International Journal of Chemical Studies

P-ISSN: 2349–8528
E-ISSN: 2321–4902
www.chemijournal.com
IJCS 2021; 9(4): 354-358

Received: 13-05-2021
Accepted: 15-06-2021

Sonam Arya
College of Applied Education and Health Sciences, Meerut, Uttar Pradesh, India

Rajat Choudhary
Division of Genetics, IARI, New Delhi, Delhi, India

Vishakha
Sardar Vallabh Bhai Patel University of Agriculture and Technology, Meerut, Uttar Pradesh, India

Reeva
Sardar Vallabh Bhai Patel University of Agriculture and Technology, Meerut, Uttar Pradesh, India

Vaishali
Sardar Vallabh Bhai Patel University of Agriculture and Technology, Meerut, Uttar Pradesh, India

Corresponding Author:
Sonam Arya
College of Applied Education and Health Sciences, Meerut, Uttar Pradesh, India

Molecular characterization of tomato germplasm using EST-SSR

Sonam Arya, Rajat Choudhary, Vishakha, Reeva and Vaishali

Abstract
Tomato is an important vegetable having high economic and medicinal values for human beings. There have been great variation have observed and show great genetic diversity. Among different molecular maker SSR were very highly reproducible for determing the polymorphism. In this research the insilico analysis were used for the identification of the eSSR markers from the EST sequence of Tomato available in NCBI. This approach is easiest, requires less time consuming, cost effective and also provide molecular mapping which leads an opportunity for gene discovery which shows linkage with a trait of interest. The main objective of this research is to develop the efficient eSSR marker using ESTs of tomato and molecular characterization of tomato can be done to determine the genetic diversity present in selected genotypes. This will possess a significant specificity and high degree of conservation. In future it will be considered as a potential tool for various genotyping, applications including studying cross-transferability and phylogenetic relationships and comparative genome mapping in crop species.

Keywords: EST, SSR, genetic diversity, *in silico* analysis

Introduction
Tomato ranks first in the world and grown in both temperate and tropical regions of the world which accounts about 14% of world vegetable production. In India it is produced mostly in Andhra Pradesh, Odisha, Karnataka, Madhya Pradesh, West Bengal, Bihar. Tomato (*Lycopersicon esculentum* L.) belongs to family Solanaceae and significant vegetable crop of special economic importance (Wang *et. al*., 2005) [28], because of high economic value and having rich source of vitamins A, B and C and various micronutrients which are very beneficial for human beings (Kaushik *et al.*, 2011) [10] and also have high antioxidant content, including rich concentration of lycopene (Fraser, 1994). Tomato has good medicinal value as its beneficial for the patients suffering from higher uric acid problems, bronchitis, asthma, also helps in curing tropid liver, dyspepsic, promoter for gastric juices and blood purifier (Giovanucci *et al.* 1995; Grant 1999) [5, 6].

Detection and assessment genetic diversity have been used for analysis of discrete molecular trait. DNA molecular markers technology, provide powerful tools for cultivar identification and seed quality control in various crops (Mongkolporn *et al.*, 2004; Garg *et al.*, 2006 and Liu *et al.*, 2007) [12, 11]. Molecular markers are indispensable for genomic study among various marker systems Simple Sequence Repeats (SSRs) have occupied a pivotal place because of their reproducibility, multiallelic nature, codominant inheritance, relative abundance and good genetic coverage (Singh *et al.*, 2013) [22].

Bioinformatics approaches are increasingly being used for molecular marker development since the sequences from many genomes are made freely available in the public databases (Kantety *et al.*, 2002; Varshney *et al.*, 2002) [9, 26]. Expressed sequence tag (EST) is presently the most widely accepted nucleotide sequence for marker based study as it represents transcribed part of the genome. A large amount of EST database present on public domain (NCBI) can be exploited for development of SSR markers and functional annotation by using software based *in silico* analysis. The development of SSR markers by this method is more preferable today (Scott *et al.*, 2000; Temnykh *et al.*, 2000) [21, 25] as they are cost effective, need less time to development and more informative in comparison to conventional method (Zane *et al.*, 2002) [29]. The EST-SSR markers not only help in molecular mapping but also provide an opportunity for gene discovery when they show linkage with a trait of interest (Thiel *et al.*, 2003).

