

*Full Length Research Paper*

# Genetic variation via SSR polymorphic information content and ecological distribution of Nigerian sesame

**Olusola Thomas ODUOYE\***, Opeyemi Ayokunnu OLUWASANYA, Oluwabukola Omowumi ARIKAWE, Abiodun SUNDAY, Omowumi Abidemi AYEKUN and Oyejide Suara OYENPEMI

National Centre for Genetic Resources and Biotechnology (NACGRAB), PMB 5382, Moor Plantation, Ibadan, Oyo State, Nigeria.

Received 7 October 2019; Accepted 10 March 2020

**Sesame (*Sesamum indicum* L.) of the family Pedaliaceae is an important and old oil crop that is cultivated mainly in the tropical and subtropical regions of Asia and Africa for several economic values. Several molecular markers have been employed to study the genetic diversity of this important crop. The study focused on the genetic diversity through polymorphism information content (PIC) by the use of Simple Sequence Repeat markers among randomly collected 22 Sesame germplasm across 2 ecological zones stored in NACGRAB seed Genebank, Nigeria. The extraction procedure followed Cetyltrimethylammonium Bromide (CTAB) and the recovered DNA were good with average concentration of 337.00 ng/μL and average quality of 1.795. 30 primers were designed but only 12 with highest genome coverage were used to analyzed the genetic data with NTSYS pc ver.2.02 and Power Marker ver.3.5. The PIC ranged between 0.36 in Primer OTO2 and 0.76 in Primer OTO5. The evolutionary relationship was constructed based on the polymorphic primers and according to their ecological locations. The populations were divided to 2 major clads. Of the 2 ecological zones, the more diverse ecological zone is the derived Savana with 13 accessions, while the less diverse is the humid forest with 9 accessions.**

**Key words:** Sesame, genetic diversity, polymorphism information content (PIC), ecological distribution, microsatellite.

## INTRODUCTION

Sesame (*Sesamum indicum* L.,  $2n = 26$ ), belonging to the Pedaliaceae family with 60 species organized into 16 genera (Singh et al., 2015) is one of the oldest cultivated plant, considered important for its edible oil (Anilakumar et al., 2010). Myanmar takes the lead position in sesame production closely followed by India and China (Koppelman

et al., 2015). Sesame is very important to human being in so many ways, especially in dietary preparation and besides direct consumption, sesame seeds are also used as an active ingredient in cosmetic industry (Dossa et al., 2016), decorative elements, antiseptics, bactericides, viricides (Bedigian, 2010), disinfectants, moth repellants

\*Corresponding author. E-mail: [solaoduoye@gmail.com](mailto:solaoduoye@gmail.com).

(Bedigian 2011) and antitubercular agents because they contain natural antioxidants such as sesamin, sesamol and sesamolol (Badril et al., 2014; Elleuch et al., 2007).

The presence of genetic diversity is a desirable requirement for any breeding program (Dossa et al., 2017a). Nigeria, being one of the most important Sesame producing countries aims to make systematic efforts to characterize and document the genetic variability of Sesame in the country. The insight of genetic variation between and within populations gives pivotal information which is indispensable for the formulation of management strategy channeled towards crop improvement program and conservation of biodiversity (Pandey et al., 2015). Morphological and agronomic characteristics as well as isozyme and molecular marker analysis have been used to determine genetic diversity in crop species (Alemu et al., 2013). The use of morphological and agronomic characteristics is prone to strong influence of environmental interference. Molecular markers offer a reliable means of identification and to understand the genetic variability in crops as it overcomes the limitation of environmental influence (Wu et al., 2014) and sesame research has witnessed a rapid development of genetic tools particularly molecular markers and their application in genetic diversity studies and marker assisted breeding (Dossa et al., 2017b; Uncu et al., 2017).

