BIODIVERSITAS
Volume 13, Number 3, July 2012
Pages: 114-117

ISSN: 1412-033X
E-ISSN: 2085-4722
DOI: 10.13057/biodiv/d130302

In Silico chloroplast SSRs mining of *Olea* species

ERTUGRUL FILIZ[1,♥], IBRAHIM KOC[2]

[1]Department of Crop and Animal Production, Cilimli Vocational School, Duzce University, Duzce, Turkey. Tel. +90 380 681 7312, ext.7406, Fax: +90 380 681 73 13, ♥email: ertugrulfiliz@gmail.com

[2]Department of Molecular Biology and Genetics, Gebze Institute of Technology, Kocaeli, Turkey

Manuscript received: 4 June 2012. Revision accepted: 21 June 2012.

ABSTRACT

Filiz E, Koc E. 2012. In Silico chloroplast SSRs mining of *Olea* species. *Biodiversitas 13: 114-117.* Simple sequence repeat (SSR) markers are highly informative and have been widely applied as molecular markers in genetic studies. The purpose of present study is to analyze the occurrence and distribution of chloroplast SSRs in genic and intergenic regions from *Olea* species viz., *Olea europaea, Olea europaea* subsp. *cuspidate*, *Olea europaea* subsp. *europaea*, *Olea europaea* subsp. *maroccana*, *Olea woodiana* subsp. *woodiana* by using bioinformatics tools. We identified 1149 chloroplast SSRs (cpSSRs) in all genome and a total of 340 (29.6%) was located in genic regions. It was observed that the most abundant repeat types were found mononucleotide SSR (66.7 %) followed by trinucleotide SSR (28.3 %), dinucleotide (2.7%), tetranucleotide (1.5%) and pentanucleotide (0.8%). cpSSRs located in genic regions were identified only mono- and trinucleotide motifs, the most abundant of which was trinucleotide (16.2%) followed by mononucleotide (14.3%). All types of repeat motif (mono-, di-, tri-, tetra- and pentanucleotide) were detected except hexanucleotide motifs. According to SSRs analysis, the most abundant observed motifs were identified for mono-, di-, tri-, tetra- and pentanucleotide cpSSRs A/T, AT/TA, AAG/CTT, AAAG/AGTTT, and AATCC/ATTGG respectively. This study results provided scientific base for phylogenetics, evolutionary genetics and diversity studies on different *Olea* species in future.

Key words: *Olea*, olive, chloroplast SSRs, SSR mining, *in silico* analysis

INTRODUCTION

Olea is a genus of about 40 species in the family *Oleaceae* and it is distributed warm temperate and tropical regions of southern Europe, Africa, southern Asia and Australasia. *Olea* known as olive (*Olea europaea* L.), is the most important oil crop in the Mediterranean region and has been used for pharmaceutical, industrial and consumer purposes (Hannachi et al. 2010). Olive is also the second most important oil fruit crop cultivated worldwide after oil palm (Baldoni and Belaj 2009). *Olea* species were evaluated by using various genetic markers such as SSRs genotyping (Taamalli et al. 2008; Ercisli et al. 2011, Gomes et al. 2009, Carriero et al. 2002); RAPDs genotyping (Bogani et al. 1994; Belaj et al. 2004); AFLPs genotyping (Owen et al. 2005; Baldoni et al. 2006), SNPs (Reale et al. 2006), Ribosomal DNA polymorphisms (Hess et al. 2000, Besnard et al. 2007), and organelle DNA polymorphisms (Besnard et al. 2002; Intrieri et al. 2007).

Simple sequence repeats (SSRs) or microsatellites are tandem repeated motifs which are 1-6 nucleotide length and located in all prokaryotic and eukaryotic genomes (Zane et al. 2002). Microsatellites are being used for mapping and tagging of genes, marker assisted selection (MAS), genome mapping and functional genomics (Kalia et al. 2011). SSRs are associated with coding and noncoding regions and can be found nuclear, chloroplast and mitochondrial genome (Provan et al. 2001; Rajendrakumar et al. 2007). The chloroplast genome includes about 120-130 genes and usually ranges in size from 120-200 kb (Sugiura 1992) and is characterized by haploidy and a lack of recombination and uniparental inheritance. Therefore, chloroplast markers (cpSSRs) are more effective indicator of population genetic structure than nuclear SSRs markers (Birky 1995; Petit et al. 1995). In the last decade, chloroplast SSRs (cpSSRs) have been widely used for plant taxonomic and phylogenetic studies, diversity analysis and population genetics (Provan et al. 2001). The main objective of this study was to analysis SSRs in chloroplast genome (cpDNA) of *Olea* species for their occurrence and distribution in both coding and non-coding regions.