The objective of this research is to develop the efficient eSSR marker using ESTs of tomato and molecular characterization of tomato can be done to determine the genetic diversity

present in selected genotypes. The eSSR markers possess significant specificity and high degree of conservation, they are considered to be potential tool for various genotyping applications including studying cross-transferability and phylogenetic relationships and comparative genome mapping in crop species.

Material and Method

The whole experiments were conducted on 47 genotypes of tomato in the molecular biology lab of Department of Agricultural Biotechnology, College of Agriculture, S.V.P.U.A.T, Meerut. The genomic DNA from the leaves of tomato plants were isolated using the standard protocol as described by Doyle and Doyle, 1990, with slight modification. In this method Cetyl Trimethyl Ammonium Bromide (CTAB) was used as a detergent to lyse the wall of cells for release of DNA.

The isolated genomic DNA was dissolved in the TE buffer and was taken for quantification by recording absorbance at λ_{260}nm in a UV VIS spectrophotometer (Eppendorf, Germany). The total amount of DNA was calculated considering that OD of 1.0 at 260 nm is equivalent to 50µg/ml of double standard DNA. Agarose gel electrophoresis of the isolated genomic DNA was performed to carry out by making gel of 0.8% used for the qualitative estimation of Genomic DNA of tomato.

EST mining and EST-SSR primer designing

EST sequences for designing the EST SSR primers were searched from the sequences available at public domain at (http://www.ncbi.nlm.nih.gov). All tomato ESTs were downloaded from NCBI and downloaded in FASTA format. Downloaded ESTs sequences were assembled in the form of contigs using an online available EGassembler program.

Detection of SSRs

The contigs were processed using SSRIT online program to find out the simple sequence repeat motifs. This tool finds all perfect possible SSR present in sequence submitted. The result of the SSRIT shows the sequence ID, motif (repeat) type, no. of repeats, SSR start and end.

Primer designing using Primer 3 software

Primers for identified SSR sequences were designed using online available software PRIMER3 (http://frodo.wi.mit.edu) using the ideal parameters of primer. The designed primers were synthesized from Bangalore Genei and Imperial Biotech (IDT), India as 0.025 µmol in dry form. These primers were diluted in 0.1X Tris-EDTA buffer solution for making primer stocks and later further diluted to appropriate working concentrations using ion free double distilled water.

Polymerase chain reaction amplification

A set of 30 SSR primers synthesized were suspended in 100µl of TE buffer. The working was made using 10µl of standard primer solution with 90µl of TE buffer. DNA amplification reaction for SSR primers were performed a total of 20µl 47 genotypes of tomato. The annealing temperature for each primer is calculated as Tm = (A+T) x 2 + (G+C)x 4. The Tm used in these experiments was 3-5ºC lower than the calculated Tm. The amplified PCR products for EST-SSR markers were resolved on 2% agarose gel in 1X TAE buffer. The banding pattern was photographed in a gel documentation unit (Alpha Innotech) visualized under UV Tran illuminator for further analysis.

Data analysis

Gene Diversity and Resolving Power of SSR primers

In order to assess the ability of primers to resolve the different varieties the resolving power (Rp) for each primer was calculated following Prevost and Wilkinson's (1999) [13] method as Rp = Ib (band information). Resolving Power is calculated as 1-[2 x (0.5-p)], p being the proportion of the 40 varieties containing the bands and Gene Diversity is calculated as $1-\sum pi^2$ (Anderson *et al.*, 1993) [1].

The bands were scored as present (1) or absent (0) for each DNA sample with the all 20 SSR for drought specific primers. Amplification was performed twice and only reproducible amplifications products were included in the data analysis. Similarity matrix using the similarity coefficient of Jaccard (1908) [8] was constructed from the whole data. Pair wise distances between DNA accessions were calculated and analysed using the Unweighted Pair Group Method Arithmetic average (UPGMA) (Sneath and Sokal, 1973) [24]. Clusters were analysed using the computer program NTSYS-PC, version 2.11s (Rohlf, 2000) [24].

