Vol.2 Iss.2 (2026); pp.34-40

Academians Journal of Biology

Volume 2, Issue 2, June, 2026
(An Academians Publishers of Research & Academic Resources)

ISSN: 3121-6404

OPEN ACCESS

Research Article

Genome-Wide Identification and Analysis of Lycopene Gene Family in Tomato (Solanum lycopersicum L.)

Lema A. M[1], Gumi A. M[2], Aliero A. A[3], Dauran N.S[4] and Sarkinfulani H. A[5]
Department of Biology, Usmanu Danfodiyo University, Sokoto State, Nigeria
Department of Plant Science, Usmanu Danfodiyo University, Sokoto State, Nigeria
Corresponding Author Email: lemaaisha990@gmail.com

Abstract: Tomato (Solanum lycopersicum L.) is an important vegetable crop and a major dietary source of lycopene, a naturally occurring carotenoid responsible for the characteristic red colour of ripe tomato fruits and recognized for its potent antioxidant properties. Despite the nutritional and biological importance of lycopene, there is limited information on the genomic organization and evolutionary characteristics of the lycopene-related gene family in tomato. In this study, an in-silico genome-wide analysis was conducted to identify and characterize the SlLYC gene family in S. lycopersicum. Lycopene-related genes were identified through genome-wide sequence analysis using Arabidopsis thaliana LYC gene sequences as queries, followed by verification of the identified proteins using InterPro. Chromosomal distribution, gene structure, subcellular localization, phylogenetic relationships, and conserved protein motifs were subsequently analyzed using appropriate bioinformatics tools. A total of four non-redundant SlLYC genes were identified and mapped to chromosomes 4, 6, 10, and 12 of the tomato genome. Subcellular localization analysis predicted that three SlLYC proteins were localized in the chloroplast, while one was predicted to be localized in the nucleus. Gene structure analysis revealed variation in exon-intron organization, with two of the identified genes being intronless. Phylogenetic analysis grouped the identified SlLYC proteins into three major groups in relation to LYC proteins from Arabidopsis thaliana. In addition, three conserved motifs were identified across the SlLYC protein sequences, indicating functional conservation within the gene family. These findings provide useful information on the genomic organization, evolutionary relationships, and structural characteristics of the SlLYC gene family and may support future molecular studies, genetic improvement, and breeding strategies aimed at enhancing lycopene content and nutritional quality in tomato.

Keywords: *Tomato; Lycopene; SlLYC gene; Biosynthesis; Genome evolution*

1. Introduction

Tomato (*Solanum lycopersicum* L.), a member of the Solanaceae family, is the second most significant fruit or vegetable crop globally after potato (*Solanum tuberosum* L). According to FAO (2019), around 182.3 million tons of tomato fruits are harvested annually across 4.85 million hectares of land. Asia is the leading producer, contributing 61.1% of global tomato output, followed by Europe (13.5%), the Americas (12.4%) and Africa (11.8%). Tomato serves as one of the most extensively studied model plants, particularly valuable for scientific research focused on fruit development and quality traits (Karlovę *et al*., 2014; Kim *et al*., 2018; Liu *et al*., 2018). Both cultivated tomato and its wild relatives are diploid species, possessing 12 chromosomes (2n=2x=24), which were first identified by Barton in 1950. These chromosomes are structurally divided into pericentric heterochromatin and distal euchromatin regions. The tomato genome is relatively compact, with an estimated size of about 950 mega bases (Mb). Repetitive sequences, which constitute roughly 30% of the genome, are predominantly found in the heterochromatic regions (Van der Hoeven *et al.,* 2002). The genome encodes approximately 35,000 genes, most of which are densely arranged in the euchromatic regions.

