Plant Omics Journal POJ

POJ 10(02):64-77 (2017) ISSN:1836-3644
doi: 10.21475/poj.10.02.17.pne385

Selective Amplification of Start codon Polymorphic Loci (SASPL): a new PCR-based molecular marker in olive

Alsamman M. Alsamman[1*], S. S. Adawy[1], S. D. Ibrahim[1], B. A. Hussein[2] and E. H. A. Hussein[2]

[1]Agricultural Genetic Engineering Research Institute (AGERI), ARC, Giza, Egypt
[2]Department of Genetics, Faculty of Agriculture, Cairo University, Giza, Egypt

***Corresponding author: smahmoud@ageri.sci.eg**

Abstract

The Selective Amplification of Start Codon Polymorphic Loci (SASPL) has been developed as a new PCR-based molecular marker. SASPL was validated for the analysis of varietal diversity on ten olive varieties. Validation included *in vitro* comparison against RAPD, SCoT and SAMPL markers. Assessment of these techniques included primer selectivity, genome coverage and the ability to target genic regions through *in silico* PCR analysis. Candidate PCR fragments were further sequenced to annotate non-identified genes in olive. Eight SASPL primers were compared to 24 RAPD, 39 SCoT and 12 SAMPL primers. The TA produced by the RAPD, SCoT, SAMPL and the eight SASPL primers were 359, 642, 571 and 269 amplicons, respectively. The highest average number of TA was revealed by SAMPL (47.6), average of PA (18.1) and genetic similarity (GS) (96%) among the olive varieties. On the other hand, SASPL analysis provided higher average number of TA (33.6), average of PA (16.2) and GS (93%) than SCoT and RAPD. The highest average of (PIC) (0.2909) was exhibited by SASPL analysis and the lowest average (0.2038) was revealed by SCoT. The highest number of UB (111) was revealed by SCoT and the lowest UB (43) was obtained by SASPL. Across the four marker types, variety Maraki was characterized by the highest number of unique markers (74). Meanwhile, the variety Manzanillo was characterized by the lowest number of unique markers (8). In addition, in the *in silico* analysis SASPL exhibited the highest chromosomal coverage (0.59%) and targeted genes (1090) using the lowest number of primers. Additionally, the average area covered by the SASPL primers (354kb) was larger than SCoT and SAMPL. RAPD analysis provided the lowest potential, chromosomal coverage (0.04%) and number of targeted genes (17) compared to SASPL, SCoT and SAMPL analysis. The total coverage of the genome, revealed by combined data was higher (1.21%) than that of each technique separately. Meanwhile, the difference between the actual and the total genomic regions covered by the combined data was about 652kb. Our results suggested that the newly developed SASPL marker is the most adequately and each of the studied marker target different genomic areas, while some areas are shared. Two SCoT and one RAPD fragment were sequenced and showed a high similarity to genes of high physiological functions; such as cyclic plant-specific DNA-binding transcription factor, SANT domain and Copia-type retrotransposon.

Keywords: Olive, RAPD, SCoT, SASPL, SAMPL, *in vitro*, *in silico.*
Abbreviations: RAPD_Random Amplified Polymorphic; DNA-SCoT_Start Codon-targeted; SAMPL_Selective Amplification of Microsatellite Polymorphic Loci; TA_total number of amplicons; PA_polymorphic amplicons; SCoT_start-codon targeted; %P_ percentage of polymorphism of calculated amplicons; PIC_polymorphism information content; GS_genetic similarity; UB_unique bands; UPGMA_unweighted pair group method with arithmetic mean; TSA_Transcriptome Shotgun Assembly; NGST_next generation sequencing technologies.

Introduction

The growing field of molecular markers research over the last decades provided new opportunities for genetic characterization and biodiversity studies in plants. Application of molecular marker techniques helped improving the genetic gain of agronomic traits and enhancing the development of marker-assisted selection (MAS). Different molecular marker platforms have different properties including reliability, reproducibility, coverage, cost and automation (Agrawal et al., 2008).Random amplified polymorphic DNA (RAPD) marker was widely used in olive (Sanz-Cortés et al., 2001; Gemas et al., 2004 and Sesli and Yegenoglu, 2015), date palm (Adawy et al., 2002), potato (Gorji et al., 2011), sugar beet (Grimmer et al., 2007) and walnut (Nicese et al., 1998). Start codon-targeted (SCoT) marker is a new relatively dominant, simple, low-cost, highly polymorphic PCR-based technique, where primers were designed based on plant universal gene composition (Gorji et al., 2011). Advantages of SCoT were validated in grape (Zhang et al., 2011), mango (Luo et al., 2010), tomato (Shahlaei et al., 2014), potato (Gorji et al., 2011) and date palm (Adawy et al., 2014).Selective amplification of microsatellite polymorphic loci (SAMPL) was developed to provide a marker system with high multiplex ratio combining the advantages of microsatellites and AFLP markers (Morgante and Vogel, 1999). Therefore, it is more applicable when low level of diversity is expected (Sarwat et al., 2008). SAMPL helped unravel genetic diversity in wheat (Roy et al., 2004), grapevine (Cretazzo et

al., 2010), citrus (Biswas et al., 2011), walnut (Kafkas et al., 2005), seabuckthorn (Raina et al., 2012). Given that, it was used to generate SCAR markers in olives (Busconi et al., 2006). *In silico* PCR is a computational tool estimating theoretical results of PCR amplicons amplified from sequenced genome or transcriptome (Lexa et al., 2001). This technique can improve primer selectivity needed to target genes (Mokhtar et al., 2016). Additionally, *in silico* PCR analysis was used to improve the effectiveness of RAPD analysis by selecting and redesigning primers for target sequence(s) of genes in oil palm (Premkrishnan and Vadivel, 2012). To overcome the noise in AFLP, *in silico* PCR was used to develop similarity coefficients and weight the AFLP bands according to their band length distribution probabilities (Koopman and Gort, 2004). This helped to improve the analysis of AFLP and provide more informative and reliable results.In addition, *in silico* PCR software can predict PCR products and perform other analyses such as BLAST to identify genes with/near the PCR primers (Boutros and Okey, 2004).

Olive, *Olea europaea L*., is one of the oldest cultivated tree and one of the most economical crops in the Mediterranean basin (Adawy et al., 2015).Olive has a juvenile period ranging from 10 to 15 years (Bracci et al., 2011) and a huge number of variety synonyms (Calzada et al., 2015). Therefore, there is an urgent need for an efficient identification technique to eliminate any mislabeling if present in olive varieties and provide genetic markers to help olive breeding programs in Egypt.

In current study, selective amplification of start Codon polymorphic loci (SASPL) marker was developed. This was performed by replacing the *Eco*RI primer with a SCoT primer in the selective amplification step of the AFLP procedure. The newly developed SASPL marker was validated *in vitro* by testing the genetic diversity and varietal discrimination among ten olive varieties. Then, results were compared with that of RAPD, SCoT and SAMPL markers. Assessment also included primer selectivity, genome coverage and ability to target genic regions using *in silico* PCR analysis.Finally, candidate PCR fragments were sequenced to confirm identification of anonymous genes of olive.

Results and Discussion

Polymorphism among the olive varieties as detected by the different molecular markers using in vitro PCR analysis

The comparison between the newly developed SASPL marker and the three other types of markers revealed that, all of the four marker types (RAPD, SCoT, SAMPL , and SASPL) provided reproducible amplicons and uniquely characterized each of the used olive varieties (Figs. 2 and 3). The total number of amplicons (TA) produced by the 24 RAPD, 39 SCoT, 12 SAMPL and the eight SASPL primers were 359, 642, 571 and 269 amplicons, respectively. The highest average of amplicons/primer (47.6) was revealed by SAMPL, where the range of TA was from 30 (SAMPL-01) to 79 (SAMPL-09) (Tables 1 and S3). RAPD analysis exhibited the lowest average of amplicons/primer (14.95), where the range of TA was from five (G12) to 24 (A16) (Tables1 and S1). In this respect, six SAMPL combinations were applied to characterize 28 pistachio accessions, and resulted in a total of 182 amplicons with an average of 30.33 per primer pair (Karimi and Kafkas, 2011). Similarly, using 21 RAPD primers on 32 olive cultivars, Belaj et al. (2003) recorded a TA of 201 amplicons with an average of 9.57 per primer. In mango, 33 SCoT primers were tested and generated a total of 273 amplicons with an average of 8.27 per primer (Luo et al., 2010).In the present study, SCoT and SASPL analysis revealed an average of 16.46 and 33.6 amplicons/primer (Table 1), where the amplicons number ranged from 7 (SCoT-02) to 30 (SCoT-31) and from 22 (SASPL-05) to 42 (SASPL-01), respectively (Tables S2 and S4).