Result

The genomic DNA from all 47 genotype of tomato was isolated and qualitatively analysed using 0.8% agarose. A total of 66,027 ESTs sequences were found related to tomato genome available on NCBI. From these ESTs 15,484 contigs were obtained after removing the redundancy by using EGassembler. The EST contigs were submitted to SSRIT software for the identification of SSR regions for primer designing. A total 611 SSRs were identified from the EST contigs, among them 336 were dinucleotide repeats, 252 were trinucleotide repeats, 17 were tetranucleotide repeats and 06 were pentanucleotide repeats. No hexa- or hepta-nucleotide repeats were obtained in assembled EST contigs of tomato.

From these obtained 611 EST-SSR only 30 pair of EST-SSR primer pair were taken on the basis of parameters selected were GC content ranging from 45 to 60%, SSR repeats were marked as target region, product size ranges from 300 to 500 bp, primer length from 18 to 25 nucleotides and melting temperature of 50 to 65^{0}C for the synthesis.

47 genotypes of tomato which were normally grown under field condition were characterized at molecular level using EST-SSR marker and were used for analyzing the genetic diversity of tomato (L*ycopersicum esculentum*). A total of 30 SSR primers were used to amplify the genomic DNA of 47 tomato genotypes, twenty pairs of SSR primers generated clear, reproducible banding patterns were chosen for the genetic diversity analysis of tomato. A total of twenty EST-SSR primers resulted in scorable, and reproducible result hence considered for the analysis of genetic diversity of tomato genotypes.

Gene diversity was calculated for all the 20 EST-SSR polymorphic primers, which varied from 0.04 to 0.96 value with a mean diversity of 0.50. The highest gene diversity 0.96 was recorded with the primer14. On the other hand the lowest gene diversity 0.04 was shown by primer 19. The higher mean PIC value indicated the informative ness of the primers pairs in detecting genetic diversity which can be used in future studies in the field of taxonomical and genetic resource management.

Resolving power of the twenty EST-SSR primers ranged from 0.08 to 1.72 with an average 0.90. The highest resolving power 1.72 was recorded for the primer 13. On the other hand the lowest resolving power 0.08 was recorded with the primer14. Thus the significant value of resolving power

indicated the ability of primers to resolve the different closely related genotypes of tomato.

SSR data were used to make pair wise comparison of the accessions based on shared and unique amplification products to generate a similarity matrix with NTSYS-PC (version 2.11s). Based on the distance matrix expressed as similarity coefficient a dendrogram was generated by the UPGMA method. Similarity value for all the 47 genotypes ranged from 0.21 to 1.00. The minimum similarity exhibited by genotype NDT-3 and KS-229. Whereas, the maximum similarity was shown by genotype NDT-2 and KS-208.

The cluster analysis based on the Unweighted Paired Group Method of Arithmetic Means (UPGMA) with 20 SSR primers allowed the discrimination of cultivars. The UPGMA based clustering of 47 tomato genotypes grouped the tomato genotypes in one group with one member stay away from the group and located at one end of the cluster at 50% similarity coefficient. The genotype NDT-3 is diverse enough and did not grouped in any cluster. The rest of the 46 tomato genotypes further grouped into four major clusters. The cluster I grouped 5 genotype namely TYPE-1, KS-254, KS-7, KS-208 and EC677067. The cluster II grouped two genotype EC638516 and EC638517. The major cluster III grouped 8 genotype namely NDT-5, NDT-8, EC638513, EC654692, EC654694, ANGOORLATA, EC654692 and EC644398. The major cluster IV grouped 31 genotypes which is further subdivided into three subclusters. The subcluster IV-a composed of 5 genotypes Arka Saurabh, EC677077, IC395352, EC677079 and EC608435. The subcluster IV-b comprised of 12 genotypes namely KS-229, EC610643, Cherry Tomato, S-22, Arka Vikas, PS2626, EC596746, EC608438, EC608441, EC608433, EC610642 and EC610639. Of them the genotypes KS-229 and EC610643were separately grouped at one end of the subcluster. The subcluster IV-c includes 14 genotypes namely NDT-1, EC677071, EC654696, NDT-4, NDT-7, IC399669, AZADT-2, AZADT-3, AZADT-8, TYPE-1, KS-7, EC638512, EC677068, EC610647.

