Physiol Mol Biol Plants (September–October 2018) 24(5):929–937
https://doi.org/10.1007/s12298-018-0558-8

RESEARCH ARTICLE

GinMicrosatDb: a genome-wide microsatellite markers database for sesame (*Sesamum indicum* L.)

Supriya Purru[1] · Sarika Sahu[1] · Saurabh Rai[1] · A. R. Rao[1] · K. V. Bhat[2]

Received: 14 July 2017 / Revised: 8 March 2018 / Accepted: 21 May 2018 / Published online: 14 July 2018

Abstract Molecular breeding in sesame is still at infancy due to limited number of microsatellite markers available and the low level of polymorphism exhibited by them. Therefore, whole genome sequencing was used for development of microsatellite markers so as to ensure availability of substantial number of polymorphic markers for use in marker assisted breeding programs. Whole genome sequencing of sesame variety 'Swetha' was done using Illumina paired-end sequencing and Roche 454 shotgun sequencing technologies (GCA_000975565.1 in GenBank). 'GinMicrosatDb', a genome-wide microsatellite marker database has been developed using the whole genome sequence data of sesame variety 'Swetha'. The database consists of microsatellites localized on both linkage groups and scaffolds with their genomic co-ordinates. It provides five sets of forward and reverse primers for each of the microsatellite loci along with the flanking sequences, primer GC content, product size and melting temperature etc. The distribution of microsatellites can be viewed and selected through a genome browser as well as through a physical map. The newly identified microsatellite markers are expected to help sesame breeders in developing marker tags for traits of economic importance thereby bringing about greater efficiency in marker-assisted selection programs.

Keywords Linkage group · Microsatellites · Physical map · Sesame

Introduction

Sesame, (*Sesamum indicum* L.) is an ancient diploid (2n) dicotyledonous oil seed crop belonging to family Pedaliaceae. It is cultivated in the tropical and sub-tropical regions of the world by poor and marginal farmers (Janick 2008). Sesame, as a source of high quality oil is valued for its stability, nutritional value and resistance to rancidity and is often referred to as the "Queen of oil seeds" (Bedigian 2003, 2010). Sesame seeds are an important source of oil (44–58%), protein (18–25%) and carbohydrates (13.5%) (Bedigian et al. 1985). Natural antioxidants sesamin, sesamolin, tocopherol etc. make its economic value very high and these can be used effectively against the microorganisms (Wang et al. 2012). The worldwide diversity centres for sesame have been identified as India, Central Asia, China, Near East and Abyssinia (Laurentin and Karlovsky 2006). The sesame is cultivated by more than 50% of world population. Amongst the countries, India ranks first in area and production of sesame. Despite the low yield, sesame can improve the water percolation of the soil and can grow under high temperature and low rainfall.

The plant has tremendous nutritive value however, intensive studies on genetics and molecular mechanisms responsible for yield and adaptability traits are lacking. Hence, it is essential to explore the sesame plant with state of the art of sequencing technology to improve the quality

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12298-018-0558-8) contains supplementary material, which is available to authorized users.

✉ K. V. Bhat
KV.bhat@icar.gov.in; kvbhat2001@yahoo.com

[1] ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India

[2] Division of Genomic Resources, ICAR-National Bureau of Plant Genetic Resources, New Delhi 110012, India

and yield. Thus, the whole genome sequencing of sesame may revolutionize the sesame genomics by generating enormous sequence data, this may in turn help in exploring the full genetic potential of crop. Moreover, sequencing of whole genome may help identify novel genes controlling stress tolerance, metabolic pathways and traits of agronomic importance capable of changing the sesame crop architecture.

Although whole genome sequences of major crops like rice, maize, sorghum, tomato, melon, potato, grapes, pigeon pea etc. are available in the public databases; the sequence information about major oil seed crops is lacking. Genome sequence information is available for oil seeds like soy bean (Schmutz et al. 2010), jatropha (Sato et al. 2011), *Brassica rapa* (Wang et al. 2011) and castor (Chan et al. 2010). A draft genome assembly of Sesame was also reported by Zhang et al. (2013) and Wang et al. (2014).

The recent advances in genome sequencing especially the evolution of next generation sequencing techniques have resulted in the availability of millions of molecular markers covering the entire genome (Imelfort et al. 2009). The highly polymorphic, co-dominant, locus specific microsatellites have been the markers of choice for trait tagging, gene mapping, diversity analysis and cultivar identification. Molecular marker studies conducted on diversity of sesame germplasm revealed the low level of polymorphism existing in the present day cultivars (Isshiki and Umezaki 1997), using RAPD (Venkataramana Bhat et al. 1999), AFLP (Ali et al. 2007; Laurentin and Karlovsky 2006; Uzun et al. 2003), ISSR (Kim et al. 2002), SRAP (Zhang et al. 2010) etc. Although it has been tedious to identify and generate such markers using conventional microsatellite enriched genomic library approaches (Ostrander et al. 1992; Paetkau 1999), these markers become the ideal tool for linkage mapping and MAS. Hence, it is essential to develop microsatellite markers which ultimately lead to the development of a saturated linkage map facilitating Marker Assisted Selection (MAS) and introgression of genes and QTLs.

In the present study, we identified genome-wide microsatellite markers, localized them on to linkage groups and developed a database to facilitate selection of SSRs for various applications in sesame improvement including genetic diversity analyses, trait tagging, molecular mapping etc. This database gives the information of simple as well as compound microsatellites, their positions in the genome, primers, flanking sequences and the physical map.

Materials and methods

Genome sequencing and assembly

Whole genome sequencing of sesame variety ‘Swetha’ was performed using two DNA sequencing platforms namely, Roche 454 FLX and Hiseq1000 from Illumina/Solexa genome Analyser. The high quality Illumina mate-pair and paired-end reads obtained after quality filtration (adaptor trimming and removal of low quality reads—average quality value less than 20) were assembled with Kmer size 63 using Velvet short read de novo assembly program. The mate-pair library of 3, 5 kb and Roche 454 shotgun datasets were assembled using the CLC Genomics Workbench with default parameters. Finally, all the Scaffolds generated using Velvet (Zerbino and Birney 2008) pipeline and CLC genomics workbench were merged for hybrid assembly, which were further assembled using CAP3 (Huang and Madan 1999) assembler for obtaining the final draft assembly. The final draft assembly (Bioproject: PRJNA219369, Biosample Number SAMN02357081) was subjected to gap closure using Gap-Closer program which improved the genome coverage statistics.

