Genomics Data 14 (2017) 82–90

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Computational deciphering of biotic stress associated genes in tomato (*Solanum lycopersicum*)

G. Tandon[a,b], S. Singh[b,*], S. Kaur[a,b], Sarika[a], M.A. Iquebal[a], A. Rai[a], D. Kumar[a]

[a] Centre for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, New Delhi 1100 12, India
[b] Departments of Computational Biology & Bioinformatics, Jacob School of Biotechnology & Bio-Engineering, SHUATS, Allahabad 211007, India

A B S T R A C T

Tomato (*Solanum lycopersicum*) is one of the major vegetable plant and a model system for fruit development. Its global importance is due to its lycopene pigment which has anti-oxidative and anti-cancerous properties. Though > 1.5 M biotic stress associated ESTs of tomato are available but cumulative analysis to predict genes is warranted. Availability of whole genome de novo assembly can advantageously be used to map them over different chromosome. Further, available 0.14 M catalogued markers can be used to introgress specific desirable genes in varietal improvement program. We report here 57 novel genes associated with biotic stress of tomato along with 50 genes having physical location over different chromosomes. We also report 52 cis-regulating elements and 69 putative miRNAs which are involved in regulation of these biotic stresses associated genes. These putative candidate genes associated with biotic stress can be used in molecular breeding in the endeavor of tomato productivity along with its sustainable germplasm management.

1. Introduction

Plants have developed gradually to live in an environment where they are generally exposed to different types of stresses in combination. Being sessile, they have developed various mechanisms which allow them to detect very minute changes and respond to those stress conditions thus minimizing damage [1]. These mechanisms activate expression of various genes expressions thus enabling them to maximize the chances of their survival [2]. Prolonged exposure of plants to biotic and abiotic stresses lead to reduction in fitness and ultimately in productivity [3]. Biotic stress induces a strong pressure on plants due to the attack of pathogens [4]. The attacking pathogens can be bacteria, fungus, Viruses or Nematodes. Bacteria enter the cell via gas or water pores (stomata and hydathodes) or any kind of wounds whereas; Fungi can enter the plant cell via extended hyphae on top or in between the plant cells. Nematodes and aphids obtain its feed by inserting a stylet directly into a plant cell [5]. Plants recognize the invaders by sensing evolutionary conserved microbial molecular signatures, known as pathogen-associated molecular patterns (PAMPs) or microbe-associated molecular patterns (MAMPs) by plant pattern recognition receptors (PRRs) [6–8]. After identifying PAMPs, PRRs triggers an immune response known as PAMP-triggered immunity (PTI), which imparts protection against non-host pathogens and limits disease caused by virulent pathogen [9]. Pathogen of the host plants release effector proteins which targets PTI components for inhibiting plant defense [10–14]. These effector proteins are recognized by Resistance proteins released by resistance genes (R-gene). Thus plants respond to these effector proteins by effector triggered immunity (ETI) which is highly specific and is assisted by hypersensitive response (HR) and systemic acquired resistance (SAR) [15,16].

Many major plants, including tomato (viz. *Solanum lycopersicum*), have been targeted by various pathogens which responds to their attacks by ETI [17]. Tomato, which belongs to Solanaceae family, is among the major vegetable plants and is the model system for fruit development [18]. It is a short lived herb and a protective food as it contains important nutrients such as lycopene, beta-carotene, flavonoids, Vitamins A, B & C and hydroxycinnamic acid derivatives [19–21]. From last few years, it has attained significant popularity because of its lycopene's anti-oxidative and anti-cancerous properties [22].

Tomatoes can be infected by around 200 diseases, which causes reduction in its production [23]. Many resistant genes (R-genes) in this plant has been revealed, whose proteins identify the avirulent proteins of pathogens and initiate the defense mechanisms. Some of the identified R-genes are Cf for *Cladosporium fulvum*, Ve for *Verticillium dahlia* and Cmm for *Clavibacter michiganensis* [24]. For preventing diseases in susceptible varieties various chemical, physical and biological methods are followed by the farmers to prevent the disease [25]. Since these

* Corresponding author.
E-mail address: satendra.singh@shiats.edu.in (S. Singh).

http://dx.doi.org/10.1016/j.gdata.2017.09.003
Received 15 November 2016; Received in revised form 20 September 2017; Accepted 23 September 2017
Available online 27 September 2017

methods cost high, are time consuming and causes health hazards so there raised the need for genetically modified crops. Thus, for the species prone to diseases, transgenic varieties have been developed. Be2 R-gene from pepper (which is a close relative of tomato) has been introduced into tomato which has been effective in providing resistance against bacterial spot disease [26].

Though > 1.5 million ESTs of tomato are available through various publications, but cumulative analysis to predict evidently existing gene model along with its number are yet to be attempted. Since whole genome based de novo assembly of tomato is available, thus these EST based gene prediction can be more realistic than genome based. Biotic stress resistant trait is one of the most important trait in tomato for productivity, thus its associated genes must be mapped over different chromosomes of tomato. Physical location of such genes can be of great advantage as > 0.14 M markers [27] are already catalogued chromosome-wise. Selection of flanking region markers can facilitate the introgression of specific desirable genes in varietal improvement program.

2. Materials and methods

2.1. ESTs collection and its assembly

In our study, firstly ESTs of various diseases of tomatoes were collected from NCBI [28]. These ESTs were for several bacterial, fungal, nematode and viral diseases. The gathered sequences were further checked for duplication by CD-HIT Software. CD-HIT compares all the sequences, separate the duplicates and cluster them into groups [29]. In CD-HIT 0.95 (95% identity) was used as clustering threshold and word size was taken as 5. The rectified sequences were then assembled into contigs using an assembly program EGassembler and Velvet programs. EGassembler is a programmed and user-customized analysis tool for cleaning, repeat masking, vector trimming, organelle masking, clustering and assembling of ESTs and other genomic fragments [30]. Velvet is a package that deals with de novo genome assembly and short read sequencing alignments [31]. For EGassembler minimum percent identity for an alignment was taken as 96% and overlap percent identity cutoff as 80. As far as Velvet is concerned, minimum contig length was settled at 100 and K-mer length was taken as 21.

