World's first database based on tomato whole genome for gene mapping and variety identification

Available at: http://cabindb.iasri.res.in/tomsatdb

Availability: http://cabindb.iasri.res.in/tomsatdb

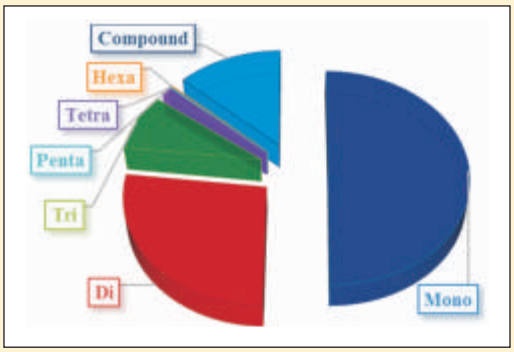
Tomato (*Solanumlycopersicon* L.), a new world solanaceous plant, is an excellent model for plant genomic research. The genus *Solanum* is one of the largest angiosperm genera and the genome has 35,000 genes spread over 12 chromosomes. The cultivated tomato is second most consumed vegetable of the world and is an important part of a diverse and balanced diet as a rich source of vitamins, minerals, phenolic antioxidants and antioxidant lycopene having anti-cancer properties.

Tomato genomics and its deciphering of biochemical pathway of fruit ripening can add more knowledge in understanding the exact mechanism in terms of genes involved, which is very critical information for further research linked to cold chain management affecting the economics of fruit productivity of various other horticultural crops. In tomato genome, chromosome wise DNA marker with their absolute physical distance was not reported but needed for fine mapping work for discovery of genes and QTL. Since traditionally, characterization of varieties is based on phenotypic observation and it is very difficult to distinguish varieties with very similar morphological characteristics, such DNA markers are needed for variety identification supplementing DUS test and product traceability in global market especially in upcoming IPR regime. Here, more than one million DNA markers are discovered and world's first database based on tomato whole genome for gene mapping and variety identification along with ready to use markers for global research community for direct genotyping.

Marker discovery and database development

The chromosome wise tomato whole genome data available in public domain, which has been sequenced by International Tomato Genome Consortium in 2012 partnering India was extracted in FASTA format. Using PERL script and MISA, information on STR numbers, motifs, repeat number, length and size of the repeat, repeat type, GC content, start and end position of the repeat and STR sequence were generated. An online relational database based on "three-tier architecture", cataloguing information of markers in MySQL has been created and user-friendly interface has been developed using PHP. Along with many flexible features, an automated primer designing tool, Primer3 has been integrated to obtain primers at desired location on selected chromosome.

Salient research findings



World's first whole genome based microsatellite DNA marker database of tomato, TomSatDB (Tomato MicroSatellite Database) contains more than 1.4 million markers. It was found that 87% and 13% of the STR markers were of simple and compound type, respectively. The "mono" repeat type (50.17%) was dominant followed by "di" (26.69%).The number of "hexa" repeat type (153) was found to be minimum (0.10%). Further, the length of STRs between 9-16 was most occurring (58.08%) followed by 5-8 and >16 as 22.06% and 6.63% respectively of total markers. It was also observed that chromosome 1 is having the highest number of markers while chromosome 6 exhibits minimum. Chromosome 8 shows highest density (177.8 markers/MBp) of markers and chromosome 2 reports minimum density of markers (128.4 markers/MBp), while the relative density of the tomato whole genome is 154.3 markers/MBp, showing that these markers are ubiquitously distributed with homogeneity in terms of distance, which is inherent attribute of microsatellite to be used as marker of choice.

Implications of this work and use of database

Identification of the cultivars accurately is essential for maintaining cultivar integrity and Plant Breeders' Rights. In era of IPR the new variety can be identified based on allelic variation among varieties supplementing DUS test and product traceability.

A total of 146602 STR markers are reported for the first time using whole genome in the database. These markers are expected to pave the way of germplasm management over abiotic and biotic stress as well as improvement through molecular breeding, leading to increased tomato productivity in various parts of the world. Apart from abiotic stress, there are more than 200 tomato diseases caused by pathogenic fungi, bacteria, viruses and nematodes affecting tomato productivity as biotic stress. In order to manage the germplasm in abiotic and biotic stress with desired productivity, such closely linked DNA markers are imperatively needed. Further economically and commercially important genes can be used for markers assisted introgression especially in new variety development programme.

Chromosomes	Simple						Compound	Total
	Mono	Di	Tri	Tetra	Penta	Hexa		
Chromosome 1	8995	4632	1476	194	35	19	2311	17662
Chromosome 2	6401	2746	867	95	15	6	1363	11493
Chromosome 3	6637	3370	1107	150	18	21	1572	12875
Chromosome 4	6332	3268	1076	115	24	13	1499	12327
Chromosome 5	5714	3300	1028	128	17	11	1632	11830
Chromosome 6	5105	2652	861	102	8	9	1219	9956
Chromosome 7	5801	3097	1045	137	22	11	1580	11693
Chromosome 8	5974	3052	966	136	19	11	1556	11714
Chromosome 9	5886	3275	1098	118	20	17	1613	12027
Chromosome 10	5369	3041	910	140	15	12	1417	10904
Chromosome 11	5005	2640	916	91	16	3	1360	10031
Chromosome 12	5365	3034	995	126	20	18	1687	11245
Chromosome 0	972	1018	231	33	3	2	586	2845
Total	**73556**	**39125**	**12576**	**1565**	**232**	**153**	**19395**	**146602**

These findings can be of immense use in tomato genomics research for tomato improvement and variety management at global level.

Publication: Iquebal MA, Sarika, Arora Vasu, VermaNidhi, Rai Anil and Kumar Dinesh (2013). First whole genome based microsatellite DNA marker database of tomato for mapping and variety identification. *BMC Plant Biology*, **13**:197.

Contact

Head
Centre for Agricultural Bioinformatics
IASRI, Library Avenue, PUSA
New Delhi - 110012, India
Email: hd.cabin@iasri.res.in

Director
Indian Agricultural Statistics Research Institute
Library Avenue, PUSA
New Delhi - 110012, India
Email: director@iasri.res.in

Developed by: Sarika, Mir Asif Iquebal, Anil Rai and Dinesh Kumar