Microsatellite discovery

The significant SSRs were located in linkage groups using MISA tool (http://pgrc.ipk-gatersleben.de/misa/) with the search criteria being the minimum frequency of repeats as 10, 6, 4, 3, 3 and 3 for mono, di, tri, tetra, penta and hexa nucleotide repeats respectively to confirm that the extracted repeats were real microsatellites (Asp et al. 2007; Kapil et al. 2014). Compound SSRs were located by considering 75 nucleotide intervening sequences between two microsatellite markers.

Primer design

In-house developed *perl* scripts were used to extract the flanking sequences and 200 bases of SSR flanking regions were considered for primer designing. Batchprimer3 (http://probes.pw.usda.gov/cgi-bin/batchprimer3/batchprimer3.cg) was used for designing primers with the criteria of GC content between 20 and 80%, primer length of 18–30 bp and annealing temperature between 57 and 63 °C.

Database development

A database on microsatellite markers of sesame crop named as GinMicrosatDb has been developed with several options like selection of markers with different search criteria, primers, flanking sequences, physical map and a genome browser. The web pages of GinMicrosatDb were built using Java Server Pages. This is a 3-tier application which completely follows Model-View-Controller (MVC) architecture. The user interface was developed using HTML, CSS, JSP and JavaScript. Server side application layer was developed using Servlet and JSP which run on a windows system and are powered by Apache Tomcat 7.0 server. All data were stored in a MySQL 5.1.34 database. Hibernate framework was used to interact with the database. MySQL database was used to connect with the Java application. Few design patterns were used viz., Singleton Design Pattern, Factory Design Pattern, DAO Design Pattern, Proxy Design pattern to solve the recursion problem and increase the performance of application.

Physical map and genome browser

We also developed physical map of SSRs based on physical distances between markers for each linkage group using MapChart (Voorrips 2002). A linkage group chart consists of a vertical bar on which the map positions and SSRs are indicated. A Lightweight Genome Viewer (LWGV) tool (Faith et al. 2007) has been used to develop genome browser that visualizes microsatellites on the linkage groups and color-coded as tracks on rectangular bars of linkage groups. The detailed information about each SSR has been shown by "mouse-over" facility.

Results and discussion

Data analysis

The high quality reads obtained from Roche 454 (~ 509 Mb) and Illumina (~ 22 GB) were used for assembly and the resulted contigs were further assembled using CAP3 assembler which resulted in the final draft assembly. The final draft assembly was subjected to gap closure using GapCloser program which improved the genome coverage. The total number of scaffolds generated in the draft assembly was 76,029. After the gap closure, the total genome length including the gaps was observed to be 340,643,414 bp while without gaps was 339,724,207 bp. Average scaffold size with gaps was found to be 4480 bp and without gaps was 4468 bp. Scaffold N50 value is 22,205 bp. Maximum scaffold size is 263,158 bp whereas minimum scaffold size is 927 bp. The assembly statistics are given in S1 Table. Finally, these scaffolds were reduced to 16 linkage groups by taking Chinese Sesame whole genome (Zhang et al. 2013) as reference.

Assembly comparison with the reported genome assembly

The genome sequencing of Yuzhi 11 variety of *Sesamum indicum* (Zhang et al. 2013) resulted in 98 Gb of data and provided 276× coverage of the estimated genome whereas the Swetha variety resulted in 30 Gb data and provided 85× coverage using different NGS technologies. The Yuzhi 11 genome assembly reported was of length 293.7 Mb, with an estimated coverage of 82.9% whereas the genome length covered by Swetha was ~ 340 Mb, along with estimated genome coverage of 96.32%. The scaffold N50 observed for Yuzhi 11 was 22.6 kb and that of Swetha was 22.2 kb.

Microsatellite mining

All 16 linkage groups were examined for the presence of microsatellites and microsatellites were found to be present on every linkage group. The total number of SSRs identified from the sequenced sesame genome was 118,004; of which 21,704 were present in compound form.

Overall, the distribution of di, tri, tetra, penta and hexa nucleotide repeats was 33, 16, 23, 5, and 3% respectively. The distribution of microsatellites in linkage groups is shown in Fig. 1 which is developed through Circos software (Krzywinski et al. 2009). The highest number of SSRs were located in linkage group 3 with the distribution of 38, 20, 31, 7, and 4% for di, tri, tetra, penta and hexa nucleotides respectively. The least number of SSRs were found in linkage group 16 with the distribution of 40, 18,