Lycopene as a phytonutrient and powerful antioxidant, is a naturally occurring carotenoid responsible for the vibrant red colour in tomato, watermelon and pink grapefruit (Alnokkari, 2023). Tomato-derived lycopene has been shown to possess the strongest antioxidant capacity and singlet oxygen-quenching potential among all dietary carotenoids (Imran *et al.,* 2020; Przybylska,

Received: 21 March, 2026; Received in revised form: 10 April, 2026; Accepted: 08 May, 2026; Available Online: 10 June, 2026
Corresponding Author: Lema A. M

2020). Given its health benefits, regular consumption of lycopene through tomatoes is considered a promising strategy for the prevention of chronic diseases such as cancer, heart disease, and diabetes (Alda *et al.,* 2009; Camera *et al.,* 2013). Furthermore, lycopene is known to enhance intercellular communication, influence hormonal and immune responses, and regulate various metabolic pathways (Amer *et al*., 2020). Lycopene concentration in tomatoes varies by species and typically increases as the fruit matures and ripens (Yoo *et al.,* 2017). The noticeable shift in colour during the ripening process is primarily due to the breakdown of chlorophyll and the subsequent accumulation of lycopene, and acyclic carotenoid in both the fruit skin and pulp (Seymour *et al*., 2013; Borghesi *et al*., 2016; D´Ambrosio *et al*., 2018). Lycopene is produce via carotenoids biosynthesis pathway which begins with the production of phytoene, a colourless compound synthesized from geranylgeranyl diphosphate through the enzymatic action of phytoene synthase (PSY). Subsequently, three key enzymes' phytoene desaturase (PDS), ζ-carotene desaturase (ZDS), and carotenoid isomerase (CRTISO) work sequentially to convert phytoene into lycopene via intermediates such as phytofluene and ζ-carotene. After the formation of lycopene, the carotenoid pathway branches into two directions, producing α- and β-carotenoids through the catalytic actions of lycopene cyclase and lycopene β-cyclase enzymes (Cazzonelli *et al.,* 2010; Shi *et al.,* 2014).

Despite the importance of lycopene in determining tomato fruit and nutritional value, there is a limited understanding of the lycopene gene family in tomato. This research provides an in-depth analysis of the lycopene gene family in tomato. This can contribute by providing insight into plant breeding and genetic by supporting genomics- guided selection and the identification of key genes for trait improvement.

2. Materials and Methods

Identification and Characterization of Lycopene Gene in *Solanum lycopersicum*

Lycopene related genes were identified using the *S. lycopersicum* whole genome released by ensembl plant (http://plant.ensambl.org) via blast search, using the Arabidopsis gene sequence as queries. Arabidopsis gene sequences were obtained from the TAIR10 genome assembly released by NCBI (http://www.ncbi.nlm.nih.gov). All the protein sequences of *S. lycopersicum* for LYC gene were further verified using interpro 102.0: http://www.ebi.ac.uk\interpro (Paysan *et al.,*2022). Additionally, all LYC protein sequence of *Arabidopsis thaliana* was retrieved from plant ensembl and verified using interpro.

Chromosomal location, Gene structure and Subcellular localization

Chromosomal locations of the identified SlLYC genes were displayed using the mapgene2chrom 2.1 software (http://mg2c.iask.in). Tandem duplicate genes were defined as adjacent homologous gene on a single chromosome with no more than one intervening gene (Zhao *et al.,*2014). For gene structure, exon-intron structure of SlLYC genes were examine by comparing CDS sequence and corresponding genomic sequence using online Gene Structure Display Sever-GSDS 2.0: http://gsds.cbi.pku.edu.cn (Hu *et al*.,2015). The subcellular localization of SlLYC protein was predicted using plant subcellular localization integrative predictor (PSI predictor) http://bis.zju.edu.cn\psi according to method of Liu *et al*. (2013) using P-value of <0.05.

Phylogenetic Analysis and Identification of Conserved Motifs

The amino acid sequence of SlLYC were imported in to MEGAX and multiple sequence alignment were performed using ClustalW Omega with a gap and gap extension penalties of 10 and 0.1 respectively (Tamura *et al*., 2011). The aligned sequences were used to construct an unrooted phylogenetic tree based on the neighbor-joining (NJ) method using bootstrap values of 1,000 replicates (Saitou and Nei, 1987) and the tree was displayed using Interactive Tree of Life- iTOL: http://itol.embl.de/index.shtml (Letunic and Bork, 2011). To deduce the evolutionary relationship of LYC proteins among Tomatoes species (*Solanum lycopersicum*) and *Arabidopsis thaliana,* additional LYC proteins were downloaded from plant ensembl database as described above and used to construct an unrooted phylogenetic tree (Jin *et al.,*2017). Evolutionary analyses were conducted in MEGA X and the number of bootstrap replicates will be 1,000. For conserved motifs identification, multiple EM for motif elicitation (MEME) suites- http://meme-suite.org/ was used (Bailey, 2009). To identify motifs within the SILYC subfamily, with the following parameters: any number of repetitions, a maximum of 10 mismatches, and an optimum motif size of 6–70 amino acid residues.

