



Bioscene

Bioscene

Volume- 21 Number- 03

ISSN: 1539-2422 (P) 2055-1583 (O)

www.explorebioscene.com

A Web Database & Similarity Search Tool for Genomic & Transcriptomic Markers of Horsegram (*Macrotyloma Uniflorum*)

Rakesh Kumar Chahota & Pawan Kumar Sharma

Distributed Information SubCentre, Department of Agricultural Biotechnology, Himachal Pradesh Agricultural University, Palampur, Himachal Pradesh, India

Corresponding Author: **Pawan Kumar Sharma**

Abstract

Problem: Horsegram is one of the least exploited legumes for research purpose and considered as a highly drought tolerant crop. It is considered to be an astounding source for elucidation and understanding the genetic basis of responses to drought tolerance. **Approach:** The genomic and transcriptomic SSR markers serve as wonderful resources for linkage mapping and detection of quantitative trait loci & most widely used markers for genotyping of plants over the recent past due to their properties viz. codominant, multi-allelic, experimentally reproducible and cross transferability among related crop species. Therefore, to provide such information in a single platform to the research community, we created an open web database and tool for finding SSRs through sequence input named Horsegram SSR markers database and Horsegram SSR finder respectively, available at (hillagric.ac.in:1005). **Findings:** The database contains about 9200 SSR markers of horsegram out of which 2341 are genomic and 6762 are of transcriptomic origin, developed from sequencing data available with DISC- Distributed Information Sub-Centre, Department of Agricultural Biotechnology, Chaudhary Sharvan Kumar Agricultural University, Palampur, Himachal Pradesh, India. The Horsegram SSR markers database contains the information of markers which can be viewed by creating a login ID through the website (hillagric.ac.in:1005), it will retrieve the all SSR genomic and transcriptomic sequences under separate tabs with detailed information such as sequence of the marker, name of the marker, repeat unit, product size and melting temperature. **Conclusion:** This sequence information can be easily accessed via website and surely provide a better search option at single platform rather than finding the data from different websites and to generate SSRs through different softwares and tools.

Keywords: Horsegram, database, SSR, genomic, transcriptomic.

Abbreviation: T_m= melting temperature, SSR= simple sequence repeat

Introduction

Genomic databases store datasets related to the genomic sequences of different organisms & gene annotations. Contrarily from gene databases, containing only coding DNA sequences, genomic databases contain also non-coding intergenic sequences. Genomic databases are listed among the data resources useful in systems biology. We have developed a marker database for horsegram a highly drought tolerant legume. Therefore, horsegram is considered to be a wonderful source for elucidation and understanding the genetic basis of responses to drought tolerance (Bhardwaj et al., 2013; Aditya et al., 2019). Furthermore, till date there is no data related to global size genomic, transcriptomic or protein biology studies on horsegram. Information over horsegram genetic resources is scarce compared to other plants (Bhardwaj et al., 2013). Transcriptome is coding region of the mRNA set derived from a genome (Bouck & Vision, 2007). Massively parallel sequencing includes next generation sequencing techniques like 454 pyrosequencing/Roche, Illumina/SolexaGAIIx, ABI/SOLiD, Pac Biosciences/PacBioRS & Helicos Biosciences/tSMS & DRS (Santos et al., 2012). These techniques do not require prior erudition of genomic sequence and are much advanced in terms of time, cost, labour, amount of data produced, data coverage, sensitivity and accuracy as compared to the orthodox sequencing methods (Wang & Messing 2011; Fang et al., 2012). Genomic and transcriptomic SSR markers are key to develop the linkage maps and to detect QTLs or genomic regions, which could be targeted for improvement and desired manipulations or to reveal the possible mechanism for a particular trait in the crops under study (Shirasawa et al., 2014; Mardis, 2008; Ellegren, 2008). To address fundamental questions like mechanisms involved or related to a particular trait like drought tolerance, it becomes essential to draw conclusion based on a comparative study (Ashraf, 2010). The development of genetic resource for depauperate plants like horsegram facilitated functional characterization of transcripts responsive to induced drought stress conditions. Creating genetic resources regarding GC content, SSR markers, genes, pathways and transcription factors associated with horse gram would boost the related research programs.