Different molecular marker systems have been developed, with its merit and demerits in terms of ease of use and degree of information to study genetic diversity. The universal markers include the randomly amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), sequence-related amplified polymorphism (SRAP), inter-simple sequence repeats (ISSR), specie-specific marker such as simple sequence repeats (SSR), diversity arrays technology (DArT), restriction-site associated DNA sequencing (RADseq), single-nucleotide polymorphism (SNP), specific-locus amplified fragment sequencing (SLAFseq), random selective amplification of microsatellite polymorphic loci (RSAMPL) (Dossa et al., 2017a) and expressed sequence tags-SSR (EST-SSR). The choice of marker is determined by the kind of study to be undertaken. A suitable marker is characterized by high polymorphism, co-dominant inheritance, frequency of occurrence and even distribution throughout the genome (Welsing et al., 2005), selective neutral behavior, easy access, easy and fast assay, low cost and high throughput, high reproducibility, and transferability between laboratories, population and/or species (Woldesenbet et al., 2005).

Microsatellites or simple sequence repeat (SSR) is considered a good choice because of high abundance and reproducibility, easy scoring, low cost, extensive coverage, high polymorphic information content (PIC) and its co dominant nature (Wei et al., 2014; Pandey et al., 2015; Dossa et al., 2017a).

The efficiency of SSR is not faultless as its application was limited in sesame and the pioneer findings on sesame specific primers reported polymorphism in 10 SSRs in sesame (Dixit et al., 2005) compared to more than 1000 SSR loci in soybean (Xin et al., 2012; Song et al., 2004). This paradigm has however shifted recently since the completion of genome sequencing for this oil rich crop. The combination of agro-morphological and molecular markers is the best choice to characterize germplasm as it gives the opportunity to comparatively analyze the phenotypes from field experiments with molecular phenotypes and genotypes from laboratory studies (Pham et al., 2011).

Sesame research has witnessed three different periods: Germplasm collection era, classical breeding and genetics era and presently the Omics era. Advancement in Omic era has projected sesame research to a higher level of seed improvement (Dossa et al., 2017b). The completion and availability of the full nuclear genome sequence gives unlimited access to reference genome information for the study of genetic traits and comparative genomics (Wang et al., 2014).

The whole sesame genome was sequenced with the help of high-throughput next-generation sequencing and informative SSRs were identified to be distributed throughout the whole genome sequence (Dossa et al., 2017a). Botstein and others (Nagy et al., 2012) published one of the first reports where PIC became the popular scientific approach for calculation of genetic variabilities through a widely acceptable formula to measure the information content of molecular markers. PIC determines the value of a marker for detecting polymorphism within a population, depending on the number of detectable alleles and the distribution of their frequency (Wei et al., 2014). PIC gives an evaluation of the ability of molecular markers to discriminate a given population (Vilas, 2014).

This research was conducted to determine, using SSR markers, the extent of genetic variation or distance among twenty two accessions of sesame cultivars collected from the forest and savannah agro ecological zones in Nigeria. They are now conserved in the gene bank of the National Centre for Genetic Resources and Biotechnology.

## MATERIALS AND METHODS

### Sample collection

About 22 Sesame accessions from NACGRAB's seed genebank that their field collections were within eight States of the Federation (Nigeria) but restricted to only two ecological zones were collected with their Accession Numbers for this study as recorded in Table 1. The samples were potted in glass house and allowed to grow for 3 weeks before collection for analysis.