The total number of polymorphic amplicons (PA) was 170 (RAPD), 382 (SCoT), 217 (SAMPL) and 130 (SASPL) (Table 1). The highest average of PA was recorded for SAMPL (18.1) and the lowest average of PA (7.1) was for RAPD (Table 1). The range of PA was from nine (SAMPL-08) to 29 (SAMPL-07) for SAMPL and from zero (G13) to 19 (A16) for RAPD (Tables S1 and S3).The average numbers of PA for SCoT and SASPL were 9.7 and 16.2, respectively (Table 1). Briefly, the range of PA was from one (SCoT-16) to 22 (SCoT-31) for SCoT and from three (SASPL-07) to 27 (SASPL-01) for SASPL (Tables S2 and S4).In addition, the highest average percentage of polymorphism (%P) was 59.5% for SCoT and the lowest average was 38%for SAMPL (Table 1).Thus, the range of %P was from 12.5% (SCoT-16) to 86.70% (SCoT-30) for SCoT, and from 17.00% (SAMPL-08) to 62.2% (SAMPL-12) for SAMPL (Tables S2 and S3). On the other hand, the %P revealed by SASPL and RAPD was 48.3% and 47.4%, respectively (Table 1). Meanwhile, the range was from 10.3% (SASPL-07) to 64.3% (SASPL-01) for SASPL and from zero (G13) to 84.6% (B13) for RAPD (Tables S4 and S1). In this regard, the average of PA in RAPD analysis of some Russian olive varieties was 15.7, and the range of percentage of polymorphism was from 50% to 95.45% (Asadiar et al., 2012). Another study using RAPD analysis showed a range of PA from eight to 17 for 122 cultivars and wild accessions of olive (Parra-Lobato *et al.*2012).While, in SCoT analysis of 10 tomato cultivars, the PA reached 36.14% (Shahlaei et al., 2014). Meanwhile, the analysis of 20 peanut accessions using 18 SCoT primers revealed a PA of 38.22% (Xiong et al.,2011). On the other hand, Azizi et al.(2009) visualized 558 scorable bands by using 7 SAMPL primer combinations, where the percentage of polymorphism across the oregano accessions ranged from 85% to 97% with an average of 92%. While, Raina et al.(2012) and Biswas et al. (2011) reported that the average number of bands per primer combination was 33 and 22.43, respectively.

Moreover, in the present study the polymorphism information content (PIC) which represents a measure of allelic diversity at a locus was estimated for the different types of markers. The highest average of PIC (0.2909) was exhibited by SASPL analysis and the lowest average (0.2038) was revealed by SCoT (Table 1), where the PIC ranged from 0.2371 (SASPL-05) to 0.3302 (SASPL-06) and from 0.1036 (SCoT-02) to 0.2846 (SCoT-31) (Tables S4 and S2).

The PIC average exhibited by RAPD and SAMPL analysis was 0.2423 and 0.2778, respectively (Table 1), where the PIC value ranged from 0.097 (A16) to 0.3162 (B17) and from 0.2072 (SAMPL-01) to 0.3628 (SAMPL-09), respectively (Tables S1 and S4). In this regards, Botstein et al. (1980) reported that the PIC index can be used to evaluate the level of gene variation. Therefore, our results could suggest that the newly developed SASPL marker is the most informative marker type, among the studied markers. Interestingly, it is worth to note that although in SASPL a SCoT primer has been used in attempt to target genic sequences, however the average PIC value of SASPL was much higher than SCoT. This could suggest that the combination of the advantages of

Table 1. The total (Tot.), average (Aver.), highest (High) and lowest (Low.) for: (A) the *in vitro* PCR metrics ,i.e., number of primers (NP) , total number of amplicons (TA), band molecular weight (BW) , polymorphic amplicons (PA) , percentage of polymorphism (%P) , polymorphism information content (PIC),genetic similarity(GS), unique positive (Pos.) and negative (Neg.) bands. (B) the *in silico* PCR metrics, i.e., number of primers (NP), total number of the *in silico* amplimers(TIA), total covered genomic area (TCGA), the percentage of total covered area (GCP%), the average of coverage per primer (ACP), the average of IA per primer (IA/P), genic *in silico* amplimers(GIA) and the number of actual detected genes (TG) as revealed by RAPD, SCoT, SAMPL, SASPL and combined data.

	A														B							
	In vitro PCR														*In silico* PCR							
	NP	TA		BW		PA		%P	PIC	GS		Unique Bands			NP	TIA	TCGA	GCP%	ACP	IA/P	GIA	TG
	Tot.	Tot.	Aver.	Low	High	Tot.	Aver.	Aver.	Aver.	Low.	High	Neg.	Pos.	Tot.	Tot.	Tot.	Tot.	Tot.	Tot.	Tot.	Tot.	Tot.
RAPD	24	359	14.95	100	1500	170	7.1	47.40%	0.2423	87%	93%	44	2	46	22	1024	572395	0.04%	26018	46.5	29	17
SCoT	39	642	16.46	150	3000	382	9.7	59.50%	0.2038	83%	90%	103	8	111	39	15366	4612460	0.35%	209657	394	2027	584
SAMPL	12	571	47.6	70	1000	217	18.1	38%	0.2778	86%	96%	74	1	75	9	9410	3587332	0.27%	163061	1045	1055	494
SASPL	8	269	33.6	70	1600	130	16.2	48.30%	0.2909	84%	93%	39	4	43	8	17632	7806865	0.59%	354858	2204	2112	1090
Combined	83	1841	22.2	-	-	899	10.8	-	-	85.20%	92.70%	260	15	275	78	43432	15926082	1.21%	723913	556.8	5223	2185

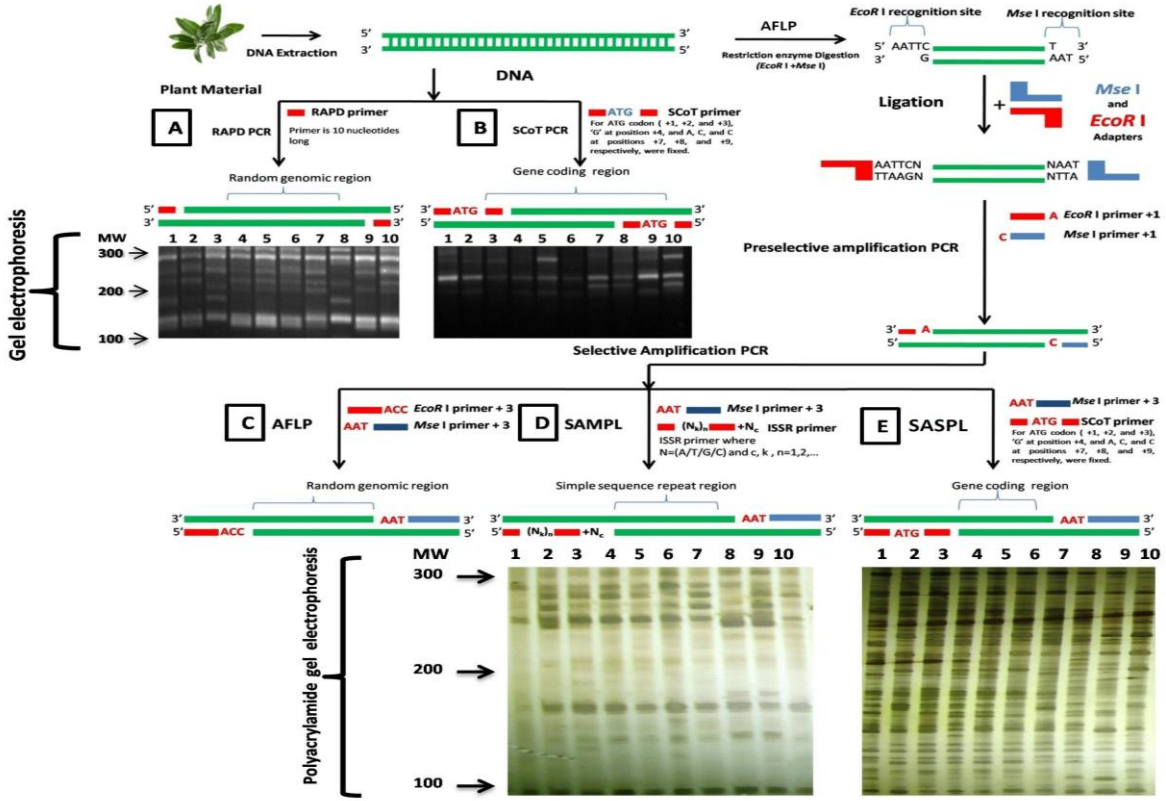

Fig 1. A schematic figure shows the different stages of SASPL marker system (E) compared to RAPD (A), SCoT (B) and AFLP (C) and SAMPL (D).

Table 2. Variety name (VN), total number of negative markers per cultivar (TNM), total number of positive markers per cultivar (TPM) and the grand total markers (GT) for the 10 olive varieties across the four marker types (RAPD, SCoT, SAMPL and SASPL).

VN	TNM	TPM	GT
Kronaki	32	2	34
Tofahi	15	0	15
Calamata	27	6	33
Manzanillo	5	3	8
Ojazi Shami	20	0	20
Koratina	26	0	26
Dolci	20	1	21
Maraki	72	2	74
Pekoal	10	1	11
Khodari	33	0	33
Total	260	15	275

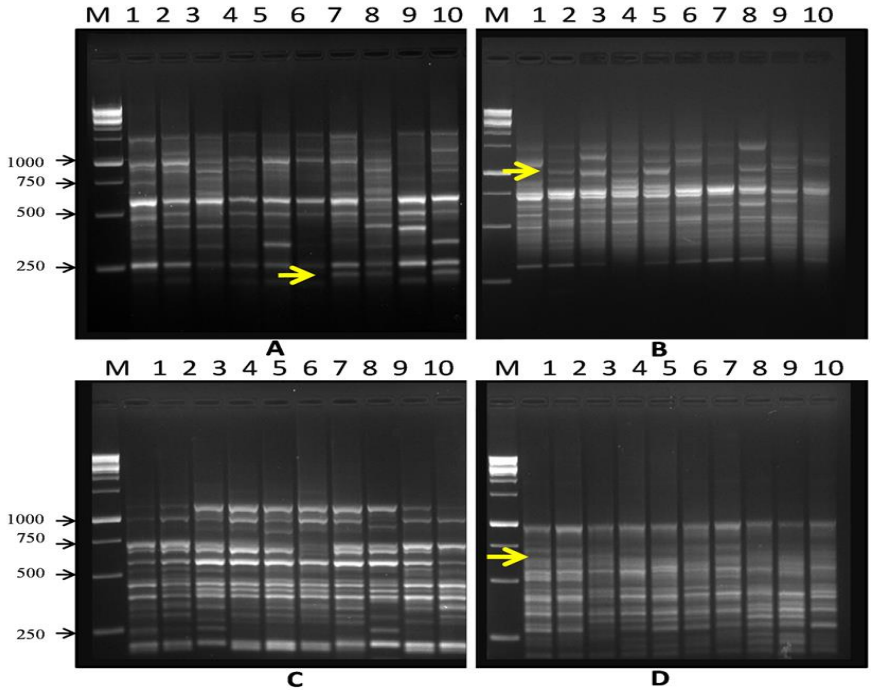

Fig 2. SCoT and RAPD profiles of the 10 olive varieties as revealed by primers SCoT-13(A), SCoT-33 (B), A-12 (C) and A-13 (D). Lanes 1 to 10 represent: Dolci, Pekoal, Manzanillo, Kronaki, Tofahi, Koratina, Maraki, Ojazi Shami, Calamata and Khodari. M: DNA molecular weight marker (1000 bp Ladder).Yellow arrows depict the PCR bands selected for sequencing.