The name *Macrotyloma* is derived from the Greek words macros=large, tylous = knob, & loma = margin, in reference to knobby structures on the pods (Blumenthal and Staples, 1993). Horsegram is a diploid plant species, chromosome number varies: 2n = 20, 22 & 24. Size of horsegram genome is about 400 Mbps (Bhardwaj et al., 2013). There are nearly 25 species of

Genus *Macrotyloma* indigenous to Africa & Asia, out of those var. *uniflorum* is the only cultivated species (Allen and Allen, 1981). In India, horsegram is cultivated in Andhra Pradesh, Karnataka, Tamil Nadu, north-western and central Himalayan regions of Himachal Pradesh, Jammu and Kashmir, Uttarakhand along with Punjab, Bihar, Uttar Pradesh, Madhya Pradesh, Rajasthan, Maharashtra and Gujarat during summer season. Applications involves the use of crop as fodder, rich in protein content & free from digestive inhibitors; widely used as a feed for milch animals and horses (Prakash et al., 2008). It is highly nutritious, medicinal important and indomitable pest resistant therefore it is a sustainable source of food, fodder, fuel supplement and green manure. Cultivated horsegram seeds exhibit protein content in a range of 16.9– 30.4% (Prakash et al., 2008). It has high lysine content, an essential amino acid. Horsegram seeds are a rich source of phosphorus, iron and vitamins such as carotene, thiamine, riboflavin, niacin and vitamin C (Ramesh et al., 2011). It contains many medicinal and therapeutically active components, therefore considered as an Ayurvedic medicine for a variety of human ailments such as edema, piles and renal stones. It exhibits high antioxidant properties and contains molybdenum, which regulates calcium and iron, which helps in transporting oxygen to cells and forms part of haemoglobin in blood (Sudha and Saral, 2023; Murthy et al., 2012). Horsegram also helps in lowering blood cholesterol level (Mehra and Upadhyaya, 2013). Chaitanya et al., (2010) proved that the seeds of *M. uniflorum* are endowed with significant anti-urolithiatic activity.

Materials & methods

The Horsegram Marker Database & SSR finder tool was created using Hewlett Packard 2012 R2 Window based Server as the computer operating system. The MySQL (dev.mysql.com) system, which is a relational database management system, was employed for management of the database contents.

Database for 9200 Genomic & transcriptomic SSR markers developed at DISC- Distributed Information Sub-Centre, Department of Agricultural Biotechnology, Chaudhary Sarvan Kumar Agricultural University, Palampur, Himachal Pradesh, India & can be accessed through DISC website www.hillagric.ac.in:1005. The database developed in MySQL using HTML & PHP version 7.2 in & apache 2.4 based server running on localhost via WampServer3.1. Similarly, SSR finder tool has been developed using the above mentioned techniques along with JavaScript functions. Validation has been carried out in PHP through Client-Side Validation & Server Side Validation. The interface of the database named Horsegram SSR markers

database involves a sign up form for new users & login form for registered users. The database menu shows Change password, Genomic SSRs & transcriptomic SSRs buttons. The search menu needs a query in marker format i.e. ATGCCCGTG of at least 12 bases for finding SSRs in the database.

