

Review Article

<https://doi.org/10.20546/ijcmas.2020.903.235>

A Decade of Tomato Transcriptomics: Status and Perspectives

Kinjal Kulshrestha* and Akarsh Parihar

Centre of Excellence in Biotechnology, Department of Agricultural Biotechnology,
Anand Agricultural University, Anand-388110, Gujarat, India

*Corresponding author

ABSTRACT

Tomato is widely consumed vegetable in the world. It is used as raw and processed form. Tomato flowers are self as well as controlled pollinated. The crop has more than thirteen species including cultivated and wild types with huge diversity and is susceptible to diverse stress conditions. Recent advances in the biotic and abiotic stress in tomato transcriptomics have been reviewed here. Transcriptome is a particular set of RNA expressed during a stress, developmental stages or a condition in a group of cells. Transcriptomics is a promising technology for understanding molecular interactions during disease, interpretation of functional elements and molecular constituents of cell. Transcriptomics allows indexing all transcribed species such as mRNA & small RNA; to determine gene structure: start site, 5' and 3' ends, transcriptional modification, splicing patterns; and to quantify differential gene expression in stress condition. There are many techniques of transcriptomics such as RNA sequencing, Microarray, EST and SAGE. Transcriptomics by RNASeq is frequently used system as being cost effective. Sample preparation is a critical step for sequencing procedure. High quality RNA is essential. As transcriptomics yield huge amount of datasets, bioinformatics tools are required for proper data analysis. Gene ontology is obtained as molecular function, cellular components and biological processes. Panel of genes such as WRKY and phenyl propanoid were found to be expressing during stresses in tomato. Evaluation of differentially expressed genes paves the way for candidate gene selection. These genes may be focussed in constructive breeding programs. Researchers recommend the reference tomato genome enrichment and bioinformatics workforce surge.

Keywords

Solanum lycopersicum, transcriptomics, tomato genome, RNA Seq., bioinformatics

Article Info

Accepted:
15 February 2020
Available Online:
10 March 2020

Introduction

Tomato (*Solanum lycopersicum*) is one of the most indispensable vegetable which is cultivated and consumed throughout the world. Tomato is preferred raw as salad and in the processed form such as ketchup, paste, soup, puree etc. by all the age groups. Tomato is a self-pollinated crop in which controlled

cross pollination is easily implemented. It belongs to the family *Solanaceae*, bears $2n=2x=24$ chromosome number and berry type fruit (Anonymous, 2014, Table 1). A substantial amount of synteny is observed between the genomes of cultivated and wild tomato species (Chetelat and Ji, 2007). Tomato includes more than thirteen inter-fertile species some of which may interbreed

(Table. 2). After compound selection events, the present form of tomato has been evolved (Simmonds, 1976). Tomato is a fruit originated in the Andes region, South America (Fig. 1, Simmonds, 1976).

This region includes Peru, Bolivia, Ecuador and Chile these days. The first tomato cultivar to be domesticated is *S. lycopersicum* var. *cerasiforme* (Fig. 2). Multiple disease resistance sources are packed in wild species like *S. pimpinellifolium*, *S. peruvianum*, *S. hirsutum* and *S. habrochaitis* (Rick, 1990). *S. penillii*, *S. chilense* and *S. peruvianum* are the sources for abiotic stress tolerance.

Tomato ranks second, world-wide, in terms of consumption. China, India, USA, Turkey, Iran and Italy are the leading tomato producers in the world. In India, Himachal Pradesh, Uttar Pradesh, Andhra Pradesh, Maharashtra, Madhya Pradesh, Gujarat, Telangana, and Karnataka are the leading tomato producing states in India (Table 2).

The area, production and productivity of tomato is 5.02 Mha, 170.75 MMT and 33.99 MT/ha for the world whereas 0.80 Mha, 19.96 MMT and 24.34 MT/ha for India (Anonymous, 2018, Table 3). Tomato juice and pulp are mild aperients and blood purifiers. Tomato is a rich source of vitamin A, C, E and lycopene. Tomato plants are found with determinant type as well as indeterminate type growth habit. Tomato flower is a cymose type inflorescence.

The flower petals shed away upon fertilization. Pear shaped fruit starts developing which ultimately gains size and become red. The breeding objectives in tomato are as follows:

- a) Abiotic stress tolerance
- b) Biotic stress resistance
- c) Quality improvement
- d) Earliness

Abiotic stress includes drought and salt tolerance etc. Biotic stress includes different disease and insect-pest whereas fruit quality improvement includes fruit firmness, early maturity, lycopene content, total soluble sugar content and total soluble solids etc.

Major diseases in tomato are Early blight, Late blight, Fusarium wilt, Bacterial wilt, Damping off, Tomato Leaf Curl virus, Tomato Yellow leaf curl virus, Tomato Mosaic virus, Bacterial Spot, Bacterial Canker, Tomato Bunch Top Virus, Tomato Spotted Wilt Virus and Powdery Mildew (Anonymous, 2014).

The continuous selection process of cultivated species and its domestication has made tomato species susceptible to many stresses. Introgression from the wild species is recognized (Sahu and Chattopadhyay, 2017). Transcriptomics is a boon to spot the resistant gene for the gene transfer in the cultivated lines from the wild type plants and resistant germplasm for crop improvement.

Transcriptomics for crop improvement

Transcriptomics is the study of transcriptome which is a complete set of RNA transcripts that are produced by the genome, under specific circumstances or in a specific cell. Gene expression pattern reflects the molecular mechanism which governs the final biology and physiology of plant.

In this way, transcriptome is critical to be understood as it reveals the complex disease traits. Though we have many methods for studying the transcription processes like RNA Microarray, ESTs, SAGE etc. but RNA sequencing has emerged as the most exploited technology among all of them (Lowe *et al.*, 2017). RNA Seq is a unique technology as it has whole-transcriptome coverage. Traditional methods such as qPCR and

microarray have limited range, less sensitivity and above all, they rely on prior information of the genome. On the other hand, RNA Seq. offers studying novel transcripts as well as synteny. RNA Seq shows insights of functional pathways based on the particular gene expression during critical conditions. It delivers high sensitivity and accuracy along with lower cost per sample. The final number of sequences analyzed depends on the objective of the research work.

High throughput sequencing technology enables the researchers to locate the gene function quantitatively. Sequencing run allows us to zoom in a particular region of the genome. The disease resistance may be governed by single genes or group of genes. Transcriptomics is the platform to identify the candidate gene or genes which are up regulated or down regulated during certain stress. These genes can be studied for additional information and can be introgressed in the susceptible genotype through conventional or modern methods after complete evidence.

In Microarray, the diseased samples are hybridized with the probes on a chip and differential gene expression is studied (Table 4). Microarray is parallelism, miniaturization, automation and multiplexing process. There are two stages to design the chip namely by probe hybridization and measuring the gene expression of controlled and diseased samples. RNA Seq is used for studying GO (Gene Ontology), KEGG (Kyoto Encyclopedia of Genes and Genomes) and differential gene expression (Bolger *et al.*, 2017). Comparison of transcriptome allows the identification of genes that are differentially expressed in distinct cell populations or in response to different treatments or during diseases. To carry out transcriptomics, extreme genotypes like resistant and susceptible, sufficient selection

pressure, adequate infrastructure, appropriate software, platforms and databases are required.

Advantages of RNA Seq

- 1) *De novo* sequencing.
- 2) Qualitative as well as quantitative transcriptome analysis.
- 3) High sensitivity and accuracy.
- 4) Low cost in comparison to arrays.
- 5) Multiple conclusions can be drawn using correct bioinformatics tools.
- 6) Unaffected by the background noise and saturation of the signals.
- 7) Captures subtle gene expression.
- 8) Quantifies the digitally discrete read counts.
- 9) Highest numbers of DEGs may be netted.

Applications of transcriptomics in crop improvement

Facilitating a breeder by gene identification for disease resistant traits.

To explore the distinctive gene in wild species.

Searching the candidate genes for resistance and explore its biosynthesis.

Identification of the genetic basis of disease and resistance.

Annotation of the function of previously unannotated genes.

Identification of single nucleotide polymorphism (SNPs).

To carry out gene expression studies for non-model crops.

Identification of previously unknown protein coding regions in existing genome database.

To enrich the public data bases with the annotated sequences.

To get information about putative proteins.

Check points as challenges and way to combat them

RNA degradation

RNA being sensitive to RNase, utmost care should be taken while crushing the sample, dissolving the pellet and handling the aliquotes of RNA because RNase present on the working platform, instruments or old gloves will affect the final RNA quantity. Autoclaving of mortar-pestle, chloroform treatment of tips, pipettes and DEPC (Diethyl pyrocarbonate) water is recommended during the whole procedure.

Presence of genomic DNA in the sample

The presence of genomic DNA should be minimized by treating the sample with DNase and selection of proper RNA extraction method. DNA contamination may give high nucleic acid content during quantification but lower amount of actual RNA bind to adaptors. When the sample is loaded in agarose gel wells, DNA contamination gives a clear bright and heavy band while RNA appears in a smear form.

Quality check

Nucleic acids have absorbance of 260 nm but EDTA, phenols and carbohydrates absorb wavelength of 230 nm. The RNA with 260/230 ratio of 2, near 2 or above is advisable for transcriptomics studies. If it is lower than 2, it indicates presence of phenols. Lower 260/280 ratio (less than 2) indicates the presence of protein in the sample. The pure nucleic acid has higher 260/230 ratio when compared with 260/280. The ratio gives an idea about protein or DNA contamination.

Techniques under transcriptomics

The extracted RNA is subjected to cDNA fragments with adapters either on one side or

both the sides. Then each molecule is sequenced by single end sequencing or paired end sequencing. The read length varies between 30-400 bp based upon the technology used. The proper technique should be selected based on the aim of the study. Samples under a gradient of stress can be compared by this technique. Novel transcripts are also mapped in RNA Seq while microarray requires predefined probes for the DNA fragment hybridization. RNA Seq reveals precise location of transcription factors and sequence variation. RNA Seq has very low background noise. Splicing diversity is also explored by RNA Seq during splice junction analysis.

Validation

Validation gives the strength to the research as it validates the gene expression in the samples. Validation is done by reverse transcription polymerase chain reaction (RT-PCR). Real-time PCR technology records the amplification in real time. Real-time PCR reveals the expression in terms of absolute and real quantification. Absolute quantification concludes the input copy number by using standard curve. Relative quantification of target genes is done by comparing the expression with a reference gene or control group.

Tomato genome

The tomato genome was published in Nature on 31st May, 2012. Its genome size is 950MB. The Tomato Genome Consortium 2012, consists of over a hundred of researchers from 14 countries. IARI, New Delhi also contributed in the research. Initially, the gene rich regions of 12 chromosomes were sequenced using BAC-by-BAC sequencing approach. In 2008, the whole genome sequencing approach was implemented. Inbred tomato cultivar 'Heinz 1706' from Pittsburgh, PA, USA was used for sequencing

using Sanger and 'next generation' technologies. In 2008, Shotgun sequencing was performed using Sanger sequencing method. The predicted genome size was 900 Mb from which 760 Mb assembled into 91 scaffolds aligning to 12 tomato chromosomes. Base accuracy of one substitution error per 29.4 kilobases and one indel error per 6.4kb was retained.

Genome of *S. pimpinellifolium* LA1589 was assembled *de novo* which yielded 739 Mb of data sets. Divergence between wild and domesticated tomato genomes was 0.6% hence, 5.4 million single nucleotide polymorphisms (SNPs) were distributed along the genome. Tomato chromosome has pericentichromatin and distal euchromatin region. The genome was found to be highly syntenic with other *Solanaceae* members such as potato, eggplant, pepper and nicotiana. Comparative genomics study has reported 34,727 and 35,004 protein coding genes in tomato and potato respectively.

From the above genes 31,741 and 32,056 genes respectively show synteny to *Arabidopsis* genes. All the protein coding genes of tomato, potato, *Arabidopsis*, rice and grape were clustered in 23,208 gene groups and 8,615 gene groups were common to all. Eudicot genomes (tomato, potato, grape and *Arabidopsis*) has 1,727 gene groups while 727 was common in fleshy fruits (tomato, potato and grape). Small RNA sequencing predicted 96 and 120 conserved miRNA genes in tomato and potato respectively.