2.2. Functional annotation and gene ontology analysis of contigs

The assembled contigs were further functionally annotated. Thus functional characterization and gene ontology study was carried out by using Blast2GO 2.8 [32]. In Blast2GO, Blastx program, with e-value cut off as 1.0E − 3 and number of blast hits as 20, was used to search against the protein database (Refseq Protein Database) for all six translated reading frames of generated contigs for each disease separately [33]. Mapping and Interpro functions of Blast2GO were used to describe the exact GO terms associated with annotations obtained from BLAST result. Result from blastx exhibited, the homologous sequences present in protein database corresponding to the contigs. These sequences were further categorized into gene ontology categories viz. Molecular Function, Biological Process and cellular components. Contigs with zero hits and no functional information were further carried for gene prediction each.

2.3. Gene prediction

Unannotated contigs from Blast2GO were further loaded in FGENESH program of Molquest software version 2.4.5 [34]. Molquest is a software package of various gene prediction programs. Genes were predicted using reference genome of *Solanum lycopersicum* and *Nicotiana tabacum*. As, *Solanum lycopersicum* and *Nicotiana tabacum* belong to the same family Solanaceae [35], and in Molquest (with version 2.4.5), genomic information of only these two plants of that family is present, so *Nicotiana tabacum* was also used as reference so that maximum genes of that family can be mined.

2.4. Promoter analysis of novel candidate genes

PLant Cis regulatory DNA Elements (PLACE) database [36] was used to validate and identify regulatory regions like cis regulatory elements, transcription factor, promoter regions in all the candidate genes predicted from FGENESH program.

2.5. Chromosome mapping

For the predicted genes, there locations on different chromosomes of *Solanum lycopersicum*, were identified with the help of Blast. Thereafter, these genes were mapped with the help of MapViewer available at NCBI (http://www.ncbi.nlm.nih.gov/mapview/).

2.6. MicroRNA prediction

MicroRNAs (miRNAs) are small RNA species that play important regulatory roles in various biological processes in plants. miRNAs are known to regulate the expressions of many stress-related genes [37]. miRNAs were obtained from miRBase (release 21) database and were used for identifying their target genes among predicted genes. miRBase database is an interface for annotated miRNA sequence data [38]. For target identification, miRANDA software was used with energy value as − 4 kJ/mol as cut off. The miRanda algorithm is based on a comparison of miRNAs complementarity to 3′UTR regions [39].

3. Results and discussion

3.1. ESTs collection and its assembly

The ESTs of several diseases affecting tomatoes, obtained from NCBI, were categorized broadly Bacterial, Fungal, Nematodes and Viral. Total 8581, 26,277, 41,832 and 1,441,496 were collected for bacterial, fungal, nematodes and viral diseases respectively. After removing duplicates number of ESTs reduced to 7101 for bacterial, 6611 for fungal, 21,541 for nematodes and 933,034 for viral diseases. Assembly was carried out with both EGassembler as well as Velvet assembler. Assembled contigs with EGassembler and Velvet for all the diseases are given in Table 1. It was observed that contigs assembled with EGassembler were larger in number as compared to those assembled with Velvet.

3.2. Functional annotation and gene ontology analysis of contigs

Contigs obtained from Velvet and EGassembler were further submitted for functional annotation. After completing blast, it was perceived that Contigs obtained from EGassembler had 120,136 hits, while from Velvet had 4382 hits. So, Contigs of Velvet assembler and EGassembler without blast hits were 5200 and 339 respectively. Disease wise details for with and without blast hits (by two different tools viz., EGassembler and Velvet) are given in Table 2. Thereafter mapping and annotation was done for the contigs for which blast hits

Table 1
ESTs and assembled contigs.

Disease	ESTs/ nucleotide collected	ESTs after repeat masking	Contigs generated (EGassembler)	Contigs generated (Velvet)
Bacterial	8581	7101	333	23
Fungal	26,277	6611	119	8
Nematode	41,832	21,541	3874	183
Viral	1,441,496	933,034	116,149	9371

Table 2
Disease wise details for with and without blast hits.

Disease	EGassembler		Velvet	
	Contigs without blast hits	Contigs with blast hits	Contigs without blast hits	Contigs with blast hits
Bacterial	10	323	3	20
Fungal	5	114	0	8
Nematode	209	3665	38	145
Viral	115	116,034	5159	4212

were obtained. These annotated sequences were categorized into three types — cellular components, biological processes and metabolic functions. For Velvet on an average 62.25% of the contigs were participating in biological processes, 67% for metabolic functions and 35.25% as cellular components. Whereas in case of EGassembler Contigs 66.25% were involved in biological processes, 64.5% for metabolic functions and 47.5% for cellular components on an average. A detail of this functional characterization is given in Table 3. Various pie charts for functional characterization of the annotated contigs are given Fig. 1.

3.3. Gene prediction and novel gene identification

The unannotated Contigs were used as input in FGENESH program of Molquest. It was observed that when *Solanum lycopersicum* was used as reference genome total 58 genes were predicted for Velvet Contigs and 38 genes for Contigs of EGassembler. But when *Nicotiana tabacum* (most phylogenetically related species to *Solanum lycopersicum* present in Molquest) was used as reference genome then 88 genes were predicted for Contigs of Velvet and 48 genes for EGassembler Contigs. After comparing the predicted genes, 50 novel genes were identified for viral disease, were there and for nematode there were 14 genes and for fungal disease there was one novel gene and for bacterial diseases there was one novel genes (with two isoforms) were identified.

3.4. Promoter analysis of novel candidate genes

Promoter analysis was performed to identify the probable cis-acting DNA sequences which may be liable in regulating candidate gene expression. The cis-acting sequences are actually a part of the gene and act as the regulatory sequences which influence the expression of the gene which contains them. Although, these sequences mostly found just upstream of the TSS but they can also be present much further upstream, or on the 3′ end of the gene, or even within the introns and exons of a gene [40]. Thus, the cis regulatory elements in the promoter region of the candidate genes were acquired from PLACE which is a database of motifs found in plant cis-acting regulatory DNA elements from all previously published reports. For the novel genes, cis regulating elements that regulate the biological process in diseased conditions were identified. To analyze this, PLACE database was used with all novel genes sequences as query sequences. Analysis of cis-regulatory elements describes the candidate gene expression and corresponding functional transcription factor. It was found that many diseases share the common transcription factors during pathogen attack. Cis-regulatory motifs and functioning transcription factor for each predicted candidate gene is given Supplementary sheet I. Among all the candidate gene, transcription factors like DOF, bZIP, WRKY, RAV, MYC and MYB expressed majorly. Details of transcription factors for genes predicted are given in Supplementary sheet I.