The molecular characterization as generated by SSR markers and correlation pattern of different characters showed that these forty seven varieties are differ enough and significantly diverse. A total of 10 tomato genotypes *viz*. NDT – 2, Azad T-6, K -208, Angoorlata, Saurabh, Vikash, EC 610642, EC 677079, EC 677068 and EC 608438 were selected as they are the member of different cluster, hence significantly diverse. Therefore these tomato genotypes were considered for their further evaluation of the research.

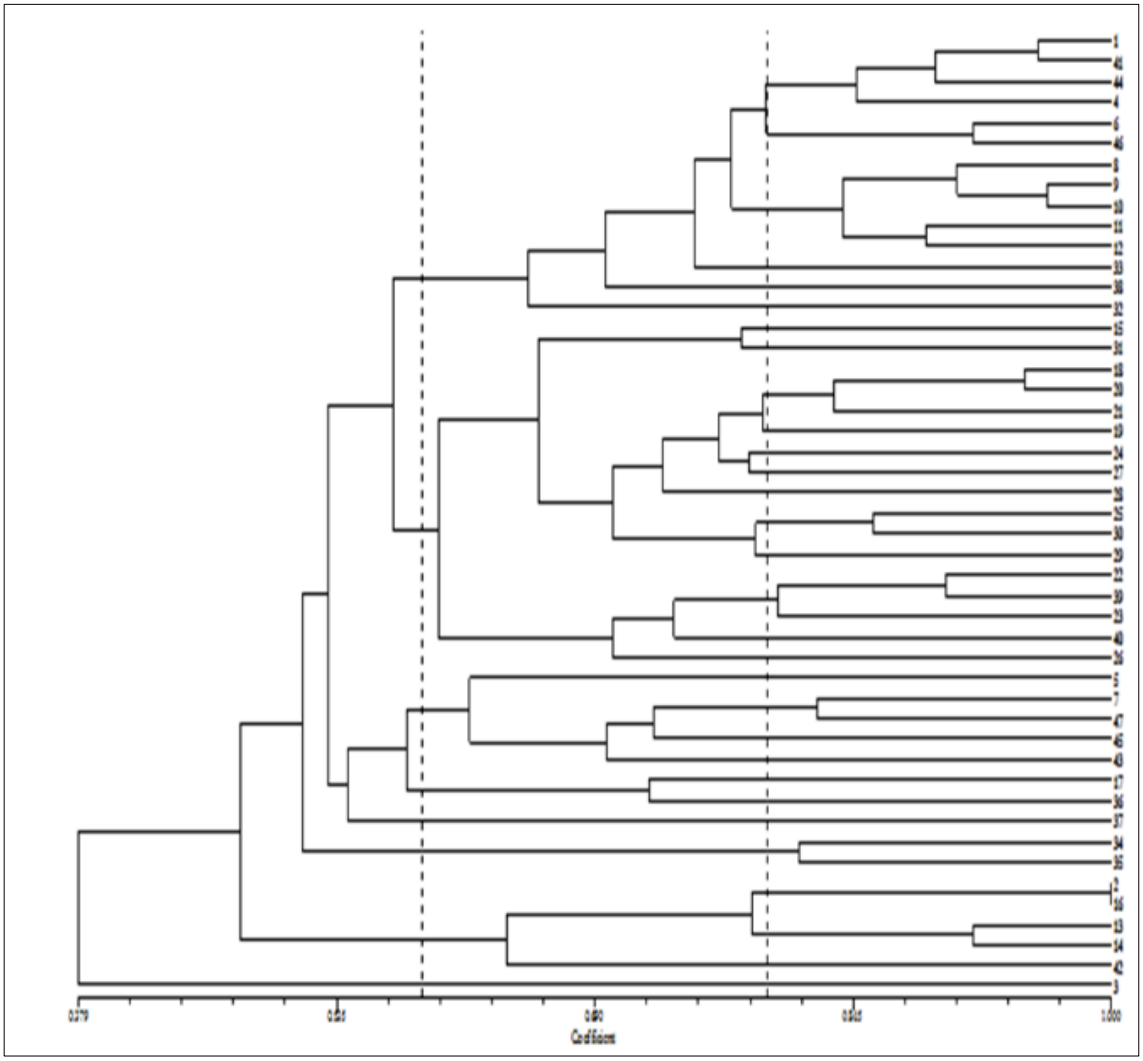

Fig 1: Dendrogram of Tomato genotypes on the basis of SSR profiling. 1. NDT1, 2. NDT2, 3. NDT3, 4. NDT, 5. NDT5, 6. NDT6, 7. NDT7, 8. NDT5, 9. AZADT2, 10. AZADT3, 11. AZADT6, 12.AZADT8, 13. KS7, 14. KS208, 15. KS209, 16. KS254, 17. TYPE1, 18. AGOORLITA, 19. CHEERYTOMATO, 20. PS2626, 21. S22, 22. ARKASAURABH, 23. ARKAVIKAS, 24. IC395352, 25. EC399667, 26. EC596747, 27. EC608433, 28. EC608435, 29. EC608438, 30. EC608439, 31. EC608441, 32. EC610639, 33. EC610642, 34. EC610643, 35. EC610647, 36. EC638512, 37. EC638513, 38. EC638516, 39, 40. EC654692, 41. EC654694, 42. EC654696, 43. EC677067, 44. EC677068, 45. EC677071, 46. EC677079, 47. EC677080.

Discussion

The genetic diversity of 47 tomato genotypes was assessed using 30 SSR primers. To estimate the genetic diversity for different tomato genotype will help to future breeding programme and may help to screen the right genotypes for future drought research programmes. For the present research a total 611 SSRs were identified from the EST contigs, among them 336 were dinucleotide repeats, 252 were trinucleotide repeats, 17 were tetranucleotide repeats and 06 were pentanucleotide repeats. Genetic diversity assessed form molecular markers can be a material basis for crop improvement (Habash *et al.*, 2009) [7]. It is desirable to have large genetic diversity for the creation of new genotypes. Gene diversity was calculated for all the 20 EST-SSR polymorphic primers, which varied from 0.04 to 0.96 value with a mean diversity of 0.50. The higher gene diversity value can be used in future studies in the field of taxonomical and genetic resource management. The significant value of resolving power indicated the ability of primers to resolve the different closely related genotypes of tomato. Information from present study will be helpful in the study of genetic relationships, representation of genetic diversity and morphological evaluation.

Conclusion

In the present research it was concluded that the development of SSR markers using database searching is more cost effective and cheap in compare to the isolation of the same from genomic libraries and cross- species amplification. Bioinformatics approach produces good and more informative microsatellite markers in a very short span of time. There is a plenty number of crops which are playing very important role to meet our food security but genetic study on the development of SSR marker is lagging in such crops. Hence, these in-silico methods are playing very important role in contributing to the development and progress in the field of science and agriculture.