Total 34 miRNA families were identified among which 10 are highly conserved in plant kingdom. Tomato-potato orthologous regions approve nine large inversions based on cytological or genetic studies. Total 18,320 orthologous tomato-potato gene pairs were identified accurately. Alignment of 71MB of euchromatic tomato genome revealed 8.7% of

nucleotide divergence from potato genome whereas heterochromatic sequences showed 30% of nucleotide divergence.

Comparative mapping of 31,760 'Heinz 1706' genes and *S. pimpinellifolium* revealed 7,378 identical genes, 11,753 synonymous changes and 12,629 non-synonymous changes. 'Heinz 1706' carried introgression from *S. pimpinellifolium* detectable on chromosomes 4, 9, 11 and 12.

Comparative transcriptome of tomato and grape supports the hypothesis that a whole-genome triplication affected the rosid lineage in common eudicot ancestor (The Tomato Genome Consortium, 2012).

NGS platforms used were Roche/454, SOLiD and Illumina sequencing which provided 21 GB, 64 GB and 82 GB data respectively. Newbler and CABOG programs were used to generate the independent assemblies (Sato and Tabata, 2016).

Recent advances of transcriptomics in tomato

Recent advances of transcriptomics in tomato found to be distributed as the following areas of work

Comparative transcriptome analysis of cultivated and wild species of tomato

Comparative transcriptome analysis of tomato cultivars

Transcriptomic analysis during abiotic stresses in tomato (drought & salt tolerance)

Transcriptomics during biotic stresses in tomato (nematode, TYLCV, fungi and early blight)

Transcriptomic analysis of pollen developmental stages of tomato

Transcriptomic analysis of hormone effect on tomato

Comparative transcriptome analysis of cultivated and wild species of tomato

The section Lycopersicon includes about 13 tomato species. The cultivated tomato is *S. lycopersicum* whereas *S. pennellii*, *S. pimpinellifolium*, *S. cheesmaniae*, *S. galpagensis*, *S. chemelewskii*, *S. arcanum*, *S. neorickii* and *S. huaylasense* are wild tomato species. Domestication is affected by continuous selection along with human migration. Suitable wild accessions are selected according to phenotype, adaptability and nutritional value. Better yielding loci are selected recurrently. Positive selection of some loci leads to lack of diversity. Domestication of tomato was oriented towards increase in yield and fruit size. These selection events can be studied during comparative high throughput sequencing. The present cultivated tomato was domesticated from *S. pimpinellifolium*. First domesticated cultivar is represented by *S. lycopersicum* var. *cerasiforme*. Tomato was re-introduced in America by Europeans about 300 years ago. Wild tomato varieties have wide genetic phenotypic base but the cultivated tomato is with selected genes. Transcriptome analysis of cultivated and wild tomato species seems to be informative for understanding the difference at molecular levels. Wild tomato species are sources of resistance for certain diseases. During resistance, the pattern of transcription factor plays an important role.

SRA sequences (Koenig *et al.*, 2013) of *S. lycopersicum* and *S. pennellii* were taken as cultivated and wild species respectively, which offered different gene expression due to natural and artificial selection. Seven different tissue types were studied namely root, stem, leaf, flowers, fruits, seedlings and vegetative portions. Total 62,676 transcripts were generated. The quality of RNA-Seq data was studied by FastQC software. Alignment of RNA Seq data was done by Tophat and

transcripts were assembled by Cufflinks. Cuffmerge was used for merging the assemblies. Differentially expressed genes, transcripts and differential splicing were recognized by cuffdiffat False discovery rate FDR<0.05 with Python script. Long non-coding RNA was selected on the basis of size (length \geq 200bp), ORF filter was \leq 100 and the coding potential calculator (CPC) prediction was used for non-coding sequence detection. Gene ontology was done by AgriGO was used for gene ontology. Gene ontology was enriched by Hypermetric extract test at FDR<0.05 (threshold). Pathway analysis was done by KOBASS. Total genes expressed in cultivated tomato were 18,719 and in wild were 18,609. Most of the genes were found to be co-expressed in both (17,202 genes). Genes expressed only in cultivated type were, 1,517 and in wild were 1,407. Differential genes were found to be expressed in floral, fruit, vegetative and root tissues. Total 618 genes were expressed in roots. Root samples had gene ontology (GO) for cellular composition (CC) and molecular function (MF). Fruits had GO for biological processes (BP). The pathways obtained were for heat and salt stress tolerance. The differences in gene expression were found mostly in carbon metabolism, sucrose metabolism, amino acid biosynthesis and gluconeogenesis. Differentially expressed long non coding RNA were 554 however coding were 426 (Dia *et al.*, 2017).

Genome-wide sequence variation between wild and cultivated tomato species by whole genome sequence mapping has been studied (Kumar and Chattopadhyay, 2017). Tomato sequences were taken from The European Bioinformatics Institute (EBI) database. These sequences were submitted by Aflitos *et al.*, (2014). SNPs were detected and the effects of them were analyzed by SnpEff. The bcf files of whole data were converted to vcf file by samtools, then from vcf to hapmap.

Hapmap diversity was analyzed by Tassel, Pi, Theta, Fst and Tajima's D values. gff annotation file plotted the stress resistant genes. SNP and InDel mining was analyzed for 29 accessions of 12 wild and 40 accessions of *S. lycopersicum*. All the wild species have unique set of SNP which was absent in the cultivated types. Low sequence variation was observed among red fruited species like *S. pimpinellifolium*, *S. cheesmaniae* and *S. galpagenses* when compared to cultivated tomato sequences. The green fruited species like *S. chemelewskii*, *S. arcanum*, *S. neorickii* and *S. huaylasense*, showed much variation when compared to *S. lycopersicum* sequences. Common SNPs were least in *S. pennellii* and lower in *S. arcanum*, *S. huaylasense* and *S. habrochaites*. The strategy concluded that, there is a wide genetic variation among the wild accessions. *S. peruvianum* and *S. chilense* had 50% variation: lower SNPs. *S. arcanum* had highly divergent accessions. The ratio of genic SNPs to total SNPs, increased with increase in phylogenetic distance. *S. pimpinellifolium* have highly dense SNPs in the chromosome 1, 3, 4 and 8. Chromosomes 5 and 7 had the lowest SNP numbers within the same species. *S. cheesmaniae* and *S. galpagense* are very similar but entirely different from *S. pimpinellifolium*. *S. cheesmaniae* and *S. galpagense* are known as morphotypes originated on the Galpagos Islands. *S. pimpinellifolium* was originated in Andese. Arcanum subsection species, *S. chmelewskii*, *S. neorickii* and *S. arcanum* appear to be similar on the basis of SNP. The first two are closer and are originated on Andese mountain ranges. The third species was originated on Andean coast.

The genetic variation by sequencing 84 tomato accessions and related wild species was discovered (Aflitos *et al.*, 2014). Groups involved in the study were Lycopersicon, Arcanum, Eriopersicon and Neolycopersicon.

The full genetic diversity of tomato crop was explored. DNA was isolated by standard DNA isolation protocol (Van der Beek *et al.*, 1992). Qubit 2.0 was used for DNA quantification. DNA sequencing was done by IlluminaHiSeq 2000 for mapping of *S. lycopersicum* cv. Heinz v 2.40. The maximum insert size was 750 bp. Samtools was used for variant calling with minimum gap distance of 5 bp, minimum alignment quality of 20 and minimum depth equal to 4. Bowtie was used to check the contamination of Escherichia coli, human, insect, mouse, bacteriophage, yeast and phytoviral genomes. *De novo* sequencing of *S. arcanum*(LA2157), *S. habrochaites*(LYC4) and *S. pennellii*(LA0716) was done by using Illumia HiSeq2000. All the data for *S. pennellii*, *S. habrochaites*, *S. arcanum* and *S. lycopersicum* was assembled using AllPaths-LG. *S. arcanum* was also assembled by CLC workbench v 7 with bubble size 300, minimum contig length 200 and a word size of 64. The scaffolds obtained by AllPaths-LG were further scaffolded using 454FLX data, Scarpascaffolder. Resultant *de novo* assembly statistics were compared with tomato reference genome, *S. lycopersicum* cv. Heinz version SL 2.40. *S. arcanum* version 1.0, *S. habrochaites* version 2.0 and *S. pennellii* version 2.0 were used for mapping of 84 accessions. The sequence diversity was assessed by BLASTN hits of CLC assembled contigs and ITGA v 2.4 annotation. Sequences were aligned by Clustal W. SNP calling was done by quality based variant detection algorithm of CLC. BWA and samtools were used to generate vcf files. All vcf files were processed by SNPEFF 3.4 and ITGA annotation was done. Parameters such as synonymous and non-synonymous SNPs, heterozygosity levels, the number of SNPs per 1 Mbp bins and location of SNP were assessed. JBrowse 1.10.12 was employed to detect the structural variants. The assembly of SL2.40 and ITGA 2.32 genome annotation was laden with the vcf files of 84 accessions.

The sequences were deposited at European Nucleotide Archive with the accession numbers PRJEB5226 (*S. arcanum* LA2157), PRJEB5227 (*S. habrochaites* LYC4), PREB5228 (*S. pennellii*) and PRJEB5235. The whole genome sequencing coverage was found 36 fold and SNPs in wild species were more than 10 million. Allogamous self-incompatible wild species showed the highest heterozygosity. The habitat and the origin of species influenced the phylogenetic relationships. The mapped reads for *S. habrochaites*, *S. arcanum* and *S. pennellii*, were 78%, 73% and 53% respectively. This showed the presence of interspecies variation. Reconstruction of additional reference genomes for tomato was recommended. There was 1% each of synonymous and non-synonymous SNPs. A segment of 2.2Mb on chromosome 6 was found to be introgressed in *S. lycopersicum* lines namely LA2838A, LA2706 and CGN15820 and *S. pimpinellifolium* (LYC2798) was found to be the closest donor for that.

RNA sequencing for identification of changes in the DNA sequences and gene expressions in cultivated and wild species of tomato was performed (Koenig *et al.*, 2013). RNA sample of *S. lycopersicum*, *S. pennellii*, *S. habrochaites*, *S. pimpinellifolium*, *S. chmielewskii* and *S. galpagensis* were used. All the species selected were diverse in their habitat as well as the genetic content. RNA was extracted using Trizol and RNeasy kit. Sequencing was done by Illumina GA II and HiSeq 2000, SNP detection, indels between samples and reference were studied by Biopearl script.

Statistical analysis was done by R-statistical programming environment. The samples aligned upto 67.4% to the reference genome. The reads covered 54% of annotated genes and 34 novel transcripts were found which differed from the published *S. lycopersicum*

genome. Cultivated accessions had <1 SNP/kb. Total 51 genes showed statistical significance ($P < 0.05$) for evolution through positive selection. In the centromeres, there was low amount of gene expression when compared to the whole chromosome. This chromosomal part had high gene density. *S. pennellii* had the highest expression variability among all the species. *S. lycopersicum* showed divergence in the expression.

It was concluded that redox pathway had been evolved as it was having much higher expression in the wild species i.e., *S. pennellii*, *S. habrochaites* and *S. pimpinellifolium*. The pathway may be useful for these species as they belong to tough environmental conditions. *S. pennellii* governed many genes which are responsible for abscisic acid pathway, salt, drought, oxidative damage and heat stress etc. *S. pennellii* lineage displayed the highest proportion of expression changes. This proved that *S. pennellii* is highly diverse when compared to all other species in the study. Connective genes were identified among *S. pennellii* and *S. lycopersicum* in the form of modules. Few genes were highly connected to *S. pennellii*, though they were present in both the species. The genes for photosynthetic tissue specific expression were correlating *S. pennellii* and *S. lycopersicum*.