3.5. Chromosome mapping

Chromosome mapping helps in locating the genes of interest on the genome of an organism. When the chromosome mapping of predicted genes, done with the help of Blast and Mapviewer, the map shown in Fig. 2 was obtained (detailed locations of the all the genes are given in Supplementary sheet II. It was observed that out of 55 genes 12 were located on chromosome number 1. Chromosome numbers 4 and 11 possessed equal no of genes i.e. 8. Details of no. of genes on each chromosome are given in Table 4.

3.6. MicroRNA prediction

For all the predicted genes, microRNAs were predicted using miRANDA and miRBase data. It was found that 6 miRNAs were predicted for bacterial genes and 3 miRNAs for fungal genes. For genes related to nematodes had 24 miRNAs while that of genes related to viral diseases had 60 miRNAs. Details of the all the miRNAs for every gene is given in Supplementary sheets III. For total 50 genes, 69 unique miRNAs were identified.

4. Discussion

Development of stress tolerant varieties is one of the biggest challenges in plant breeding. For this the pathways related to various biotic stresses need to be deciphered and for this interpretation the genes should be known, so that the connectivity of these genes can be cracked. So keeping in view this point, new genes were identified from the available EST data at NCBI. These ESTs were assembled into contigs and these contigs were then annotated. From the annotation, it was observed that maximum number of contigs played role in metabolic processes and cellular process for bacterial, fungal, nematode and viral contigs.

Further, from unannotated contigs, 50 novel genes were predicted. For these genes promoter analysis was carried out. Many Transcription factors were identified for all the genes. The highly expressed transcription factors correspond to their role in biotic stress. It has found in previous studies that Dof transcription factor, with highly conserved Dof domain, play major role in two biotic stresses (watermelon mosaic virus and downy mildew) in cucumber [41]. Studies describe that Dof is a plant specific transcription factor and contains conserved C2-C2 zinc finger which help it in binding DNA. Thus it is known as DNA-binding One Zinc Finger [42]. bZIP transcription factor, also known as Basic Leucine Zipper, is present in all eukaryotes and is known to regulate processes corresponding to pathogen attack in many crops like Arabidopsis, cotton and maize [43]. Among the highly expressed WRKY transcription factor is participating in many stress (whether biotic or

Table 3
Functional characterization of the annotated contigs.

Disease	EGassembler			Velvet		
	Biological processes (in %)	Molecular functions (in %)	Cellular components (in %)	Biological processes (in %)	Molecular functions (in %)	Cellular components (in %)
Bacterial	71	68	66	90	95	45
Fungal	63	60	45	75	88	50
Nematode	62	61	38	60	61	34
Viral	69	69	41	24	24	12

abiotic) signaling pathways. Thus, it acts as a good candidate for various stress tolerance mechanisms [44]. MYB transcription factor represents a family of proteins that comprise a conserved domain, the MYB DNA-binding domain. It has been reported that in Arabidopsis, MYB encodes an activator of the hypersensitive cell death program in response to pathogen attack [45]. Thus presence of various cis regulatory elements on the promoter region of the novel genes indicates their possible involvement in biotic stress pathways in Tomato.

To find the location of the novel genes, chromosome mapping was done. It was observed that all the predicted genes related to viral pathogen are located on chromosome no 1 so it can be said that chromosome no 1 is majorly involved in pathways related to attack by viral pathogens in plants. Later, using the gene data, when miRNAs were predicted it was observed that 69 unique miRNAs were there which were involved in various processes in tomatoes.

5. Conclusion

Tomato is a very significant fruit crop and is being widely used due to its good nutritive value. It is affected by various pathogens as a result its production goes down. So being an important plant, there is a need to develop new resistant varieties which can survive under the stress

