Assessment of Genetic Diversity in Cultivated Tomato (Solanum lycopersicon L.) Genotypes using Molecular Markers

Supneet Kaur*, A.K. Singh and Sreshti Bagati

School of Biotechnology, Sher-e-Kashmir University of Agriculture Sciences and Technology, Chatha, Jammu-180009, India

*Corresponding author

ABSTRACT

An experiment consisting of 24 genotypes of Tomato was conducted during the year 2016 at the Research Farm and Molecular Biology Laboratory of School of Biotechnology, SKUAST-J, Chatha. Seeds were sown in pots and 21 days old leaflets were used for DNA extraction. For molecular characterization, 24 genotypes of tomato were subjected to banding profiling by using 9 SSR and 13 RAPD markers. A total of 45 alleles were detected with an average of 5.00 allele per genotype for SSR markers. Out of 45 alleles 43 were found to be polymorphic. Similarly, RAPD marker detected 203 alleles, out of which 191 alleles were polymorphic with different product sizes. Maximum number of bands were produced by A1773078 (8 bands), AW03747 (8 bands) and Y09371 (8 bands) for SSR markers and for RAPD marker highest number of band was observed by OPO-20(41 bands). Polymorphic information content (PIC value) ranged from 0.16 (A1773078 and A1998183) to 0.50 (SSR 241) with an average of 0.303 for SSR and 0.243 (OPC-04) to 0.432 (OPU-16) with an average of 0.29 for RAPD markers. High PIC value for SSR than RAPD concludes that SSR markers are more efficient for genetic diversity studies. Further, clustering analysis based on UPGMA grouped these genotypes into 2 main cluster with various sub cluster. Cluster analysis done using SSR markers grouped 20 genotypes in cluster A and 4 genotypes in cluster B. Similarly, for RAPD markers, cluster analysis grouped 19 genotypes in cluster A and 5 genotypes in cluster B. Based on clustering analysis of genotypes using SSR, RAPD and comparative clustering of SSR and RAPD markers, Utkal pragyan and Angha, Utkal pragyan and NDT-4 and Utkal pragyan and NDT-4, respectively were found to be most diverse and hence based on cluster analysis the identified diverse tomato lines can be efficiently selected for carrying out various breeding and crop improvement programmes.

Keywords
Genetic diversity analysis, Molecular characterization, RAPD, SSR

Introduction

Tomato is one of the significant vegetable crops owing to economic importance in the horticulture industry and originated in South America. Fleshy berry of cultivated tomato has been ranked third in global production, after potato and sweet potato (Hsueh-Li tan et al., 2010). It is the most popular garden vegetable belonging to the genus Lycopersicon, the resemblance between leaves and flowers of potato and tomato plants seems to certify this taxonomic grouping (Wang et al., 2005 and Shidfar et al., 2011).
Tomatoes rank first in their nutrient contribution to the diet due to high pro-vitamin A and vitamin C content. Tomato fruits are a rich source of ascorbic acid and flavour of fruits is controlled by various volatile components and balance of sugar: acid ratio. It is extensively used in salad, for culinary purposes, in various processed forms like pastes, sauces, pulps, juices, ketchup and as flavouring ingredients in soups, meat or fish dishes (Gosselin and Trudel, 1984). The fruit contains significant amounts of lycopene, beta-carotene, magnesium, iron, phosphorus, potassium, riboflavin, niacin, sodium and thiamine. It also possesses antioxidant properties (Zhang et al., 2009).

Genetic diversity analysis of tomato is essential to enhance the genetic yield potential with good nutritional properties. Molecular characterization, by itself or in conjunction with other data (phenotypic traits or georeferenced data), provides reliable information for assessing, among other factors, the amount of genetic diversity (Perera et al., 2000), the structure of diversity in samples and populations (Shim et al., 2000 and Figliuolo et al., 2004) rates of genetic divergence among populations (Maestri et al., 2002) and the distribution of diversity in populations found in different locations (Ferguson et al., 2004 and Perera et al., 2000). Several studies in this context have been carried out in various crops like rice, barley, common bean, tomato and Indian mustard (Sharma et al., 2015; Kumar et al., 2016; Zargar et al., 2016; Rai et al., 2016 and Visalakshi et al., 2013).