### DNA extraction procedure

About 2 g of leaves from 3 weeks old of potted sesame was ground

**Table 1.** List of sesame accessions, nucleic acid concentration and purity recovered.

S/N	Sesame accession number	State of collection/ ecological zone	Nucleic acid conc. (ng/μl)	260/280 purity
1	NGB00421	Imo/ HF	170.30	1.78
2	NGB00400	Kwara/ DS	68.60	1.84
3	NGB00402	Edo/ HF	777.00	1.76
4	NGB00368	Oyo/ DS	92.30	1.69
5	NGB00406	Nassarawa/ DS	86.70	1.74
6	NGB00420	Kogi/ DS	44.60	1.79
7	NGB00382	Ondo/ HF	175.50	1.75
8	NG512162	Anambra/ HF	143.60	1.88
9	NGB00379	Ondo/ HF	1056.30	1.89
10	NG512165	Edo/ HF	811.00	1.77
11	NG612166	Imo/ HF	690.10	1.82
12	NG512167	Kogi/ DS	165.80	1.76
13	NG512171	Anambra/ HF	1131.60	1.73
14	NG512173	Kogi/ DS	29.80	1.77
15	NG512174	Ondo/ HF	372.90	1.84
16	NGB00422	Nassarawa / DS	94.80	1.90
17	NGB00405	Kogi/ DS	49.60	1.76
18	NGB00425	Kwara/ DS	144.80	1.88
19	NGB00428	Nassarawa/ DS	155.10	1.86
20	NGB00433	Oyo/ DS	193.70	1.75
21	NGB01409	Kwara/ DS	490.20	1.80
22	NGB00419	Kogi/ DS	589.00	1.74

DS: Derived Savanna vegetation. HF: Humid forest vegetation.

after being surface sterilized with ethanol and 700μl of freshly prepared modified CTAB extraction buffer (200mMTris, pH 7.5; 50mM EDTA, pH 8.0; 2M NaCl; 2% CTAB; 1% beta-mercaptoethanol (just before use)) was added in mortar and pestle. The resultant mixture was homogenized and incubated in 65°C water bath for 20min. It was allowed to cool for 7 min and 600μL chloroform: isoamylalcohol (24:1) was added in the tubes. The mixture was placed in an orbital shaker at 450 rpm for 1 h. The supernatant recovered after centrifugation at 10,000 rpm for 7 min was transferred into new tubes and up to 600μL Isopropanol was added and kept in -20°C freezer overnight for DNA precipitation. The pellet was collected by centrifugation at 10,000 rpm for 2 min and washed with 70% ethanol up to 3 times. Pellet was air-dried in hood until no further trace of ethanol. An average of 150 μL nuclease free water was added to elute the DNA.

#### DNA quantitation

Nanodrop spectrophotometer 2000 model was used to check the quality and quantity of the extracted DNA prior to amplification of the gene regions.

#### Gene amplification

Primers were designed from 3 gene regions of *S. indicum* with the following accession numbers: XM\_011086313; XM\_011086585 and XM\_020696983. Up to 30 primers were designed and tested but only 12 most polymorphic, with widest genome coverage were used

for the analysis (Table 2). The good amplification result was produced and retried for reproducibility in total reaction volume of 20 μL (2.5 μL of 10x PCR buffer, 1.0 μL 25 mM MgCl<sub>2</sub>, 1.0 μL of 10μM forward primer, 1.0 μL 10μM reverse primer, 1.0 μL DMSO, 2.0 μL 2.5Mm DNTPs, 0.1 μL Taq 5u/ul, 3.0 μL 10ng/μl DNA, 8.4 μL H<sub>2</sub>O) for the following PCR reactions: 94°C for 5min incubation, 94°C for 15 s denaturation; 58°C for 20 s annealing; 72°C for 25 s extension; 72°C for 7 min final extension and held finally at 4°C till samples taken out of PCR machine. The product was checked on gel electrophoresis and viewed with 100 bp ladder in trans-illuminator.