MATERIAL AND METHODS

All the chloroplast genome sequences of olive (*Olea europaea, Olea europaea* subsp. *cuspidate*, *Olea europaea* subsp. *europaea*, *Olea europaea* subsp. *maroccana*, *Olea woodiana* subsp. *woodiana*) were downloaded in FASTA format from GenBank (ftp://ncbi.nlm.nih.gov/genbank/genomes/) (Table 1.). The identification of chloroplast microsatellites was carried out by MISA perl script (http://pgrc.ipk-gatersleben.de/misa/). The minimum motif repeat size were set to 8 for mononucleotide, 5 for dinucleotide, 3 for trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide in locating the microsatellites with maximum differences two SSRs was

100. SSRs were searched in full chloroplast genome as well as separate coding and non-coding regions for each species. Also, GC content was calculated.

Table 1. Details of chloroplast genomes in *Olea* species

Plant species	Plastome size (bp)	Accession Number	G+C content (%)
Olea europaea	155888	NC_013707	37.80%
O. europaea subsp. *cuspidata*	155862	NC_015604	37.81%
O. europaea subsp. *europaea*	155875	NC_015401	37.81%
O. europaea subsp. *maroccana*	155896	NC_015623	37.81%
O. woodiana subsp. *woodiana*	155942	NC_015608	37.79%

RESULTS AND DISCUSSION

Abundance of SSRs

A total of five *Olea* species chloroplast complete genome sequences were evaluated for chloroplast SSRs and we found 1149 SSRs, of which 340 (29.6%) were localized in genic regions and the 809 (70.4%) cpSSRs were localized in intergenic regions (Table 2.). G+C content of the species are closely similar frequency ranging from 37.79 to 37.81% (Table 1.)

Table 2. The number of the genic and intergenic cpSSRs based on motif size for each species

	Mono		Di		Tri		Tetra		Penta		Hexa		
	G	I	G	I	G	I	G	I	G	I	G	I	Total
Olea europaea	33	122	0	6	35	31	0	4	0	2	0	0	233
Olea europaea subsp. *cuspidata*	33	119	0	6	35	30	0	3	0	2	0	0	228
Olea europaea subsp. *europaea*	33	120	0	6	35	30	0	3	0	2	0	0	229
Olea europaea subsp. *maroccana*	33	120	0	6	35	30	0	3	0	2	0	0	229
Olea woodiana subsp. *woodiana*	32	121	0	7	36	28	0	4	0	2	0	0	230
Total													1149

Note: G: genic, I: intergenic

Total number of cpSSRs in the chloroplast genomes ranged from 228 to 233 and the density of microsatellites ranged from 1.46-1.49 cpSSR per kb. *Olea europaea* subsp. *cuspidate* chloroplast has the least number of cpSSRs (228) while *Olea europaea* chloroplast had the most abundant cpSSRs (233). The density of microsatellites were found for *Olea europaea*, *Olea europaea* subsp. *cuspidate*, *Olea europaea* subsp. *europaea*, *Olea europaea* subsp. *maroccana* and *Olea woodiana* subsp. *woodiana* 1.49, 1.46, 1.47, 1.47 and 1.47 cpSSR per kb respectively. An average frequency was 1.47 cpSSR per kb which was higher than some cereal species cpSSRs (1.36 cpSSR per kb) (Melotto-passarin et al. 2011), *Solanaceae* species cpSSRs (1.26 cpSSR per kb) (Tambarussi et al. 2009), *Solanum lycopersicum* EST-SSRs (1.3 SSR per kb) (Gupta et al. 2010). However, an average frequency of SSRs in *Olea* species in present study (1.47 SSR per kb) is lower than found in loblolly pine EST-SSRs (42.9 SSR per kb), some cereal species EST-SSRs (6 SSR per kb) (Varshney et al. 2002), palm EST-SSRs (4.4 SSR per kb) (Palliyarakkal et al. 2011).

Distribution of cpSSRs

The investigation of different types of SSR repeats showed that the percentage of occurrence of mononucleotide SSR (66.7 %) was the highest followed by trinucleotide SSR (28.3%), dinucleotide (2.7%), tetranucleotide (1.5%) and pentanucleotide (0.8%) (Figure 1.).