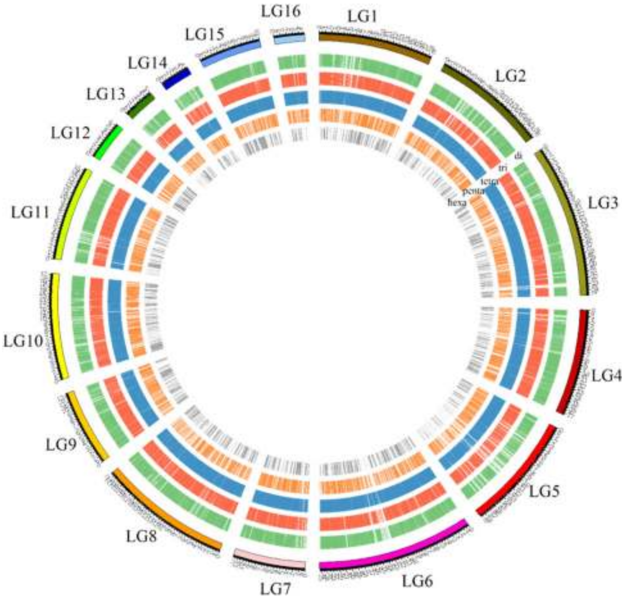


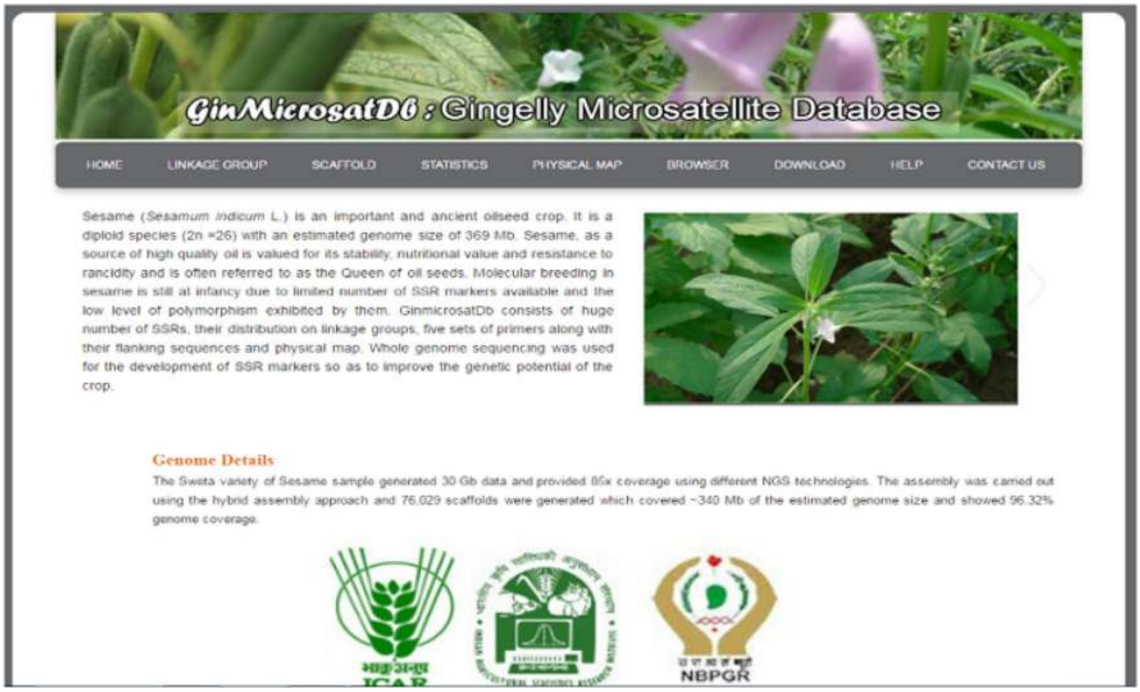
Fig. 1 Distribution of microsatellites in linkage groups

30, 9, and 3% for di, tri, tetra, penta and hexa nucleotides respectively.

However, the pattern of distribution of different types of SSRs is almost similar in all linkage groups ranging from 35–40%, 18–21%, 30–32%, 7–9%, 3–6% for di, tri, tetra, penta and hexa nucleotide repeats respectively. The frequency of occurrence of di, tri, tetra nucleotide repeats is more abundant in linkage group 3 whereas penta and hexa nucleotide repeats were more frequent in linkage group 6. Among the different repeat types, di-nucleotides were most common (39,248) followed by tetra and tri-nucleotides repeats (27,499 and 19,090 respectively). The least frequent were hexa (3648) and penta (6815) nucleotide repeats. Di-nucleotide repeats were found to be more abundant in other plant species like Arabidopsis, peanut, sugar beet, cabbage soy bean, pea, sunflower and grape (Kumpatla and Mukhopadhyay 2005) while in cereals like rice, wheat and barley (La Rota et al. 2005) tri-nucleotide repeats were predominant. Among di-nucleotide repeats, AT (36%) repeat is more prevalent followed by TA (26%) and CG is least repeated. The length of AT repeat ranges from a minimum of 12 to a maximum of 56 nucleotides with the frequency of 6 and 28 respectively. The ATT repeat was more abundant with length ranging from 12 to 63. The repeats AAAT, AAAAT, AAAAAT were most frequent among tetra, penta, hexa nucleotides with the length ranging from 12–24, 15–25, and 18–30 respectively. Most of the SSRs (99%) have length less than 50 base pair while very few (0.23%) have length greater than 100 base pair. Compound SSRs account for nearly 12.7%. Among compound repeats, the highest number of repeats were located in linkage group 3 followed by linkage group 6 and 8.

Database utility

User can retrieve the information according to requirement based on the various options available in the database (Fig. 2). User can search SSRs using multiple searching criteria (Fig. 3). User can type LG1 to LG16 in order to see SSRs present on particular linkage group. The user can give repeat length from 1 to 6 (mono to hexa) and frequency of repeats as required by the user and can even type a particular repeat of his/her interest in the space provided. The database consists of information of SSRs viz., the location of SSRs on linkage groups, the frequency of repeats, primers and flanking sequences (Fig. 4). Five sets of primers are provided in the database for the convenience of user. The primer information consists of forward and reverse primer sequences, GC content, melting temperature, length of primer and product size (Fig. 5). The flanking sequences of SSRs are also provided in the database. The primer information consists of SSR id, the coordinates of flanking sequence on linkage groups/scaffolds, left and right flanking sequences (Fig. 6). Genome browser (Fig. 7) shows the distribution of SSRs in all 16