3. Results

Identification and Characterization of Lycopene Gene in *Solanum lycopersicum*

Arabidopsis gene and ensemble plant (http:\\plant.ensambl.org) were used to comprehensively identify the SlLYC gene. In total we identified 4 non-redundant LYC gene member which were further named according to the name of their SlLYC orthologs.

Chromosomal location, Gene structure and Subcellular localization

To examine the genomic distribution of the identified SlLYC gene, their approximate location on chromosome were analyzed. All the identified SlLYC gene were mapped on the chromosome of *S lycopersicum* named as 4,6,10 and 12 (Figure 1). To investigate the evolution of

SlLYC gene the intron-exon architecture was analyzed by comparing the CDS sequence and genomic sequence of each SlLYC gene (Figure 2). Among the 4 of the SlLYC gene, 1 having 11 exon and 10 introns (1, 25%), 1 with 6 exon and 5 introns (1, 25%) while the remaining 2 are intron less with only exon (2, 50%). 1 of S1LYC gene having upstream\downstream. 2 are up to 1.5kb, 1 up to 6kb and 1 ≥ 6kb. The subcellular localization analysis showed 3 SlLYC gene were localized in the chloroplast of the cell while 1SlLYC gene was in the nucleus.

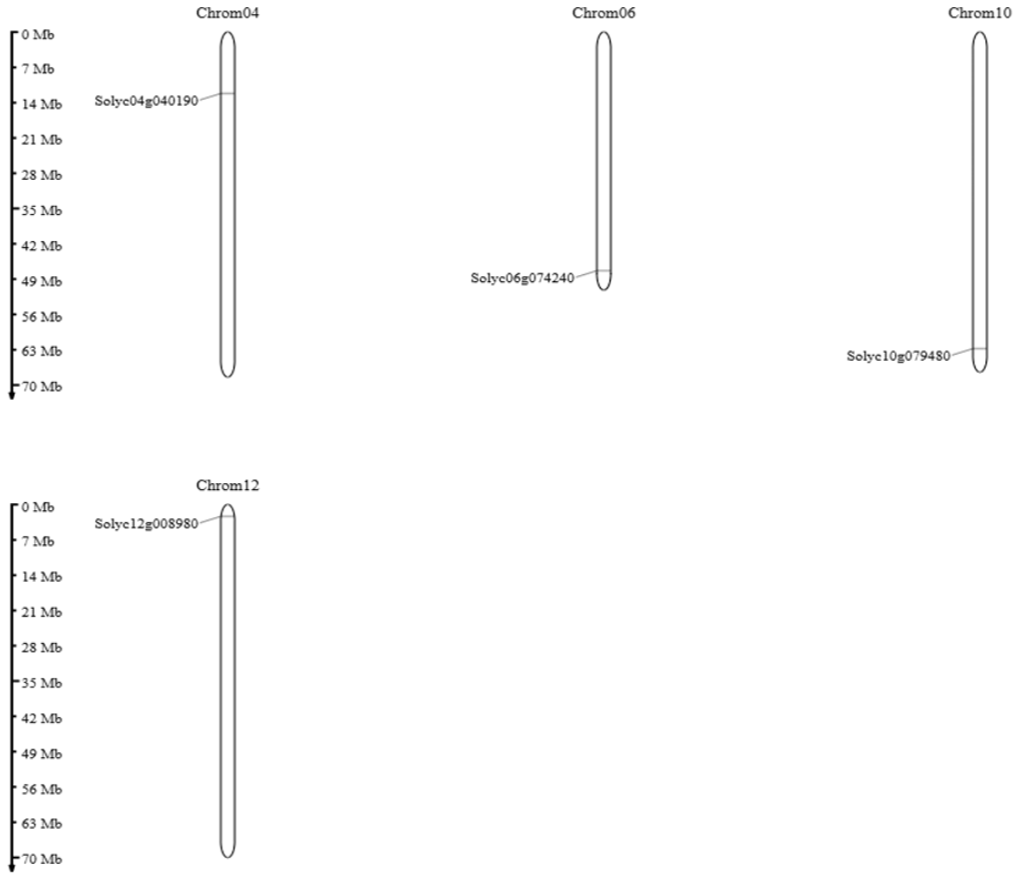