Comparative transcriptome analysis of tomato cultivars

The cultivars are selected on the basis of desired characteristics and are maintained carefully for breeding and selection. The common names of the cultivars are usually inspired by region or the developers. Due to difference in the origin, worldwide, tomato cultivars have few genetic differences. AT3, SL120, Hissar Lalit, Arka Vishal, Rajshree, Rupalietc are popular tomato cultivars in Indian.

Mnichal (*S. lycopersicum*) as normal (N) and stunted (S) cherry tomato plants was sequenced (Rahim *et al.*, 2018). RNA was extracted by RNeasy mini kit. TrueSeq RNA library preparation kit (Illumina) was used for library preparation. IlluminaHiSeq 2000 was used as a genome sequencer. The transcripts were mapped to Ensemble database with TopHat v.2.1.1. Assembly was examined by Swiss Prot database and for gene ontology. Cufflink gave the expression pattern. DEGs were found by DESeq at $p=0.005$ and $q=0.05$. GO and KEGG were done by WebGestalt and DAVID. qRT-PCR was done by LightCycler 96, Roche Life Science, Germany and the data was analyzed by LightCycler 96 software. There were total 117.99 Million paired reads. Clean reads were 115.45 Million. Mapping of Mnichal (N) matched 97.6% and of cherry tomato (S), 98.0% with *S. lycopersicum* genome. There were 35,216 transcripts and 35,216 genes in N and S respectively. Seventy two percentage of the sequence of N had 90-100% coverage, while for S, it was 71%. At Q20, 96.60% sequences were represented for N and 96.81% for S. Total DEGs in N and S were 661 and total 420 genes were up-regulated whereas 214 were down-regulated, 32 genes expressed only in S and 108 only in N.

The gene ontology showed genes for metabolic process, response to stimulus and biological regulation under biological processes (BP), whereas for membrane and nucleus under cellular components (CC), and protein binding, ion binding, nucleic acid binding and hydrolyase activity under molecular function (MF). The enriched GO terms gave the genes for catalytic activity and metabolic process. The functional annotation clustering showed 22 clusters with score in the range of 0.02 to 2.37. Among them, six clusters recorded enrichment score greater than 1.0. Highest enriched terms were for steroid biosystem, WRKY transcription

factor, DNA damage/repair, tetratricopeptide repeat, MADS-box TF and mitogen activated protein kinase. Transcripts for RPA3B, RPA2B, XRCC3 and RPA1E are related to homologous recombination; *Adenylateisopentyl transferase-3*, related to cytokinin biosynthesis and *cytokinin oxidase 3 (CKX3)* catalyzing the degradation of cytokinine, were up-regulated in S, cherry tomato. Genes 3BETAHSD/D2, DWF5 and DIM are related to steroid biosynthesis and were down-regulated in S. Genes of auxin signaling pathway namely *IAA14*, *AX6B_SOYBN*, *AXX15_SOYBN* and *12KD_FRAAN* were found to be up-regulated however *AIR12* was down-regulated in stunted tomato plants. Ethylene biosynthetic genes viz. *1-aminocyclopropane-1-carboxylate oxidase 1 (ACO1)* and *1-aminocyclopropane-1-carboxylate oxidase 3 (AOC3)* have a higher expression in S. *ERF* (ethylene responsive genes), specifically *ERF13* was down-regulated in S. *ERF003* and *ERF13* exhibited nil expression in N. WRKY TF genes viz. *WRKY 40*, *WRKY41*, *WRKY50* and *WRKY51* found to be up-regulated in S. MADS box TF genes including *AGL36* and *SEPALLATA 2* were up-regulated while *SVP* and *AGL19* were down-regulated in S. *YDA* (MAPK) (cell expansion) was down-regulated in S and the TRP like genes, namely *FKBP65*, *LPA1* and *NOXY38* were down-regulated while *ATSD11* was up-regulated in S.

The transcriptome of tomato (*S. lycopersicum*) cultivars Ailsa Craig (AC) and HG-6-61 was evaluated at seven stages viz. 7DAF (Days After Flowering), 14DAF, 21DAF, 28DAF, 35DAF, 42DAF and 49DAF by Ye *et al.*, (2015). The RNA was extracted by Green spin RNA quick extraction. Sequencing was done by Illumina AnalyzerIix. The clean reads were mapped to SL2.40 version of tomato genome from SOL Genomics, using TopHat (v 2.0.4). Differential expression was analyzed by

EdgeR and the GO enrichment was done by DAVID (Database for Annotation, Visualization and Integrated Discovery). Primers for qRT-PCR were designed by Primer 3. The ascorbic acid, carotenoid and flavonoid content were estimated using HPLC. Sequences of Heinz and *S. pimpinellifolium* were downloaded from TFGD (Tomato Functional Genomic Database). Co-expression analysis was done by CORREL, Excel 2003. cDNAs of transcription factors of MYB, NAC and ZIF were amplified. Lycopene (carotenoid) and naringenin-chalcone (flavonoid) were found to be higher in AC earlier to HG-6-61. Ascorbic acid was higher in HG-6-61 at 49 DAF when compared to AC. Though, carotenoid content was increasing with maturity in both the samples but rate of accumulation was faster in AC. All the flavonoids increased with maturity except chlorogenic acid and rutin. The RNASeq yielded 9.5 M reads per sample. For AC, 95.88% and for HG6-61, 96.03% of the sequences were uniquely mapped to *S. lycopersicum*. Average CDS for AC was 77.31% although for HG6-61 it was 76.65%. Total 26,392 genes were found to be expressed, which is 76% of 34,727 genes in the reference genome. This showed a saturated coverage of expression. Total 14,758 genes were expressed in all the stages and five of these were most expressed in AC at all the stages with RPKM>4000. Two of these genes, *Solyc05g0530702* and *Solyc05g0540902* are located on the chromosome number 5 and encode a protein with unknown function. *Solyc01g110700.2* is located on chromosome number 1 and codes for pre-mRNA splicing factor ATP dependent RNA helicase. *Solyc01g109660.2* is also located on chromosome 1 and codes for CCR 2 glycine-rich RNA binding protein. *Solyc11g008510.1* codes for 60s ribosomal protein. K-mean cluster analysis of 26,397 genes' expression pattern gave 20 groups. The largest group was of 3,523 genes (13.2%)

which had a stable expression at 14DAF, 28DAF, 35DAF, 42DAF and 48DAF. This group has genes for cell wall, protein modulation and RNA regulation. The second largest group had 2,763 genes (10.5%), stable at all the stages and codes for protein and development categories. The third largest group had 2,130 genes (8.1%), with declining activity from 7DAF to 42DAF but activity level arose at 49DAF. AC had 3,531, 5,576, 7616, 8,645, 8,968 and 9,878 differentially expressed genes at 14DAF, 21DAF, 28DAF, 35DAF, 42DAF and 49DAF, respectively. There were 20 transcription factors for flavonoid metabolic pathways among them 12 were positively correlated and eight had negative correlation. Total 37 transcription factor were there for carotenoid biosynthesis pathway. MADS box showed high correlation with carotenoid synthesis but the ascorbic acid metabolic pathway had positive correlation with oxidized ascorbate accumulation.

Transcriptomic analysis during abiotic stresses in tomato

Stress affects the plant's metabolism, growth and root development in numerous ways. Response towards the stress is governed by the interactions at molecular level. Abiotic stress is caused by non-living entities like salt, water logging or drought.

Salt stress

Various accessions of *S. pimpinellifolium* are reported to possess high salinity tolerance and thus are the reliable source of salt tolerant alleles. Though many QTLs (quantitative trait loci) are also identified in *S. pimpinellifolium* for biotic stress (Chen *et al.*, 2013), abiotic stress (Cheng *et al.*, 2010), fruit quality traits (Capel *et al.*, 2015) and other agronomic traits. *S. pimpinellifolium* has been evolved in challenging environmental conditions and

carries phenotypic robustness while the cultivated species of tomato has lost those traits during domestication. Salt stress is economically important factor in plant growth. This type of stress is reported to disturb 20% of irrigated land and reduce the food production by one-third (Machado and Serralheiro, 2017). Ion toxicity, oxidative stress and nutritional disorder are created throughout the salt stress. The water up-take is affected due to high concentration of sodium chloride in cytosol and cell organelles. These all cause unstable production. The combined application of NaCl and CaCl₂ has been reported to increase antioxidant activity in plants. NaCl is known to delay the tomato seed germination time, seedling length and vigor. Calcium supplement to the seedlings alleviated the effect of salt.

The sequencing of *S. pimpinellifolium* (LA 0480) was done and 15 genes for salt tolerance were elucidated (Razali *et al.*, 2018). *S. pimpinellifolium* is the closest species to the cultivated tomato i.e., *S. lycopersicum* (Tomato Genome Consortium, 2012). Total 25,134 protein coding genes were annotated. *S. pimpinellifolium* has small red fruits and is facultative autogamous. The species is well distributed in costal Peru, Equador and Chile. So it has genes for salt tolerance due to exposure to the brakish ground water. The sequencing was done by IlluminaHiSeq 2000, with the depth of 197x which gave data of 811Mb (N50, 75,736 bp). RNA was extracted by ZR Plant RNA mini preparation kit, Zymo. Library was prepared by NEBNext Ultra RNA library preparation kit. Trimmomatic was used for removing adaptor sequences. Trinity v 2.0.6 was the assembler. TransRate (reference free quality assessment tool) removed all the low quality transcripts. BUSCO (Benchmarking Universal Single Copy Orthologs) searched the completeness of the genome assembly and

protein annotation was done using SwissProt. Functional domain, protein signature and associated gene ontology was found using InterProscan. OrthoMCL was employed to read orthologous and paralogous protein relationships between different species in the study. BWA was the aligner while CIROS was used for circular plot. The SNP calling was done by samtools. Inositol was measured by K-INOSL assay kit. DEAP was developed and used for annotating protein coding genes. Inositol-3-phosphate synthase and phosphatase genes were significantly higher in number in *S. pimpinellifolium*. So it was concluded that inositol pathway may play a role in salt tolerance of LA 0480. Genes which had orthologs in *S. pimpinellifolium* and *S. lycopersicum* are *AtCIPK24* (Osmotic signaling stress signaling), *AtHKT1* and *AtDREB2A* (Ion exclusion from shoot), *AtSOS*, *SINHX1*, *AtNHX3* (Tissue tolerance vacuolar Na⁺ compartmentation); *AtVPI.1*, *PcMIP*, *tomPRO2* (tissue tolerance increased proton pumping), *AtTPS1* (Tissue tolerance-Synthesis of compatible solutes), *SIAPX*, *AtAPX1*, *SIGST*, *AvSOD* and *AtMDAR1* (Tissue tolerance- degradation of reactive oxygen species).

Drought

Plants are called to be drought stress when the transpirational demand is higher than the root capacity. The reasons of drought are high temperature, windy environment and rapid growth. High humidity also reduces the transpirational water movement. The drought phenotype resembles late blight in tomato. So it is crucial to identify the correct cause of plant health retardation by molecular biology tools. Under drought stress, net photosynthesis, stomatal conductance, intercellular CO₂ concentration, gross O₂ uptake evolution and gross O₂ uptake decline. Tomato plant avoid photodamage by downregulating PSII activity, emitting

substantial portion of light as heat, declined CO₂ assimilation, reduced stomatal conductance and using oxygen as alternative electron acceptor in photorespiration or Mehler reaction.

Tomato samples under drought and re-watered condition were compared by RNA sequencing (Lovieno *et al.*, 2016). The cultivar selected was M82 (LA3475) which was from TGRC, California. RNA was extracted by Trizol method. cDNA libraries with IlluminaTruseq RNA kit were prepared and sequencing was done by IlluminaHiSeq 1500 platform. Raw reads were trimmed by TrimGalorepackage (www.bioinformatics.babraham.ac.uk). Cutadapt was used to remove the adapters. Quality was checked by FastQC. Mapping was done by Bowtie (v 2.2.0) and TopHat (v 2.0.8). Stomatal conductance, CO₂ assimilation, chlorophyll fluorescence, ABA, proline content and genes for photosystem were measured in the samples. Many DEGs were found to be down-regulated during the drought stress. Drought samples namely Dr1 and Dr2 were compared with watered plants (WW and RW).