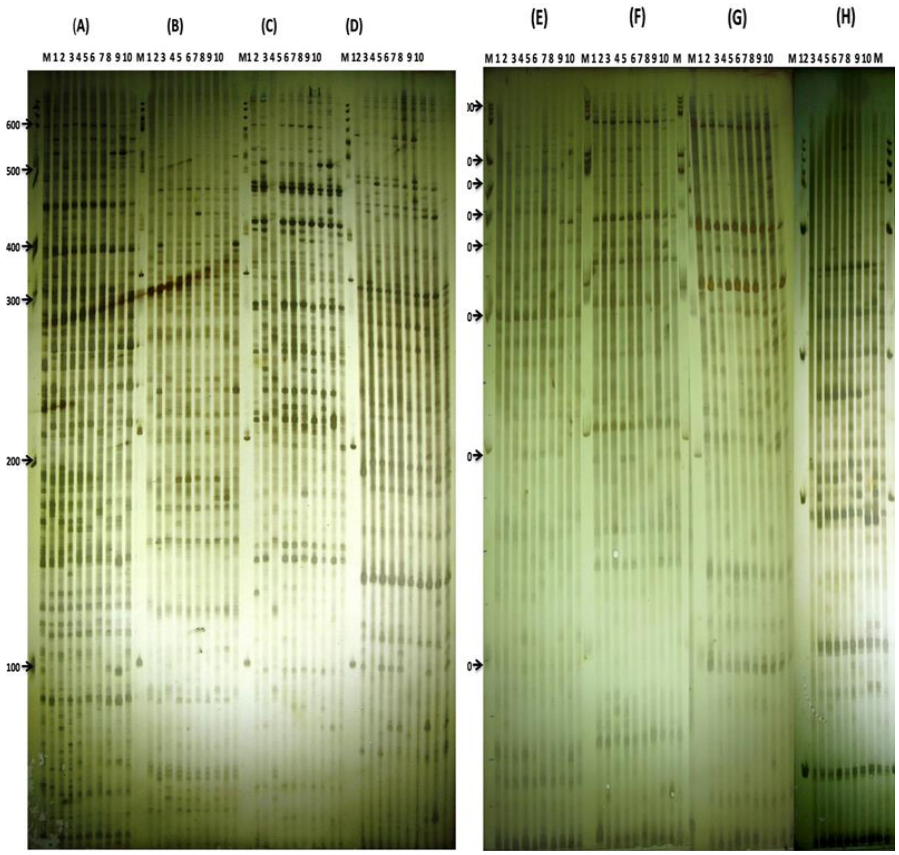

Fig 3. SAMPLs and SASPLs profiles of the 10 olive varieties as revealed by primer combinations SAMPL-09 (A), SAMPL-10 (B), SAMPL-11 (C), SAMPL-12 (D) , SASPL-03 (E), SASPL-02 (F), SASPL-08 (G) and SASPL-01 (H). Lanes 1 to 10 represent: Dolci, Pekoal, Manzanillo, Kronaki, Tofahi, Koratina, Maraki, Ojazi Shami, Calamata and Khodari. M: DNA molecular weight marker (100 bp Ladder).

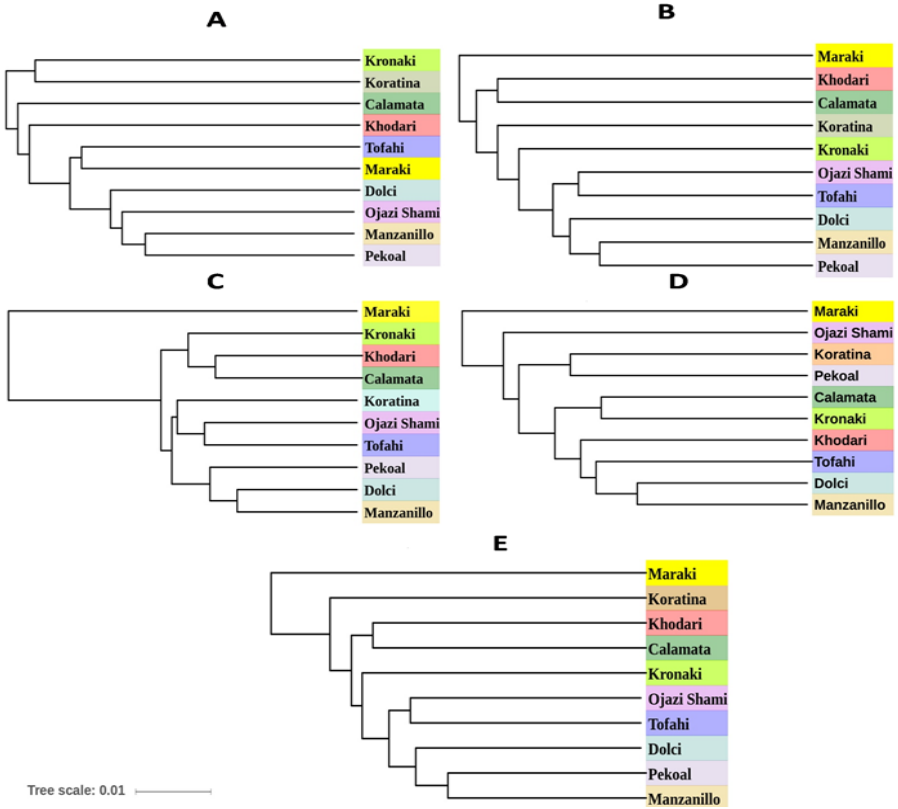

Fig 4. The 10 olive varieties phylogenetic tree constructed from the RAPD (A), SCoT (B), SAMPL (C), SASPL (D) and combined (E) data using UPGMA and similarity matrices computed according to Dice coefficient**.**

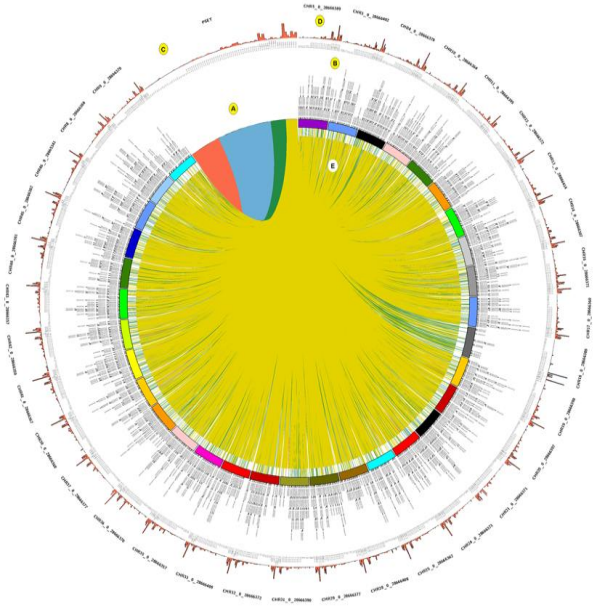

Fig 5. The *in silico* PCR results for the RAPD (red), SCoT (blue), SAMPL (green) , SASPL (yellow) (A) against olive genome, revealing the possible adjoined genes (B), primer total genome coverage percentages statics (C) , each primer total chromosome coverage percentages statics (D) and the position of possible PCR amplimers with genes (extended lines) or without (short lines) (E).

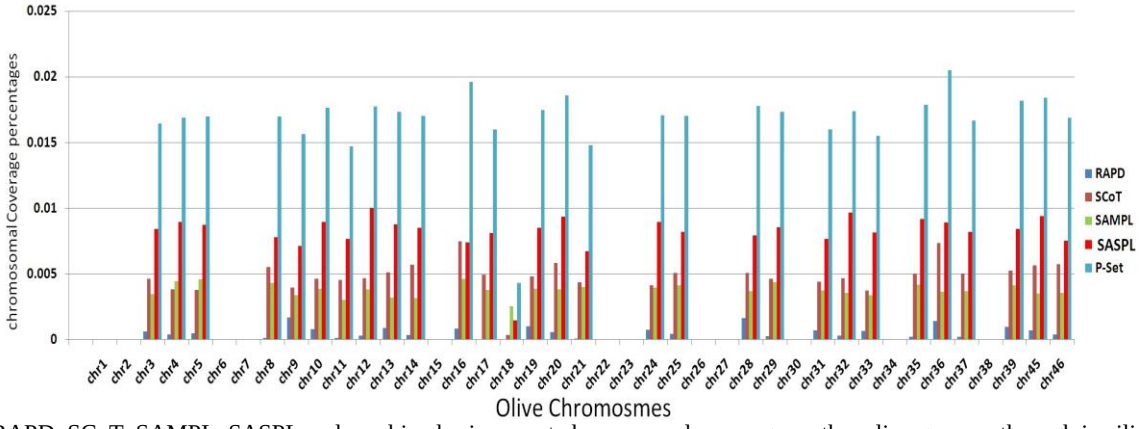

Fig 6. RAPD, SCoT, SAMPL, SASPL and combined primers set chromosomal coverage on the olive genome through in silico PCR analysis.