There are more than 7500 tomato varieties which are successfully bred and grown worldwide. Genetic analysis of tomato is essential for the enhancement of the genetic yield potential and maximum utilization of the desirable characters for synthesis of any ideal genotypes (Kumar et al., 2003). Measurements of genetic diversity can be used in breeding programs to increase the genetic variation in base populations by crossing cultivars with a high level of genetic distance as well as for the introgression of exotic germplasm. Study of phenotypic and genetic diversity in landrace collections is important for germplasm conservation. In addition, the characterization of much diversified materials with molecular markers offers a unique opportunity to define significant marker-trait associations of biological and agronomic interest. Microsatellite or simple sequence repeat (SSR) marker has been used in plant diversity analysis; the popularity of these markers is due to their ease of amplification by polymerase chain reaction (PCR), their co-dominant nature and their typically high levels of allelic diversity at different loci.

There are numerous reports suggesting the usefulness of microsatellite markers for measuring the genetic variability in a wider taxonomic range (Ryberg et al., 2002; Li et al., 2007; Chan et al., 2008; Banhos et al., 2008). RAPD is an efficient method for varietal identification, study of polymorphism, gene mapping, biodiversity, genetic map construction, hybridization and phylogenetic relationship in tomato varieties (Salunke et al., 2012; Singh et al., 2014; Rai et al., 2016). Moreover, the main advantages of RAPD over other molecular methods are the low sample DNA requirements, high frequency of detectable polymorphic DNA bands and independent from the effects of environmental factors (Kuras et al., 2004).

The present study has been conducted to assess the genetic diversity within different tomato genotypes by using molecular markers and development of phylogenetic tree by using bio informatics tools. Estimation of genetic diversity and relationships between germplasm collections are important for facilitating efficient germplasm collection, evaluation and utilization (Rafalski, 2009).
The use of molecular markers serves as a modern and suitable approach to cultivar and variety identification, since it is more rapid and cost-effective (Korir et al., 2012).

Information from present study will be helpful in the study of genetic relationships, representation of genetic diversity and morphological evaluation. Genetic diversity analysis at molecular level by SSR markers that cover the whole genome would be helpful to identify the diverse parents. In the present study we employed two different marker systems i.e. SSRs and RAPD to evaluate the efficiency of these markers in diversity analysis of tomato genotypes along with various parameters to elucidate genetic diversity among tomato genotypes.

**Materials and Methods**

**Genotypes**

The material for the present study consists of 24 genotypes obtained from IIVR, Varanasi (Table 1) and were initially grown in pots and 21 day old seedlings were collected for DNA extraction. The methodology adopted is as under:

**Genomic DNA extraction**

The total genomic DNA was isolated from each genotype by CTAB method (Clarke, 2009) with certain modifications. Fresh and young leaves from twenty one days old genotypes were used for genomic DNA extraction.

The DNA was then quantified by using agarose gel electrophoresis (0.8%) and nanodrop at 260 and 280 nm. The concentrated DNA samples were diluted to a working concentration of 25ng/μl. An equal amount of DNA from 24 genotypes were taken for PCR amplification.

**Molecular analysis**

**RAPD genotyping**

A set of 13 arbitrary random 10-mer primers were selected for use in DNA amplification of genomic DNA. Based on earlier studies, RAPD primers were selected. They were got synthesized from IDT (Integrated DNA Technologies, USA). DNA amplification was carried out in a PCR tubes containing 25 μL reaction mixture. Reaction mixture contained 3 μL of template DNA (25 ng/ μL), 2.5 μL of 10X PCR Buffer, 1.5 μL MgCl₂ (50 mM/ μL), 2 μL (2.5 mM/ μL of each dNTPs (dTTPs, dGTPs, dCTPs, dATPs), 5 μL primer (5 pico mol), 0.5 μL (5 u/ μL) Taq DNA polymerase. PCR tubes containing master mix and DNA template were thoroughly mixed and subjected to the thermal profile. The amplification reaction was carried out in a gradient Thermocycler. An initial denaturation step of 4 minutes was programmed in the thermocycler, followed by a loop of 35 cycles each consisting of denaturation (at 94°C for 1 minute), annealing (at 36-38°C for 1 minute) and extension (at 72°C for 2 minutes). The final extension was performed at 72°C for 10 minutes. The PCR products were then stored at 4°C for 10 minutes. Products were separated on 1.5% agarose gel along with standard molecular weight marker (100 bp ladder) (Sigma Aldrich, USA). The gel was visually examined under UV and documented using gel documentation system. The list of RAPD primers used is detailed in Table 2.

**SSR genotyping**

DNA amplification was carried out in 0.2ml PCR tubes containing 25 μl reaction mixture. The reaction mixture contained primer 2.5 μl of each forward and reverse primers (10 pico mole), 3μl of template DNA (25ng/μl), 2.5 μl of 10x PCR buffer without MgCl₂, 1.5 μl of MgCl₂ (mM/ μl), 2 μl (2.5 Mm/ μl) of each
dNTPs (dTTPs, dGTPs, dCTPs, dATPs), 0.5 µl (5 U/µl) Taq DNA polymerase. PCR tubes containing master mix and DNA template were thoroughly mixed and subjected to the thermal profile. The amplification reaction was carried out in a gradient thermocycler.