Total 119 DEGs were common in all of them. Genes for histone, cell modifying enzymes, heat shock proteins and chlorophyll were found in the groups. Gene ontology showed that genes for photosynthetic light harvesting, chlorophyll a/b binding and cell wall modification (Pectinases, *Solyc09g075350*) were down regulated in Dr1 and Dr2 samples. Heat shock proteins were up regulated in Dr1 and Dr2 however Histones and chlorophyll genes were down regulated in Dr1 and Dr2. Pyrroline-5-carboxylate synthase (*P5CS*) and 9-cis-epoxycarotenoid dioxygenase (*NCED*) coding genes, *Solyc08g043170.2.1* and *Solyc07g056570.1* respectively, were validated by qPCR. Sample Dr1 had high expression of *P5CS* and Dr2 had for *NCED*.

Transcriptomics during biotic stresses in tomato

Biotic stress is caused by biological agents. Bacterial, fungal, nematode infections are common obstacles for plant health. After a stress, the stress response is initiated at cellular level which activate signal transduction pathways that transmit information within the individual cell and throughout the plant. This phenomenon leads changes in many gene networks. Biotic stress activated R proteins and results in hypersensitive response (HR) and plant immunity. There is H₂O₂ accumulates at the site of pathogen attack (oxidative burst) induction of programmed cell death (PCD). HR, PCD and systematic acquired resistance (SAR) are characterized by higher expression of several genes which codes for PR proteins. Mitogen-activated protein kinase (MAPK) cascade is present in higher plants which plays an important role in signal transduction in response to hormone, biological signals, pathogen attack and environmental stress (Gorovits and Czosnek, 2007).

Nematode

Nematodes are obligate endoparasite of plants. They damage the host root system severely. Nematode stress affect the nutrient and water uptake of plants hence plant growth is retarded. Due to gall formation, coarse roots increases. Fine root system is indispensable for healthy plant but root galls reduce the hydraulic conductivity of the root system. Nematode infected plants have lower leaf water potential hence lower stomatal conductivity, transpiration and photosynthesis. The affected cells are the nematode feeding sites known as giant cells. Two oesophageal gland types namely two subventral glands and one dorsal gland are promoting effector production. Plant parasitic nematode has several cuticle proteins which

are crucial for parasitism. Plant defense system, is suppressed during the nematode infection (Iberkleid *et al.*, 2015).

The differentially expressed genes in susceptible and resistant tomato during nematode infection were identified (Shukla *et al.*, 2018). The nematode susceptible cultivar Pusa Ruby (PR) and Moneymaker (MM) as a nematode resistant samples were taken. Nematode infected root samples were taken at different infection stages. Stage 1: 1, 2, 3 dpi (days post nematode infection), stage 2: 5, 6, 7 dpi; stage 3: 13,14, 15 dpi; stage 4: 18, 19, 20 dpi and stage 5: 26, 27, 28 dpi for Pusa Ruby. Money Maker had only two stages of the above as 1 and 2. The RNA was extracted by Trizol method. HiSeq 2000, Illumina was used as sequencing platform. Demultiplexing of data was done by CASAVA whereas TopHat was used for the alignment while annotation was done by PANTHER database and transcription factors were identified by BLASTX. DEGs were recognized by DESeq2. Gene enrichment was done by AgriGO. Functional categorization was done by using MapMan. The sequencing data had 1,154,560,291 pair-ended reads for replicate 1 and 537,461,341 single-end reads for replicate 2 of 100bp in length. Genes mapped to *S. lycopersicum*, assembly SL2.50 were 72-92% in all the samples. Total reads mapped to *Meleidogyne incognita* were 0.1-6.3%. The resultant raw data and the processed data were submitted in GEO (Gene Expression Omnibus) repository of NCBI (National Center for Biotechnology Information) with accession number GSE88763 and SRA accession number SRP091567. In PR, 24,411 genes were found to be expressed, out of which, 1,827 were significantly differentially expressed. In stage 2, 18 DEGs; stage 3, 905 DEGs; stage 4, 1,054 DEGs and stage 5, 1,308 DEGs were found. In MM, 23,393 genes were expressed. There was absence of DEGs at stage 1 in MM but stage 2 had 25

DEGs. Genes found to be differentially expressed were for cell wall degradation: (43 genes), cell wall modification (29 genes), cell wall protein (5 genes) and cell wall synthesis under differential regulation for altered cell wall architecture. For developmental genes; cell cycle (3 genes), cytoskeletal organization (2 genes), root cap proteins (7 genes), transcription factor controlling developmental processes (5 genes) were found to be differentially expressed. Genes involved in ethylene and jasmonic acid biosynthesis were also found to be differentially expressed. Solyc01g080500.2 and tubulin alpha chain: Solyc08g006890.2 displayed uniform expression during all the stages and hence, were used as controls.

Tomato Yellow Leaf Curl Virus

Tomato Yellow Leaf Curl Virus (TYLCV) is a monopartite geminivirus complex which belongs to the genus *Begomovirus* of the family Geminiviridae. EPPO code for the virus is TYLCV0. White fly *Bemisia tabaci* is the causal organism for TYLCV but potential seed transmission of TYLCV is also noted in tomato. It is among 'top ten' pathologically important plant virus. TYLCV has a single-stranded circular DNA genome of 2.8 kb which is encapsidated in a twinned icosahedral virion. TYLCV cause plant stunting, leaf curling and yellowing hence severe yield losses. Other than tomato, TYLCV hosts pepper (*Capsicum* species) common beans (*Phaseolus vulgaris*) cucurbit (*Cucumins* species) and estoma (*Eustoma grandiflora*). TYLCV was first reported in Middle East in 1931. TYLCV is prominent in tropical and subtropical regions (Kil *et al.*, 2016).

The 2787 nucleotide long genome of TYLCV encodes two large open-reading frames (ORF) on viral strand namely VI and V2, four on complementary strand namely C1 to C4. V1

encodes coat protein and V2 encodes proteins for virus movements and suppression of host defense response. C1 encodes replication associated protein, C2 transcriptional activator protein, C3 a replication enhancer and C4 a movement determinant. TYLCV has been reported to cause 5-100% infectious tomato plants in India, 20% tomato yield loss in USA, 30-100% in the Caribbean Islands, Mexico, Central America, Venezuela and 100% in Jamaica (CABI, 2020).

The transcriptome of TYLCV resistant *S. habrochaites* (R) and susceptible cultivars *S. lycopersicum* (S) was analyzed by Sade *et al.*, (2013, 2015). The resistant & susceptible samples responded differently during the diseased condition. The resistant cultivar showed proper fruit development and low virus content. The susceptible cultivar was found devoid of fruits and showed disease symptoms. Four loci which are linked to the TYLCV resistance namely Ty-1/Ty-3 and Ty-4 in *S. chilense*, Ty-2 in *S. habrochaites*, Ty-5 in *S. peruvianum* were identified. Total 69 genes were found to be expressed in TYLCV R plant before and upto 7 days after infection (Eybishtz *et al.*, 2009). The transcriptome of S and R plants were compared to analyze the changes in *LeHT1*, the hexose transporter gene. The transcriptome of *LeHT1* silenced plants was similar to the TYLCV resistant plants. This revealed that *LeHT1* confers upon the TYLCV resistance all together with other genes rather than responding directly to the stress. R-software was used to carry out statistical analysis (www.r-project.org). Homologous of resistant and susceptible genes in tomato were identified in *Arabidopsis* at $p \leq 0.05$ and fold change of ≥ 1.3 . Here, 441 putative metabolites were found to be matched with the databases like KEGG, CHEBI and KNAPSACK.

Transcriptional changes in a TYLCV resistant (CLN2777A) and a susceptible (TMXA48-4-

0) breeding line of tomato during TYLCV infection were highlighted (Chen *et al.*, 2013). RNA of these samples was extracted and the cDNA libraries were sequenced on HiSeq 2000 sequencer, Illumina. The raw sequencing data was submitted to NCBI Sequencer Read Archive under accession number SRP028618. These raw reads were preprocessed by Fast QC to remove the low-quality reads and then mapped to the tomato genome using spliced read mapper Tophat v 2.0. Transcript abundance and differential gene expression were calculated with the program Cufflinks. Gene fragments were normalized with fragments per kilo base of exons per million mapped reads (FPKM) values. Five genes were validated by RT-PCR. These genes were associated with pathogen resistance. The defense response of resistant and susceptible lines was unique as 209 and 807 genes differentially expressed in the resistant and susceptible lines respectively. The proportion of upregulated DEGs was higher in resistant line (58.37%) than in susceptible line (9.17%). However, 38 DEGs were common in both the lines but all down-regulated in susceptible line and 11 genes were up-regulated in the resistant lines. Gene ontology showed higher expression of genes responsible for catalytic activity. The functional classes of DEGs were analyzed through Blast2GO gene ontology software. Resistant and susceptible lines yielded 67.46% and 63.69% of DEGs, respectively. There were 30.50% of total DEGs involved in cellular functions in susceptible line while only 23.04% in resistant lines. Higher number of DEGs in resistant lines were observed for developmental processes and multicellular organismal processes than in susceptible lines. Hence, the resistant line has certain gene interaction which enables to compete against the stress. There were few DEGs involved in cell killing, cell proliferation, nitrogen utilization and growth in susceptible lines and rhythmic processes in resistant lines.

In the molecular function class many DEGs were found for catalytic activity. DEGs for nucleic acid binding transcription factor activity and antioxidant activity were specific to susceptible lines only. DEGs for protein binding transcription factor activity were found only in resistant lines. Several DEGs were found exclusively for symplast and cell junction in susceptible lines. There were 16 WRKY genes identified as down-regulated DEGs in susceptible lines whereas seven WRKY genes found up-regulated in resistant lines.

Fungi

The impacts of early and mild biotic stress in Biogenic Volatile Organic Compound (BVOC) emissions from tomato were investigated (Kasal-Slavik *et al.*, 2017). Tomato was exposed to fungus *Botrytis cinerea* and *Oidium neolycopestici*. Aphid *Myzus persicae* was also used as biotic stress agent. Then the tomato plants were treated with methyl jasmonate to identify BVOC emissions related to jasmonic acid (JA) signaling pathway activation. Tomato plant was found to activate JA pathway upon infection with *B. cinerea* as changes in BVOC emissions were observed.

Functional gene changes in tomato genotypes resistant and susceptible to *Verticillium dahlia* through RNA-Seq was studied by Tan *et al.*, (2015) studied for roots. RNA was extracted from the tomato roots at two days after infection of *V. dahlia*. The mRNA libraries were sequenced on Illumina sequencer and 2GB data was processed. High quality reads i.e., with Q30 (80%) were used for further processing. All the clean reads were aligned to the genome of tomato by TopHat (V 2.0). Cufflinks program was used to look for transcript abundance and differential gene expression. FDR threshold ≤ 0.01 and Fold change ≥ 2 were selected. Total 1,985 DEGs

were identified, among which 1,953 (98.39%) were assigned by Nr (Non-redundant), 1,579 (79.55%) by Swiss-Prot, 1,739 (87.71%) by GO (Gene Ontology), 862 (43.43%) by COG (Cluster Orthologous Group) and 380 (19.14%) by KEGG. The data obtained here was deposited in the NCBI SRA database (SRX1022130) in the form of high quality data. 'G+C' content was above 40%, 89.15% sequences of control and 71.04% of treated samples were aligned to the tomato genome. Total 87 KEGG pathways were assigned and were found associated with phenylpropanoid metabolism and plant pathogen interaction.

Early Blight

Early Blight is caused by *Alternaria* sp. which is air borne and soil inhabiting fungi. The disease on foliage is leaf blight, on stem is collar rot and on fruit is fruit rot. Thus, all plant developmental stages are crucial for this fungus. Early Blight occurs through-out the world. Warm climate, long nights and dew affects are favorable for the fungi. The spores may enter a plant system through wounds. Spores of Early Blight survive on plant debris and are spread by wind and rain. Tomato yield loss incurred due to Early Blight is 15-100%. Conventional methods of prevention include preventing long duration leaf wetness, sanitation and application of fungicides. The transcriptome to the infected crop may reveal the disease at an early stage. Hence, transcriptomics of Early Blight infected tomato is very significant.