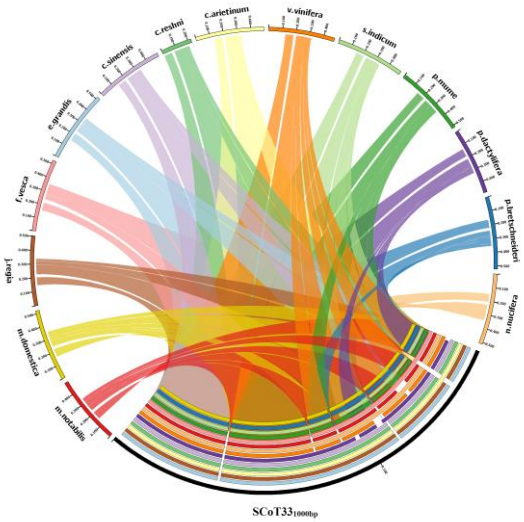

Fig 7. SCoT33$_{1000bp}$ BLAST results against the NCBI non-redundant database showing coverage comparison with different *dof* genes belonging to different plant species.

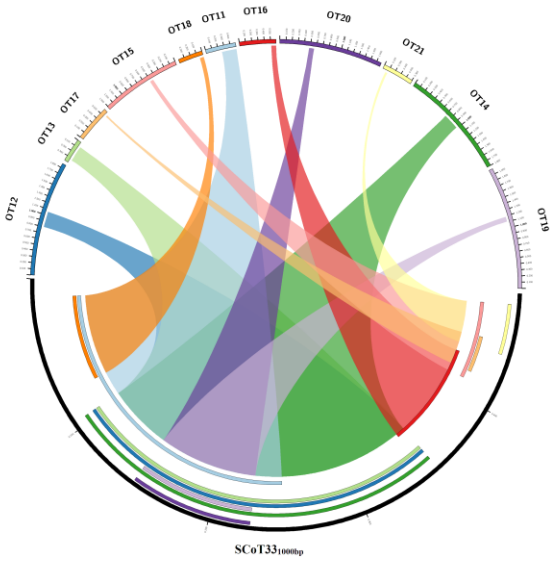

Fig 8. SCoT33$_{1000bp}$ BLAST results against the Transcriptome Shotgun Assembly (TSA) database showing coverage comparison with different published *Olea* EST genes.

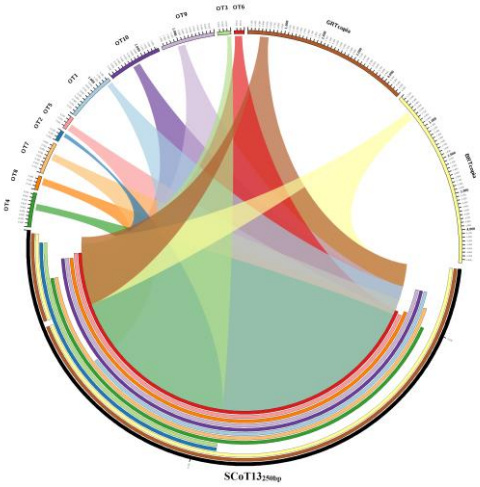

Fig 9. SCoT13$_{250bp}$ BLAST results against the NCBI non-redundant and TSA databases showing coverage comparison with different published *Olea* genes.

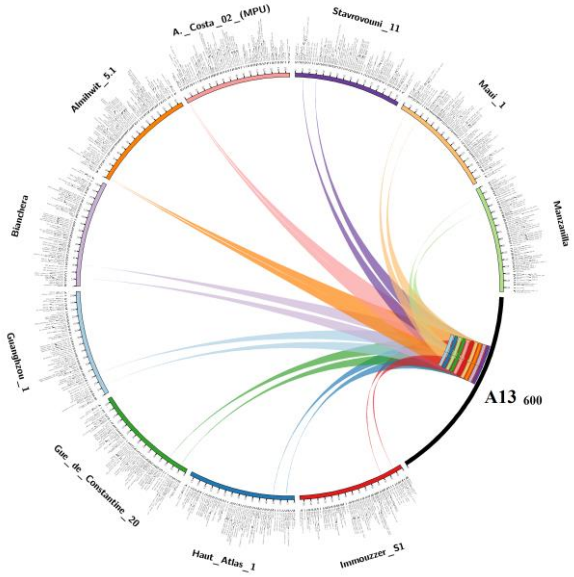

Fig 10. A13$_{600bp}$ BLAST results against the NCBI non-redundant database showing coverage comparison with different chloroplast genomes belonging to different olive varieties.

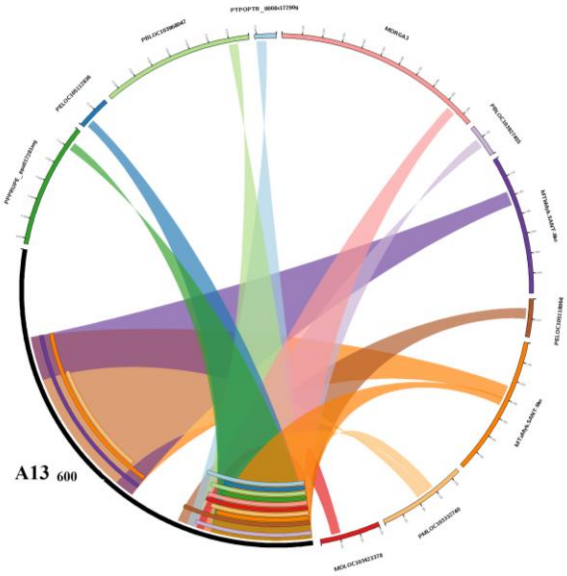

Fig 11. $A13_{600bp}$ BLASTX results against the NCBI non-redundant and TSA databases showing coverage comparison with different published *Olea* genes and different SANT genes belonging to different plant species.

AFLP and SCoT technology could have a higher potentiality in detecting the genetic diversity and hence could be more informative for linkage analysis.

Using RAPD markers, the average value of PIC recorded in castor genotypes was 0.88 (Gajera et al., 2010), 2.47 in Indian bitter gourd accessions (Behera et al., 2008), 0.243in rice landraces (Muthusamy et al., 2008) and 0.241 in *Vanilla* genus (Verma et al.,2009). In addition, the average number of PIC for SCoT was 0.142 using 10 primers on 10 tomato cultivars (Shahlaei et al., 2014) , 0.78 as detected 36 primers on 60 medicinal plants (Bhattacharyya et al., 2013) and 0.82 using 36 primers on 64 grapes varieties (Guo et al., 2015). Furthermore, by employing two SAMPL primer combinations on 55 bread wheat genotypes, the PIC average was 0.221 and the PIC values ranged from 0.035 to 0.499 with an average of 0.264 (Roy et al., 2002).

Identification of olive varieties by unique markers

In the present study the four marker types, successfully produced unique positive and/or negative markers and succeeded to characterize the ten olive varieties (Tables S5, S6, S7, and S8). The highest number of unique bands (UB) (111) was revealed by SCoT (8 positive and 103 negative) and the lowest UB (43) was obtained by SASPL (4 positive and 43 negative markers) (Table 1). Across the four marker types, variety Maraki was characterized by the highest number of unique markers (74). Meanwhile, the variety Manzanillo was characterized by the lowest number of unique markers (8). The varieties Tofahi, Ojazi Shami, Koratina and Khodari showed no positive UB. Interestingly, the variety Calamata was characterized by six positive unique markers (Table 2).

RAPD analysis produced a total of 46 unique markers (2 positive and 44 negative markers). One positive marker was exerted by each of Maraki and Dolci. The highest number of negative unique markers (9) was detected in Khodari (Table S5). In this respect, Besnard et al. (2001) successfully identified an olive cultivar by three different unique RAPD markers. Also four specific RAPD markers, were each corresponding uniquely to three different olive varieties (Parra-Lobato et al. 2012). Seven ISSRs unique bands were used as variety specific markers to distinguish seven olive varieties (El Saied et al.2012). SAMPL analysis produced a total of 75 unique markers (one positive and 74 negative). The single unique positive SAMPL marker was detected in Calamata variety in addition to seven negative markers. The highest number of negative unique markers (38) was observed in the variety Maraki (Table S7).

From Table (S7) it could be concluded that some primers are more informative than others in identifying the olive genotypes such as SAMPL-04, SAMPL -05 and SAMPL -06.

Genetic relationships among olive varieties

Across the four types of markers the highest genetic similarity (GS) (96%) was revealed by SAMPL, followed by 93% in SASPL and RAPD, then 90% in SCoT (Table 1). Meanwhile, the lowest GS for SCoT, SASPL, SAMPL and RAPD were 83%, 84%, 86% and 87%, respectively. Additionally, the GS revealed by the combined data analysis ranged from 85.2% to 92.7% (Table 1). In this regards, Luo et al. (2012) reported that the range of GS based on SCoT analysis in mango was from 66.2% to 94.2%. While in peanut the SCoT-based GS was from 87% to 100% (Xiong et al., 2011). Likewise, based on RAPD, the GS was from 54% to 79% in Portuguese olive (Martins-Lopes et al., 2007). However, it was from 36% to 62% in some Russian olive varieties (Asadiar et al., 2012). While, El Saied et al., (2012) used ISSR to estimate the genetic diversity among 22 olive varieties (Egyptian and foreign) and revealed that, the genetic similarity ranged from 81% to 99%.