An initial denaturation step of 4 minutes was programmed in the thermocycler, followed by a loop of 35 cycles each consisting of denaturation (at 94°C for 1 minute), annealing (49°C-55°C for 1 minute) and extension (at 72°C for 2 minutes). The final extension was performed at 72°C for 7 minutes. The PCR products were then stored at 4°C for 10 minutes. PCR products were separated on agarose gel electrophoresis. The 25µl PCR products plus loading dye (Bromophenol blue) were loaded into 3% agarose gel in 1XTBE buffer (10mM Tris borate, 1mM EDTA) containing 5µl of EtBr. 100bp DNA ladder were also loaded as a molecular marker. Electrophoresis were carried out at 80 V for 3-4 hours and then viewed under Biometra gel documentation unit. The list of SSR primers used is detailed in Table 3.

Data analysis

The SSR and RAPD PCR bands were examined under ultraviolet transilluminator and photographed under gel documentation unit. The profile developed by each marker was scored (1) for the presence and (0) for the absence of a band for each genotype. In order to compare the efficiency of these two marker systems in genotype identification, differentiation and diversity analysis, we considered the following parameters for each assay unit (U).

Number of bands (NB);
Number of polymorphic bands (NPB);
Number of monomorphic bands (NMB);
Number of unique bands (NUB);
Percentage of polymorphic bands (PPB);
Polymorphic information content (PIC) according to Powell et al., (1996): PIC= 2fi (1-fi);
Marker index (MI) according to Powell et al., (1996): MI = PIC · b · a;
Resolving power (RP) according to Prevost and Wilkinson (1999): RP=Rib

Scored data were used for the estimation of Jaccard’s similarity coefficient using NTSYSpc version 2.02e (Rohlf, 1998) package to compute pair-wise Jaccard’s similarity coefficient (Jaccard, 1908) and this similarity matrix was used in cluster analysis using the unweighted pair-group method with arithmetic averages (UPGMA) and dendrogram was constructed using SAHN method.

Results and Discussion

Allele diversity in tomato using two different marker systems

Both the marker techniques (RAPD and SSR) proved to be highly effective in discriminating the 24 genotypes. Results obtained are summarized in Tables 2–4. 13 RAPD and 9 SSR primers used in the present study amplified 191 and 43 polymorphic bands for RAPD and SSR respectively. An average number of 14.69 polymorphic bands per assay unit were identified for RAPD, whereas in SSR it was 5.00. The utility of a given marker is a balance between the level of polymorphism it can detect, and its capacity to identify multiple polymorphisms (Powell et al., 1996). Marker index is a feature of a marker which elucidates the discriminatory power of a marker and therefore it was calculated for all the markers. Due to high multiplex ratio component (14.69) for RAPD, higher marker index value was observed for RAPD (4.12) in comparison to SSR (1.04). For SSR markers only five alleles per locus
are considered, however for RAPD an average of 15.62 alleles per locus, ranging from 6 (OPN-15) to 41 (OPO-20) was observed. The results of this study are in close conformity with the results of Sharifova et al., (2013). Since a higher number of genotypes as well as RAPD primers were used in this study and that can be the reason for higher number of alleles per locus observed in case of RAPD marker. PIC is an important feature of a primer which indicates its potential to differentiate various individuals. An average PIC of 0.290 was observed for RAPD where as it was 0.300 for SSR markers (Table 2 and 3). Highest PIC was observed for primers OPU-17 (0.37), OPN-15 and OPQ-20 (0.300) in RAPD assay (Table 2). The results are in close conformity with the findings of Dhaliwal et al., (2009) and Thamir et al., (2014).

The highest PIC was observed for primer SSR 241(0.50) in SSR assay (Table 3). Shah et al., (2013), Benor et al., (2008) Zhou et al., (2015) and Korir et al., (2014) also reported similar results while studying the PIC of genomic SSRs and ESTT-SSR markers in various tomato lines. In the present study, SSR and RAPD markers detected medium locus polymorphism among the 24 tomato genotypes, indicating that both markers are of great utility for genetic diversity studies of tomatoes and can further be utilized in strengthening tomato breeding programmes. Further resolving power/discriminatory power of a marker, which indicates the discriminatory potential of the primer to distinguish the genotypes or individuals, was estimated for each primer. An average resolving power of 13.58 was observed for RAPD whereas for SSR it was 1.83. Highest resolving power of 44.58 was observed for primer OPA-02 among RAPD markers while as highest resolving power of 1.68 was observed for primer AW037347 among the SSR markers. Percentage of polymorphic bands was 100% for SSR markers (AI778183, AW037347, Y09371, SSR 241, X90770, SSR 74, X90937 and SSR 9) and for RAPD markers it ranged from 50% (OPA-02), 66.67% (OPY-15), 81.82% (OPA-13), 96.42% (OPC-14) and 100% (OPA-07,OPA-19, OPC-04,OPC-20,OPN-15,OPO-20,OPQ-20,OPU-16 and OPU-17).