Differentially expressed miRNA and mRNA during Early blight through next generation sequencing was studied (Sarkar *et al.*, 2017). Illumina TrueSeq RNA library protocol was followed. The quality check was done by SeqQC. The sequence was deposited in GEO database, accession number GSE75922 and GSE75923. Tools like TopHat-2.0.72 and Cufflinks-2.0.14 were used for the transcript

assembly. Cuffdiff was employed to find the significant changes in the transcript expression i.e., DEGs. Gene Ontology was extracted by BiNGO of Cytoscape v3.2.1 FDR ≤ 0.05 . The miRNA sequences were aligned to tomato unigene library by using psRNATarget (plant small RNA Target Analysis Server). Total 52 miRNA were protruding among them, 39 were down regulated and 13 were up regulated. Seven novel miRNA were significant as well as differentially expressing ($P \leq 0.05$). Results showed that 3,154 genes in 17 major molecular functions, 62 genes for three different cellular components and 57 genes for biological process. KEGG analysis showed 5,080 genes for 334 different pathways. Here, 24 pathways were found to be significant for *Alternaria* stress. Differentially regulated genes were found for biosynthesis of ethylene, salicylic acid, jasmonic acid, abscisic acid, ROS activity, NB-LRRs and WRKY transcription factors.

Transcriptomic analysis of pollen developmental stages of tomato

Pollen carries haploid genetic material to the next generation. The developmental stages of pollen are among important stages of a plant. Tetrad (microspore) formation takes place after meiosis I and II of pollen mother cell. Bicellular pollen results after mitosis I. Plant reproductive organs are often more prone to damage at high temperature or stress than vegetative organs. Tomato fruit development is impaired at high temperature.

The most sensitive flowering phases are meiosis and fertilization. Heat stress during meiotic phase may lead to reduction in pollen fertility and reduction in germination and non viable pollen tube. Though, heat sensitivity differs among tomato cultivars. Pollen should withstand stress and drying to ensure the healthy plant. Transcriptome and proteome of

pollens from tomato flower buds at tetrad (T), post-meiotic (PM) and mature (M) stage under control (C) and heat stressed (HS) conditions were evaluated (Keller *et al.*, 2018). RNA was isolated with Macherey-Nagel Nucleospinmi RNA isolation kit. RNAs >200nt were selected for the further analysis. MACE (Massive analysis of cDNA ends) libraries were prepared according to protocol of Bokszczanin *et al.*, (2015). cDNA was synthesized using Life science technologies kit. Bioruptor was used for fragmentation of the cDNAs. Sequencing was done using Illumina HiSeq2000. Sequences were aligned by Next Gen Map (version 0.4.12). Quantification of the transcripts was done by gff files of tomato with HTSeq Python. PCA was derived by R-package Facto MineR (version 1.33). The LC-MS/MS libraries were screened against SGN (Sol Genomics Network) tomato proteome and were quantified by using MaxQuant. Transcripts for T, PM, M, C and HS were 11,000, 9,000, 5,000, 12,606 and 12,684 respectively. In control samples, 4,538 genes were common in all the pollen stages and for heat stress the number was 4,332. There were 957 protein groups in C and 1,062 in HS.

During PCA analysis, PC1 showed 52.3% of the variance inherent in the data which was an indicator of separation among the stages. PC3 showed slight separation for C and HS, during tetrad stage. In LCMS/MS library, PC1 carried 24.5% of the variance. Here, PC2 revealed no better separation of the developmental stages in transcriptome as well as proteome study. In overlapped matrix, two translation modes were displayed namely, direct translation and delayed translation. 'Direct translation' means increased transcript levels along with increased protein levels in the same developmental stage. 'Delayed translation' implies increase in protein levels is postponed by one developmental stage but there is increase in the transcript level.

Total 54 and 108 genes were found in direct and delayed translation, respectively, in T (tetrad). Delayed translation showed 223 genes in PM and direct translation showed 53 genes in M. KOG annotation stated that T had expressions for chaperones, protein turnover, Hsp90s, Hsp70 (BIP4), sHsp (Hsp23.8-MI), carbohydrate metabolism, lipid metabolism, nucleic acid metabolism and secondary metabolites. Many genes showed 'delayed translation' in tetrad, namely genes for Hsp100, Hsp70, chromatin structure, disulfide isomerases and few subtilisin-like proteases. Genes for carbohydrate metabolism, energy metabolism, lipid metabolism, secondary metabolites and nucleotide metabolism possessed 'direct translation'. Delayed translation of genes for chromatin structure indicates the chromatin remodeling in post-meiotic stage of pollen.

Transcriptomic analysis of hormone effect on tomato

Leaf development depicts the overall plant vigour. Hormones are regulators for leaf development. Hormone governs the physiological activities in plants. The coordination between different hormones and plant parts is required for a healthy plant. They are produced in small amount and in some part of the plant to regulate the plant responses. Auxin regulates the leaf phyllotaxis, initiation from shoot apical meristem, serration, margin formation and lobe formation. Mislocalization or misexpression of auxin causes simplified leaves in tomato. Gibberellins (GA) promote seed germination, stem elongation, trichome development, pollen maturation and flowering. GA controls cell proliferation, expansion and leaf complexity. Cytokinin (CK) promote morphogenesis and senescence. Exogenous CK has little effect on morphogenesis while endogenous application is vital in tomato. Jasmonic acid, abscisic

acid, ethylene and strigolactones are also imperative hormones in tomato plant growth (Shwartz *et al.*, 2016).

The transcriptome of tomato plants treated with IAA, GA24 and IAA+GA24. The RNA quality was analyzed (Zhan *et al.*, 2018). Sequencing was done using IlluminaHiSeq 2500 platform. Mapping was done by HISAT. Expectation maximization was used for gene expression level quantification which includes RSEM and FPKM (Fragment Per Kilobase per Million mapped). Differential expression of genes was done by NOI Seq (R/Bio C Package). Gene ontology was studied by Blast2GO. Functional annotation of all the unigenes was made by WEGO. All the metabolic pathways for all the treatments were drawn by using KOBAS. The protein interaction was correlated by using STRING Database, Cytoscape. Primers were designed by Primer Premier 5 software. Total numbers of raw reads were 564.8 million. Total mapped reads were 93.9% and uniquely mapped were 79.1%. Clean reads were found at Q20 and above. Sum of 88.6% were found mapped to exons, 3.6% to introns and 7.9% were mapped with intergenic region. Total DEGs in IAA treated tomato samples were 2,326, of which 1,185 were up-regulated and 1141 down-regulated. DEGs in GR24 were 260, of which 168 were up-regulated and 92 down-regulated. DEGs for IAA+GR24 were 1,379, whereas 1,063 were up-regulated and 316 down-regulated. Genes found to be up-regulated only in IAA treated samples were 480, only in GR24 treated samples were 1,021 and in IAA+GR24 treatment were 427. Down-regulated genes during IAA were 852, GR24=23 and IAA+GR24=49. Gene Ontology in IAA treated samples was for Biological processes (single organism process, response to stimulus and microtubule based process); for Cellular component (cell periphery, plasma membrane and cell wall); for Molecular function (oxidoreductases

activity, acting on glycosyl compounds and hydrolyzing O-glycosyl compounds). GO for GR24 treatment showed negative regulation of catalysis and MF. GO for IAA+GR24 treatment has genes for BP (response to stimulus, chemical and organic substances); for CC (cell periphery, plasma membrane and extracellular region); and for MF (Oxidoreductases, sequence specific DNA Binding and Nucleic acid binding).

Network analysis of DEG showed 19 network internodes for IAA, 13 for GR24 and four for IAA+GR24. IAA treated samples showed 27, GR24 treatment: 12 and IAA+GR24:23 KEGG pathways. Six pathways were common in all the three samples. *CrtQ* gene was found to be down-regulated in IAA and IAA+GR24 treated samples. *CrtH* was down-regulated in IAA samples. Sixteen genes were used for the process of validation.

The transcriptome of the tomato leaves in response to ABA treatment was examined (Wang *et al.*, 2013). ABA treated and controlled tomato RNA samples were subjected to deep sequencing using IlluminaHiSeq 2000. The sequencing gave 20.95 GB of clean data. Adapters were removed by SeqPrep (github.com/jstjhon/SeqPrep).

Sequence mapping was done by Tophat and assembled by Cufflinks. The sequence alignment was done by SOAPaligner/soap2 software (soap.genomics.org.cn) and 81.97% sequences matched to unique (36.53%) and multiple (45.44%) genomic locations. The sequences were merged with Cuffmerge. Total 90.02% of transcripts possessed ORF.

The data obtained had high homology with NCBI non-redundant protein database. There were genes for salinity, drought and cold tolerance. The largest group with regulatory protein was the 'transcription factors'.

Blast2GO was used for transcript classification. Gene Ontology gave total 31,107 transcripts, 23 groups for biological processes, 19 for cellular components and 15 for molecular functions.

In controlled samples, 38,626 transcripts were expressed and in ABA treated those were 37,989. The transcription factors (TF families) identified were, *bZIP* (111 transcripts), *Bhlh* (146), *MYB* (212), *AP2/ERF* (84), *NAC* (63) and *WRKY* (81). 41 *Heat shock factors (HSFs)* and 61 *MADS-boxes* were also detected. Differential expression was studied using edgeR. Gene ontology was done by Blast2GO. COG was done by Blastx 2.2.24+STRING 9.0. Metabolic pathways were studied using KEGG.

Genes related to ABA pathway were ABA signaling transduction (PYP/PYL, PP2C, SnRK2, ABF), bZIP(ABRT), BHLH (DREB), MYB (MYC), AP2 (AP2/ERF), NAC, WRKY, HSF and MADS-box. MYB had the highest up-regulation (66), BHLH had 47, bZIP 40, WRKY and NAC recorded 33 and 23 respectively. All the genes related to heat shock proteins were Hsp 90, Hsp 70, Shsp, CAT and SOD; for ROS scavenging system: GLR, APX, MDAR, HAR, GST, GPX, POD, Txr and Prxr; for pathogens resistance were PAL, PPO, GLU, chitinase, TGA, PR1, JAR1, JAZ, ETR and ERF/EREBP.

Interactive transcriptome of abiotic and biotic stress in tomato

Tomato root transcriptome under water stress, fungal infection and nematode invasion was studied (Balestrini *et al.*, 2019). Arbuscularmycorrhizal (AM) fungus was *Rhizophagus intraradices* and nematode was *Meloidogyne incognita*. Tomato cultivar namely 'San Marzano nano' (*S. lycopersicum*) was taken under the study. Tomato roots were

compared as non-fungal colonized (C, control), fungal colonized (AM), fungal colonized and water stressed (AM_WS), nematode infected (RKN), nematode infected and fungal colonized (RKN_AM). Sequencing gave $187,7 \times 10^6$ reads. Reference genome used for mapping was SL2.40.26 (*S. lycopersicum*). Higher numbers of genes were differentially up-regulated in AM (12%) than in RKN (10%). Gene ontology for AM_WS and RKN_AM was done for response to oxidative stress, peroxidase activity and heme binding.

Molecular functional genes in AM_WS were found to be up-regulated for transcription regulatory activity while metabolic process related genes in RKN_AM were up-regulated for protein ubiquitination and protein amino acid phosphorylation. Cellular component was over-represented by gene for microtubule based moment. Five out of total six genes for putative ripening related proteins (RRP) were highly up-regulated in AM colonized (unstressed) and water stressed tomato samples.

Blue copper proteins, germin-like proteins, glutathione-S-transferase and cell wall related genes were significantly up-regulated in AM fungal infected samples. Blue copper binding genes are considered markers for AM colonization in plants. AM symbiosis was found completely function in water stressed tomato roots (AM_WS). Genes for cytochrome P450 (CYPs) were highly up-regulated in AM_WS tomato root samples. CYPs are known for sterol biosynthesis and production of secondary metabolites.