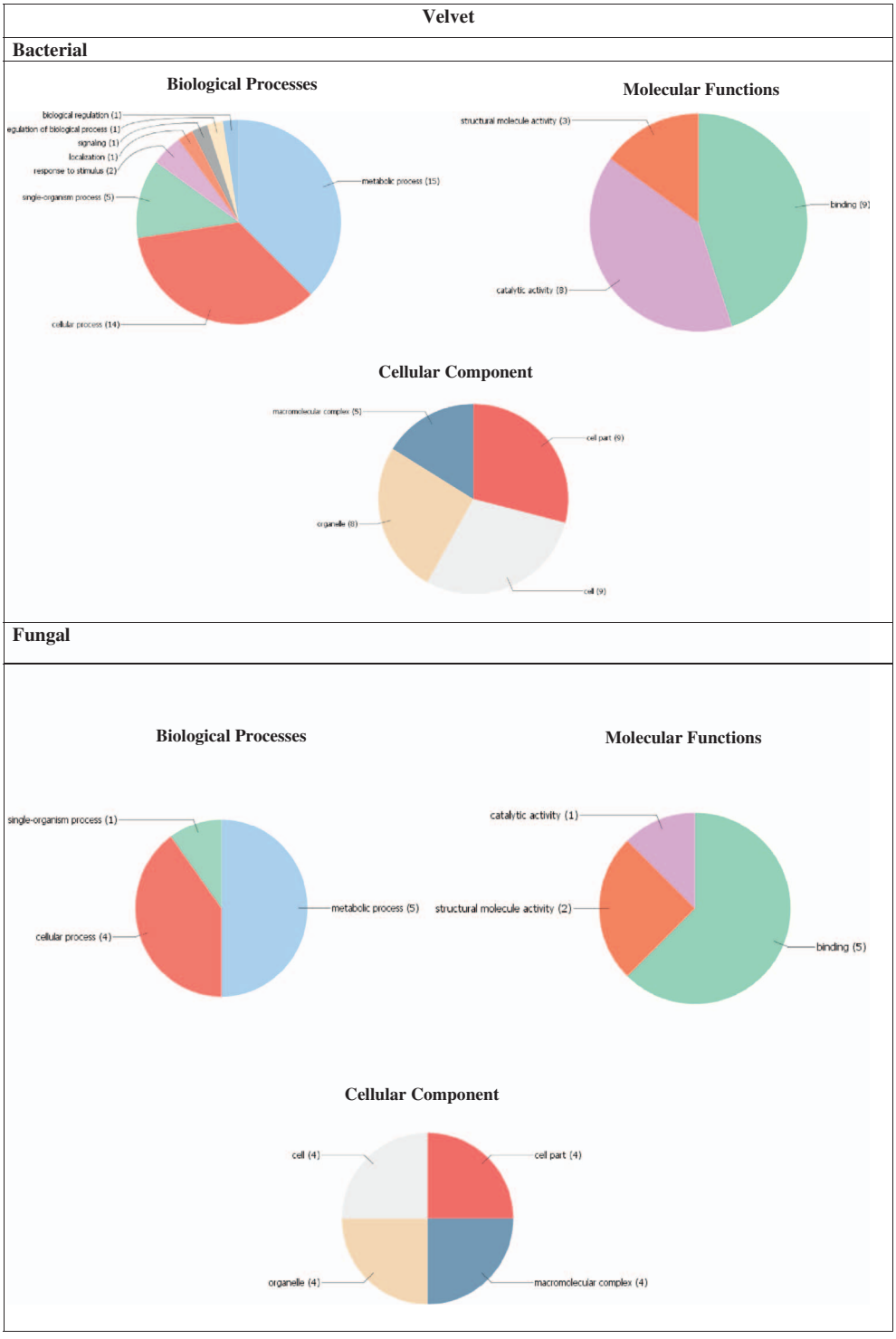
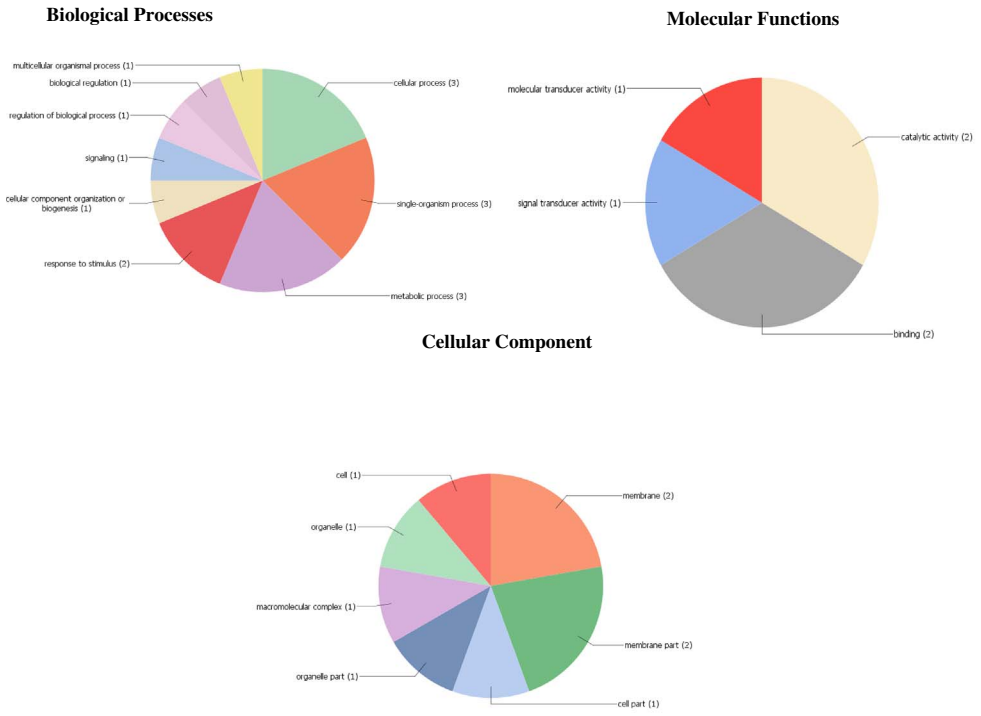


Fig. 1. Functional characterization of the annotated contigs (represented as pie charts).

Viral



Nematode

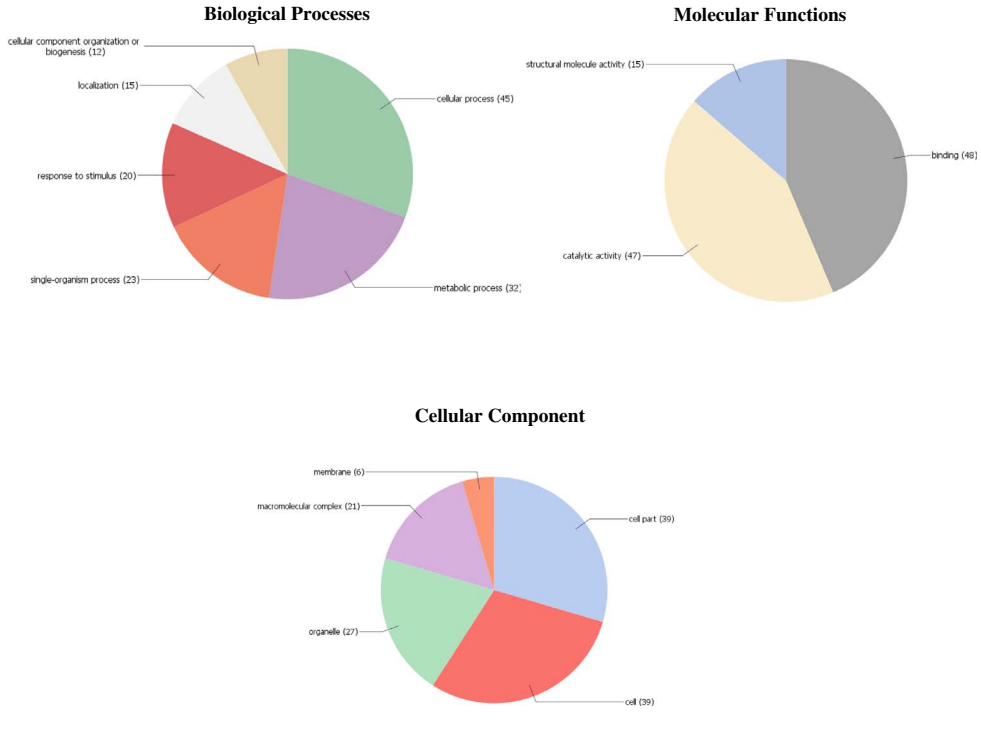
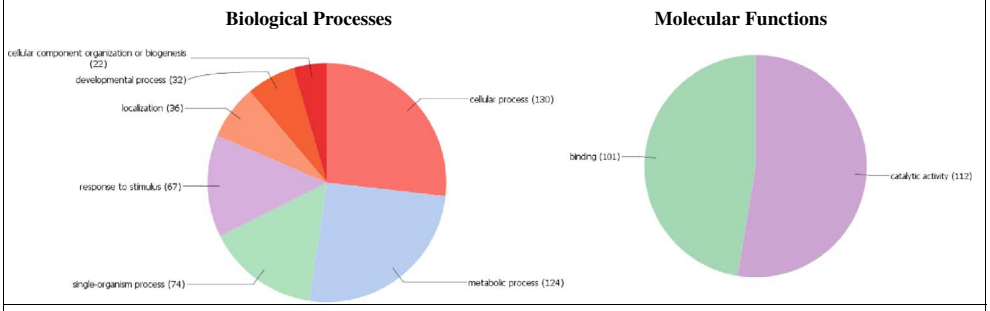
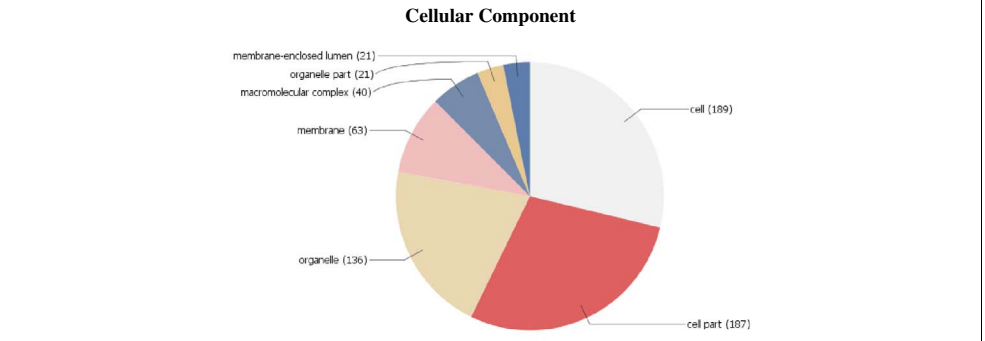