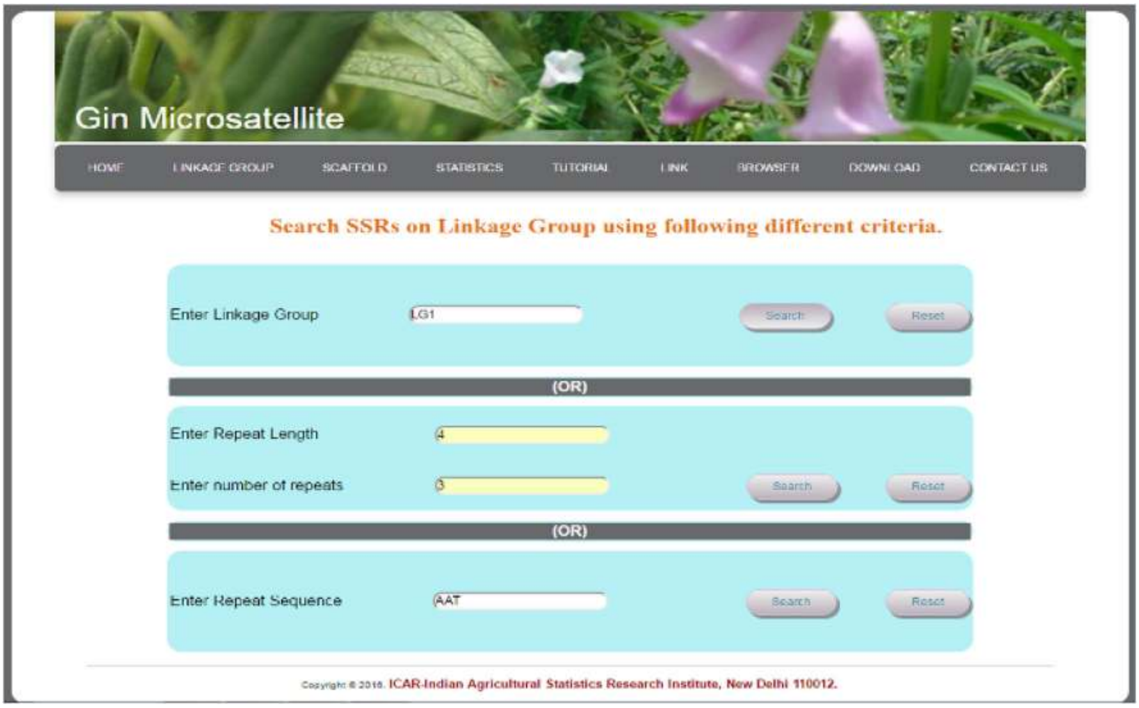


Sesame (*Sesamum indicum* L.) is an important and ancient oilseed crop. It is a diploid species (2n =26) with an estimated genome size of 369 Mb. Sesame, as a source of high quality oil is valued for its stability, nutritional value and resistance to rancidity and is often referred to as the Queen of oil seeds. Molecular breeding in sesame is still at infancy due to limited number of SSR markers available and the low level of polymorphism exhibited by them. GinmicrosatDb consists of huge number of SSRs, their distribution on linkage groups, five sets of primers along with their flanking sequences and physical map. Whole genome sequencing was used for the development of SSR markers so as to improve the genetic potential of the crop.

Genome Details

The Sweta variety of Sesame sample generated 30 Gb data and provided 85x coverage using different NGS technologies. The assembly was carried out using the hybrid assembly approach and 76,029 scaffolds were generated which covered ~340 Mb of the estimated genome size and showed 96.32% genome coverage.

Fig. 2 The home page of GinMicrosatDb. It provides access to different information viz. microsatellites localized on linkage groups and scaffolds, physical map, browser and also tutorial



Search SSRs on Linkage Group using following different criteria.

Enter Linkage Group LG1

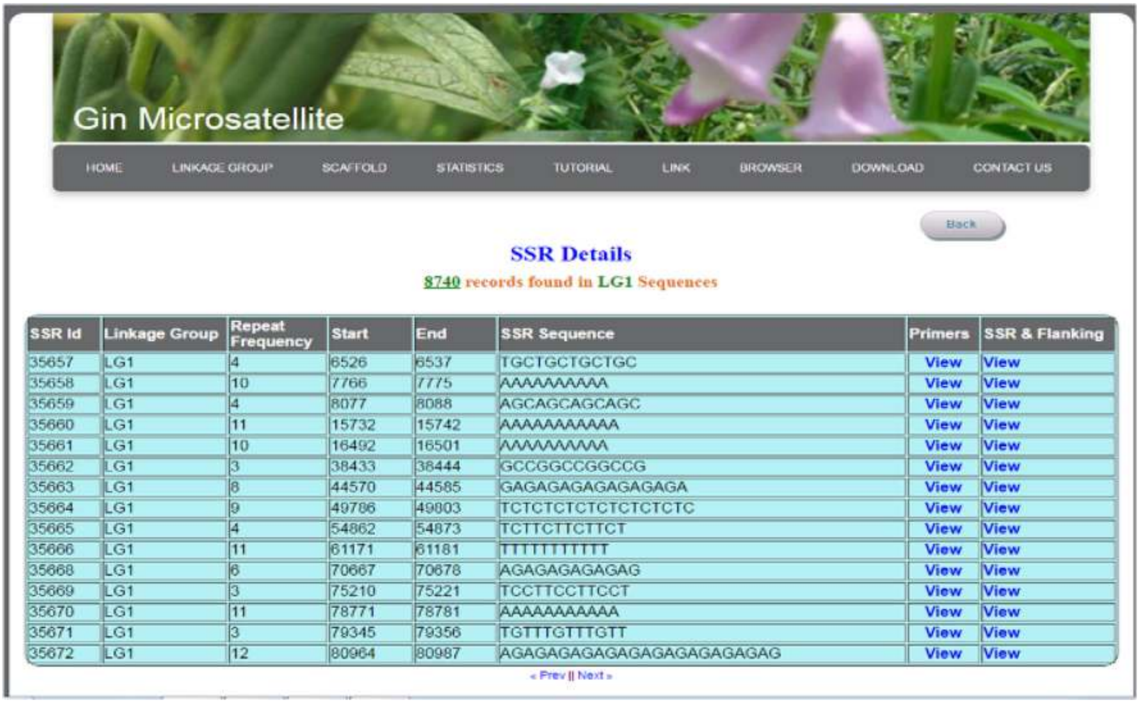
Enter Repeat Length 4

Enter number of repeats 3

Enter Repeat Sequence AAT

Copyright © 2016. ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110012.