Pan-transcriptome of tomato

Pan-transcriptome of 399 tomato accessions by analysing differentially expressed genes (DEGs) and global expression presence/absence variation (ePAV) was

constructed (Dan *et al.*, 2020). Genes that are expressed only in a specific group are called as ePAV. The study contained 26 accessions of *S. pimpinellifolium* (PIM), 114 accessions of *S. lycopersicum* var. *cerasiforme* (CER) and 259 accessions of big fruited *S. lycopersicum* (BIG). The reference genome was SL3.0. The average mapping rate was 93.35% however, the mean mapping rate for PIM, CER and BIG was 90.26%, 93.03% and 93.80%, respectively. The lower mapping rate in PIM group shows that there is lack of genes in Heinz 1706 (SL3.0) genome when compared to *S. pimpinellifolium* (PIM). Thus few genes have been eliminated during the selection procedure.

Unmapped genes in all the accessions were assembled *de novo*. This assembly gave 82,617; 1, 48,156 and 1, 60,347 transcripts in PIM, CER and BIG, respectively. Total 907 (PIM), 2,328 (CER) and 3,946 (BIG) high-confidence RTAs were absent from Heinz 1706. Genes which were expressed in PIM group but not in CER or *vice versa* were ePAV genes during tomato domestication.

Hence, genes expressed in CER group but not in BIG or *vice versa* are were ePAV genes during tomato breeding. These genes are identified by orthologous gene clustering. Total 6,123 RTAs orthologous clusters were identified among which 5,261 were single-copy clusters.

Here, 2,678 ePAV RTAs clusters were identified (2,706 ePAV genes, 640 genes lost expression, 2,066 gained expression) during domestication. Similarly, 4,970 ePAV RTAs clusters were identified (4,992 ePAV genes, 1,692 genes lost expression, 3,300 gained expression) during tomato breeding. Total 3,629 genes were found to be significant during tomato selection process, and 19 of these were associated with reduced fruit total soluble solid (TSS) in recent cultivars.

“Next gen” RNA-Seq provides opportunity for genome wide survey in unbiased manner. Though per-sample cost is high at the beginning but the depth is much more in comparison to micro-array. Real-time quantitative PCR requires primer designing and micro-array is dependent on probes. While RNA-Seq is free from all these limitations. RNA-Seq sequencing is supplied by Illumina, Nanopore, Applied Bioscience etc whereas Micro-array chip can be customized by companies like Affymetrix, Agilent Technologies, BD Bioscience, BioRad, Biogenex, Illumina, Life technologies and Qiagen. Life science, Quagen AB gene, Applied Biosystems, Biogen, BioRad, Eppendorf and Invitrogen are reliable companies for real-time PCR system.

Variants for RNA-Seq are HiSeq, MiSeq and QuartzSeq. Species specific as well as stress specific microarray chips can be designed using pre-defined probes. Chemistry for RNA-Seq is sequencing by synthesis while microarray works on the principle of hybridization of complementary sequences. TaqMan probe, TAMRA, ROX, Scorpion and SYBR Green chemistry are exemplary in real-time PCR (Table 4).

According to Boston University, Medical Campus Microarray and Sequencing Resource, transcriptome technologies though microarray chip and RNA sequencing can be grouped in to low (100-300\$), medium (300-700\$) and high (above 700 \$) price range products, per application. Low price range products includes Illumina TrueSeq, NEBNext and Kapa RNA HyperPrep. Medium price range products are miRNA 4.0, Human transcriptome, Human Exon, Mouse transcriptome, Rat transcriptome and *Arabidopsis* array. The highest price range is for Illumina NextSeq 500 (150 & 300 cycles) and high output (75, 150 and 300 cycles) (Table 5).

Manual RNA extraction protocol viz. with Trizol as well as RNA isolation by kit such as Macherey-Nagel, RNAOut, Green spin, RNeasy, and ZR are constantly uses by the researchers. The literature shows that Illuminasequencins is the most exploited tool for sequencing. Though few researchers have also used microarray also. Most of the primers are designed by Primer Premier 5, Primer 3, QuantPrime and PrimeQuest. Latest real-time cyclers in use are AB 7300 & 7500 realtime PCR system, ABI 7900HT, LightCycler 96 Roche Life Science and CFX connect real-time system BioRad. The most expedient bioinformatics tools are Cofflinks, TopHap, Cuffmerg, Cifffdiff, Bowtie, CLCBlast2GO, Trinity and R-Software (Table 6&7).

- High-throughput sequencing methods give an insight to understand the plant molecular mechanisms during stress condition.
- Next generation sequencing is an efficient method over microarray.
- For the expression under different stages (developmental, infection), transcriptomics reveals biochemical signaling pathways which are related to the plant stress response.
- Genes are categorized in Molecular Function, Cellular Components and Biological Processes.
- WRKY and phenyl propanoid genes impart resistance in tomato plants towards various stresses.
- Differentially expressed transcripts may promote targeting of a candidate gene, which may be used in the breeding programs for resistance or tolerance to certain type of stress.
- Gene expression profiling, genome annotation, miRNA, Transcription Factors and discovery of non-coding RNA, etc. are also employed using transcriptomics.

Limitations for RNA-Seq

Transcriptomics is cost effective in long run and with multiple samples. The transcripts found require proper analysis and interpretation through bioinformatics and statistical tools. The human resources in RNA sequencing technology are comparatively less in numbers.

Library preparation

Library preparation involves many steps to be followed during specified time duration. Following all the steps critically is very significant. The products in-between and after the last step have peculiar storage condition. Micro RNAs, Piwi-interacting RNAs, short interfering RNA are sequenced after adaptor ligation but higher RNA (total or mRNAs) molecules are fragmented by hydrolysis or nebulization before sequencing. The cDNA is biased towards sequence identification from 3' end, thus the same end is more informative (Wang *et al.*, 2009).

Proper selection of bioinformatics tools

Diverse bioinformatics tools direct the proper interpretation of data. A set of tools are required to store, retrieve and process the data generated. Data analysis includes removal of adapter sequence, quality enhancement, mapping, annotation, gene enrichment, pathway identification, gene ontology and differential gene expression. Identification of unique splicing events which are occurring between two distant sequences and sequences flanked by exons of diverse genes is a decisive task. Short reads with high copy number and long repetitive sections demand more efforts.

Cost and coverage

Sequencing depth is proportional to the

coverage. Depth allows identification of rare variants. Deep sequencing has larger genome, complex transcriptome and higher coverage data and hence, higher cost. Though, the cost is competent over arrays.

Way ahead

The era is demanding an easy though specific approach to transcriptomics. The institutions with a full facilitated laboratory should design abundant projects on abiotic and biotic stress resistance for model crops as well as for novel crops. The developing countries should be focused more and more as they have saturating population rise. Easy access to various bioinformatics and statistical tools may pave the way for informal approach of transcriptomics. As these days, researchers are having huge data sets but they may require proper application of bioinformatics tools. Organizing workshops for these tools and generating technically well qualified human resources may be fruitful. Targeted genes after differential expression should be directed for gene editing transformation. The comparison of transcripts of wild and cultivated species is the treasure of resistant genes.

Compliance with ethical standards

Funding: The manuscript is a review paper hence funding is not applied.

Conflict of interest: Author1 declares that she has no conflict of interest.

Author 2 declares that he has no conflict of interest.

Ethical approval: This article is a review and does not contain any studies with human or animal participants performed by any of the authors.

Table.1 Botanical classification of tomato

Kingdom	<i>Plantae</i>
Clade	Angiosperms, Eudicots, Asterids
Order	<i>Solanales</i>
Family	<i>Solanaceae</i>
Genus	<i>Solanum</i>
Species	<i>Solanum lycopersicum</i> L
Chromosome no.	2n=2x=24
Ploidy level	Diploid
Genome size	950 Mb

Table.2 Domesticated varieties, wild varieties and major producers of tomato

Domesticated species	<i>Solanum lycopersicum</i>
Wild species	<i>Solanum galapagenes</i> , <i>S. pimpinellifolium</i> , <i>S. arcanum</i> , <i>S. cheesmanie</i> , <i>S. chilense</i> , <i>S. chmielewskii</i> , <i>S. habrochaites</i> , <i>S. neorickii</i> , <i>S. pennellii</i> , <i>S. peruvianum</i>
Major producers in the world	China, India, USA, Turkey, Egypt
Leading producer states in India	AP, Karnataka, MP, Telangana, Gujarat, Bihar, WB, Maharashtra, Chhatishgarh and HP

Table.3 Area, production and productivity of tomato (Anonymous, 2018)

	Area (Mha)	Production (MMT)	Productivity(MT/ha)
World	5.02	170.75	34.01
India	0.78	19.37	24.84

Table.4 Comparison between Real-time PCR, Microarray and “Next gen” RNA Seq

Sr. No.	Keys	Real-time quantitative PCR	Microarray	“Next gen” RNA- Seq.
1	Technique	Real time, quantitative measurement of expression	Quantitative measurement of expression	Genome wide survey is done
2	Way	Gene by gene study	Genome wide survey is required	Completely unbiased, universal compatibility
3	Cost effectiveness	Cost effective	Cost effective if targeted specifically	Per sample cost is high but the depth of the data is there
4	Prior requirements	Primer designing	Prior requirements: Probes	Excludes such requirements, able to start from the scratch, <i>Arabidopsis</i> genome or any other genome may be used for mapping and annotation, if done for a novel crop
5	Companies supplying the kit	Life Science, Qiagen, ABgene, Applied Biosystems, Biogene, BioRad, Eppendorf, Invitrogen	Affymetrix, Agilent Technologies, BD Biosciences, Bio-Rad, Biogenex, Illumina, Life technologies, Qiagen	Illumina, Nanopore, Applied Biosystem
6	Variants of the machine	One step, two step, dye binding, fluorescent probe	Customized chips, species/condition specific	HiSeq., MiSeq. QuartzSeq
7	Chemistry	TaqMan probe (FRET), TAMRA, ROX, Scorpion, SYBR Green (dsDNA binding dye)	Hybridization with the complementary	Sequencing by synthesis
8	Specificity	Amplification of targeted DNA molecule	Targeted genes’ expression is studied	Unbiased for expression studies, expression of coding and non-coding RNA
9	Bioinformatics Tools	Optical module, few Computation	Scanner, few computation	Require diverse tools
10	Property	Quantitative expression	Compares control and infected	<i>De novo</i> expression, absolute expression
11	Entities present	Primer, probe, oligonucleotides, Taq polymerase, target DNA	Probe, Blocking agent, Mask	Sample in flow cell

Table.5 Boston University, Medical Campus Microarray and Sequencing Resource (Curtsey: <http://www.bumc.bu.edu/microarray/pricing/>)

Sr. No.	Applications	Internal pricing (\$)	External Academic pricing(\$)
1	IlluminaTrueSeq Standard mRNA	250	280
2	NEBNext Ultra II RNA	225	255
3	IlluminaTrueSeq Standard Total RNA	300	330
4	Kapa RNA HyperPrep	290	320
5	IlluminaTrueSeq Small RNA	320	350
6	NEBNext Small RNA	270	300
7	IlluminaNextera XT (Small Genome)	135	155
8	IlluminaNextSeq 500 150 cycles (130M)	1600	1700
9	IlluminaNextSeq 500 300 cycles (130M)	2300	2400
10	High Output 75 cycles (400M)	1900	2000
11	High Output 150 cycles (400M)	3250	3350
12	High Output 300 cycles (400M)	5100	5200
13	Ion Torrent PGM 314 (400-700K)	630	-
14	Ion Torrent PGM 316 (2-3.5M)	790	-
15	Ion Torrent PGM 318 (3-5M)	950	-
16	miRNA 4.0 (Multispecies) (Affymetrix)	335	355
17	Human Transcriptome 2.0 Array (Aff.)	500	530
18	Human Exon 1.0 ST Array (Aff.)	665	695
19	Mouse Transcriptome 1.0 Array (Aff.)	500	530
20	Mouse Exon 1.0 ST Array (Aff.)	665	695
21	Rat Transcriptome 1.0 Array (Aff.)	500	530
22	Arabidopsis 1.0ST Array (Aff.)	415	445
23	C. elegans 1.0ST Array (Aff.)	415	445
24	Drosophila 1.0ST Array (Aff.)	415	445
25	Rhesus 1.0ST Array (Aff.)	415	445
26	Zebrafish 1.0ST Array (Aff.)	415	445
27	Yeast Genome 2.0 Array (Aff.)	440	470
28	E. coli Genome 2.0 Array (Aff.)	475	505
29	AribidopsisGenome 2.0 Array (ATH1) (Aff.)	540	570
30	Drosophila Genome 2.0 Array(Aff.)	515	545