Genetic relationship among tomato genotypes

Both the marker systems showed a high degree of similarity in the topology of their respective dendrograms. Although some differences in positioning of some genotypes was observed. However, all the dendrograms reflected similar pattern of relationship among most of the genotypes (Fig. 1A-C). In order to find out the genetic relationship among the tomato genotypes, analysis was done separately as well as in combination for RAPD and SSR data sets. The similarity coefficient for RAPD based diversity analysis ranged from 0.39 to 0.77 (Fig. 1A), whereas for SSR it ranged from 0.11 to 0.89 (Fig. 1B). Further, Similarity coefficient ranged from 0.37 to 0.72 for the combined RAPD and SSR based data sets (Fig. 1C). The dendrogram generated from RAPD data grouped genotypes in two main clusters, cluster A and cluster B, as represented in Fig. 1A. Cluster A was further divided into two sub clusters i.e. sub cluster A1 and sub cluster A2. Sub cluster A1 had 13 genotypes and (Utkal Pragyan, Hisar Lalit, DCT-1, Kashi Hemant, CO-3, Arka Abhay, ANGHA(L-E415), Dhanshri, Punjab Ratta, PANT-T-5, Hisar Anmol, FEB-2 and Swarna Sampada) and sub cluster A2 had 6 genotypes (AZAD-T-2, PT-11, NDUTUR-73, BT-136, SEL-12 and NDT-9). Cluster B consisted of further two sub clusters B1 and B2. Sub cluster B1 had 2 genotypes (ANGHA-1 and ANGHA) while Sub cluster B2 had 3 genotypes (NDT-1, AZAT-T-2 and NDT-4). The dendrogram showed genetic variation among the 24 genotypes of Tomato.
Table 1 List of genotypes of tomato used for study

| S. No | NAME OF GENOTYPES      | S. No | NAME OF GENOTYPES      |
|-------|------------------------|-------|------------------------|
| 1.    | UTKAL PRAGYAN          | 13.   | HISAR ANMOL            |
| 2.    | HISAR LALIT            | 14.   | AZAD-T-2               |
| 3.    | KASHI HEMANT           | 15.   | PT-11                  |
| 4.    | FEB-2                  | 16.   | NDTUR-73               |
| 5.    | DCT-1                  | 17.   | BT-136                 |
| 6.    | CO-3                   | 18.   | SEL-12                 |
| 7.    | ARKA ABHAY             | 19.   | NDT-9                  |
| 8.    | SWARNA SAMPADA         | 20.   | ANGHA-1                |
| 9.    | ANGHA(L-E415)          | 21.   | ANGHA                  |
| 10.   | DHANSHRI               | 22.   | NDT-1                  |
| 11.   | PUNJAB RATTA           | 23.   | ANAND TOMATO-3         |
| 12.   | PANT –T-5              | 24.   | NDT-4                  |

Table 2 RAPD primers with various parameters used for revealing the discriminatory power of each primer