Fig. 3 Search for SSRs by linkage group or repeat length and frequency or repeat sequence



SSR Details

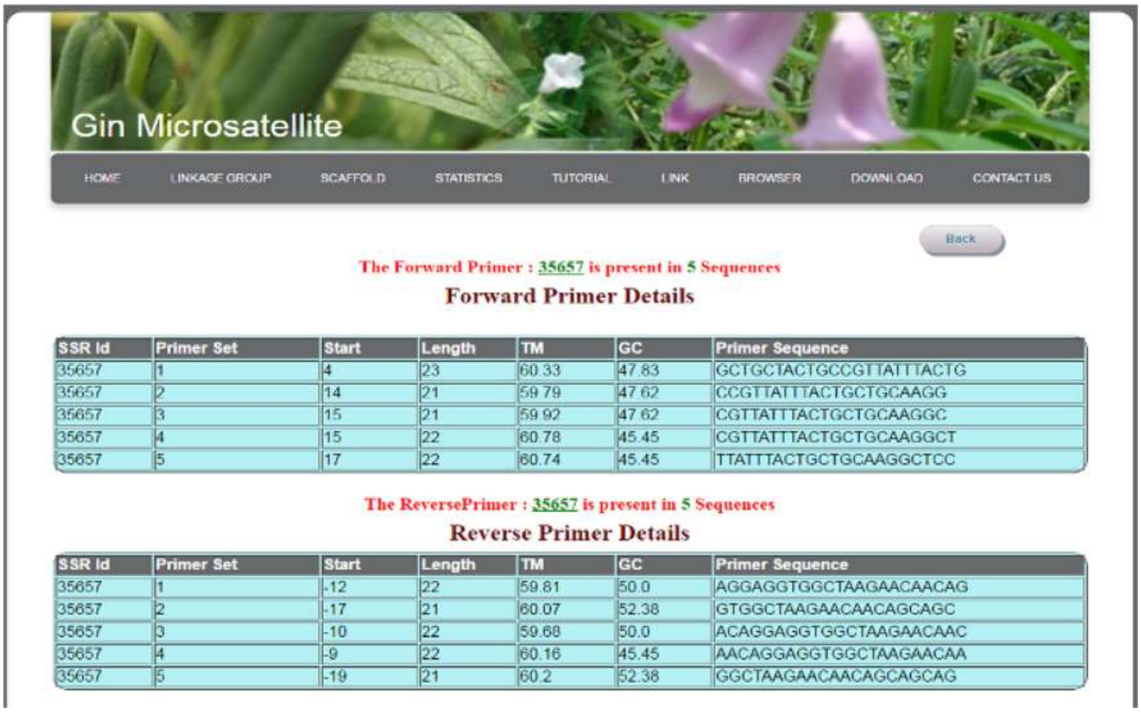
8740 records found in LG1 Sequences

SSR Id	Linkage Group	Repeat Frequency	Start	End	SSR Sequence	Primers	SSR & Flanking
35657	LG1	4	6526	6537	TGCTGCTGCTGC	View	View
35658	LG1	10	7766	7775	AAAAAAAAAA	View	View
35659	LG1	4	8077	8088	AGCAGCAGCAGC	View	View
35660	LG1	11	15732	15742	AAAAAAAAAAA	View	View
35661	LG1	10	16492	16501	AAAAAAAAAA	View	View
35662	LG1	3	38433	38444	GCCGGCCGGCCG	View	View
35663	LG1	8	44570	44585	GAGAGAGAGAGAGAGA	View	View
35664	LG1	9	49786	49803	TCTCTCTCTCTCTCTCTC	View	View
35665	LG1	4	54862	54873	TCTTCTTCTTCT	View	View
35666	LG1	11	61171	61181	TTTTTTTTTTT	View	View
35668	LG1	6	70667	70678	AGAGAGAGAGAG	View	View
35669	LG1	3	75210	75221	TCCTTCCTTCCT	View	View
35670	LG1	11	78771	78781	AAAAAAAAAAA	View	View
35671	LG1	3	79345	79356	TGTTTGTTTGTT	View	View
35672	LG1	12	80964	80987	AGAGAGAGAGAGAGAGAGAGAGAG	View	View

Fig. 4 SSR details. This page shows the details of microsatellites present in particular linkage group viz. frequency of the particular repeat, their genomic coordinates, SSR sequence and also provides access to view primers and flanking sequences

linkage groups and user can access the details of SSRs with the link provided.

The earlier transcriptome sequencing studies in sesame resulted in generation of sequences of length 47.99 Mb (Zhang et al. 2011) and 54.25 Mb (Wei et al. 2011). The present study is first of its kind to our limited knowledge that resulted in the discovery of 118,004 SSR markers in 16 linkage groups in which frequency of occurrence of SSRs