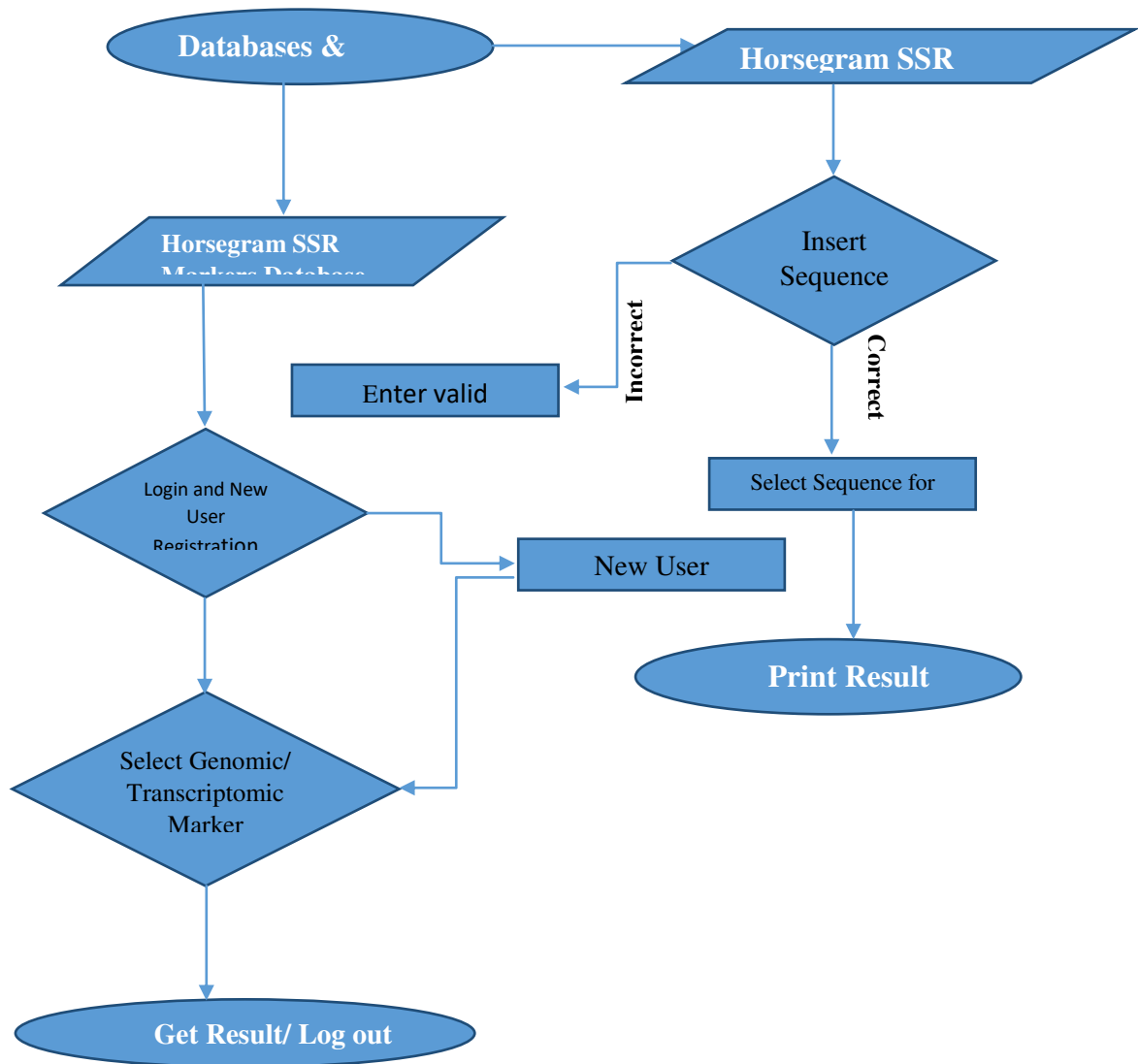
Results

The SSR markers can be classified into two categories, genome-SSR & expressed sequence tag (EST)-SSR. The genome-SSR markers are developed from random genome sequences from, for example, SSR enrichment genomic libraries, while EST-SSR markers are from sequences of cDNAs. *In silico* analyses, SSR motifs are identified from the sequence data with the SSRIT (Temnykh et al., 2001), MISA (Theil et al., 2003), &/or SciRoKo (Kofler et al., 2007) programs, & PCR primers designed on the flanking sequences of the SSR motifs using the Primer3 program (Rozen & Skaletsky, 2000; Kaldate et al., 2017). The identified SSR motifs, repeat numbers of the motif, PCR primers, & expected amplicon sizes are also available from the database.

How to use Horsegram Marker Database?

The home page of the DISC, CSKHPKVwebsite representssix buttons viz. Home, About, Staff, Gallery, Database & Tools and Contact Us. By clicking Database and Tools section-Horsegram Marker Database represents Genomic & Transcriptomic SSR markers registered in this database (see flowchart).

Horsegram Database Flowchart



Users can click “Horsegram SSR Markers Database”, to access page, register to create a login ID & password. After logging in, the main page of the database appears representing Genomic & Transcriptomic marker heads along with image icons. User can access the database by clicking either the image or options present in Menu bar. Under the Genomic SSRs or Transcriptomic SSRs button/ images the table shows S. no., sequence ID, repeat unit, forward primer, Tm, reverse primer, Tm, & product size in base pairs can be accessed or searched as described in the below section & Table 1,2.