Figure 1: Chromosomal location of identified of SlLYC gene. 4 SlLYC gene were mapped on chromosome (4, 6,10 and12). A scale on the left shows the chromosome length (Mb)

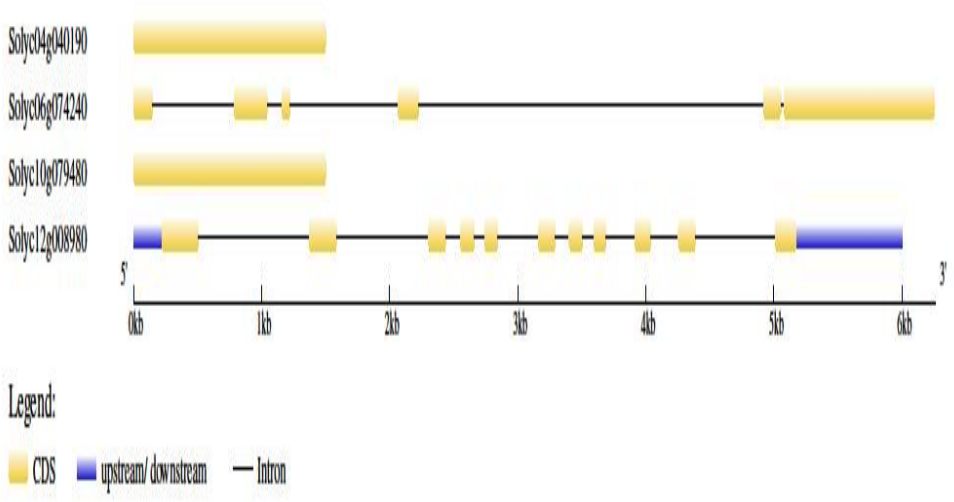

Figure 2: Exon-intron structure of each *SlLYC* gene is proportionally displayed in terms of the scale at the bottom. Golden bars represent exons, black lines represent introns and blue bars represent upstream\dowmstream. Most of the *SlLYC* genes are intronless consisting only of exons and are <1,500 bp in length.

Phylogenetic Analysis and Identification of Conserved Motifs

To explore the phylogenetic relationship of identified LYC protein in *S. lycopersicum*, a phylogenetic tree of SlLYC and *Arabidopsis* LYC protein was constructed using the neighbor-joining method. According to the phylogenetic result SlLYC protein was grouped in 3, group 1 consist of 2 SlLY protein and 1 *Arabidopsis* LYC protein, group 2 consist of 1 SlLYC protein only and group 3 consist of 1 SlLYC protein and 1 *Arabidopsis* LYC protein. This revealed that the LYC protein *S. lycopersicum* is higher than that of *Arabidopsis.* (Figure 3). Motif Conserved analysis, a total of 3 conserved motif were identified in the SlLYC protei using meme software. The 3 motifs occur in all the identified SlLYC protein (Figure 4).