To examine the genetic relationships among the 10 olive varieties based on RAPD, SCoT, SAMPL and SASPL results, the scored data were analyzed using the Dice coefficient to compute the similarity matrices. These similarity matrices were used to generate a dendrogram using the UPGMA method. The phylogenetic relationships generated by SCoT, SASPL, SAMPL and the combined data separated the variety Maraki from all other varieties. This suggests that the variety Maraki was genetically distinct from the other Egyptian olives. Manzanillo and Pekoal were clustered together in the dendrograms of RAPD, SCoT and the combined data. However, in SAMPL, the variety Dolci

was added to the group of Manzanillo and Pekoal. Khodari and Calamata were clustered together in the dendrograms of SCoT, SAMPL and the combined data. Meanwhile, in the dendrogram of RAPD each variety was separated in a singular cluster and SASPL substituted Khodari with Kronaki. Tofahi and Ojazi Shami were clustered together in the dendrograms of SCoT, SAMPL and the combined data (Fig. 4). These discrepancies in the GS revealed by the different marker types could be attributed to the different mechanism of detecting the polymorphism and genome coverage by the different markers .Therefore, the GS based on the combined data could be more representative of the genetic relationships. Such findings were previously reported by Qian et al. (2001), McGregor et al. (2000), Belaj et al. (2003) and Costa et al. (2016).

In silico PCR analysis of the olive genome as revealed by the different molecular marker techniques

To compare between different molecular marker techniques, a variety of metrics must be optimized such as expected heterozygosity, ratio of multiplexing, and effectiveness of kinship between accessions.This was applied in soybean (Powell et al., 1996),melon (Garcia-Mas et al.,2000), tetraploid potato (McGregor et al.,2000) and white yam (Mignouna et al., 2003). However, in the current study, the assessment of these techniques included primer selectivity, genome coverage and ability to target genic regions in the olive genome through *in silico* PCR analysis. Out of the 83 primers used in this study only five primers (A13, B13, SAMPL-05, SAMPL-06 and SAMPL-07) did not show any *in silico* amplimers matching our criteria. Meanwhile, the total number of *in silico* amplimers detected by the other 78 primers was 43,432 (Fig. 5 and Table 1). The highest number of *in silico*-amplimers (IA)(17,632) was revealed by SASPL with an average of 2,204 IA per primer (IA/P).While the lowest IA (1,024) was obtained by RAPD, with an average of 46.5 IA/P (Table 1).The total coverage of genomic area using the combined data was 15.9 Mbp (1.21%) of the olive genome. The largest genomic area (7.8 Mbp) was covered by SASPL (0.59%), while the smallest area (0.57 Mbp) was covered by RAPD (0.04%) (Table 1). SCoT and SAMPL covered 4.6Mbp (0.35%) and 3.5 Mbp (0.27%), respectively, of the olive genome.

In contrast to SASPL technique that revealed the highest chromosomal coverage, RAPD revealed the lowest across all olive chromosomes (Fig. 6).The distribution of IA revealed by SCoT, SASPL, SAMPL and RAPD was illustrated in Figure 5.The primer SASPL-02 covered the largest genomic area, while SAMPL-08 covered the lowest on chromosome 42 (Fig. 5).Only 5,223 amplimers, out of the total 43,432, were genic *in silico* amplimers (GIA) (close/adjacent to genes).The highest number of GIA (2,112 amplimers) was produced by SASPL .This was followed by SCoT (2027) ,then SAMPL (1055) and the lowest number (29 amplimers) was exerted by RAPD. This finding is quite reasonable since in SASPL a SCoT primer is introduced in the selective amplification thus directing the high resolution of AFLP towards the genic regions and consequently increasing the efficiency of the SCoT analysis. Since the same gene can be targeted by more than one marker, the actual number of targeted genes (TG) was 2,185. SASPL produced the highest number of TG (1,090 genes) and RAPD the lowest (17 genes) (Table 1). In SCoT, the number of primers (39) was higher than SAMPL (12) and SASPL (eight). However, the average area covered by primers (ACP) in SASPL was larger than SCoT and SAMPL. This could be due to the large genomic area shared by the different SCoT amplimers, which was apparent by comparing the GIA value (2,027) and the TG value (584).

Comparison between the in silico and in vitro PCR analysis results

Much higher number of amplimers was detected by the *in silico* PCR compared to the *in vitro* PCR. Noticeably, in the *in silico* PCR many could be of the same fragment size (bp). For example SASPL-02, SAMPL-01, SCoT-06, and B-12, produced 8, 7, 10, and 24 amplimers with similar fragment sizes (281, 639, 287, and 892 bp, respectively).Such fragments cannot be scored as different markers using the conventional gel electrophoresis techniques. This observation was previously reported for *in vitro* AFLP amplimers that produced different sequences with the same fragment sizes (Koopman and Gort, 2004).Additionally; amplimers might be slightly different in fragment size (1 or 2 bp) and consequently scored as one amplicon in the *in vitro* PCR. Herein, the newly developed SASPL marker provided higher efficiency in both *in vitro* and *in silico* PCR compared to the other marker types. It provided the following attributes: 1) higher averages of TA and PIC than RAPD and SCoT; 2) highest TCGA, GCP, TG and chromosomal coverage by using the lowest number of primers; and 3) overcomes the disadvantage of similarity in SCoT primer sequences.

RAPD analysis provided the lowest average of TA and PA, TCGA, GCP%, TG, and chromosomal coverage which suggests its lower potentiality to target genic regions (Table 1 and Fig. 6). SCoT and SAMPL both demonstrated higher number of unique bands, GCP%, TG, and chromosomal coverage than RAPD. The dendrogram based on the combined data of the four molecular markers, showed some similarity to other dendrograms constructed separately for RAPD, SCoT, SAMPL and SASPL. It also provided better understanding of the relationships among the olive varieties. Additionally, the value of total genome coverage (TCGA) revealed by the combined data was higher than that of each technique separately (Fig. 6).The difference between the actual covered genomic areas (15,926,082) and the total area covered by the combined data (16,579,052) was about 652 kb. This may suggest that these markers targeted different genomic areas however other areas were common. Therefore, the analysis of variability may require more than one DNA-based marker (Costa et al., 2016). Given that, our results support that AFLP-based (SASPL and SAMPL) markers had more relative efficiency than other marker systems , similar results were reported in SAMPL (Sarwat et al., 2008) and SSAP (Tam et al., 2005).

Sequence analysis of some polymorphic PCR bands

The Basic Local Alignment Search Tool (BLAST) at the NCBI website was used to find regions with local similarities then compare our queries of nucleotide sequences against public databases. Two SCoT ($SCoT13_{250}$ and $SCoT33_{1000}$) and one RAPD ($A13_{600}$) polymorphic PCR products were successfully sequenced. These two bands were separated using gel electrophoresis, cloned, and sequenced then trimmed at low quality reads. Interestingly, products of both SCoT-33 and SCoT-13 primers were adjacent to genes (11 and 28 genes, respectively) of important biological processes (Fig. 5).Using BLASTN search against the NCBI non-redundant database, $SCoT33_{1000}$ showed similarity with proteins belonging to cyclic plant-specific DNA-binding transcription *(dof)* factor (Fig. 7).

The biological importance of Dof proteins was studied in different plant growth stages such as regulation of flowering time, responses of photoperiod sensitivity and vascular development in the roots of *Arabidopsis thaliana* (Fornara et al., 2009), *Jatropha curcas* (Yang et al., 2011) and *Ipomoea batatas* (Tanaka et al., 2009).The highest percentage of identity (47%) was revealed for *dof* 2-like gene of *Malus domestica* with query coverage of 42%. Meanwhile, the lower percentage of identity (29%) was revealed for *Nelumbo nucifera* with the highest query coverage of 59%. A comparison was presented in Fig. (7) for the percentage of coverage of different *dof* gene sequences in different plant species.

Although BLAST did not reveal any *Dof* genes in *Olea* species, when applying the $SCoT33_{1000}$ against the available *Olea* transcriptome sequences at the NCBI TSA database (Fig. 8); some hits resulted with high percentage of identity (94%). Furthermore, high percentage of coverage was exhibited by OT11 (GBKW01094999), OT12 (GCJV01039897), OT13 (GABQ01051714), and OT14 (GBKW01052744) sequences (Fig. 8). These sequences were reported in association with the transcriptional changes occurring during cold acclimation in olive leaves (de la O Leyva-Pérez et al., 2014 and Guerra et al., 2015). This may elucidate the role of $SCoT33_{1000}$ sequence in the developmental stages of plants.

Retrotransposons (Copia-type) are common in the plant kingdom, and often contain a percentage of DNA that may play a role in genome organization and evolution of plants (Stergiou et al., 2002; Khaliq et al., 2012).The $SCoT13_{250}$ showed high percentage of identity and query coverage with two sequences of *copia*-type retrotransposons of two different olive sequences (94595-B and 58259-G) (Fig. 9) along with unannotated sequences belonging to *Vitis vinifera* and different plant species. Using TSA database, the sequence of $SCoT13_{250}$ revealed high similarity (97%) with RNA sequences of Olea species and high percentage of query coverage was revealed by OT5 (GABQ01049086), OT6 (GBKW01110591), OT9 (GCJV01005145), OT10 (GBKW01129684), and OT8 (GABQ01051812) (Fig. 9). These sequences were reported in the annotation of functional olive transcriptome (Muñoz-Mérida et al., 2013; de la O Leyva-Pérezet al., 2014 and Guerra et al., 2015).