#### Data analysis

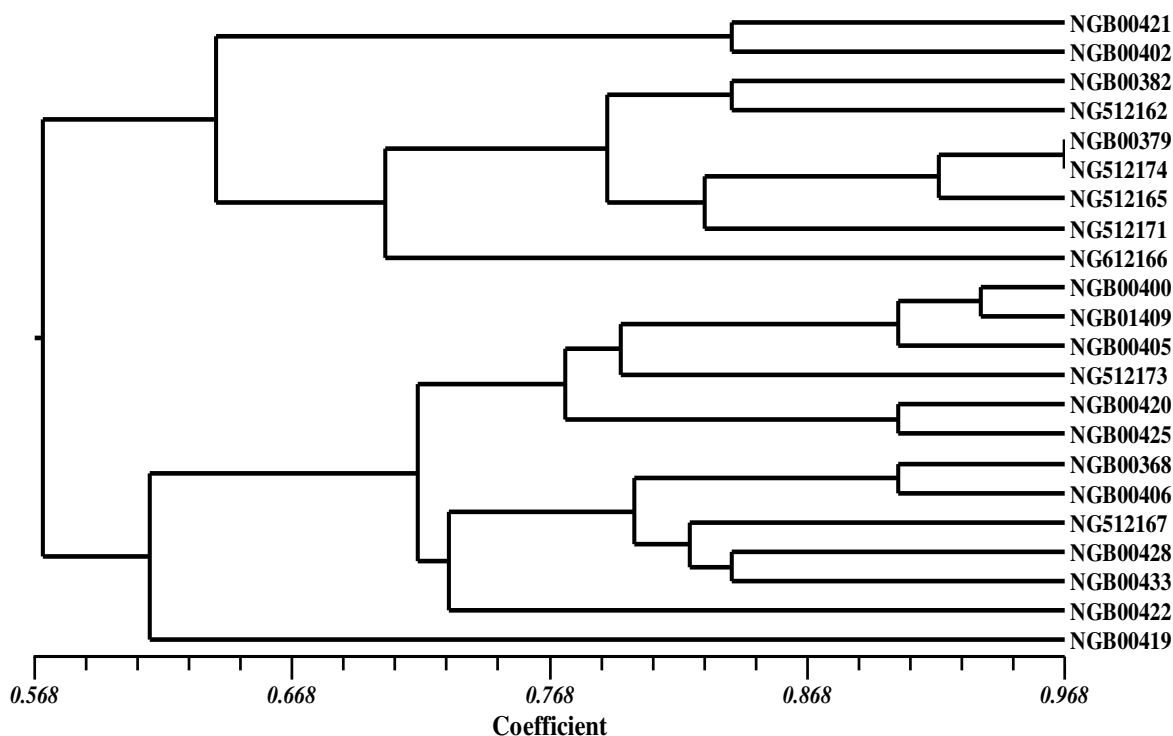
The data generated from the product of amplification and gel-electrophoresis were subjected to analysis using NTSYS pc ver.2.02 (Figure 1) and Power Marker ver.3.5 (Table 3).

## RESULTS AND DISCUSSION

The mean value for DNA purity of 1.79 recovered (Table 1) is close to the acceptable figure of 1.80 which is a prerequisite for downstream analysis (Desjardins and Conklin, 2010). The 12SSR primers (Table 2) revealed polymorphic motifs by producing variable number of alleles which are different in sizes (Table 3).

**Table 2.** The 12 polymorphic SSR primers used for the analysis of 22 sesame accessions.

S/N	Forward	Reverse	Product size	Tm
OTO1	5'ATCCACGACAGCACACTCAG 3'	5'CCTCACTTTAGCCGGAAGT 3'	155	57.5
OTO2	5' ATCTTTAGCCCCGTTCTGGT 3'	5'TGCTCTGCCCTATCCTTCAT 3'	160	59
OTO3	5'TGTGATCTTTAGCCCCGTTCTG 3'	5'TGCTCTGCCCTATCCTTCAT 3'	118	58
OTO4	5'TGATCTTTAGCCCCGTTCTG 3'	5'TGCTCTGCCCTATCCTTCAT 3'	129	58
OTO5	5'TCAACAACAGCAGACCAGGG 3'	5'AAGTTTGAGACCCCGCAACA 3'	154	58
OTO6	5'TGGTCAACAACAGCAGACCA 3'	5'AAGAAGTTTGAGACCCCGCA 3'	160	57
OTO7	5'AGCATTATGGTCAACAACAGCA 3'	5'CTGCTGCGAGAAGTCTCT 3'	144	58
OTO8	5'CAGACCAGGGCGCCTCAC 3'	5'TTGAGACCCCGCAACAGGC 3'	156	60
OTO9	5'CTCGGCTTGCCAAGGAACA 3'	5'CATGTGCTCGTTCCTCCAGT 3'	146	59
OTO10	5'GACCTGAAACTGGCCAGAT 3'	5'GATGCTTGCTTTGCCCGTG 3'	155	59
OTO11	5'CAGGTGGCAATCTGGAATCT 3'	5'TTGGAGGGGTTCTTCTTTT 3'	150	59
OTO12	5'CTTTTCTCAACGATGCCACA 3'	5'CTCGTCTGGCTCGACATACA 3'	155	57

**Figure 1.** The phylogenetic analysis constructed based on SSR markers for 22 sesame accessions collected from 2 different ecological zones in Nigeria.