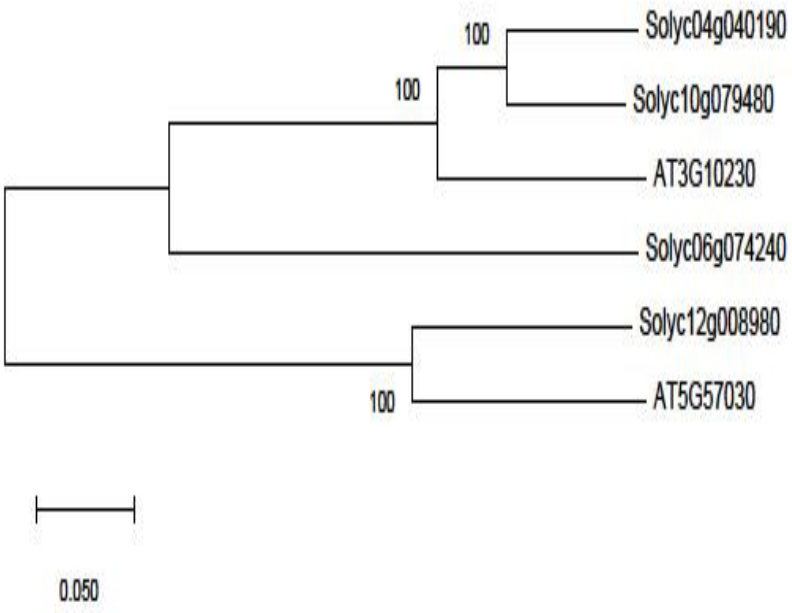

Figure 3: Phylogenetic analysis of LYC proteins between *Solanum lycopersicum* and *Arabidopsis thaliana*. A total of 6 (4 SlLYC and 2 AtLYC) conserved protein domains were aligned using ClustalW, and the phylogenetic tree was constructed by MEGA X using the neighbour-joining method with 1,000 bootstrap replicates.

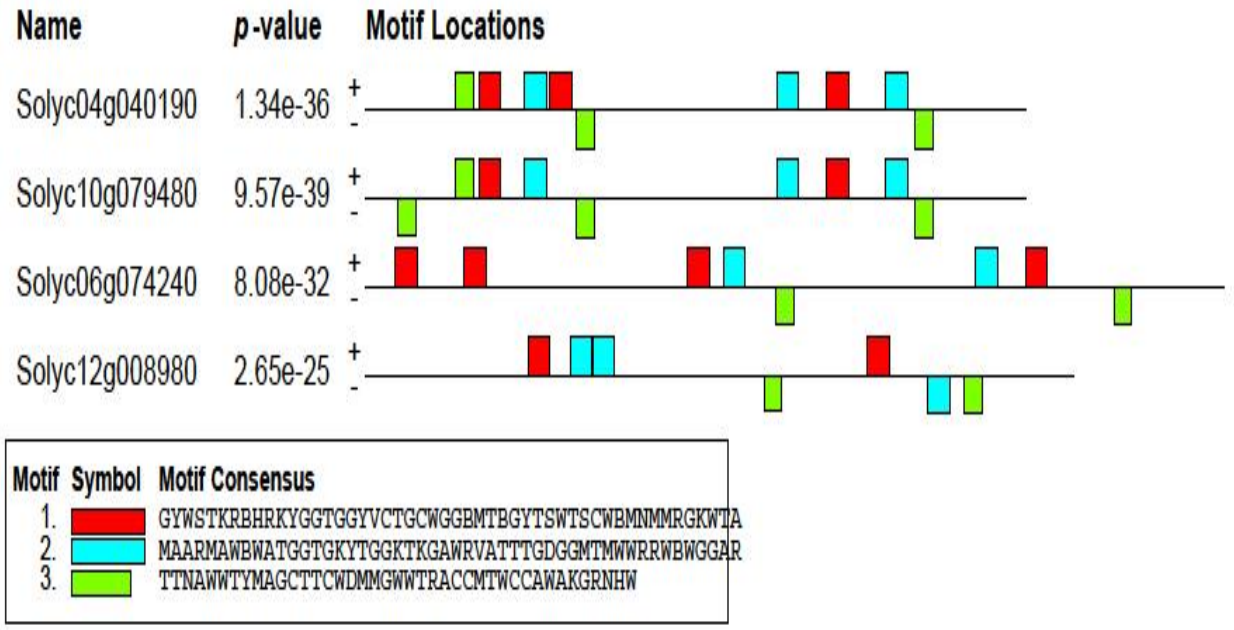

Figure 4: Conserved motif of SlLYC protein. Different coloured boxes represent different conserved motifs and their consensus sequences were shown below.