Fig. 1. (*continued*)

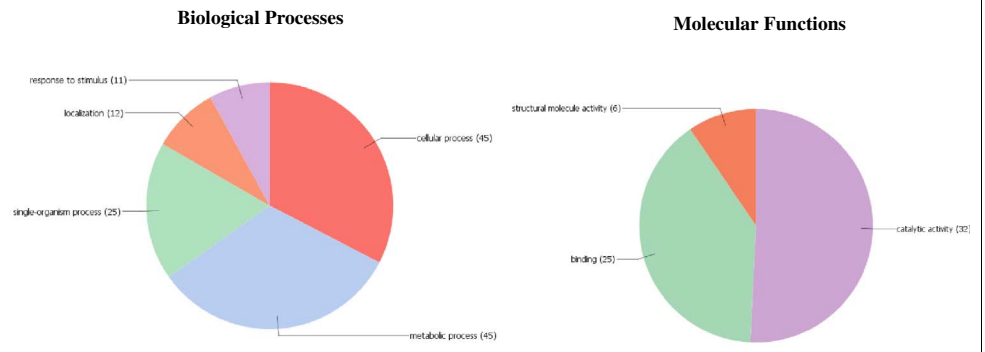
EGassembler

Bacterial





Fungal



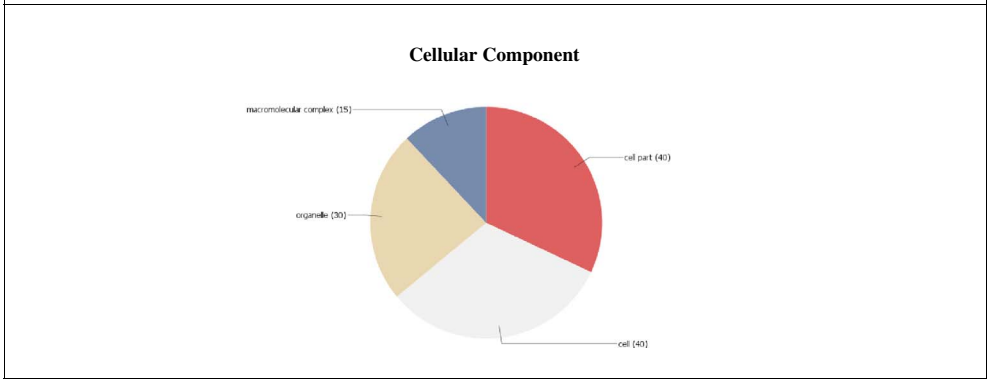
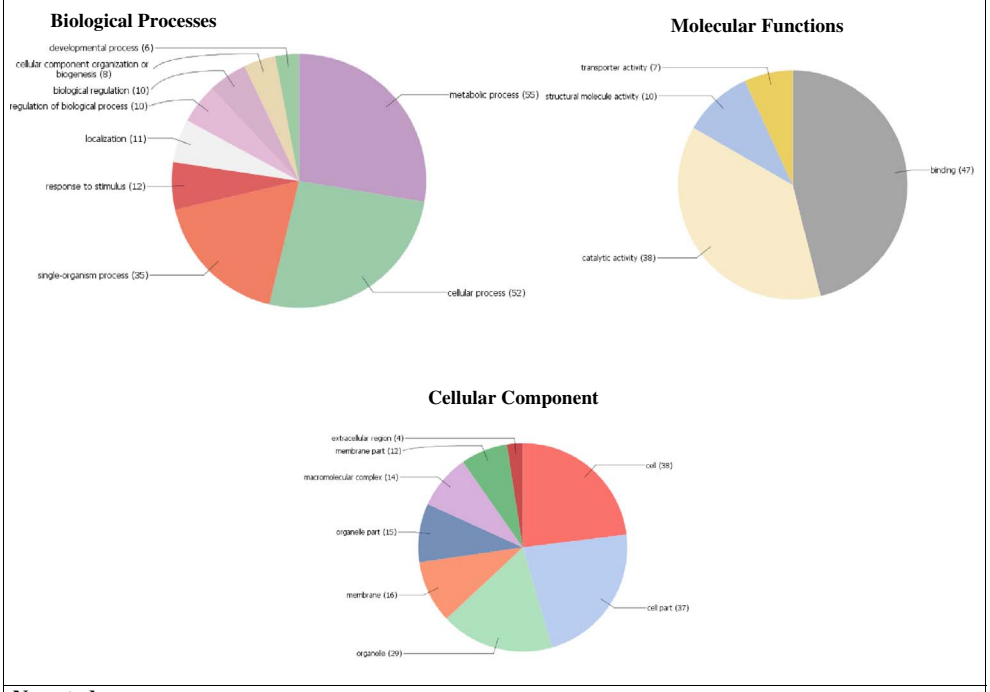