The Forward Primer : 35657 is present in 5 Sequences

Forward Primer Details

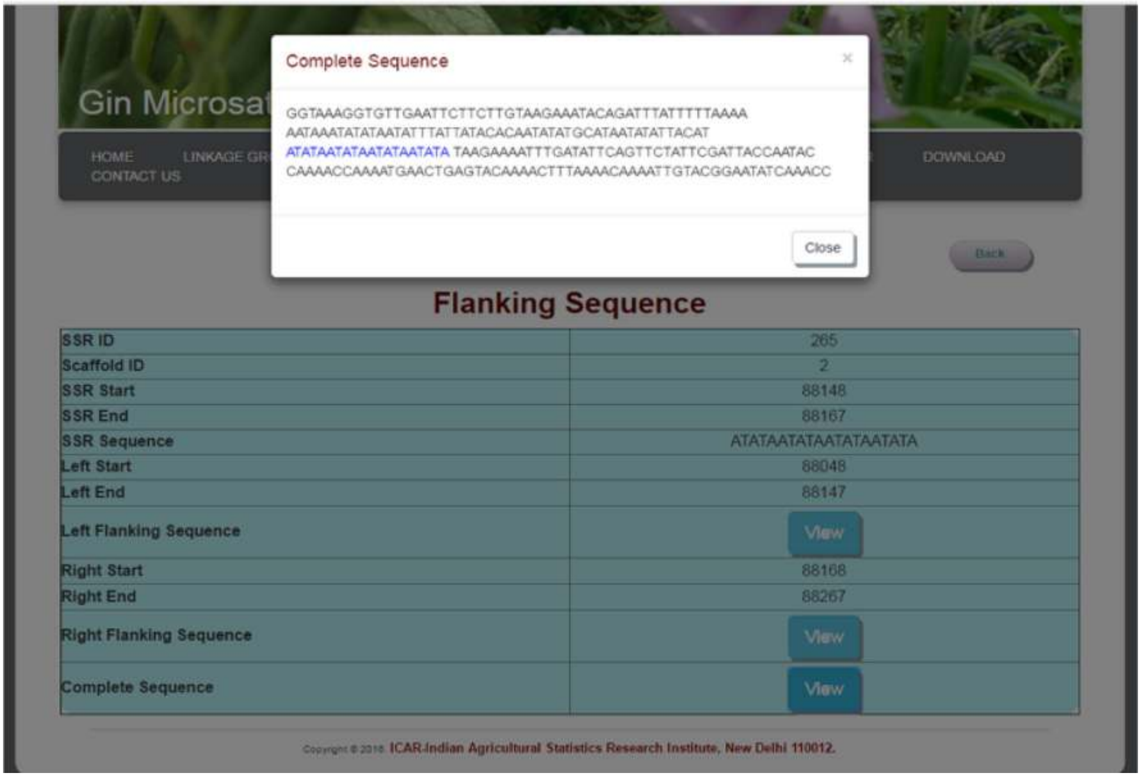
SSR Id	Primer Set	Start	Length	TM	GC	Primer Sequence
35657	1	4	23	60.33	47.83	GCTGCTACTGCCGTTATTTACTG
35657	2	14	21	59.79	47.62	CCGTTATTTACTGCTGCAAGG
35657	3	15	21	59.92	47.62	CGTTATTTACTGCTGCAAGGC
35657	4	15	22	60.78	45.45	CGTTATTTACTGCTGCAAGGCT
35657	5	17	22	60.74	45.45	TTATTTACTGCTGCAAGGCTCC

The ReversePrimer : 35657 is present in 5 Sequences

Reverse Primer Details

SSR Id	Primer Set	Start	Length	TM	GC	Primer Sequence
35657	1	-12	22	59.81	50.0	AGGAGGTGGCTAAGAACAACAG
35657	2	-17	21	60.07	52.38	GTGGCTAAGAACAACAGCAGC
35657	3	-10	22	59.68	50.0	ACAGGAGGTGGCTAAGAACAAC
35657	4	-9	22	60.16	45.45	AACAGGAGGTGGCTAAGAACAA
35657	5	-19	21	60.2	52.38	GGCTAAGAACAACAGCAGCAG

Fig. 5 This page shows the details of primers viz. GC content, Melting temperature, primer sequence



Complete Sequence

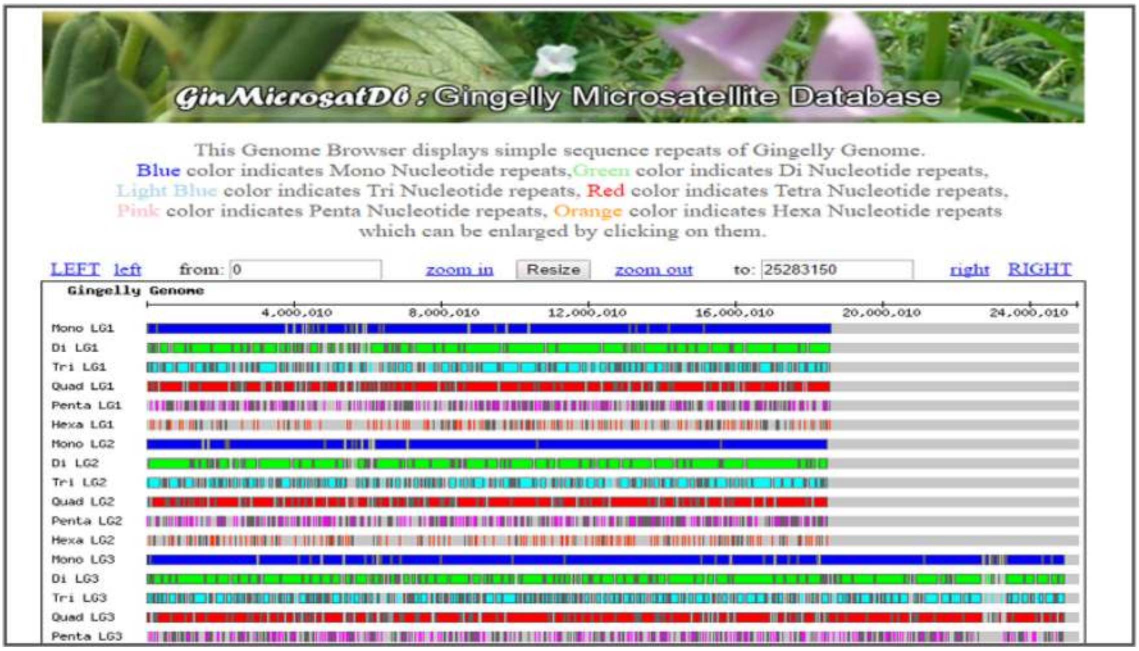
GGTAAAGGTGTTGAATTCTTCTTGTAAGAAATACAGATTTATTTTTAAAA
AATAAATATATAATATTTATTATACACAATATATGCATAATATATTACAT
ATATAATATAATATAATATA TAAGAAAATTTGATATTCAGTTCTATTCGATTACCAATAC
CAAAACCAAAATGAACTGAGTACAAAACTTTAAAACAAAATTGTACGGAATATCAAACC

Flanking Sequence

SSR ID	265
Scaffold ID	2
SSR Start	88148
SSR End	88167
SSR Sequence	ATATAATATAATATAATATA
Left Start	88048
Left End	88147
Left Flanking Sequence	View
Right Start	88168
Right End	88267
Right Flanking Sequence	View
Complete Sequence	View

Copyright © 2016 ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110012.