4. Discussion

In this study, the total number of LYC genes in *S. lycopersicum* is higher than that of *Arabidopsis thaliana* (2) and lower than the Carotenoid metabolism related gene in grapevine fruit {54} (Xiangpeng, 2017) and cabbage {67}(Wenxue, 2025) and Among the studied species, the genome size of *S. lycopersicum* (950mb) is larger than that of *Arabidopsis thaliana* (135 Mb), grapevine fruit (500 Mb) and cabbage (530 Mb) the diversity in genome size of plants may influence the number of carotenoid gene in plant species and a larger genome may allow for more genetic diversity and variation in carotenoid biosynthesis. Lycopene can subsequently undergo cyclization to form other carotenoids; α and β carotenoid through the action of two lycopene cyclase and lycopene β-cyclase (Cazzonelli *et al*., 2010; Shi *et al*., 2014). Carotenoids participate in a wide range of physiological processes, including plant growth, development, and responses to environmental stimuli. Besides their functions as pigments and nutrients, carotenoids are also the precursors of many important volatile flavor compounds in plants, conferring the sensory attribute that can be detected by consumers (Vogel *et al*., 2010).

To examine the genomic distribution of the identified SlLYC gene, their approximate location on chromosome were analyzed. All the identified 4 SlLYC gene were mapped on the chromosome *S. lycopersicum* number 4,6,10 and 12. 46 out of the 62-carotenoid gene in cabbage are located on the chromosome C01-C09, respectively while 16 on the scaffold. In our study, the intron-exon architecture of the 4 LYC gene revealed that 1 having 11 exons and 10 introns (1,25%), 1 with 6 exons and 5 introns (1,25%) while the remaining are intron less, (2, 50%) having only exons. The existence of introns in a gene leads to alternative splicing of the gene product. The process of removing immature mRNA by splicing occurs in a majority of eukaryotic protein-coding genes, producing variant transcripts of the same gene (Filichkin *et al*., 2015). Alternative splicing affects transcript export, protein translation, and degradation (Reddy *et al*., 2013). Though alternative splicing in plants is changeable during the growth, developmental process, and environmental stresses (Staiger and Brownx, 2013). Motif analysis among SlLYC protein sequence indicated a conserved array of the motif across all identified member of LYC subfamily.

The phylogenetic analysis revealed the presence of 4 LYC protein *S. lycopersicum*, divided into three major groups. Group 1 included 2 SlLYC protein, while group 2 and 3 contained 1 each. *Arabidopsis thaliana* possessed only one LYC protein in group 1 and 3. This disparity in the number of LYC homologs between tomato and Arabidopsis suggests a possible gene duplication event in tomato leading to the expansion of LYC gene family. Evolutionary changes have helped plants adapt to both abiotic and biotic stresses by

controlling a network of certain genes through the modulation of specific transcription factors (Nan and Goe, 2019). The subcellular location tool predicted 3 SlLYC are chloroplast protein while the remaining 1is a nucleus protein. Carotenoids are formed and sequestered at very high levels in plastids, mainly chloroplasts and chromoplasts (Egea *et al*., 2010).

5. Conclusion

Genomic analysis identified 4 non-redundant lycopene gene distributed across chromosomes 4, 6, 10 and 12. Gene structure analysis revealed variation in exon-intron architecture with some gene being intron less. Subcellular localization indicates that most SlLYC gene are chloroplast localized, aligned with their role in pigment biosynthesis. SlLYC gene were grouped in to 3 based on phylogenetic analysis. Conserved motif in SlLYC protein shown functional conservation with in the gene family. The finding provides valuable insight for breeding programs aimed at enhancing lycopene content and improving nutritional quality.

References

Alda, L. M., Gogoasa, I., Bordean, D. M., Gergen, I., Alda, S., & Moldovan, C. (2009). Bioavailability of all-trans and cis-isomers of lycopene. *Journal of Applied Polymer Technology, 15*(4), 540–542.