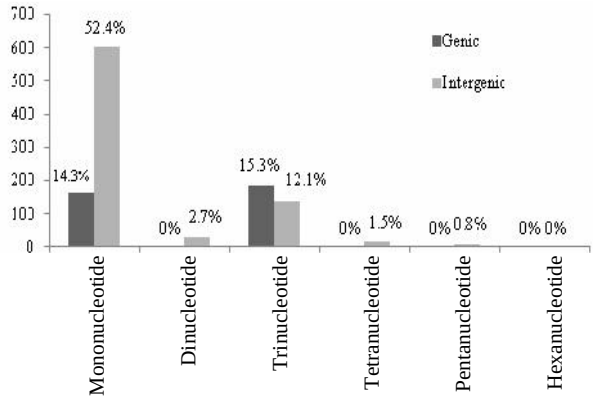

Figure 1. Frequencies (%) of different repeats in genic and intergenic regions.

The most abundant genic cpSSRs type was trinucleotide (16.2%) followed by mononucleotide (14.3%). This finding supports that triplet SSR repeats can be located easily within coding regions (Hancock and Simon 2005). Our results revealed that A/T repeats (98%) were found to be more abundant than the G/C (2%) motifs. These results were consistent with SSRs analysis of major cereal organelle genome data (Rajendrakumar et al. 2008) (Figure 2.).

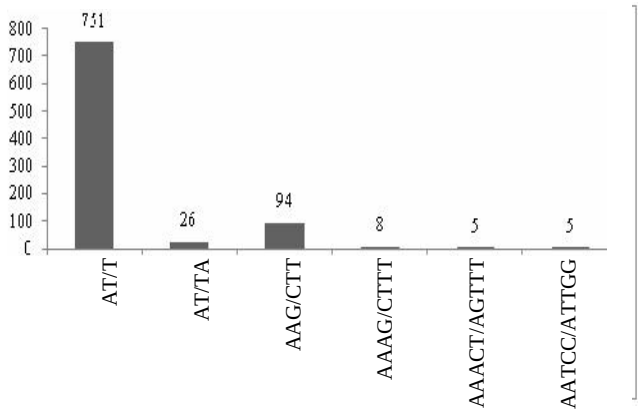

Figure 2. The most abundant cpSSRs motifs in all chloroplast genomes.

For dinucleotide repeats, AT/TA motif was the most common dinucleotide repeat with a frequency of 83.9%. These results are in agreement with previous findings (Gandhi et al. 2010; Rajendrakumar et al. 2007; Kuntal et al. 2012). The higher AT/TA frequencies in chloroplast genomes can be explained to be the conclusion of the high A/T content of the genomes. Among the trinucleotide repeats, AAG/CTT motif was the most common (28.9%) followed by AAT/TTA (24.9%) and AAC/GTT (16%) and this finding was not consistent with earlier studies results (Rajendrakumar et al. 2007; Melotto-passarin et al. 2011; Kuntal et al. 2012). In tetranucleotide SSR motifs, the maximum frequency of 47.1% was showed by AAAG/CTTT followed by AAAT/ATTT (29.4%). In pentanucleotide SSR motifs, frequency of AAACT/ AGTTT and AATCC/ATTGG motifs were found to be equal (50%). Interestingly, there are no any hexanucleotide repeats in all chloroplast genomes. A total of 340 cpSSRs (30.5%) were found to be in genic regions while a total of 809 cpSSRs (69.5%) were found to be in intergenic regions and genic cpSSRs were identified only mono and trinucleotide motifs. In general, intergenic cpSSR (69.5%) in *Olea* species were more abundant than genic cpSSR (29.6%) and this result is consistent with earlier studies Asteraceae (Timme et al. 2007), Fabaceae (Saski et al. 2005), and Solanaceae (Daniell et al. 2006), *Saccharum* (Melotto-passarin et al. 2011). According to the analysis, the most abundant genic cpSSRs were found for trinucleotide motifs (57.2%). The SSRs in genes show a higher mutation rate (instability) than nonrepetitive regions in genome. Also, SSRs variations affect gene expression, inactivation of gene activity, and/or a change of function (Li et al. 2004). Thus, our results corroborate this hypothesis that the most number of genic trinucleotide cpSSRs may cause dynamic evolution and mutational forces in exons and exhibit genetic and phenotypic variations. The number of plastome genes with cpSSR were identified (Table 3.) and we found *Olea europaea*, *Olea europaea* subsp. *cuspidate*, *Olea europaea* subsp. *europaea*, *Olea europaea* subsp. *maroccana* with 49 genes while *Olea woodiana* subsp. *woodiana* has 46 genes.