Fig. 6 This page shows the details of flanking sequence, their coordinates on genome



GinMicrosatDb : Gingelly Microsatellite Database

This Genome Browser displays simple sequence repeats of Gingelly Genome.
Blue color indicates Mono Nucleotide repeats, Green color indicates Di Nucleotide repeats,
Light Blue color indicates Tri Nucleotide repeats, Red color indicates Tetra Nucleotide repeats,
Pink color indicates Penta Nucleotide repeats, Orange color indicates Hexa Nucleotide repeats
which can be enlarged by clicking on them.

Fig. 7 The genome browser of GinMicrosatDb. It displays microsatellites localized on linkage groups. Different types of repeats are displayed in different colors for easy identification

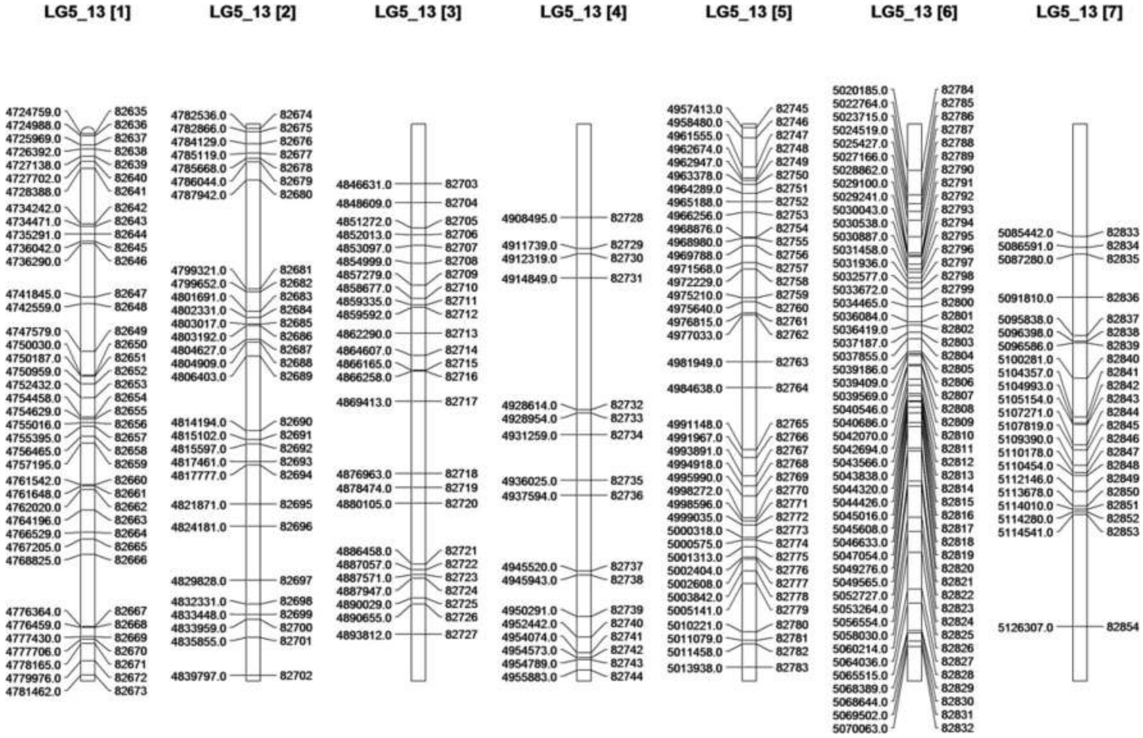


Fig. 8 A part of physical map of linkage group 5

was one for every 1.76 kb length of genome. The distribution of SSRs varies from one linkage group to other, however, the density is almost similar in all linkage groups ranging from 1.64 to 1.90 kb. Higher density was observed in linkage group 5 (Fig. 8) followed by 6 in which, one SSR was present for every 1.9 and 1.84 kb length of genome respectively. Lowest density was observed in linkage group 15 (Fig. 9) with one SSR present per 1.64 kb. However, it may be noted that there are no substantial differences in frequency of occurrence of SSRs across the linkage groups in sesame.