| S.NO  | PRIMER | SEQUENCE  | PIC  | MI  | RP  | NB  | NPB | NMB  | NUB  | PPB  |
|-------|--------|-----------|------|-----|-----|-----|-----|------|------|------|
| 1     | OPA-07 | 5'GAAACGGGTG3' | 0.25 | 2.52 | 3.25 | 10  | 10  | 0    | 2    | 100  |
| 2     | OPA-13 | 5'CAGCAACCACAC3' | 0.26 | 1.93 | 18   | 11  | 9   | 2    | 0    | 81.8 2 |
| 3     | OPA-19 | 5'CAAAACGTCGG3'  | 0.28 | 4.62 | 6.83 | 16  | 16  | 0    | 3    | 100  |
| 4     | OPA-02 | 5'TGCCGAGCTG3'   | 0.27 | 0.67 | 44.58 | 10 | 5   | 5    | 0    | 50   |
| 5     | OPC-04 | 5'CGGCATCTAC3'   | 0.24 | 3.88 | 6.25 | 16  | 16  | 0    | 2    | 100  |
| 6     | OPC-14 | 5'TGCCGTGCTTG3'  | 0.29 | 7.78 | 24.83 | 28 | 27  | 1    | 1    | 96.4 2 |
| 7     | OPC-20 | 5'ACTTCCGCCAC3'  | 0.27 | 5.52 | 7.75 | 20  | 20  | 0    | 4    | 100  |
| 8     | OPN-15 | 5'CGCCGACTGT3'   | 0.30 | 1.84 | 7.16 | 6   | 6   | 0    | 0    | 100  |
| 9     | OPO-20 | 5'ACCCGGTCAC3'   | 0.26 | 11.02 | 9.75 | 41  | 41  | 0    | 7    | 100  |
| 10    | OPQ-20 | 5'TCGCCCCAGTC3'  | 0.30 | 3.32 | 9.75 | 11  | 11  | 0    | 1    | 100  |
| 11    | OPU-16 | 5'CTGCGCCTGGA3'  | 0.43 | 6.48 | 22.17 | 15 | 15  | 0    | 0    | 100  |
| 12    | OPY-15 | 5'AGTCCGCCCTT3'  | 0.25 | 1.34 | 11.75 | 12 | 8   | 4    | 3    | 66.6 7 |
| 13    | OPU-17 | 5'ACCTGGGGAG3'   | 0.37 | 2.62 | 4.5  | 7   | 7   | 0    | 0    | 100  |
| AVERAGE|       |             | 0.29 | 4.12 | 13.5 | 15.6 | 14.6 | 0.92 | 1.77 | s91.9 2 |

* NB: number of bands, NPB: number of polymorphic bands, NMB: number of monomorphic bands, NUB: number of unique bands, PPB: percentage of polymorphic bands, PIC: polymorphism information content, MI: marker index, Rp: resolving power.
Table 3: SSR marker with various parameters revealing the discriminatory power of primer

| S. No. | PRIMER  | SEQUENCE 5′→ 3′                  | PIC | MI  | RP   | NB  | NPB | NMB  | NUB | PPB |
|--------|---------|---------------------------------|-----|-----|------|-----|-----|------|-----|-----|
| 1      | AI773078| F:GATGGACACCCCTC AATTATGTT R:TCCAAGTATCAGGC ACACCAGC | 0.16| 1.28| 1.58 | 8   | 8   | 0    | 4   | 100 |
| 2      | AI778183| F:GCGAAGAAGATGAG TCTAGAGCATAG R:CTCTCTCCCATGAGT TCTCCTCTTC | 0.26| 1.12| 1.58 | 7   | 7   | 0    | 4   | 100 |
| 3      | AW037347| F:GCCACGTAGTCATG ATATACATAG R:GCCTCGGACAATGA ATTG | 0.21| 1.68| 2    | 8   | 8   | 0    | 2   | 100 |
| 4      | Y09371  | F:TGAGAACAACGTTT AGAGGAGCTG R:CGGGCAGAATCTCG AACTC | 0.16| 1.34| 1.58 | 8   | 8   | 0    | 4   | 100 |
| 5      | SSR 241 | F:TCCTCGCTAATTGAT CCACC R:TCAACACG CATAGTG GAGGAGG | 0.50| 0.99| 1.91 | 2   | 2   | 0    | 0   | 100 |
| 6      | X90770  | F:TGTAGATAA CCTTCTCTAGGACAATTC R:ACGGACGGATGGAC AAATG | 0.32| 1.29| 2.41 | 4   | 4   | 0    | 0   | 100 |
| 7      | SSR74   | F:ACTCACCATGGCTG CTTCTT R:TTTCTTTGAAGGGTCT TTCCCC | 0.46| 0.93| 1.75 | 2   | 2   | 0    | 0   | 100 |
| 8      | X90937  | F:TGCCCATGACGTTCC ATC R:GACAGACAGAGAGA CAGATTTAGAG | 0.29| 0.59| 1.91 | 2   | 2   | 0    | 0   | 100 |
| 9      | SSR 9   | F:CCCTTTGCAAGTTCT TCTCTCA R:TTCATGAGCCAACA TAGGAGG | 0.33| 0.13| 1.75 | 4   | 4   | 0    | 0   | 100 |
|        | AVERAGE |                                                | 0.30| 1.04| 1.83 | 5.00| 5.00| 5.00 | 0.00| 100 |

* NB: number of bands, NPB: number of polymorphic bands, NMB: number of monomorphic bands, NUB: number of unique bands, PPB: percentage of polymorphic bands, PIC: polymorphism information content, MI: marker index, Rp: resolving power.