Fig. 1. (*continued*)

Viral



Nematode

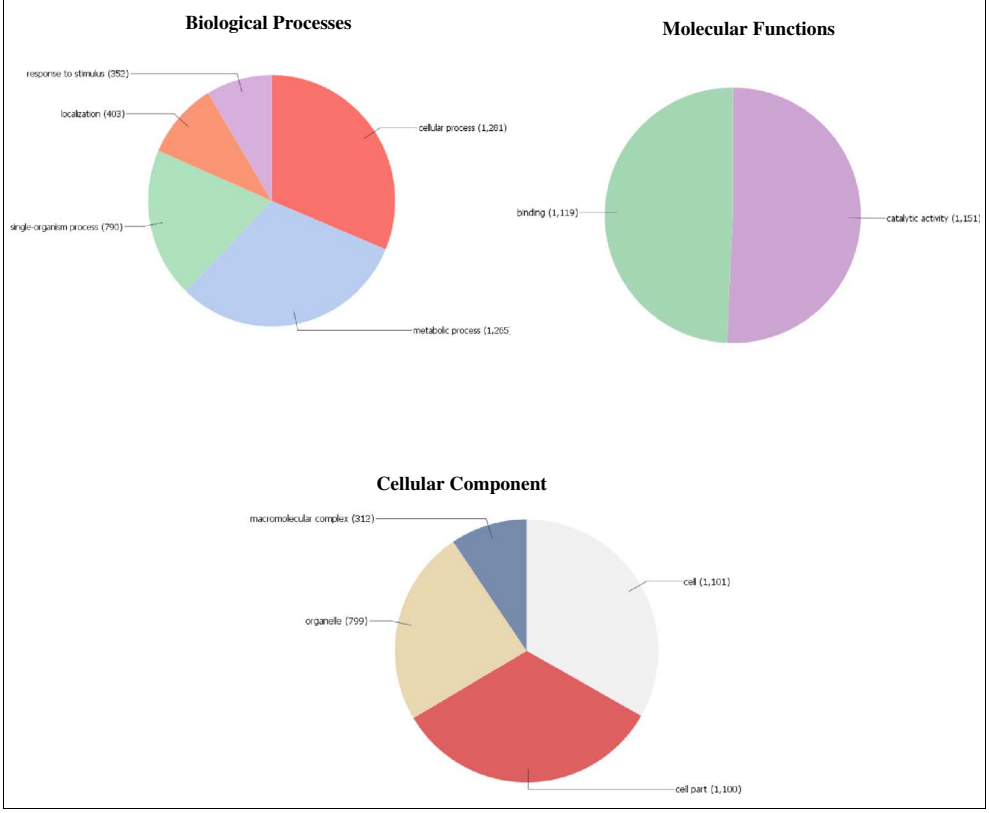


Fig. 1. (*continued*)

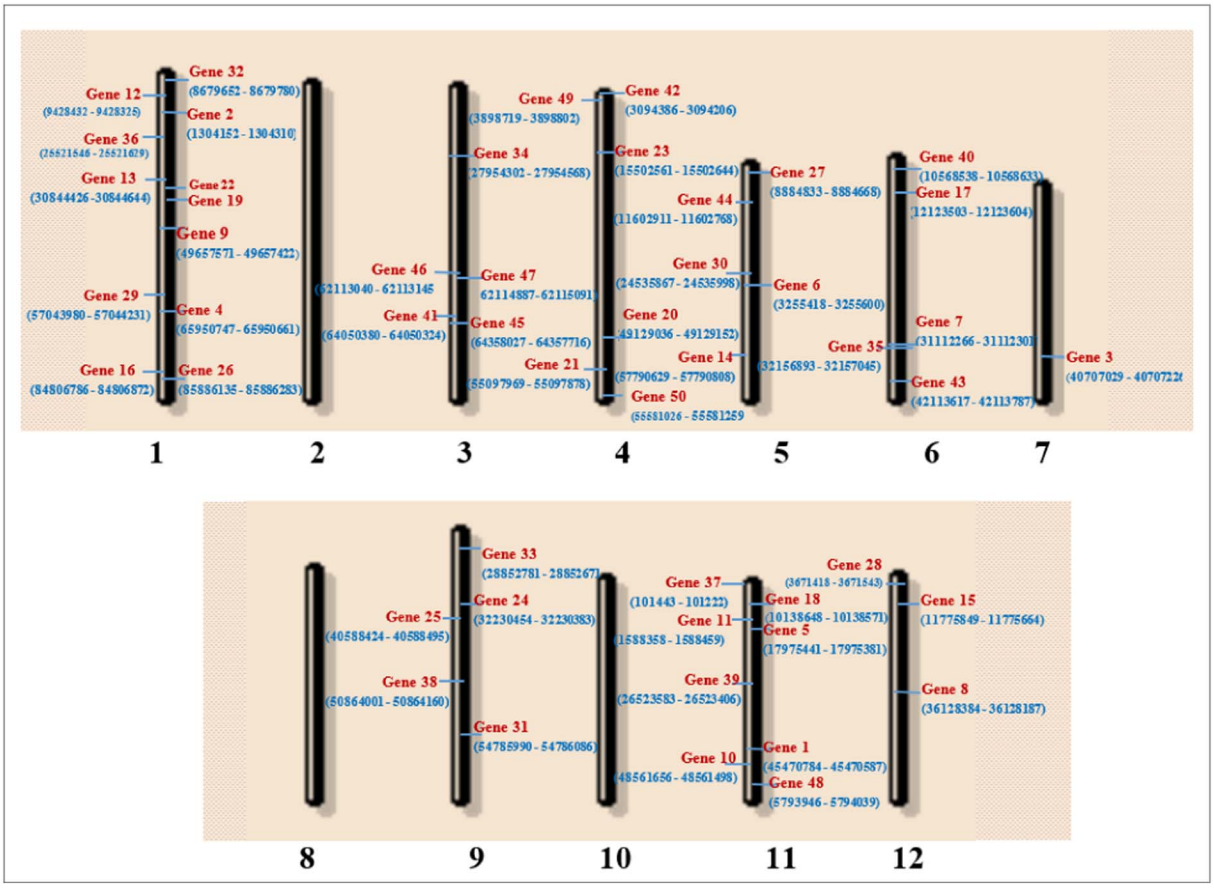


Fig. 2. Chromosome mapping of the predicted genes.