The BLASTN search of the RAPD fragment $A13_{600}$ revealed some hits with high percentage of identity (94%) and query coverage (21%). All blast results were limited to plant chloroplast genomes, where, published olive cultivars chloroplast genomes had the highest similarity. This may suggest that $A13_{600}$ sequence include common region of chloroplast genomes in olive. Therefore, we attempted to identify genes within a window of 1000 bp in 10 of these olive chloroplast genomes (Fig. 10).Based on this criteria, seven olive varieties shared the existence of tRNA-Arg, tRNA-Asp, ATP synthase CF1 α subunit, photosystem II protein M (*psbM*) and cytochrome b6/f complex subunit N genes. At the same time, two chloroplast genomes shared protein D1 of the photosystem II, tRNA-Lys and maturase K enzyme. However, one variety was distinct by the presence of tRNA-Gly.

The BLASTX search revealed that $A13_{600}$ sequence may contain the domain of SANT protein (Fig. 9).Proteins with the domains of MYB or SANT were likely to participate in the regulation of plant flowering and early-fruit development (Barg et al., 2005; Choi et al., 2007). A highest percentage of identity (41%) and query coverage (27%) were detected between our sequence and the genes of both *Trichorapa* and *Euphratica* species. The highest percentage of query coverage was detected in two domains of Myb/SANT-like DNA-binding proteins in *Medicago truncatula* (Fig. 11). Consequently, $A13_{600}$ sequence shared a high similarity with nuclear and chloroplast genes. This could be due to common repetitive sequences in chloroplast genomes (Xu et al., 2002; Li et al., 2013). Another reason could be the transposition of some chloroplast DNA to the nucleus (Daley and Whelan, 2005; de Grey, 2005).

Materials and Methods

Plant materials and DNA extraction

Ten olive varieties (Kronaki, Tofahi, Calamata, Manzanillo, Ojazi Shami, Koratina, Dolci, Maraki, Pekoal and Khodari) were provided by the Horticultural Research Institute, ARC, Egypt. The total DNA was isolated from collected fresh young leaves using DNAeasy Plant Mini Kit (Qiagen, Santa Clarita, CA).

RAPD PCR analysis

Twenty four RAPD primers were applied (Table S1). The PCR reaction was carried out following the method of Adawy et al.(2002). Reactions were performed in a total volume of 25 µl including 1X reaction buffer, 0.2 µM of dNTPs,1.5m M $MgCl_2$, 0.2 µM primer, 0.5 unit of *Taq* polymerase (Qiagen Ltd., Germany) and 50 ng of genomic DNA. The temperature profile was set as: initial denaturation cycle at 94°C/5 min followed by 40 cycles of 94°C/1min, 36°C/1 min and 72°C/2 min. The final polymerization cycle was set at 72°C/7 min and the reaction was hold at 4°C. The amplified products were resolved in 1.5% agarose gel in 1X TBE buffer and stained with ethidium bromide (0.5 mg/mL) then visualized on transilluminator UV light.

SCoT PCR analysis

A set of thirty nine SCoT primers were tested .The nucleotide sequences of primers were drived form Luo et al.(2010) and Adawy et al.(2014) (Table S2).The PCR reaction was performed as described by Luo et al. (2010).The total reaction volume was set to 25 µl of 1X reaction buffer containing 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.4µM primer; 50 ng genomic DNA and 0.5U of *Taq* DNA polymerase (Qiagen Ltd., Germany). PCR temperature profile was set as: an initial denaturation step at $94^{O}C$/3 min followed by 36 cycles of $94^{O}C$/50 s, $50^{O}C$/1 min and $72^{O}C$/2 min; then a final extension at $72^{O}C$/5 min.PCR products were visualized using the same procedure as in RAPD.

SAMPL and SASPL PCR analysis

SAMPL is an AFLP-based marker where the *Eco*RI primer is replaced by ISSR primer in the selective amplification step through the AFLP procedure. SAMPL was developed by Morgante and Vogel (1999) to provide a high multiplex ratio marker system that combines the advantages of microsatellites and AFLP markers. ISSR requires no information about the species and are arbitrary multi loci markers (Figure 1).

For the newly developed SASPL markers, the *Eco*RI primer was replaced by SCoT primer in the selective amplification step of the AFLP procedure. SCoT primers were adopted to SASPL procedure from Luo et al. (2010).These SCoT primers were designed to target genic regions following the method described by Joshi et al. (1997)

and Sawant et al. (1999). Stages of SASPL PCR system were described in comparison with other AFLP-based (AFLP and SAMPL) and non-AFLP-based (RAPD and SCoT) marker systems in Fig. 1. As in AFLP, SASPL can show three different kinds of polymorphisms: a mutation in the restriction site, a mutation in the sequence adjacent to the restriction site and complementary to the primer extensions, or a deletion/insertion within the amplified fragment. Additionally SASPL is expected to target polymorphism within the gene regions that contain genes on both plus and minus DNA strands, which is the basis of SCoT primer design integrated in the SASPL system. SASPL can be performed on any genome, regardless of its complexity and structure and without any prior sequence knowledge and hundreds of markers can be typed quickly and at low cost, offering fine-scale genome coverage. Although AFLP data have poor information contents in analyses requiring precise estimations of heterozygosity, SCoT primers target gene regions surrounding the ATG initiation codon on both DNA strands as shown in Fig.1. In the present study the analysis of SAMPL and SASPL were performed using twelve and eight primer combinations (Table S3 and S4, respectively). This was carried out according to a modified protocol of Vos et al. (1995) using AFLP® Analysis System II (Invitrogen, USA) (Cat.No.10483-022) (Fig. 1). For both SASPL and SAMPL procedures, approximately 400 ng DNA of each of the olive varieties was digested simultaneously with *EcoRI* and *MseI* at 37°C for 2hr.*EcoRI* and *MseI* adaptors were ligated to the digested DNA samples to generate template DNA for amplification. The ligation products were diluted 10 folds and were preamplified in a thermocycler for 20 cycles set at 94°C /30 sec, 56°C/60 sec and 72°C/60 sec. The amplification products were diluted 50 folds. The selective amplification was carried out with the SAMPL or SASPL primer combinations and 5µl of the diluted PCR products from the preamplification product. The PCR selective amplification temperature profile was as follows: one cycle at 94°C/30 sec, 65°C/30 sec 72°C/60 sec; followed by 12 cycles of touch down PCR in which the annealing temperature was decreased by 0.7°C every cycle until a 'touchdown' annealing temperature of 56°Cwas reached. Once reached, another 23 cycles were conducted as described above for preamplification. Two µl of the reaction product was mixed with an equal volume of formamide loading buffer (98% [v/v] formamide, 10 mM EDTA, 0.005% [v/v] of each of xylene cyanol and bromophenol blue), denatured by incubating at 92°C for 3min and quickly cooled on ice. The products were analyzed on 6% (w/v) denaturing polyacrylamide gels. The gel was silver stained according to the protocol described by the manufacturer (Promega Corp., USA, Silver Sequence DNA Staining Reagents, Lot. 171120)**.**

Molecular marker data statistical analysis

For SCoT, RAPD, SAMPL and SASPL analysis, only clear and unambiguous bands were visually scored as either present (1) or absent (0) for all samples and final data sets included both polymorphic and monomorphic bands. Then, a binary statistic matrix was constructed. Dice's similarity matrix coefficients were then calculated between varieties using the unweighted pair group method with arithmetic averages (UPGMA). This matrix was used to construct a phylogenetic tree (dendrogram) using the online construction utility DendroUPGMA (http:// genomes.urv.es/UPGMA/) (Garcia-Vallvé et al., 1999). The polymorphism information content (PIC) was calculated using the PowerMarker software (Liu and Muse, 2005).

In silico PCR analysis

Olive is a diploid (2n = 46) with a genome size of 1.38 Gb (Cruz et al., 2016) and about 47 % of its unigenes are shared with *Vitis vinifera* (Muleo et al., 2012). The full genome of olive (Cruz et al., 2016) was used as template for the *in silico* PCR analysis against all studied PCR primers (83 primers) to reveal the possible PCR amplimers.

Practical Extracting and Reporting Language (PERL) scripts were used for performing the *in silico* PCR analysis by the following criteria: the maximum length of produced amplimer ≤ 1500 bp and the minimum length ≥ 50.In addition, the maximum acceptable sequence mismatch between the primer and the DNA template for SCoT ≤ 4 bp, RAPD ≤ 1bp, SAMPL ≤ 0bp and SASPL ≤ 4bp.

For the primer genome coverage statics the overlap layout consensus algorithm was used to report each primer-covering area. This algorithm ensures that if two PCR amplimers share same genomic areas, only the collective area will be reported and redundant area will be removed. The same algorithm was used for genome sequence assembly (Li et al., 2012).

For calculating the total coverage for each PCR marker type, all *in silico* PCR amplimers belonging to each marker type were processed collectively .The same procedure was conducted in calculating the total primer set (83 primers) coverage.

Circos software package (Circos 0.66) was used for visualizing output results in a circular layout (Krzywinski et al., 2009).

Cloning and sequencing of some PCR polymorphic fragments

Eight SCoT and four RAPD PCR bands which showed variation among the olive varieties were recovered by elution from agarose gels and the DNA was purified. The DNA was inserted into pGEM-T easy vector according to the manufacturer's instructions (Promega Corporation, Cat. No. A1360). DNA sequencing was carried out by Macrogen Inc., Korea.