S. No	SEQ ID	FP	Tm (°C)	RP	Tm (°C)	Size (bp)
1	MUGSSR-551	CACATCCACCATAT CAATAGGC	59.19	CTTCATCGAGGT CATTAGTTGG	58.72	200
2	MUGSSR-552	GCCTATTCAGGTCA GTCAGGA	59.3	GATACTGTGGCA GACAAGAAGC	59.01	378
3	MUGSSR-553	ACGGAATCTGATGA TTGAGCA	60.62	CACAGAATGAGA ATGCACGTAA	58.85	359
4	MUGSSR-554	GGAAGCTTGAGAG GAAGTGTG	59.08	CCACCTGTAGGC CATTATGAA	59.83	370
5	MUGSSR-555	TTGACGGTGTTCGA TAGTTGA	59.18	CTCCACCACCTA AGCCAGTTC	61.05	288

S. No	SEQ ID	FP	Tm (°C)	RP	Tm (°C)	Size (bp)
1	sra_data_contig_13	TTCAAAGCTG GTTCTAGGTC	54.66	GGTTAGCAGTGAA AGTGAGG	55.02	156
2	sra_data_contig_46	TGTTGTTGGGT TCTTCTTCT	54.82	CTGCTCTCTCTCT CTCTCACA	55.01	158
3	sra_data_contig_46	GTTGAGAAGC ACTTCTTGGA	55.58	ACTCTGCTCCCTC TCAAACCT	55.69	142
4	sra_data_contig_46	GTTGTGAGGG TGAAATTGAG	55.55	TCTCTCTCTCTGT GCTCTGC	55.94	158
5	sra_data_contig_46	GTAAACCTAA GCCGAAGGAC	55.64	AGGGACTTCCATT GAGTGTA	54.59	148

Further, the Horsegram SSR finder tool is also available and can be found at exactly below the Horsegram SSR markers Database button along with database under the “Databases & Tools Developed” menu. By selecting

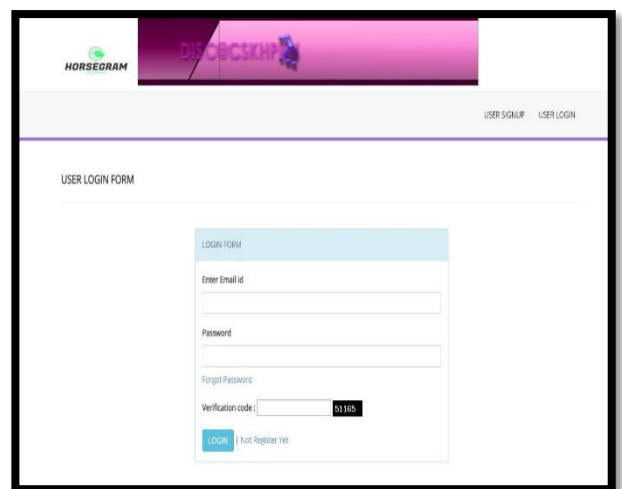
Horsegram SSR Finder the page navigates to the next page which shows a query box. One can insert his query in form of atleast 12 basepairs nucleotides viz. AATTTCGCGTTGC & then press search button. It will show a table showing forward & reverse sequence of the primer along with hyper link. If user wants to go for further details, then click on the hyperlink, it will open details of the marker information such as: sequence ID, simple sequence repeats, forward primer, predicted Tm of forward primer, reverse primer, predicted Tm of reverse primer, predicted amplicon size in base pairs.

From here, one can move either back to search or print results (screenshots attached).