The total number of alleles identified at 12 primer loci across the 22 accessions of sesame is 47 (Table 3). These alleles ranged between 3 and 6 per locus. The average number of alleles per locus is 3.92. The loci with the highest number of alleles are found in OTO4 (6) and OTO10 (6). These two SSR loci with the highest variation of polymorphic motifs are best for discriminating the 22 accessions of sesame used in this study. In the same way, the frequency of the major alleles shared across the

22 accessions at each locus are between 27% (OTO4) and 77% (OTO2) covering about 50% (11 genotypes/accessions). A moderately high level of diversity was found among the 12 loci examined ranging between 0.38 and 0.79 with average of 0.62. The implication of the number of allele diversity and their frequency among the accessions is referred to as polymorphic information content (PIC) (Shahriar et al., 2014). PIC value of each marker is estimated on the

**Table 3.** Sesame microsatellite markers showing number of alleles, gene diversity and polymorphic information content (PIC) among 12 polymorphic markers for diversity studies.

Marker	Major allele frequency	Sample size	No. of obs.	Allele No.	Availability	Gene diversity	PIC
OTO1	0.5455	22.0000	22.0000	3.0000	1.0000	0.5950	0.5262
OTO2	0.7727	22.0000	22.0000	4.0000	1.0000	0.3843	0.3619
OTO3	0.5455	22.0000	22.0000	3.0000	1.0000	0.5826	0.5076
OTO4	0.2727	22.0000	22.0000	6.0000	1.0000	0.7934	0.7620
OTO5	0.5455	22.0000	22.0000	3.0000	1.0000	0.5620	0.4762
OTO6	0.4545	22.0000	22.0000	3.0000	1.0000	0.6405	0.5669
OTO7	0.5909	22.0000	22.0000	4.0000	1.0000	0.5826	0.5332
OTO8	0.5000	22.0000	22.0000	4.0000	1.0000	0.6405	0.5804
OTO9	0.5000	22.0000	22.0000	3.0000	1.0000	0.5744	0.4838
OTO10	0.3636	22.0000	22.0000	6.0000	1.0000	0.7479	0.7086
OTO11	0.4091	22.0000	22.0000	4.0000	1.0000	0.6983	0.6436
OTO12	0.5455	22.0000	22.0000	4.0000	1.0000	0.6116	0.5549
Mean	0.5038	22.0000	22.0000	3.9167	1.0000	0.6178	0.5588

basis of its alleles. There is significance in the variation of PIC for all the studied SSR loci. In this study, the level of polymorphism among the 22 sesame accessions was evaluated by calculating PIC values for each of the 12 SSR loci. The PIC values ranged from 0.36 (OTO4) to 0.76 (OTO2) with an average of 0.59 per locus (Table 3). The SSR primer loci with highest PIC is however more informative than what we had on the number of alleles. The PIC of OTO4 (0.76) is higher than OTO10 (0.70) which clearly revealed that marker OTO4 becomes the best locus for discriminating the 22 accessions among the used markers followed by OTO10, OTO11, OTO8 and so on.