Table.6 Methodology used by various researchers for transcriptomics in tomato

Sr. No.	Author (Year)	RNA/DNA isolation and Library preparation	Transcriptome technology	Primer designing, RT-PCR	Bioinformatics tools	Reference genome and database used
1	Koeing <i>et al.</i> , 2013	Trizol, RNeasy kit	Illumina GA II, HiSeq 2000	-	R-stastical programming	Heinz, Potato
2	Keller <i>et al.</i> , 2013	Macherey-Nagel miRNA isolation kit, MACE	IlluminmaHiSeq 2000	-	Next Gen Map, HTSeq Python, Facto MineR R-package, MaxQuant, KOG	SGN, ITGA2.4 (International Tomato Annotation Group, Sol Genomics Network)
3	Wang <i>et al.</i> , 2013	Trizol, TruSeq	IlluminmaHiSeq 2000	Primer Premier 5	SeqPrep, TopHat, Cufflinks, Trinity, SOAPaligner, Cuffmerge, Blas2GO, R software	ITAG2.3, , NCBI NR, GO, STING, KEGG
4	Chen <i>et al.</i> , 2013	RNAOut kit, TrueSeq	IlluminmaHiSeq 2000	Primer Pemier5, qTOWER 2.0/2.2	FastQC, TopHat, Cufflink, Blast2GO	Heinz 1706
5	Sade <i>et al.</i> , (2013, 2015)	Tri-Reagent method	Microarray	QuantPrime, 7300 Real Time PCR System	R software	KEGG, CHEBI and KNAPSACK
6	Aflitos <i>et al.</i> , 2014	DNA isolation protocol (Van der Beek <i>et al.</i> , 1992)	IlluminmaHiSeq 2000	-	Bwotie, CLC, Clustal W, SNPEFF, SAMTOOLS, CLC, ALLPATHS, SCARPA, Clustal w	Heinz v 2.40, ITAG 2.4
7	Tan <i>et al.</i> , 2015	Trizol method	IlluminmaHiSeq 2000	DNAMAN 6, ABI 7900HT	TopHat, Cufflink, BLAST	SwissProt, KEGG, COG
8	Ye <i>et al.</i> , 2015	Green spin RNA quick extraction, mRNA Seq	IlluminaAnalyzerIIx	Primer 3	TopHat, EdgeR, DAVID, CORREL	SL2.40, SOL Genomics, Tomato Functional Genomic Database (TFGD)

9	Lovieno <i>et al.</i> , 2016	Trizol method, TruSeq	IlluminaHiSeq 1500	ABI 7900HT	Trim Galore, Cutadapt, FastQC, Bowtie, TopHat	Tomato genome v 2.40
10	Dia <i>et al.</i> , 2017	-	-	-	FastQC, TopHat, Cuffmerge, Cuffdiff, KOBASS	SRA (Koeing <i>et al.</i> , 2013)
11	Kumar & Chattopadhyay <i>et al.</i> , 2017	-	-	-		EBI (Aflitos <i>et al.</i> , 2014)
12	Sarkar <i>et al.</i> , 2017	Trizol, TruSeq	IlluminaTrueSeq	AB 7500 FAST	SeqQC, TopHat 2, Cufflinks, Cuffdiff, BiNGO, KEGG	psRNATarget
13	Rahim <i>et al.</i> , 2018	RNeasy mini kit, TruSeq	IlluminaHiSeq 2000	LightCycler 96 Roche Life Science	TopHat, Cufflink, DESeq	SwissProt, KEGG
14	Razali <i>et al.</i> , 2018	ZR Plant RNA mini kit, Zymo, NEBNext	IlluminaHiSeq 2000	-	Trimmomatic, Trinity, TransRate, BUSCO, InterProscan, OrthoMCL, BWA, CIROS, SAMtools, DEAP	SwissProt
15	Shukla <i>et al.</i> , 2018	Trizol method	IlluminaHiSeq2000	PrimerQuest, CFX connect real-time system	CASAVA, TopHat, PANTHER, BLASTX, DESeq2, AgriGO, MapMan	SL2.50, KEGG
16	Balestrini <i>et al.</i> , 2019	RNeasy Plant mini kit, TruSeq	IlluminaHiScanSQ, SBS v3 kit	-	FastQC, CLC, Edge R, AgriGO	SL2.40.26
17	Dan <i>et al.</i> , 2020	Quick RNA isolation kit	IlluminaHiSeq2000	Primer3, CFX manager	Cytoscape, MultiQuant, SOAP2, SOAPsnp, BWA, Samtools, PHYLIP, Hisat2, PEER	NCBI, European Nucleotide archive. ITAG2.4, SL3.0

Table.7 Technology used for transcriptomics by the researchers

Sr. No.	Author	Trait studied in tomato	Methodology performed for transcriptome
1	Koeing <i>et al.</i> , 2013	Cultivated vs wild species	Illumina sequencing
2	Keller <i>et al.</i> , 2013	Pollen developmental stages	Illumina sequencing
3	Wang <i>et al.</i> , 2013	Effect of hormones	Illumina sequencing
4	Chen <i>et al.</i> , 2013	TYLCV resistance	Illumina sequencing
5	Sade <i>et al.</i> , 2013&2015	TYLCV resistance	Microarray
6	Aflitos <i>et al.</i> , 2014	Cultivated vs wild species	Illumina sequencing
7	Tan <i>et al.</i> , 2015	<i>Verticillium dahlia</i> resistance	Illumina sequencing
8	Ye <i>et al.</i> , 2015	Cultivars, days after flowering	Illumina sequencing
9	Lovieno <i>et al.</i> , 2016	Drought stress	Illumina sequencing
10	Dia <i>et al.</i> , 2017	Cultivated vs wild species	Illumina sequencing
11	Kumar & Chattopadhyay <i>et al.</i> , 2017	Cultivated vs wild species	Illumina sequencing
12	Sarkar <i>et al.</i> , 2017	Early Blight	Illumina sequencing
13	Rahim <i>et al.</i> , 2018	Normal and cherry tomato	Illumina sequencing
14	Razali <i>et al.</i> , 2018	Salt stress	Illumina sequencing
15	Shukla <i>et al.</i> , 2018	Nematode resistance	Illumina sequencing
16	Balestrini <i>et al.</i> , 2019	Fungus, nematode and water stress	Illuminasequencing
17	Dan <i>et al.</i> , 2020	Pan-transcriptome of tomato	Illuminasequencing

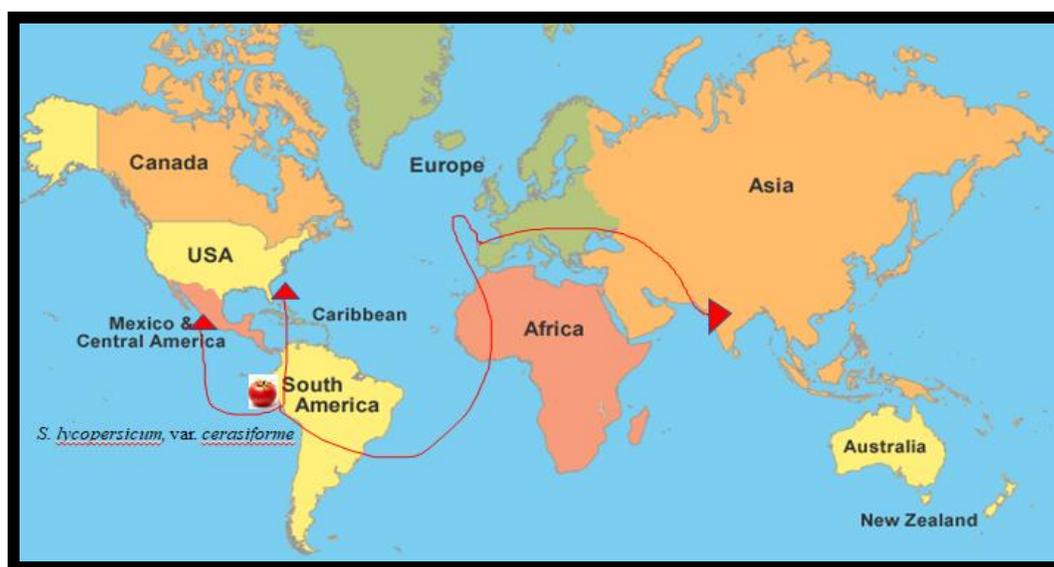


Fig.1 Center of origin of tomato (Simmonds, 1976)

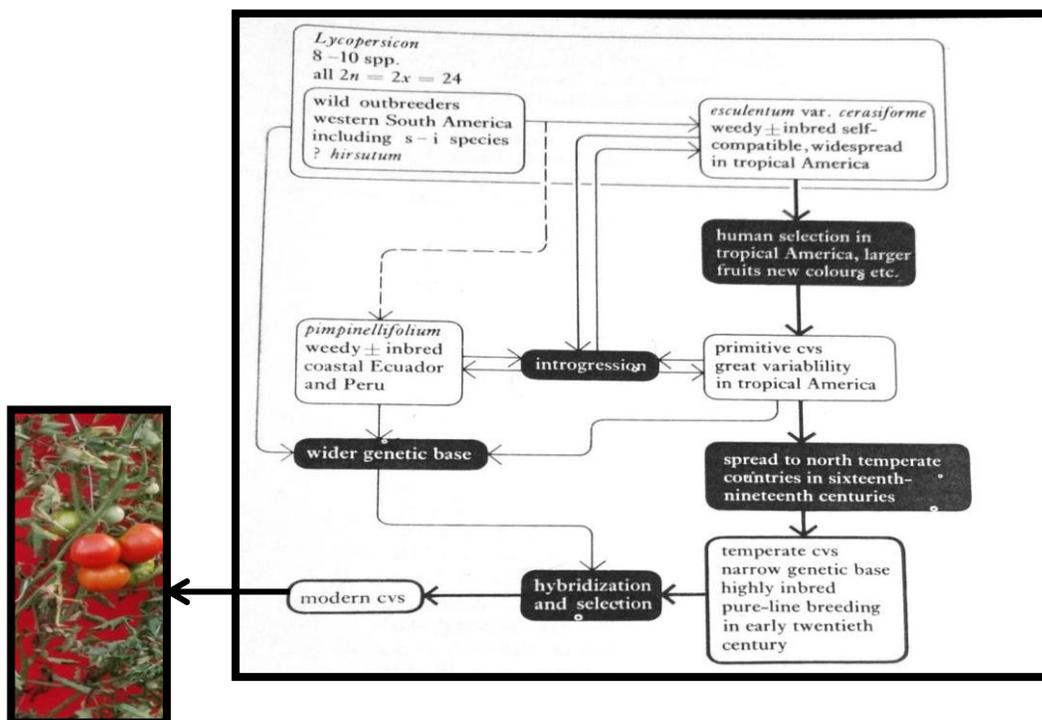


Fig.2 Evolutionary relationships of tomato (Simmonds, 1976)

