrosia apollinea*. Acta Pharm 68(2):235–242. https://doi.org/10.2478/acph-2018-0012

Schierenbeck KA (2016) Population-level genetic variation and climate change in a biodiversity hotspot. Ann Bot 119(2):215–228. https://doi.org/10.1093/aob/mcw214

Sheded MG, Pulford ID, Hamed AI (2006) Presence of major and trace elements in seven medicinal plants growing in the South-Eastern Desert, Egypt. J Arid Environ 66(2):210–217. https://doi.org/10.1016/j.jaridenv.2005.10.022

Skrovankova S, Sumczynski D, Mlcek J, Jurikova T, Sochor J (2015) Bioactive compounds and antioxidant activity in different types of berries. Int J Mol Sci 16(10):24673–24706. https://doi.org/10.3390/ijms161024673

Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. Univ Kansas Sci Bul 38:1409–1438

Talebi SM, Askary M, Khalili N, Matsyura A, Ghorbanpour M, Kariman K (2021) Genetic structure and essential oil composition in wild populations of *Salvia multicaulis* Vahl. Biochem Syst Ecol. https://doi.org/10.1016/j.bse.2021.104269

Torkey HM, Abou-yousef HM, Azeiz AZA, Hoda E (2009) Insecticidal effect of cucurbitacin E Glycoside isolated from *Citrullus colocynthis* Against *Aphis craccivora*. Aust J Basic Appl Sci 3:4060–4066

Tyagi R, Sharma V, Sureja AK, Das Munshi A, Arya L, Saha D, Verma M (2020) Genetic diversity and population structure detection in sponge gourd (*Luffa cylindrica*) using ISSR, SCoT and morphological markers. Physiol Mol Biol Plants 26(1):119–131. https://doi.org/10.1007/s12298-019-00723-y

Uma C, Sekar KG (2014) Phytochemical analysis of a folklore medicinal plant *Citrullus colocynthis* L (bitter apple). J Pharmacogn Phytochem 2:195–202

Verma KS, Ul Haq S, Kachhwaha S, Kothari SL (2017) RAPD and ISSR marker assessment of genetic diversity in *Citrullus colocynthis* (L.) Schrad: a unique source of germplasm highly adapted to drought and high-temperature stress. 3 Biotech 7(5):288. https://doi.org/10.1007/s13205-017-0918-z

Xanthopoulou A, Ganopoulos I, Kalivas A, Nianiou-Obeidat I, Ralli P, Moysiadis T, Tsaftaris A, Madesis P (2015) Comparative analysis of genetic diversity in Greek Genebank collection of summer squash (*Cucurbita pepo*) landraces using start codon targeted (SCoT) polymorphism and ISSR markers. Aust J Crop Sci 9:14–21

Xie D, Xu Y, Wang J, Liu W, Zhou Q, Luo S et al (2019) The wax gourd genomes offer insights into the genetic diversity and ancestral cucurbit karyotype. Nat Commun 10(1):5158. https://doi.org/10.1038/s41467-019-13185-3