Table 4
No. of genes on each chromosome.

Chromosome no.	No of genes
ch01	12
ch02	0
ch03	3
ch04	8
ch05	5
ch06	5
ch07	1
ch08	0
ch09	5
ch10	0
ch11	8
ch12	3

condition. So using the genomics approach, with ESTs as an input data, various novel genes have been deciphered. The advance biotechnological methods need to be revolutionized to the new predicted genes for developing new resistant breeds. Cis regulatory elements and transcription factors study provides good insight of their role in corresponding stress condition, whose validation is further warranted.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gdata.2017.09.003.

References

[1] N.J. Atkinson, P.E. Urwin, The interaction of plant biotic and abiotic stresses: from genes to the field, J. Exp. Bot. 63 (2012) 3523–3544, http://dx.doi.org/10.1093/jxb/ers100.

[2] A.E. Dorantes-Acosta, C.V. Sánchez-Hernández, M.A. Arteaga-Vázquez, B. Carroll, Biotic Stress in Plants: Life Lessons from Your Parents and Grandparents, (2012), http://dx.doi.org/10.3389/fgene.2012.00256.

[3] I. Ben Rejeb, V. Pastor, B. Mauch-Mani, Plant responses to simultaneous biotic and abiotic stress: molecular mechanisms, Plants 3 (2014) 458–475, http://dx.doi.org/10.3390/plants3040458.

[4] M. Fujita, Y. Fujita, Y. Noutoshi, F. Takahashi, Y. Narusaka, K. Yamaguchi-Shinozaki, et al., Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks, Curr. Opin. Plant Biol. 9 (2006) 436–442, http://dx.doi.org/10.1016/j.pbi.2006.05.014.

[5] J.D.G. Jones, J.L. Dangl, The plant immune system, Nature 444 (2006) 323–329, http://dx.doi.org/10.1038/nature05286.

[6] F.M. Ausubel, Are innate immune signaling pathways in plants and animals conserved? Nat. Immunol. 6 (2005) 973–979, http://dx.doi.org/10.1038/ni1253.

[7] P. Bittel, S. Robatzek, Microbe-associated molecular patterns (MAMPs) probe plant immunity, Curr. Opin. Plant Biol. 10 (2007) 335–341, http://dx.doi.org/10.1016/j.pbi.2007.04.021.

[8] T. Boller, G. Felix, A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors, Annu. Rev. Plant Biol. 60 (2009) 379–406, http://dx.doi.org/10.1146/annurev.arplant.57.032905.105346.

[9] M.J. Pel, C.M. Pieterse, Microbial recognition and evasion of host immunity, J. Exp. Bot. 64 (2013) 1237–1248, http://dx.doi.org/10.1093/jxb/ers262.

[10] R.B. Abramovitch, R. Janjusevic, C.E. Stebbins, G.B. Martin, Type III effector AvrPtoB requires intrinsic E3 ubiquitin ligase activity to suppress plant cell death and immunity, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 2851–2856, http://dx.doi.org/10.1073/pnas.0507892103.

[11] T. Boller, S.Y. He, Innate immunity in plants: an arms race between pattern recognition receptors in plants and effectors in microbial pathogens, Science 324 (2009) 742–744, http://dx.doi.org/10.1126/science.1171647.

[12] H. Cui, Y. Wang, L. Xue, J. Chu, C. Yan, J. Fu, et al., Pseudomonas syringae effector protein AvrB perturbs *Arabidopsis* hormone signaling by activating MAP kinase 4, Cell Host Microbe 7 (2010) 164–175, http://dx.doi.org/10.1016/j.chom.2010.01.009.

[13] S.R. Grant, E.J. Fisher, J.H. Chang, B.M. Mole, J.L. Dangl, Subterfuge and manipulation: type III effector proteins of phytopathogenic bacteria, Annu. Rev. Microbiol. 60 (2006) 425–449, http://dx.doi.org/10.1146/annurev.micro.60.080805.142251.

[14] J.-M. Zhou, J. Chai, Plant pathogenic bacterial type III effectors subdue host responses, Curr. Opin. Microbiol. 11 (2008) 179–185, http://dx.doi.org/10.1016/j.mib.2008.02.004.

[15] C. Denoux, R. Galletti, N. Mammarella, S. Gopalan, D. Werck, G. De Lorenzo, et al., Activation of defense response pathways by OGs and Flg22 elicitors in *Arabidopsis* seedlings, Mol. Plant 1 (2008) 423–445, http://dx.doi.org/10.1093/mp/ssn019.