Table 3. Frequency (%) of plastome genes with cpSSRs

Species	Number of plastome genes	Genes with cpSSR	Genes with cpSSR %
Olea europaea	130	49	37.7%
Olea europaea subsp. *cuspidata*	130	49	37.7%
Olea europaea subsp. *europaea*	130	49	37.7%
Olea europaea subsp. *maroccana*	130	49	37.7%
Olea woodiana subsp. *woodiana*	130	46	35.4%

Many studies revealed that large numbers of SSRs are distributed in genic regions of genomes, containing protein-coding genes and expressed sequence tags (ESTs). SSRs distributions play key role for history of genome evolution and mutational processes (Morgante et al. 2002). In this study, we investigated the distribution of different types of SSR in coding and non-coding regions of five different species of *Olea* genus by using bioinformatic tools. The location of SSR in the genome determines its functional role like gene regulation, development and evolution (Kalia et al. 2011). According to analysis, intergenic cpSSRs are predominant over genic cpSSRs and as expected, trinucleotide repeats are more common in coding regions. Except mononucleotide and trinucleotide repeats, other classes of repeats were low in number in the chloroplast genomes.

CONCLUSION

SSR markers are very informative because they are codominant and highly polymorphic. In addition, SSRs markers are highly mutable loci and they can be used for characterization of genome and a particular region can be identified in the genome. This study results can be targeted towards the diversity and evolutionary studies of *Olea* species.

REFERENCES

Baldoni L, Belaj A. 2009. Olive. In: Vollmann J, Rajean I (eds) Oil crops. Handbook of plant breeding. Vol 4. Springer, New York.

Baldoni L, Tosti N, Ricciolini C, Belaj A, Arcioni S, Pannelli G, Germana MA, Mulas M, Porceddu A. 2006. Genetic structure of wild and cultivated olives in the central Mediterranean basin. Ann Bot 98: 935-942.

Belaj A, Rallo L, Trujillo I, Baldoni L. 2004. Using RAPD and AFLP markers to distinguish individuals obtained by clonal selection of 'Arbequina' and 'Manzanilla de Sevilla' olive. HortScience 39: 1566-1570.

Besnard G, Henry P, Wille L, Cooke D, Chapuis E . 2007. On the origin of the invasive olives (*Olea europaea* L., Oleaceae). Heredity 99: 608-619.

Besnard G, Khadari B, Baradat P, Berville A. 2002. Combination of chloroplast and mitochondrial DNA polymorphism to study cytoplasm genetic differentiation in the olive complex (*Olea europaea* L.). Theor Appl Genet 105: 139-144.

Birky CW Jr. 1995. Uniparental inheritance of mitochondrial and chloroplast genes: Mechanisms and evolution. Proc Natl Acad Sci USA 92: 11331-11338.

Bogani P, Cavalieri D, Petruccelli R, Roselli G. 1994. Identification of olive tree cultivars by using random amplified polymorphic DNA. Acta Hortic 356: 98-101.

Carriero F, Fontanazza G, Cellini F, Giorio G. 2002. Identification of simple sequence repeats (SSRs) in olive (*Olea europaea* L.). Theor Appl Genet 104 (2-3): 301-307.

Daniell H, Lee SB, Grevich J, Saski C et al.. 2006. Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analyses with other Solanaceae genomes. Theor Appl Genet 112: 1503-1518.

Ercisli S, Ipek A, Barut E. 2011. SSR marker-based DNA fingerprinting and cultivar identification of olives (*Olea europaea*). Biochem Genet. 49 (9-10): 555-61.

Gandhi SG, Awasthi P, Bedi YS. 2010. Analysis of SSR dynamics in chloroplast genomes of Brassicaceae family. Bioinformation 5 (1): 16-20.

Gomes S, Martins-Lopes P, Lopes J, Guedes-Pinto H. 2009. Assessing Genetic Diversity in *Olea europaea* L. using ISSR and SSR Markers. Plant Mol Biol Rep 27: 365-373.

Gupta S, Tripathi KP, Roy S, Sharma A. 2010. Analysis of unigene derived microsatellite markers in family Solanaceae. Bioinformation 5 (3): 113-121.

Hancock JM, Simon M. 2005. Simple sequence repeats in proteins and their significance for network evolution. Gene 345: 113-118.