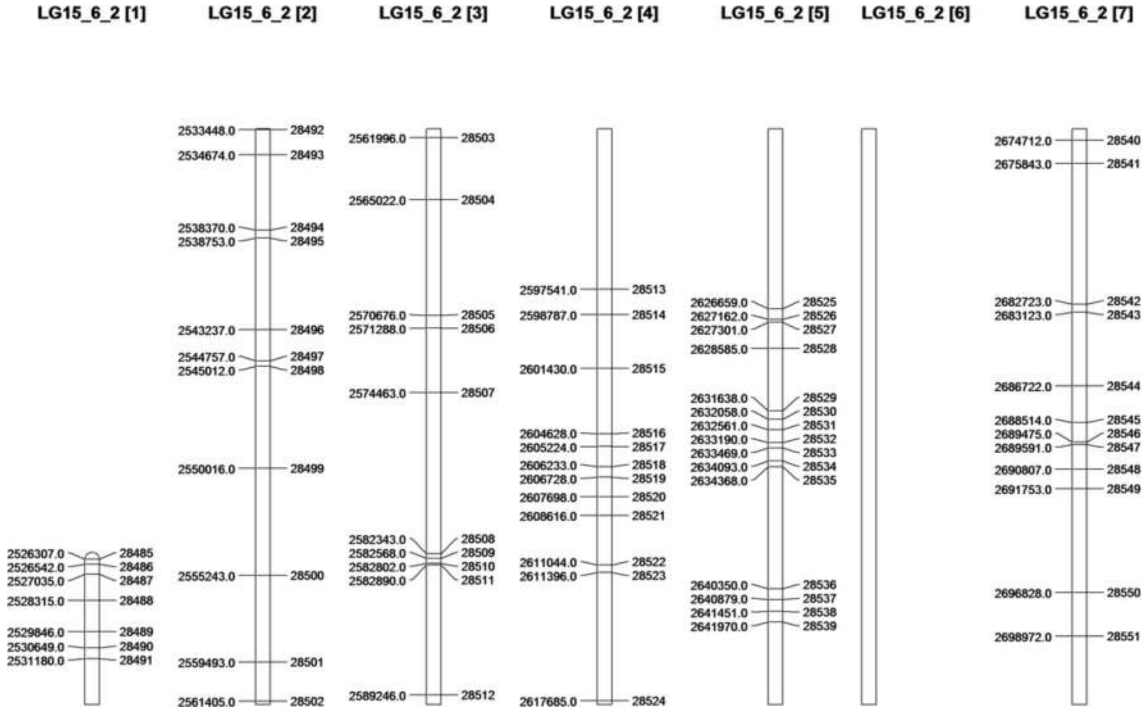


Fig. 9 A part of physical map of linkage group 15

Conclusion

GinMicroSatDb is freely available through http://backwin.cabgrid.res.in:8080/Gingelly7. The database is useful for selecting unlinked markers across linkage groups so as to ensure their random distribution which generally reduces the biases in estimation of genetic diversity and relatedness of genotypes. In future, it will be integrated with ESTs and QTL data. The database can be used effectively for identifying putative genome sequences of significance in crop improvement programs and biotechnological interventions.

Acknowledgements This study was funded by ICAR-National Agricultural Innovation Project (No. 415501-05) and Consortium Research Platform on Genomics. The facilities provided by Indian Council of Agricultural Research, NBPGR and ICAR-IASRI are acknowledged.

Compliance with ethical standards

Conflict of interest None declared.

References

Ali GM, Yasumoto S, Seki-Katsuta M (2007) Assessment of genetic diversity in sesame (*Sesamum indicum* L.) detected by amplified fragment length polymorphism markers. Electron J Biotechnol 10:12–23

Asp T, Frei UK, Didion T, Nielsen KK, Lubberstedt T (2007) Frequency, type, and distribution of EST-SSRs from three genotypes of *Lolium perenne*, and their conservation across orthologous sequences of *Festuca arundinacea*, *Brachypodium distachyon*, and *Oryza sativa*. BMC Plant Biol 7(1):36

Bedigian D (2003) Evolution of sesame revisited: domestication, diversity and prospects. Genet Resour Crop Evol 50:779–787

Bedigian D (2010) Characterization of sesame (*Sesamum indicum* L.) germplasm: a critique. Genet Resour Crop Evol 57:641–647

Bedigian D, Seigler DS, Harlan JR (1985) Sesamin, sesamolin and the origin of sesame. Biochem Syst Ecol 13:133–139

Chan AP et al (2010) Draft genome sequence of the oilseed species *Ricinus communis*. Nat Biotechnol 28:951–956

Faith JJ, Olson AJ, Gardner TS, Sachidanandam R (2007) Lightweight genome viewer: portable software for browsing genomics data in its chromosomal context. BMC Bioinform 8:344

Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. Genome Res 9:868–877

Imelfort M, Duran C, Batley J, Edwards D (2009) Discovering genetic polymorphisms in next generation sequencing data. Plant Biotechnol J 7:312–317

Isshiki S, Umezaki T (1997) Genetic variations of isozymes in cultivated sesame (*Sesamum indicum* L.). Euphytica 93:375–377

Janick J (2008) Plant breeding reviews, vol 30. Wiley, London

Kapil A, Rai PK, Shanker A (2014) ChloroSSRdb: a repository of perfect and imperfect chloroplastic simple sequence repeats (cpSSRs) of green plants. Database 2014:bau107

Kim DH, Zur G, Danin Poleg Y, Lee SW, Shim KB, Kang CW, Kashi Y (2002) Genetic relationships of sesame germplasm collection as revealed by inter simple sequence repeats. Plant Breed 121:259–262

Krzywinski M et al (2009) Circos: an information aesthetic for comparative genomics. Genome Res 19:1639–1645

Kumpatla SP, Mukhopadhyay S (2005) Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. Genome 48:985–998

La Rota M, Kantety RV, Yu J-K, Sorrells ME (2005) Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. BMC Genom 6:23

Laurentin HE, Karlovsky P (2006) Genetic relationship and diversity in a sesame (*Sesamum indicum* L.) germplasm collection using amplified fragment length polymorphism (AFLP). BMC Genet 7:10

Ostrander EA, Jong PM, Rine J, Duyk G (1992) Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences. Proc Natl Acad Sci USA 89:3419–3423

Paetkau D (1999) Microsatellites obtained using strand extension: an enrichment protocol. Biotechniques 26(690–692):694–697

Sato S et al (2011) Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. DNA Res 18:65–76

Schmutz J et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Uzun B, Lee D, Donini P, ÇAğirgan ML (2003) Identification of a molecular marker linked to the closed capsule mutant trait in sesame using AFLP. Plant Breed 122:95–97

Venkataramana Bhat K, Babrekar PP, Lakhanpaul S (1999) Study of genetic diversity in Indian and exotic sesame (*Sesamum indicum* L.) germplasm using random amplified polymorphic DNA (RAPD) markers. Euphytica 110:21–34

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:77–78

Wang X et al (2011) The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet 43:1035–1039

Wang L, Zhang Y, Li P, Wang X, Zhang W, Wei W, Zhang X (2012) HPLC analysis of seed sesamin and sesamolin variation in a sesame germplasm collection in China. J Am Oil Chem Soc 89:1011–1020

Wang L, Yu S, Tong C, Zhao Y, Liu Y, Song C, Li D (2014) Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. Genome Biol 15(2):R39

Wei W, Qi X, Wang L, Zhang Y, Hua W, Li D, Lv H, Zhang X (2011) Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. BMC Genomics 12:451

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829

Zhang Y, Zhang X, Hua W, Wang L, Che Z (2010) Analysis of genetic diversity among indigenous landraces from sesame (*Sesamum indicum* L.) core collection in China as revealed by SRAP and SSR markers. Genes Genom 32:207–215

Zhang T, Zhang X, Hu S, Yu J (2011) An efficient procedure for plant organellar genome assembly, based on whole genome data from the 454 GS FLX sequencing platform. Plant Methods 7:38

Zhang H, Miao H, Wang L, Qu L, Liu H, Wang Q, Yue M (2013) Genome sequencing of the important oilseed crop *Sesamum indicum* L. Genome Biol 14:401