References

1. Anderson JA, Churchill GA, Sutrique JE, Tanksley SD, Sorrells ME. Optimizing parental selection for genetic linkage maps. Gen 1993;36:181-186.
2. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 1987;19:11-15.
3. Fraser LG, Harvey CF, Crowhurst RN, De Silva HN. EST-derived microsatellites from *Actinidia* species and their potential for mapping. Theor. Appl. Genet 2004;108:1010-1016.
4. Garg K, Green P, Nickerson DA. Identification of candidate coding region single nucleotide polymorphisms in 165 human genes using assembled expressed sequence tags. Gen. Res 1999;9:1087-1092.
5. Giovanucci E, Ascherio A, Rimm EB, Stampfer MM, Colditz GA, Willett WC. Intake of carotenoids and retinal in relationship to risk of prostate cancer. J Natl. Cancer Inst 1995;87:1767-1776.
6. Grant WB. An ecologic study of dietary links to prostate cancer. Alternative Med. Rev 1999;4:167-9.
7. Habash DZ, Kehel Z, Nachit M. Genomic approaches for designing durum wheat ready for climate change with a focus on drought. Journal of Experi. Bot 2009;60(10):2805-2815.
8. Jaccard P. Nouvelles researches sur la distribution florale. Bulletin de la Société vaudoise des sciences naturelles 1908;44:223-270.
9. Kantety RV, Rota ML, Matthews DE, Sorrells ME. Data mining for simple sequence repeats in expressed sequence tags from barely, maize, rice, sorghum and wheat. Plant Mol. Biol 2002;48:501-510.
10. Kaushik SK, Tomar DS, Dixit AK. Genetics of fruit yield and it's contributing characters in tomato (*Solanum lycopersicom*). Journal of Agricul. Biotech. and Sustain. Develop 2011;3(10):209-213.
11. Liu Q, Xu J, Liu YZ, Zhao XL, Deng XX. A novel bud mutation that confers abnormal patterns of lycopene accumulation in sweet orange fruit (*Citrus sinensis* L. *Osbeck*). Journal of Exper. l Bot 2007;58:4161-4171.
12. Mongkolporn O, Dokmaihom Y, Kanchana-udomkan C, Pakdeevaraporn P. Genetic purity test of F1 hybrid *Capsicum* using molecular analysis. J Hort. Sci. Biotechnol 2004;79(3):449-451.
13. Prevost A, Wilkinson MJ. A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. Theot. and Appld. Gent 1999;98(1):107-112.
14. Rohlf J. NTSYSpc: Numerical taxonomy and multivariate analysis system. Version 2.1. Users Guide, Exeter. Software, Setauket, 38. New York 2000.
15. Sato S, Peet MM, Thomas JF. Physiological factors limit fruit set of tomato (*Lycopersicon esculentum* Mill.) under chronic, mild heat stress. Plant, Cell and Environ 2000;23:719-726.
16. Sairam RK, Deshmukh PS, Saxena DC. Role of antioxidant systems in wheat genotypes tolerance to water stress. Biol Plantarum 1998;41:387-394.
17. Salekdeh GH, Siopongco J, Wade LJ, Ghareyazie B, Bennett J. Proteomic analysis of rice leaves during drought stress and recovery. Prot 2002;2:1131-1145.
18. Sánchez-Rodríguez E, Rubio-Wilhelmi MDM, Ríos JJ, Blasco B, Rosales MÁ, Melgarejo R *et al.* Ammonia production and assimilation: Its importance as a tolerance mechanism during moderate water deficit in tomato plants. J Plant Physiol 2011;168:816-823.
19. Saneoka H, Nagasaka C, Hahn DT, Yang WJ, Premachandra GS, Joly RJ *et al.* Salt tolerance of glycinebetaine-deficient and – containing maize lines. Plant Physiol 1995;107:631-638.
20. Sankar B, Jaleel CA, Manivannan P, Kumar AK, Somasundaram R, Panneerselvam R. Drought induced biochemical modifications and proline metabolism in *Abelmoschus esculentus* (L.) Moench. Acta Bot. Croat 2007;66(1):43-56.
21. Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS *et al.* Analysis of SSRs derived from grape ESTs. Theor. Appl. Genet 2000;100:723-726.
22. Singh S. Water logging and its effect on cropping pattern and crop productivity in South-West Punjab: A case study of Muktsar district. Journal of Eco. and Social Develop 2013;9:1.
23. Singh M, Singh NP, Arya S, Singh B, Vaishali. Diversity analysis of tomato germplasm (*Lycopersicom esculentum*) using SSR markers. Intern. Journal of agri. Sci. and res 2014;4(4):41-48.
24. Sneath PHA, Sokal RR. Numerical Taxonomy. WH Freeman, San Francisco CA 1973.
25. Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L *et al.* Mapping and genome organization of

microsatellite sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet 2000;100:697-712.
26. Thiel T, Michalek W, Varshney RK, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor. Appl. Genet 2002;106:411-422.
27. Varshney RK, Thiel T, Stein N, Langridge P, Graner A. *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cellular and Mol. Bio. Letters 2002;7:537-546.
28. Wang J, Huang X. A method for finding single-nucleotide polymorphisms with allele frequencies in sequences of deep coverage. BMC Bioinformatics 2005;7:220.
29. Zane L, Bargelloni L, Patarnello T. Strategies for microsatellite isolation: a review. Mole. Ecol 2002;11:1-16.