Hannachi H, Breton C, Msallem M, Hadj SB, Gazzah M, Bervillé A. 2010. Genetic relationships between cultivated and wild olive trees (*Olea europaea* L. var. *europaea* and var. *sylvestris*) based on nuclear and chloroplast SSR markers. Nat Res 1: 95-103.

Hess J, Kadereit W, Vargas P. 2000. The colonization history of *Olea europaea* L. in Macaronesia based on internal transcribed spacer 1 (ITS-1) sequences, randomly amplified polymorphic DNAs (RAPD) and intersimple sequence repeats (ISSR). Mol Ecol 9: 857-868.

Intrieri MC, Muleo R, Buiatti M. 2007. Chloroplast DNA polymorphisms as molecular markers to identify cultivars of *Olea europaea* L. J Hort Sci Biotechnol 82: 109-113.

Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK. 2011. Microsatellite markers: an overview of the recent progress in plants. Euphytica 177: 309-334.

Kuntal H, Sharma V, Daniell H. 2012. Microsatellite analysis in organelle genomes of *Chlorophyta*. Bioinformation 8 (6): 255-259.

Li YC, Korol AB, Fahima T, Nevo E. 2004. Microsatellites within genes: Structure, function, and evolution. Mol Biol Evol 21 (6): 991-1007.

Melotto-passarin DM, Tambarussi EV, Dressano K, De Martin VF, Carrer H. 2011. Characterization of chloroplast DNA microsatellites from *Saccharum* spp. and related species. Genet Mol Res 10 (3): 2024-2033.

Morgante M, Hanafey M, Powell W. 2002. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 30: 194-200.

Owen CA, Bita EC, Banilas G, Hajjar SE, Sellinakis V, Aksoy U, Hepaksoy S, Chamoun R, Talhook SN, Metzidakis I, Hatzopoulos P, Kalaitzis P. 2005. AFLP reveals structural details of genetic deiversity within cultivated olive germoplasm from the Eastern Mediterranean. Theor Appl Genet 110: 1169-1176.

Palliyarakkal MK, Ramaswamy M, Vadivel A. 2011. Microsatellites in palm (Arecaceae) sequences. Bioinformation 7 (7): 347- 351.

Petit RJ, Bahrman N, Baradat P. 1995. Comparison of genetic differentiation in maritime pine (*Pinus pinaster* Ait.) estimated using isozyme, total protein and terpenic loci. Heredity 75: 382-389.

Provan J, Powell W, Hollingsworth PM. 2001. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends Eco Evol 16: 142-147.

Rajendrakumar P, Biswal AK, Balachandran SM, Srinivasarao K, Sundaram RM. 2007. Simple sequence repeats in organellar genomes of rice: frequency and distribution in genic and intergenic regions. Bioinformatics 23: 1-4.

Rajendrakumar P, Biswal AK, Balachandran SM, Sundaram RM. 2008. In silico analysis of microsatellites in organellar genomes of major cereals for understanding the phylogenetic relationships. In Silico Biology 8: 9-18.

Reale S, Doveri S, Dı ´az A, Angiolillo A, Lucentini L, Pilla F, Martı ´n A, Donini P, Lee D. 2006. SNP-based markers for discriminating olive (*Olea europaea* L.) cultivars. Genome 49: 1193-1205.

Saski C, Lee SB, Daniell H, Wood TC, Tomkins J, Kim HG, Jansen RK. 2005. Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes. Plant Mol Biol 59: 309-322.

Sugiura M. 1992. The chloroplast genome. Plant Mol Biol 19: 149-68.

Taamalli W, Guena F, Bassi D, Daoud D, Zarrouk M. 2008. SSR marker based fingerprinting of Tunisian olive (*Olea europaea* L.) varieties. J Agron 7:176-181.

Tambarussi EV, Melotto-passarin DM, Barbosa AL, Brigati JB et al.. 2009. In silico analysis of simple sequence repeats from chloroplast genomes of *Solanaceae* species. Crop Breed Appl Biotech 9: 344-352.

Timme R, Kuehl EJ, Boore JL, Jansen RK. 2007. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: identification of divergent regions and categorization of shared repeats. Am J Bot 94: 302-312.

Varshney RK, Thiel T, Stein N, Langridge P, Graner A. 2002. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 7: 537 - 546.

Zane L, Bargelloni L, Patarnello T. 2002. Strategies for microsatellite isolation: a review. Mol Ecol 11: 1-16.