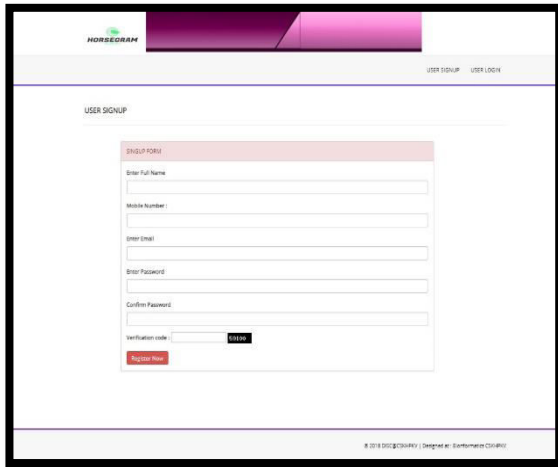
1. Database main page Form



2. User Login Form

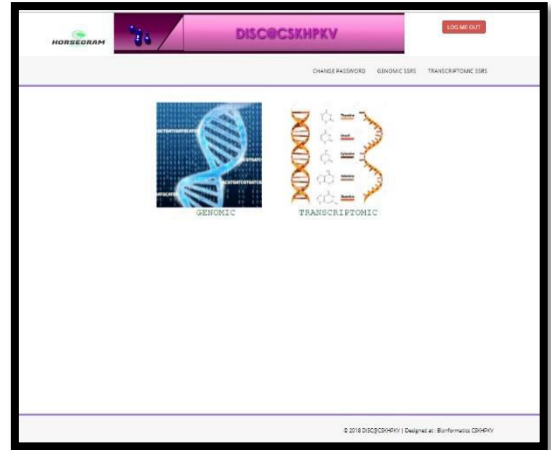


3. User sign Up



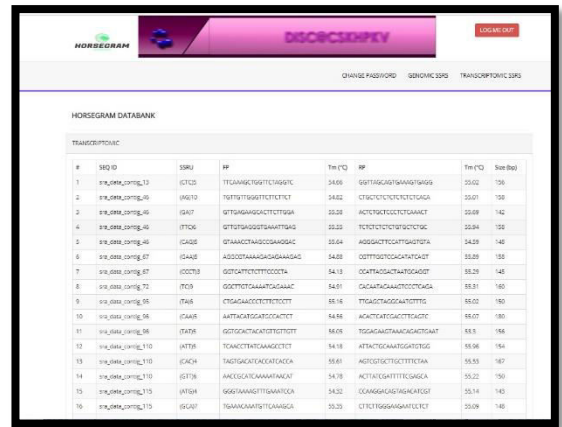
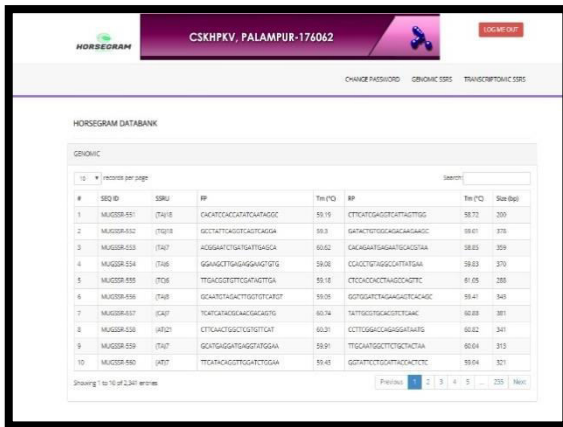
4. Genomic & transcriptomic

Markers



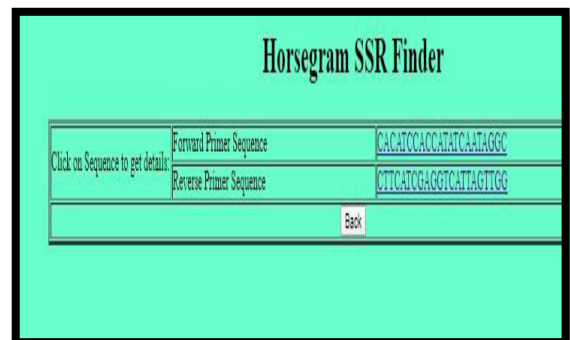
5. Genomic markers Transcriptomic markers

6.



7. Horsegram SSR sequence finder

8.



Horsegram SSR

9.

Seq. ID :- MUGSSR-551	
SSR Marker:-Genomic	
SSR Details	
Sequence ID :	MUGSSR-551
Simple Sequence Repeats :	(TA) ₁₈
Forward Primer :	CACATCCACCATATCAATAGGC
Predicted Tm (°C) Forward Primer :	59.19
Reverse Primer :	CTTCATCGAGGTCATTAGTTGG
Predicted Tm (°C) Reverse Primer :	58.72
Predicted Amplicon Size (bp) :	200

Horsegram SSR detail

Future directions:

The database will be upgraded from time to time with new available markers and linkage maps and QTL collections will be added as the work progresses further. The other legume crops will also be included in the database whenever the data will be available through experimental validation in our Department/University.

Acknowledgement:

Authors are highly thankful to BTISNet, DBT, Ministry of Science & Technology, Govt. of India for providing financial support to carry out this work & Mr. Hoshiar Sharma for providing technical assistance.

References:

1. Aditya, J.P., Bhartiya, A., Chahota, R.K., Joshi, D., Chandra, N., Kant, L., Pattanayak, A (2019). Planta, Ancient orphan legume horse gram: a potential food & forage crop of future. Volume 250: Page. 891–909.
2. Allen, O. N., and Allen, E. K (1981). *The Leguminosae: A source book of characteristics, uses and nodulation*. Madison, WI: University of Wisconsin Press, and London: Macmillan Publishers Ltd.
3. Ashraf, M (2010). Biotechnology advances, Inducing drought tolerance in plants: recent advances. Volume 28: Page.169–183.
4. Bhardwaj, J., Chauhan, R., Swarnkar, M.K., Chahota, R.K., Singh, A.K., Shankar, R., Yadav, S.K (2013). BMC Genomics, Comprehensive transcriptomic study on horsegram (*macrotyloma uniflorum*): de novo