Furthermore, the Unweighted Pair-Group Method with Arithmetic mean (UPGMA) revealed 3 major clusters and one outlier (Figure 1). Two genotypes are however very close with up to 97% similarity (NGB00397 and NG512174). Following the site of these samples collection, they both came from Ondo state, humid forest vegetation. Although, they differ phenotypically and were collected from different locations in the same ecological zone, it was inferred that they possess similar genetic characters. In addition to this, the dendrogram separated into two major clusters on the node 57% based on the different ecological distribution of the genotypes. One cluster is made of 13 genotypes and are found to be the collections from the drier vegetation, the derived Savanna. While the second cluster comprise of 9 genotypes collected from the humid forest vegetation. These studies also proved certain patterns of association between genetic similarity and geographical proximity in sesame as also reported by Dossa et al. (2017b). Furthermore, it had been reported that the frequency of cross-pollination in sesame could be as high as 60% (Dossa et al., 2016). Hence the percentage similarity of geographically-close accessions may be insignificant and non-divergent. This is a plausible reason why breeders

essentially must consider the geographical divergence of two parental genotypes if they really want to achieve any significant variation in their expected Sesame progenies which is similar to the report of Wei et al.(2014).

## CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

## ACKNOWLEDGEMENT

The authors wish to acknowledge the generous release of the germplasm of Sesame seeds from the seed genebank by the head of the Department and his staff who assisted in collection from the extremely cold vault.

## ABBREVIATIONS

**PIC**, Polymorphism information content; **CTAB**, cetyltrimethylammonium bromide; **RAPD**, randomly amplified polymorphic DNA; **EDTA**, ethylene di-amine tetra-acetate.

## REFERENCES

- Alemu A, Petros Y, Tesfaye K (2013). Genetic distance of sesame (*Sesamum indicum* L.) cultivars and varieties from northwestern Ethiopia using inter simple sequence repeat marker. East African Journal of Science 7(1):31-40.
- Anilakumar KR, Pal A, Khanum F, Bawa AS (2010). Nutritional, medicinal and industrial uses of sesame (*Sesamum indicum* L.) seeds-an overview. *Agriculturae Conspectus Scientificus* 75 (4):159-168.
- Badril J, Yepuri V, Ghanta A, Siva S, Siddiq EA (2014). Development of microsatellite markers in sesame (*Sesamum indicum* L.). *Turkish Journal of Agriculture* 38:603-614.
- Bedigian D (2010). Characterization of sesame (*Sesamum indicum* L.)