[16] M.T. Lotze, H.J. Zeh, A. Rubartelli, L.J. Sparvero, A.A. Amoscato, N.R. Washburn, et al., The grateful dead: damage-associated molecular pattern molecules and reduction/oxidation regulate immunity, Immunol. Rev. 220 (2007) 60–81, http://dx.doi.org/10.1111/j.1600-065X.2007.00579.x.
[17] C.-S. Oh, G.B. Martin, Effector-triggered immunity mediated by the Pto kinase, Trends Plant Sci. 16 (2011) 132–140, http://dx.doi.org/10.1016/j.tplants.2010.11.001.
[18] S. Sato, S. Tabata, H. Hirakawa, E. Asamizu, K. Shirasawa, S. Isobe, et al., The tomato genome sequence provides insights into fleshy fruit evolution, Nature 485 (2012) 635–641, http://dx.doi.org/10.1038/nature11119.
[19] G. Bauchet, M. Causse, Genetic diversity in tomato (*Solanum lycopersicum*) and its wild relatives, Genet. Divers. Plants, InTech, 2012, , http://dx.doi.org/10.5772/33073.
[20] C.M. Rick, The tomato, Sci. Am. 239 (1978) 76–87, http://dx.doi.org/10.1038/scientificamerican0878-76.
[21] D. Bhowmik, K.P. Sampath Kumar, S. Paswan, S. Srivastava, Tomato — a natural medicine and its health benefits, J. Pharmacogn. Phytochem. (2012) 1 www.phytojournal.com.
[22] S. Agarwal, A.V. Rao, Tomato lycopene and its role in human health and chronic diseases, CMAJ 163 (2000) 739–744 http://www.ncbi.nlm.nih.gov/pubmed/11022591.
[23] A.N. Lukyanenko, Disease Resistance in Tomato, (1991), pp. 99–119, http://dx.doi.org/10.1007/978-3-642-84275-7_9.
[24] M.R. Ercolano, W. Sanseverino, P. Carli, F. Ferriello, L. Frusciante, Genetic and genomic approaches for R-gene mediated disease resistance in tomato: retrospects and prospects, Plant Cell Rep. 31 (2012) 973–985, http://dx.doi.org/10.1007/s00299-012-1234-z.
[25] Yuliar, Y.A. Nion, K. Toyota, Recent trends in control methods for bacterial wilt diseases caused by *Ralstonia solanacearum*, Microbes Environ. 30 (2015) 1–11, http://dx.doi.org/10.1264/jsme2.ME14144.
[26] D.M. Horvath, R.E. Stall, J.B. Jones, M.H. Pauly, G.E. Vallad, D. Dahlbeck, et al., Transgenic resistance confers effective field level control of bacterial spot disease in tomato, PLoS One 7 (2012) e42036, , http://dx.doi.org/10.1371/journal.pone.0042036.
[27] M.A. Iquebal, V. Arora, N. Verma, A. Rai, D. Kumar, First whole genome based microsatellite DNA marker database of tomato for mapping and variety identification, BMC Plant Biol. 13 (2013) 1–197, http://dx.doi.org/10.1186/1471-2229-13-197.
[28] T. Madden, The NCBI Handbook, 16 (2003), pp. 1–17.
[29] Y. Huang, B. Niu, Y. Gao, L. Fu, W. Li, CD-HIT Suite: a web server for clustering and comparing biological sequences, Bioinformatics 26 (2010) 680–682, http://dx.doi.org/10.1093/bioinformatics/btq003.
[30] A. Masoudi-Nejad, K. Tonomura, S. Kawashima, Y. Moriya, M. Suzuki, M. Itoh, et al., EGassembler: online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments, Nucleic Acids Res. 34 (2006) W459–62, http://dx.doi.org/10.1093/nar/gkl066.
[31] D.R. Zerbino, E. Birney, Velvet: algorithms for de novo short read assembly using de Bruijn graphs, Genome Res. 18 (2008) 821–829, http://dx.doi.org/10.1101/gr.074492.107.
[32] A. Conesa, S. Götz, J.M. García-Gómez, J. Terol, M. Talón, M. Robles, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, Bioinformatics 21 (2005) 3674–3676, http://dx.doi.org/10.1093/bioinformatics/bti610.
[33] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410, http://dx.doi.org/10.1016/S0022-2836(05)80360-2.
[34] A.A. Salamov, V.V. Solovyev, Ab initio gene finding in *Drosophila* genomic DNA, Genome Res. 10 (2000) 516–522 http://www.ncbi.nlm.nih.gov/pubmed/10779491.
[35] S. Knapp, Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the Solanaceae, J. Exp. Bot. 53 (2002) 2001–2022, http://dx.doi.org/10.1093/jxb/erf068.
[36] K. Higo, Y. Ugawa, M. Iwamoto, T. Korenaga, Plant cis-acting regulatory DNA elements (PLACE) database: 1999, Nucleic Acids Res. 27 (1999) 297–300 http://www.ncbi.nlm.nih.gov/pubmed/9847208.
[37] N.A. Eckardt, A microRNA cascade in plant defense, Plant Cell 24 (2012) 840, http://dx.doi.org/10.1105/tpc.112.240311.
[38] S. Griffiths-Jones, R.J. Grocock, S. van Dongen, A. Bateman, A.J. Enright, miRBase: microRNA sequences, targets and gene nomenclature, Nucleic Acids Res. 34 (2006) D140–4, http://dx.doi.org/10.1093/nar/gkj112.
[39] A.J. Enright, B. John, U. Gaul, T. Tuschl, C. Sander, D.S. Marks, MicroRNA targets in *Drosophila*, Genome Biol. 5 (2003) R1, http://dx.doi.org/10.1186/gb-2003-5-1-r1.
[40] P.J. Wittkopp, G. Kalay, Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence, Nat. Rev. Genet. 13 (2011) 59, http://dx.doi.org/10.1038/nrg3095.
[41] C. Wen, Q. Cheng, L. Zhao, A. Mao, J. Yang, S. Yu, et al., Identification and characterisation of Dof transcription factors in the cucumber genome, Sci. Rep. 6 (2016) 23072, http://dx.doi.org/10.1038/srep23072.
[42] S. Yanagisawa, Dof domain proteins: plant-specific transcription factors associated with diverse phenomena unique to plants, Plant Cell Physiol. 45 (2004) 386–391 http://www.ncbi.nlm.nih.gov/pubmed/15111712.
[43] M.S. Alves, S.P. Dadalto, A.B. Gonçalves, G.B. De Souza, V.A. Barros, L.G. Fietto, Plant bZIP transcription factors responsive to pathogens: a review, Int. J. Mol. Sci. 14 (2013) 7815–7828, http://dx.doi.org/10.3390/ijms14047815.
[44] S.P. Pandey, I.E. Somssich, The role of WRKY transcription factors in plant immunity, Plant Physiol. 150 (2009) 1648–1655, http://dx.doi.org/10.1104/pp.109.138990.
[45] S. Ambawat, P. Sharma, N.R. Yadav, R.C. Yadav, MYB transcription factor genes as regulators for plant responses: an overview, Physiol. Mol. Biol. Plants 19 (2013) 307–321, http://dx.doi.org/10.1007/s12298-013-0179-1.