- assembly, functional characterization & comparative analysis in relation to drought stress. Volume 14: Page.647.
5. Bouck, A. and Vision, T (2007). Molecular ecology, The molecular ecologist's guide to expressed sequence tags. Volume 16: Page.907–924.
 6. Chaitanya, D.A.K., Kumar, S., Reddy, M., Mukherjee, A.M., Sumanth, N.S.V., Ramesh, D.A (2010). Journal of innovative trends in pharmaceutical science, Anti urolithiatic activity of *macrotyloma uniflorum* seed extract on ethylene glycol induced urolithiasis in albino rats. Volume 1: Page.216–226.
 7. Ellegren, H (2008). Molecular ecology, Sequencing goes 454 & takes large-scale genomics into the wild. Volume 17: Page.1629–1631.
 8. Fang, G., Munera, D., Friedman, D.I, Mandlik, A., *et al.* (2012). Nature biotechnology, Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. Volume 30: Page.1232–1239.
 9. Kaldate, R., Rana, M., Sharma, V., Hirakawa, H., Kumar, R., Singh, G., Chahota, R.K., Isobe, S.N., Sharma, T.R (2017). Molecular breeding, Development of genome-wide SSR markers in horsegram & their use for genetic diversity & cross-transferability analysis. 37: Page.103.
 10. Kofler, R., Schlotterer, C., Lelley, T (2007). Bioinformatics, Sciroko: a new tool for whole genome microsatellite search & investigation. 23: Page.1683–1685.
 11. Mardis, E.R (2008). Trends in genetics, The impact of next-generation sequencing technology on genetics. Volume 24: Page.133–141.
 12. Mehra, A. and Upadhyaya, M (2013). International journal of agricultural and food science, *Macrotyloma uniflorum* (lam.) traditional crop of Kumaun Himalaya & ethnobotanical perspectives. Volume 3 Number 4: Page.148–150.
 13. Murthy, S.M., Devaraj, V., Anitha, R.P., Tejavathi, D.H (2012). Recent research in science and technology, Studies on the activities of antioxidant enzymes under induced drought stress in vivo & in vitro plants of *macrotyloma uniflorum* (lam.) Verdc. Volume 4: Page.34–37.
 14. Prakash, B.G., Guled, M.B., Bhosale, A.M (2008). Karnataka journal of agricultural sciences, Identification of suitable horsegram varieties for northern dry zone of Karnataka. Volume 21: Page.343–345.
 15. Ramesh, C.K., Rehman, A., Prabhakar, B.T., Vijay, A.B.R., Aditya, S.J (2011). Journal of applied pharmaceutical science, Antioxidant potential in sprout vs. Seeds of *vigna radiata* & *macrotyloma uniflorum*. Volume 1: Page.99–103.

16. Rozen, S. and Skaletsky, H (2000). Methods in molecular biology, Primer3 on the www for general users & for biologist programmers. Volume 132: Page.365–386.
17. Santos, C.S., Pinheiro, M., Silva, A.I., Egas, C., Vasconcelos, M.W (2012). BMC Genomics, Searching for resistance genes to *bursaphelenchus xylophilus* using high throughput screening. 13: Page. 599.
18. Shirasawa K, Isobe S, Tabata S, Hirakawa H (2014). Breeding science, Kazusa marker database: a database for genomics, genetics, & molecular breeding in plants. Volume 64: Page. 264–271.
19. Sudha, S and Saral, M.A (2023). Heliyon, Studies on phytochemical, mineral content, in vitro anti-urolithiatic and anti-diabetic activities of horse gram flour extracts and its biosynthesized Ag nanoparticles. Volume 9 Number 6: e16572
20. Temnykh, S., Declerck, G., Lukashova, A., Lipovich, L., Cartinhour, S., Mccouch, S (2001). Genome research, Computational & experimental analysis of microsatellites in rice (*oryza sativa* l.): frequency, length variation, transposon associations, & genetic marker potential. Volume 11 Number 8: Page.1441-52.
21. Thiel, T., Michalek, W., Varshney, R.K., Graner, A (2003) Theoretical and applied genetics, Exploiting EST databases for the development & characterization of gene derived SSR-markers in barley (*Hordeum vulgare* l.). Volume 106: Page.411–422.
22. Wang, W and Messing, J (2011). Plos one, High-throughput sequencing of three *lemnoideae* (duckweeds) chloroplast genomes from total DNA. Volume 6: e24670.