Sequence analysis

To estimate the generated sequence traces quality, the original forward and reverse sequences were assembled using CodonCode Aligner 3.0 (CodonCode Co., USA). The NCBI online BLAST tool was used with its default parameters to align the generated sequences using BLAST algorithm (Altschul et al., 1994) against the NCBI database. PERL scripts were used to visualize BLAST results using Circos 0.66 software package (Krzywinski et al., 2009).

Conclusion

The newly developed SASPL markers proved to be a reliable tool to study the genetic variability distinguish closely related varieties and cover large genomic area. In addition, using the *in silico* PCR analysis was helpful to identify potential genes included or closes to the studied PCR primers and to test marker techniques selectivity and genome coverage. In addition, our results demonstrates the potential use of SASPL, SCoT, SAMPL and RAPD techniques to identify the investigated Egyptian olive varieties by unique markers and generate high number of polymorphic markers with high

genomic coverage. These markers could be useful for olive breeding programs and the detection of genetic diversity among the Egyptian varieties. Moreover, the sequenced amplicons $A13_{600}$, $SCoT13_{250bp}$ and $SCoT33_{1000bp}$ reveled a high similarity with olive nuclear and chloroplast genes as well as genes from different plant species, suggesting the importance of conducting further investigation on these genes structural and functional levels.

Acknowledgments

The authors would like to dedicate this work to the soul of Prof. Dr. Sami Adawy who participated in this work but passed away during preparing the manuscript. Also we would like to thank Dr. Mohamed Atia, Mr. Morad Mokhtar and Mr. Azmy Marzouk (Molecular Genetics and Genome Mapping Lab. (MGGM), Agricultural Genetic Engineering Research Institute (AGERI), ARC, Egypt for their valuable support during this study.

References

Adawy S, Hussein E, El-Khishin D (2002) Genetic variability studies and molecular fingerprinting of some Egyptian date palm (*Phoenix dactylifera* L.) cultivars II-RAPD and ISSR profiling. Arab J Biotech. 5:225-236.

Adawy SS, Jiang J, Atia MAM (2014) Identification of novel sex-specific PCR-based markers to distinguish the genders in Egyptian date palm trees. Int J Agric Sci Res. 4:45–54.

Adawy SS, Mokhtar MM, Alsamman AM, Sakr MM (2015) Development of annotated EST-SSR database in olive (*Olea europaea*). Int J Sci Res. 4:1063–1073.

Agrawal, M, Shrivastava N, Padh H (2008) Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Reports. 27(4):617-631.

Altschul SF, Boguski MS, Gish W, WoottonJC (1994) Issues in searching molecular sequence databases. Nat Genet. 6:119–129.

Asadiar LS, Rahmani F, Siami A (2012) Assessment of genetic variation in Russian olive (*Elaeagnus angustifolia*) based on morphological traits and random amplified polymorphic DNA (RAPD) genetic markers. J Med Plants Res. 6:1652–1661.

Atienza SG, de la Rosa R, León L, Martín A, Belaj A (2014) Identification of QTL for agronomic traits of importance for olive breeding. Mol Breed. 34(2):725-737.

Azizi A, Wagner C, Honermeier B, Friedt W (2009) Intraspecific diversity and relationship between subspecies of *Origanum vulgare* revealed by comparative AFLP and SAMPL marker analysis. Plant Syst Evol. 281:151–160.

Barg R, Sobolev I, Eilon T, Gur A, Chmelnitsky I, Shabtai S, Grotewold E, Salts Y (2005) The tomato early fruit specific gene *Lefsm1* defines a novel class of plant-specific SANT/MYB domain proteins. Planta. 221:197–211.

Behera TK, Singh AK, Staub JE (2008) Comparative analysis of genetic diversity in Indian bitter gourd (*Momordica charantia* L.) using RAPD and ISSR markers for developing crop improvement strategies. Sci Hort. 115(3):209-17.

Belaj A, Satovic Z, Cipriani G, Baldoni L, Testolin R, Rallo L, Trujillo I (2003) Comparative study of the discriminating capacity of RAPD, AFLP and SSR markers and of their effectiveness in establishing genetic relationships in olive. Theor Appl Genet. 107(4):736-44.

Besnard G, Breton C, Baradat P, Khadari B, Bervillé A (2001) Cultivar identification in olive based on RAPD markers. J Am Soc Hortic Sci. 126:668–675.

Bhattacharyya P, Kumaria S, Kumar S, Tandon P (2013). Start Codon Targeted (SCoT) marker reveals genetic diversity of *Dendrobium nobile* Lindl., an endangered medicinal orchid species. Gene. 529 (1):21-6.

Biswas MK, Chai L, Amar MH, Zhang X, Deng XX (2011) Comparative analysis of genetic diversity in Citrus germplasm collection using AFLP, SSAP, SAMPL and SSR markers. Sci Hortic. 129:798–803.

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am J Hum Genet. 32 (3):314-331.

Boutros PC, Okey AB (2004) PUNS: transcriptomic-and genomic-*in silico* PCR for enhanced primer design. Bioinformatics. 20(15):2399–2400.

Bracci T, Busconi M, Fogher C, Sebastiani L (2011) Molecular studies in olive (*Olea europaea* L.): overview on DNA markers applications and recent advances in genome analysis. Plant Cell Rep. 30:449–462.

Busconi M, Sebastiani L, Fogher C (2006) Development of SCAR markers for germplasm characterisation in olive tree (*Olea europea* L.). Mol Breed. 17:59–68.

Calzada D, Aguerri M, Baos S, Montaner D, Mata M, Dopazo J, Quiralte J, Florido F, Lahoz C, Cárdaba B (2015) Therapeutic targets for olive pollen allergy defined by gene markers modulated by Ole e 1-derived peptides. Mol Immunol. 64:252–261.

Choi K, Park C, Lee J, Oh M, Noh B, Lee I (2007) Arabidopsis homologs of components of the SWR1 complex regulate flowering and plant development. Development. 134:1931–1941.

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. 27:86–93.

Costa R, Pereira G, Garrido I, Tavares-de-Sousa MM, Espinosa F (2016) Comparison of RAPD, ISSR, and AFLP molecular markers to reveal and classify orchardgrass (*Dactylis glomerata* L.) germplasm variations. PloS One.11(4): e0152972.

Cretazzo E, Meneghetti S, De Andrés MT, Gaforio L, Frare E, Cifre J (2010) Clone differentiation and varietal identification by means of SSR, AFLP, SAMPL and M-AFLP in order to assess the clonal selection of grapevine: the case study of Manto Negro, Callet and Moll, autochthonous cultivars of Majorca. Ann Appl Biol. 157:213–227.

Cruz F, Julca I, Gómez-Garrido J, Loska D, Marcet-Houben M, Cano E, Galán B, Frias L, Ribeca, P, Derdak S, Gut M (2016) Genome sequence of the olive tree, *Olea europaea*. Giga Science. 5(1):1:29.

Daley DO, Whelan J (2005). Why genes persist in organelle genomes. Genome Biol. doi: 10.1186/gb-2005-6-5-110.

De Grey A (2005). Forces maintaining organellar genomes: is any as strong as genetic code disparity or hydrophobicity?. BioEssays. 27:436–446.

De la O Leyva-Pérez M, Valverde-Corredor A, Valderrama R, Valderrama R, Jiménez-Ruiz J, Muñoz-Merida A, Trelles O, Barroso JB, Mercado-Blanco J, Luque F (2014) Early and delayed long-term transcriptional changes and short-term transient responses during cold acclimation in olive leaves. DNA Res. doi:dsu033.

El Saied SH, Hegazi AA, Tawfik AA, Sayed HA (2012) Molecular characterization of local and imported olive cultivars grown in Egypt using ISSR technique. J Hortic Sci. 4(2): 148-154.

Fornara F, Panigrahi KCS, Gissot L, Sauerbrunn N, Rühl M, Jarillo JA, Coupland G (2009) Arabidopsis DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. Dev Cell. 17:75–86.

Gajera BB, Kumar N, Singh AS, Punvar BS, Ravikiran R, Subhash N, Jadeja GC (2010) Assessment of genetic diversity in castor (*Ricinus communis* L.) using RAPD and ISSR markers. Ind Crops Prod. 32(3):491-8.

Garcia-Mas J, Oliver M, Gomez-Paniagua H, De Vicente MC (2000) Comparing AFLP, RAPD and RFLP markers for measuring genetic diversity in melon. Theor Appl Genet. 101(5-6):860-4.

Garcia-Vallvé S, Palau J, Romeu A (1999) Horizontal gene transfer in glycosyl hydrolases inferred from codon usage in *Escherichia coli* and *Bacillus subtilis*. Mol Biol Evol. 16:1125–1134.

Gemas VJV, Almadanim MC, Tenreiro R, Martins A, FevereiroP (2004) Genetic diversity in the olive tree (*Olea europaea* L. subsp. *europaea*) cultivated in Portugal revealed by RAPD and ISSR markers. Genet Resour Crop Evol.51:501–511.

Gorji AM, Poczai P, Polgar Z, Taller J (2011) Efficiency of arbitrarily amplified dominant markers (SCoT, ISSR and RAPD) for diagnostic fingerprinting in tetraploid potato. Am J Potato Res. 88:226–237.

Grimmer MK, Trybush S, Hanley S, Francis SA, Karp A, Asher MJC (2007) An anchored linkage map for sugar beet based on AFLP, SNP and RAPD markers and QTL mapping of a new source of resistance to Beet necrotic yellow vein virus. Theor Appl Genet. 114(7):1151-1160.

Guerra D, Lamontanara A, Bagnaresi P, Orrù L, Rizza F, Zelasco S, Beghè D, Ganino T, Pagani D, Cattivelli L, Mazzucotelli E (2015) Transcriptome changes associated with cold acclimation in leaves of olive tree (*Olea europaea* L.). Tree Genet Genomes. 11(6):1-24.