- germplasm: A critique. *Genetic Resources and Crop Evolution* 57:641-647.
- Bedigian D (2011). Cultivated sesame and wild relatives in the genus *Sesamum* L. in sesame: The Genus *Sesamum*. Medicinal and aromatic plants-industrial profiles; Bedigian, D., Ed.; CRC Press, Taylor and Francis Group: Boca Raton, FL, USA, pp. 33-77.
- Desjardins P, Conklin D (2010). Nanodropmicrovolume quantitation of nucleic acids. *Journal of Visualized Experiments* 45:e2565 DOI: 10.3791/2565.
- Dixit A, Jin MH, Chung JW, Yu JW, Kichung H, Ma KH, Park YP, Chi EG (2005). Development of polymorphic microsatellite markers in sesame (*Sesamum indicum* L.). *Molecular Ecology Notes* 5:736-738.
- Dossa K, Wei X, Zhang Y, Fonceka D, Yang W, Diouf D, Liao B, Cissé N, Zhang X (2016). Analysis of genetic diversity and population structure of sesame accessions from Africa and Asia as major centers of its cultivation. *Genes* 7(14):doi: 10.3390/genes7040014
- Dossa K, Yu J, Liao B, Cisse N, Zhang X (2017a). Development of highly informative genome-wide single sequence repeat markers for breeding applications in sesame and construction of a web resource: *Frontiers in Plant Science* 8:1470. doi: 10.3389/fpls.2017.01470
- Dossa K, Diouf D, Wang L, Wei X, Zhang Y, Niang M, Fonceka D, Yu J, Mmadi MA, Yehouessi LW, Liao B (2017b). The emerging oilseed crop *Sesamum indicum* enters the "omics" era. *Frontiers in Plant Science* 8:1154. doi: 10.3389/fpls.2017.01154
- Elleuch M, Besbes S, Roiseux O, Blecker C, Attia H (2007). Quality characteristics of sesame seeds and by-products. *Food Chemistry* 103:641-650.
- Shahriar MH, Robin AHK, Begum SN, Hoque A (2014). Diversity analysis of some selected rice genotypes through SSR- based molecular markers. *Journal of the Bangladesh Agricultural University* 12(2):307-311.
- Koppelman SJ, Söylemez G, Niemann L, Gaskin FE, Baumert JL, Taylor SL (2015). Sandwich enzyme-linked immunosorbent assay for detecting sesame seed in foods. *Hindawi Publishing Corporation BioMed Research International Article ID 853836*, 10 pages doi:10.1155/2015/853836.
- Nagy S, Hegedus G, Poczai P, Cernak I, Gorji AM, Taller J (2012). PIC calc: an online program to calculate polymorphic information content for molecular genetic studies. *Biochemical Genetics* 50:670-672. DOI 10.1007/s10528-012-9509-1
- Pandey SK, Das A, Rai P, Dasgupta T (2015). Morphological and genetic diversity assessment of sesame (*Sesamum indicum* L.) accessions differing in origin. *Physiology and Molecular Biology of Plants* 21(4):519-529
- Pham TD, Geleta M, Bui TM, Bui TC, Merker A, Carlsson AS (2011). Comparative analysis of genetic diversity of sesame (*Sesamum indicum* L.) from Vietnam and Cambodia using agro-morphological and molecular markers. *Hereditas* 148:28-35. DOI: 10.1111/j. 1601-5223.2010.02196.
- Singh KM, Kumar DB, Kumar DS, Manorama S (2015). Assessment of Genetic Diversity among Indian Sesame (*Sesamum indicum* L.) Accessions using RAPD, ISSR and SSR Markers. *Research Journal of Biotechnology* 10(8):35-47.
- Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan P B (2004). A new integrated genetic linkage map of the soybean. *Theoretical and Applied Genetics* 109:122-128.
- Uncu AO, Gultekin V, Allmer J, Fray A, Doganlar S (2017). Genomic simple sequence repeat markers reveal patterns of genetic relatedness and diversity in sesame. *The Plant Genome* 8(2):1-12. doi: 10.3835/plantgenome2014.11.0087.
- Vilas BL (2014). Polymorphic information content, transferability to other pulses and molecular diversity analysis in chickpea (*Cicer arietinum* L) using microsatellite markers. An unpublished Thesis from Department of Plant Biotechnology University of Agricultural Sciences, Bangalore-560065.
- Wang L, Yu J, Li D, Zhang X (2014). Sinbase: an integrated database to study genomics, genetics and comparative genomics in *Sesamum indicum*. *Plant and Cell Physiology* 56:e2. doi: 10.1093/pcp/pcu175
- Wei X, Wang L, Zhang Y, Qi X, Wang X, Ding X, Zhang J, Zhang X (2014). Development of simple sequence repeat (SSR) markers of sesame (*Sesamum indicum*) from a genome survey. *Molecules* 19:5150-5162 doi:10.3390/molecules19045150
- Welsing K, Nybom H, Wolff K, Kahl G (2005). DNA fingerprinting in plants: principles, methods and application. Taylor and Francis Group, USA. 444p.
- Woldesenbet DT, Tesfaye K, Bekele E (2005). Genetic diversity of sesame germplasm collection (*sesamum indicum* L): implication for conservation, improvement and use. *Physiology and Molecular Biology of Plants* 6(2):7-18
- Wu K, Yang M, Liu H, Tao Y, Mei J, Zhaol YZ (2014). Genetic analysis and molecular characterization of Chinese sesame (*Sesamum indicum* L.) cultivars using insertion-deletion (indel) and simple sequence repeat (SSR) markers. *BMC Genetics* 15:35
- Xin D, Sun J, Wang J, Jiang H, Hu G, Liu C, Chen Q (2012). Identification and characterization of SSRs from soybean (*Glycine max*) ESTs. *Molecular Biology Reports* 39:9047-9057.