References

- Anonymous (2014).Biology of *Solanum lycopersicum* (Tomato).
- Anonymous (2018).3rd Advance Estimates of Area and Production of Horticulture Crops, Department of Agriculture Cooperation & Farmers Welfare, Krishibhawan, New Delhi.
- Balestrini R, Rosso LC, Veronico P,Melillo MT, Luca FD, Fanelli E, Colagiero M, di Fossalunga AS, Ciancio A, Pentimone I (2019) Transcriptome responses to water deficit and nematode infection in mycorrhizal tomato roots. *Front Microbiol* 10:1807. doi.org/10.3389/fmicb.2019.01807.
- Bokszczanin KL, Krezdorn N, Fragkostefanakis S, Muller S, Rycak L, Chen Y, Hoffmeier K, Kreutz J, Paupiere MJ, Chaturvedi P (2015) Identification of novel small ncRNAs in pollen of tomato. *BMC Genomics* 16:714.
- Bolger M E, Arsova B, Usadel B (2017) Plant genome and transcriptome annotations: from misconceptions to simple solutions. *BriefBioinform*135:1-13.
- CABI (2020).Invasive Species Compendium. Wallingford, UK: CAB International (www.cabi.org/isc).
- Capel C, Fernandez del Carmen A, Alba JM M, Lima-Sila V,Hernández-Gras F, Salinas M, Boronat A, Angosto T., Botella MA, Fernández-Muñoz R, Granell A,Capel J (2015) Wide-genome QTL mapping of fruit quality traits in a tomato RIL population derived from the wild-relative species *Solanum pimpinellifolium*L. *TheorAppl Genet* 128:2019https://doi.org/10.1007/s00122-015-2563-4.
- Chen T, Lv Y, Zhao T, Li N, Yang Y, Xu W, He X, Liu T, Zang B (2013) Comparative Transcriptome Profiling of a Resistant vs. Susceptible Tomato

- (*Solanum lycopersicum*) Cultivar in Response to Infection by Tomato Yellow Leaf Curl Virus. PLoS ONE, 8 (11): e80816 doi:10.1371/journal.pone.0080816.
- Cheng L, Zou Y, Ding S, Zhang J, Yu X, Cao J, Lu G (2010) Polyamine accumulation in transgenic tomato enhances the tolerance to high temperature stress. *Plant Biol* 51(5):489-499.
- Chetelat, R. T. & Ji, Y. (2007) GISH analysis of meiotic chromosome pairing in *Solanum lycopersicoides* introgression lines of cultivated tomato. *Genome* 50: 825–833.
- Chetelat RT, Pertuze RA, Faundez L, Graham EB, Jones CM (2009) Distribution, ecology and reproductive biology of wild tomatoes and related nightshades from the Atacama Desert region of northern Chile *Euphytica* <http://www.springerlink.com/content/92072848g7547w41/>.
- Dan L, Liang Y, Jin-zhe Z, Guang-tao Z, Hong-jun L, Ya-qing L, Yan-ling W, Xue C, Tian-shu S, San-wen H, Yao-yao W (2020) Domestication and breeding changed tomato fruit transcriptome. *J Integr Agric* 19(1):120-132. doi: 10.1016/S2095-3119(19)62824-8.
- Dia Q, Geng L, Lu M, Jin W, Nan X, He P, Yao Y (2017) Comparative transcriptome analysis of the different tissues between the cultivated and wild tomato. *PLoS ONE* 12(3) e0172411, doi: 10.1371/journal.pone.0172411.
- Eybishtz A, Peretz Y, Sade D, Akad F, Czosnek H (2009) Silencing of a single gene in tomato plants resistant to Tomato yellow leaf curl virus renders them susceptible to the virus. *Plant Mol Biol* 71:157–171.
- Gorovits R, Czosnek H (2007) Biotic and Abiotic Stress Responses in Tomato Breeding Lines Resistant and Susceptible to Tomato Yellow Leaf Curl Virus. doi:10.1007/978-1-4020-4769-5_13.
- Haupt-Herting S, Fock HP (2000) Exchange of oxygen and its role in energy dissipation during drought stress in tomato plants. *Physiol Plant* 110:489-95.
- Iberkleid I, Sela N, Brown MS (2015) *Meloidogyne javanica* fatty acid and retinol binding protein (Mj-FAR-1) regulates expression of lipid, cell wall, stress and phenylpropanoid-related genes during nematode infection of tomato. *BMC Genomics* 16:272 <https://doi.org/10.1186/s12864-015-1426-3>.
- Ji Y, Chetelat RT (2007) GISH analysis of meiotic chromosome pairing in *Solanum lycopersicoides* introgression lines of cultivated tomato. *Genomes* 50: 825-833.
- Kasal-Slavik T, Eschweiler J, Kleist E, Mumm R, Goldbach HE, Schouten A, and Wildt, J (2017) Early biotic stress detection in tomato (*Solanum lycopersicum*) by BVOC emissions. *Phytochem* 144:180-188.
- Keller M, Spot in consortium & Simm S (2018). The coupling of transcriptome and proteome adaptation during development and heat stress response of tomato pollen. *BMC Genomics* 19(1):447 doi: 10.1186/s12864-018-4824-5.
- Kil E, Kim S, Lee Y, Byun H, Park J, Seo H, Kim C, Shim J, Lee J, Kim J, Lee, K., Choi, H. and Lee, S. (2016). Tomato yellow leaf curl virus (TYLCV-IL): a seed-transmissible geminivirus in tomatoes. *SciRep* 6:19013. [tps://doi.org/10.1038/srep19013](https://doi.org/10.1038/srep19013).
- Koenig D, Jimenez-Gomez JM, Kimura S, Fulop D, Chitwood DH, Headland L R, Kumar R, Covington MF, Devisetty UK, Tat AV, Tohge T, Bolger A,

- Scheeberger K, Ossowski S, Lanz C, Xiong G, Taylor-Teeples M, Brady SM, Pauly M, Weigel D, Usadel B, Fernie AR, Peng J, Sinha NR and Maloof JN (2013) Comparative transcriptomics reveals pattern of selection in domesticated and wild tomato. *Proc Natl Acad Sci USA* 110 (28), E2655-E2662.
- Lovieno P, Punzo P, Guida G, Mistretta C, Van Oosten MJ, Nurcato R, Bostan H, Colantuono C, Costa A, Bagnaresi P, Chiusano ML, Albrizio R, Giorio P, Batelli G & Grillo S (2016) Transcriptomic changes drive physiological responses to progressive drought stress and rehydration in tomato. *Front Plant Sci* 7:371.
- Lowe R, Shirley N, Bleckley M, Dolan S, Shafee A (2017) *PLoS Comput Biol* 13(5):e1005457.
- Machado R, Serralheiro RP (2017) Soil salinity: Effect on vegetable crop growth, management practices to prevent and mitigate soil salinization. *Horticulturae* 3:30 doi:10.3390/horticulturae3020030.
- Peralta IE, Spooner DM, Knapp S (2005) New species of wild tomatoes (*Solanum* section *Lycopersicum*: *Solanaceae*) from northern Peru. *Syst Bot* 30(2):424–434. doi: 10.1600/0363644054223657.
- Rahim MA, Jung H, Afrin KS, Lee J, Nou I (2018) Comparative transcriptome analysis provides insights into dwarfism in cherry tomato (*Solanum lycopersicum* var. *cerasiforme*). *PLoS ONE* 13(12): e0208770 doi:10.1371/journal.pone.0208770
- Razali R, Bougouffa S, Morton MJL, Lightfoot DJ, Alam I, Essack M, Aroid ST, Kamau AA, Schmockel SM, Pailles Y, Shahid M, Michell CT, Al-Babill S, Ho YS, Tester M, Bajic VB, Sonia N (2018) The genome sequence of the wild tomato *Solanum pimpinellifolium* provides insights into salinity tolerance. *Front Plant Sci* 9(1402) doi: 10.3389/fpls.2018.01402
- Rick CM (1990) Perspective from plant genetics: the Tomato Genetics Stock Center. In: McGuire PE, Qualset CO, editors. Genetic resources at risk: scientific issue, technologies and funding policies. Genetic Resources conservation program, University of California 11-20.
- Rick C, Tanksley S (1981) Genetic variation in *Solanum pennellii*: comparison with two other sympatric tomato species. *Plant Syst. Evol.* 139:11-45.
- Sade D, Brotman Y, Eybishtz A, Cuadros-Inostroza A, Fernie AR, Willmitzer L, (2013) Involvement of the hexose transporter gene *LeHT1* and of sugars in resistance of tomato to tomato yellow leaf curl virus. *Mol Plant* 6:1707–1710.
- Sade D, Shriki O, Cuadros-Inostroza A, Tohge T, Semel Y, Haviv Y, Willmitzer L, Fernie A, Czosnek H, Brotman Y (2015) Comparative metabolomics and transcriptomics of plant response to Tomato yellow leaf curl virus infection in resistant and susceptible tomato cultivars. *Metabolomics* 11(1):81–97. doi: 10.1007/s11306-014-0670-x.
- Sahu, K. K. & Chattopadhyay, D. (2017). Genome-wide sequence variations between wild and cultivated tomato species revisited by whole genome sequence mapping. *BMC Genomics*, 18:430.
- Sarkar D, Maji RK, Dey S, Sarkar A, Gosh Z, Kundu P (2017) Integrated miRNA and mRNA expression profiling reveals the response regulators of a susceptible tomato cultivar to early blight disease. *DNA Research* 24(3):235-250 doi:10.1093/dnares/dsx003.
- Sato S, Tabata S (2016) Functional Genomics and Biotechnology in Solanaceae and

- Cucurbitaceae Crops. 70doi
10.1007/978-3-662-48535-4_1.
- Shukla N, Yadav R, Kaur P, Rasmussen S, Goel S, Agrwal M, Jagannath A, Gupta R, Kumar A (2018) Transcriptome analysis of root-knot nematode (*Meloidogyne incognita*)-infected tomato (*Solanum lycopersicum*) roots reveals complex gene expression profiles and metabolic networks of both host and nematode during susceptible and resistance responses. *Mol Plant Pathol* 19(3):615-633.
- Shwartz I, Levy M, Ori N, Bar M (2016) Hormones in tomato leaf development. *Dev Biol* 419(1):132-142
- Simmonds Norman (1976). Evolution of crop plants, Longman Group Ltd. ISBN 10: 0582444969.
- Tan G, Liu K, Kang J, Xu K, Zhang Y, Hu L, Zhang J, Li C (2015) Transcriptome analysis of the compatible interaction of tomato with *Verticillium dahlia* using RNA-sequencing. *Front Plant Sci*6:428-438
- The 100 Tomato Genome Sequencing Consortium (2014) Exploring genetic variation in the tomato (*Solanum section Lycopersicon*) clade by wholegenome sequencing. *Plant J* 80:136–148.
- Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485:635-641.
- Van der Beek JG, Verker R, Zabel P, Lindhout P. (1992) Mapping strategy for resistance genes in tomato based on RFLPs between cultivars: Cf9 (resistance to *Cladosporium fulvum*) on chromosome 1. *Theor Appl Genet* 84:1–2.
- Wang Y, Tao X, Tang XM, Xiao L, Sun JL, Yan XF, Li D, Deng HY, Ma XR (2013) Comparative transcriptome analysis of tomato (*Solanum lycopersicum*) in response to exogenous abscisic acid. *BMC Genomics* 14:841.
- Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* (PMC).
- Ye J, Hu T, Yang C, Li H, Yang M, Ijaz R, Ye Z, Zang Y (2015) Transcriptome Profiling of Tomato Fruit Development Reveals Transcription Factors Associated with Ascorbic Acid, Carotenoid and Flavonoid Biosynthesis. *PLoS ONE* 10(7): e0130885. doi:10.1371/journal.pone.0130885
- Zhan Y, Qu Y, Zhu L, Shen C, Feng X, Yu C (2018) Transcriptome analysis of tomato (*Solanum lycopersicum* L.) shoots reveals a crosstalk between auxin and stringolactone. *PLoS ONE* 13(7:e0201124).

How to cite this article:

Kinjal Kulshrestha and Akarsh Parihar. 2020. A Decade of Tomato Transcriptomics: Status and Perspectives. *Int.J.Curr.Microbiol.App.Sci.* 9(03): 2026-2056.
doi: <https://doi.org/10.20546/ijcmas.2020.903.235>