Guo DL, Zhang JY, Liu CH. Genetic diversity in some grape varieties revealed by SCoT analyses (2015) Mol Biol Rep.39(5):5307-13.

Joshi C, Zhou H, Huang X, Chiang VL (1997) Context sequences of translation initiation codon in plants. Plant Mol Biol. 35:993–1001.

Kafkas S, Ozkan H, Sutyemez M (2005) DNA polymorphism and assessment of genetic relationships in walnut genotypes based on AFLP and SAMPL markers. J Am Soc Hortic Sci.130:585–590.

Karimi HR, Kafkas S (2011) Genetic relationships among Pistacia species studied by SAMPL markers. Plant Syst Evol. 297:207–212.

Khaliq I, Khan MA, Pearce S (2012) Ty1-Copia retrotransposons are heterogeneous, extremely high copy number and are major players in the genome organization and evolution of *Agave tequilana*. Genet Resour Crop Evol. 59:575–587.

Koopman W J, Gort G (2004) Significance tests and weighted values for AFLP similarities, based on Arabidopsis *in silico* AFLP fragment length distributions. Genetics. 167(4):1915-1928.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. Genome Res.19:1639–1645.

Lexa M, Horak J, Brzobohaty B (2001) Virtual PCR. Bioinformatics .17:192–193.

Li X, Gao H, Wang Y, Song J, Henry R, Wu H, Hu Z, Yao H, Luo H, Luo K, Pan H, Chen S (2013) Complete chloroplast genome sequence of *Magnolia grandiflora* and comparative analysis with related species. Sci China Life Sci. 56:189-198.

Li Z, Chen Y, Mu D, Yuan J, Shi Y, Zhang H, Gan J, Li N, Hu X, Liu B, Yang B (2012) Comparison of the two major classes of assembly algorithms: overlap layout consensus and de bruijn graph. Briefi Funct Genomics. 11(1): 25-37.

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics. 21:2128–2129.

Luo C, He X-H, Chen H, Hu Y, Ou SJ (2012) Genetic relationship and diversity of *Mangifera indica* L.: revealed through SCoT analysis. Genet Resour Crop Evol. 59:1505–1515.

Luo C, He X-H, Chen H, Ou SJ, Gao MP (2010) Analysis of diversity and relationships among mango cultivars using Start Codon Targeted (SCoT) markers. Biochem Syst Ecol. 38:1176–1184.

Martins-Lopes P, Lima-Brito J, Gomes S, Meirinhos J, Santos L, Guedes-Pinto H (2007) RAPD and ISSR molecular markers in *Olea europaea* L.: Genetic variability and molecular cultivar identification. Genet Resour Crop Evol. 54(1): 117-128.

McGregor CE, Lambert CA, Greyling MM, Louw JH , Warnich L (2000) A comparative assessment of DNA fingerprinting techniques (RAPD, ISSR, AFLP and SSR) in tetraploid potato (*Solanum tuberosum* L.) germplasm. Euphytica. 113(2):135-44.

Mignouna HD, Abang MM, Fagbemi SA (2003) A comparative assessment of molecular marker assays (AFLP, RAPD and SSR) for white yam (*Dioscorea rotundata*) germplasm characterization. Ann Appl Biol. 142(3):269-76.

Mokhtar MM, Adawy SS, El-Assal SE-DS, Hussein EHA (2016) Genic, intergenic SSR database generation, SNPs determination and pathway annotations, in date palm (*Phoenix dactylifera* L.). PLoS One. doi:e0159268.

Morgante M, Vogel JM (1999) Compound microsatellite primers for the detection of genetic polymorphisms. U.S. Patent No. 5:955,276.

Muleo R, Cavallini A, PerrottaG, BaldoniL, Morgante M, Velasco R (2012) Olive Tree Genomic. In Tech. p. 133–148.

Muñoz-Mérida A, González-Plaza JJ, Blanco AM, García-López MC, Rodríguez JM, Pedrola L, Sicardo MD, Hernández ML, De la Rosa R, Belaj A, Gil-Borja M, Luque F, Martínez-Rivas JM, Pisano DG, Trelles O, Valpuesta V, Beuzón CR (2013) De novo assembly and functional annotation of the olive (*Olea europaea*) transcriptome. DNA Res. doi:dss036.

Muthusamy S, Kanagarajan S, Ponnusamy S (2008) Efficiency of RAPD and ISSR markers system in assessing genetic variation of rice bean (*Vigna umbellata*) landraces. Electron J Biotechnol. 11(3):32-41.

Nicese FP, Hormaza JI, McGranahan GH (1998) Molecular characterization and genetic relatedness among walnut (*Juglans regia* L.) genotypes based on RAPD markers. Euphytica. 101(2):199-206.

Parra-Lobato MC, Delgado-Martinez FJ, Gomez-Jimenez MC (2012) Morphological traits and RAPD markers for characterization and identification of minor Spanish olive cultivars from the Extremadura region. Genet Mol Res. 11:2401–2411.

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed. 2(3):225-38.

Premkrishnan BV, Vadivel A (2012) *In silico* RAPD priming sites in expressed sequences and iSCAR markers for oil palm. Comp Funct Genomics. doi:10.1155/2012/913709.

QianW, GeS, HongDY (2001) Genetic variation within and among populations of a wild rice (*Oryza granulate*) from China detected by RAPD and ISSR markers. Theor Appl Genet. 102(2-3): 440-449.

Raina SN, Jain S, Sehgal D, Kumar A, Dar T H, Bhat V, Pandey V, Vaishnavi S, Bhargav A, Singh V (2012) Diversity and relationships of multipurpose seabuckthorn (Hippophae L.) germplasm from the Indian Himalayas as assessed by AFLP and SAMPL markers. Genet Resour Crop Evol. 59:1033–1053.

Rasmussen JP, Saint CP, Monis PT (2007) Use of DNA melting simulation software for *in silico* diagnostic assay design: targeting regions with complex melting curves and confirmation by real-time PCR using intercalating dyes. BMC Bioinformatics. 8:1-1.

Roy JK, Balyan HS, Prasad M, Gupta PK (2002) Use of SAMPL for a study of DNA polymorphism, genetic diversity and possible gene tagging in bread wheat. Theor Appl Genet. 104:465–472.

Roy JK, Lakshmikumaran MS, Balyan HS, Gupta PK (2004) AFLP-based genetic diversity and its comparison with diversity based on SSR, SAMPL, and phenotypic traits in bread wheat. Biochem Genet. 42:43–59.

Sanz-Cortés F, Badenes ML, Paz S,IniguezA, Llacer G (2001) Molecular characterization of olive cultivars using RAPD markers. J Am Soc Hortic Sci.126:7-12.

Sarwat M, Das S, Srivastava PS (2008) Analysis of genetic diversity through AFLP, SAMPL, ISSR and RAPD markers in *Tribulus terrestris*, a medicinal herb. Plant Cell Rep. 27:519–528.

Sawant SV, Singh PK, Gupta SK, Madnala R, Tuli R (1999) Conserved nucleotide sequences in highly expressed genes in plants. J Genet. 78:123–31.

Schuler GD (1997) Sequence mapping by electronic PCR. Genome Res.7:541-550.

Sesli M, Yegenoglu ED (2015) Genetic dissimilarities between wild olives by random amplified polymorphic DNA (RAPD) assay. Afr J Biotechnol. 9:8970–8976.

Shahlaei A, Torabi S, Khosroshahli M (2014) Efficiacy of SCoT and ISSR markers in assesment of tomato (*Lycopersicum esculentum* Mill.) genetic diversity. Int J Biosci .5:14–22.

Stergiou G, Katsiotis A, Hagidimitriou M, Loukas M (2002) Genomic and chromosomal organization of *Ty1*-copia-like sequences in *Olea europaea* and evolutionary relationships of Olea retroelements. Theor Appl Genet.104:926–933.

Tam SM, Mhiri C, Vogelaar A, Kerkveld M, Pearce SR, Grandbastien MA (2005) Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon-based SSAP, AFLP and SSR. Theor Appl Genet.110(5):819-831.

Tanaka M, Takahata Y, Nakayama H, NakataniM, Tahara M (2009) Altered carbohydrate metabolism in the storage roots of sweetpotato plants overexpressing the *SRF1* gene, which encodes a Dof zinc finger transcription factor. Planta.230:737–746.

Verma PC, Chakrabarty D, Jena SN, Mishra DK, Singh PK, Sawant SV, Tuli R (2009) The extent of genetic diversity among Vanilla species: comparative results for RAPD and ISSR. Ind Crops Prod. 29(2):581-9.

Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Friters A, Pot J, Paleman J, Kuiper M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23:4407–4414.

Xiong F, Zhong R, Han Z, Zhong R, Han Z, Jiang J, He L, Zhuang W, TangR (2011) Start codon targeted polymorphism for evaluation of functional genetic variation and relationships in cultivated peanut (*Arachis hypogaea* L.) genotypes. Mol Biol Rep. 38:3487–3494.

Xu D, Abe J, Gai J, Shimamoto Y (2002) Diversity of chloroplast DNA SSRs in wild and cultivated soybeans: evidence for multiple origins of cultivated soybean. Theor Appl Genet. 105:645–653.

Yang J, Yang MF, Zhang WP, Chen F, Shen SH (2011) A putative flowering-time-related Dof transcription factor gene, JcDof3, is controlled by the circadian clock in *Jatropha curcas.* Plant Sci. 181:667–674.

Zhang J, Guo D, Gong Y, Liu CH, Li M, Zhang GH (2011) Optimization of Start Codon Targeted Polymorphism PCR (SCoT-PCR) system in *Vitis vinifera*. J Fruit Sci. 209:214.