Alnokkari, A. S. (2023). Optimizing analytical methods to determine lycopene levels in Syrian tomatoes from various regions. *Agricultural Science Digest, 44*, 1–7. https://doi.org/10.18805/ag.DF-550

Amer, S. A., Kishawy, A. T., Osman, A., Mahrose, K. M., Hassanine, S. I., & Rehman, Z. U. (2020). Influence of dietary graded levels of lycopene on the growth performance, muscle, cholesterol level and oxidative status of Japanese quail fed high-fat diet. *Anais da Academia Brasileira de Ciências, 92*(2), 1678–1782.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., & Clementi, L. (2009). MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Research, 37*, 202–208. https://doi.org/10.1093/nar/gkp335

Borghesi, E., Ferrante, A., Gordillo, B., Rodríguez-Pulido, F. J., Cocetta, G., Trivellini, A., Mensuali, A. S., Malorgio, F., & Heresdia, F. J. (2016). Comparative physiology during ripening in tomato rich-anthocyanins fruits. *Plant Growth Regulation, 80*, 207–214. https://doi.org/10.1007/s10725-016-0158-y

Camera, M., Cortes, M., Meta, S., Ruiz, V. F., Camerax, R. A., & Manzoor, S. (2013). *Lycopene: A review of chemical and biological activity related to beneficial health effects*. In A. Rahman (Ed.), *Studies in natural products chemistry* (Vol. 40, pp. 383–426). Elsevier. https://doi.org/10.1016/B978-0-444-59603-1.00011-4

Cazzonelli, C. I., & Pogson, B. J. (2010). Source to sink: Regulation of carotenoid biosynthesis in plants. *Trends in Plant Science, 15*, 266–274. https://doi.org/10.1016/j.tplants.2010.02.003

D'Ambrosio, C., Stigliani, A. L., & Giorio, G. (2018). CRISPR/Cas9 editing of carotenoid genes in tomato. *Transgenic Research, 27*, 367–378. https://doi.org/10.1007/s11248-018-0079-9

Egea, I., Barsan, C., Bian, W., Purgatto, E., Latche, A., Chervin, C., Bouzayen, M., & Pech, J. C. (2010). Chromoplast differentiation: Current status and perspectives. *Plant and Cell Physiology, 51*, 1601–1611.

Filichkin, S., Priest, H. D., Megraw, M., & Mockler, T. C. (2015). Alternative splicing in plants: Directing traffic at the crossroads of adaptation and environmental stress. *Current Opinion in Plant Biology, 24*, 125–135.

Food and Agriculture Organization of the United Nations. (2019). *FAOSTAT*. http://www.fao.org/faostat/

Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., & Gao, G. (2015). GSDS 2.0: An upgraded gene features visualization server. *Bioinformatics, 31*, 1296–1297.

Imran, M., Ghorat, F., Ul-Haq, I., Ur-Rehman, H., Aslam, F., Heydari, M., & Shariati, M. A. (2020). Lycopene as a natural antioxidant used to prevent human health disorders. *Antioxidants, 9*(8), 706–727.

Jin, J. P., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J. C., & Gao, G. (2017). Plant TFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research, 45*(1), D1040–D1045.

Karlová, R., Chapman, N., David, K., Angenent, G. C., Seymour, G. B., & de Maagd, R. A. (2014). Transcriptional control of fleshy fruit development and ripening. *Journal of Experimental Botany, 65*, 4527–4541. https://doi.org/10.1093/jxb/eru316

Kim, J. Y., Kim, S. K., Jung, J., Jeong, M. J., & Ryu, C. M. (2018). Exploring the sound-modulated delay in tomato ripening through expression analysis of coding and non-coding RNAs. *Annals of Botany, 122*, 1231–1244. https://doi.org/10.1093/aob/mcy134

Letunic, I., & Bork, P. (2011). Interactive Tree of Life v2: Online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research, 39*, 475–478.

Liu, Z., Kong, L., Zhang, M., Lv, Y., Liu, Y., Zou, M., Lu, G., Cao, J., & Yu, X. (2013). Genome-wide identification, phylogeny, evolution and expression patterns of AP2/ERF genes and cytokinin response factors in *Brassica rapa* ssp. *pekinensis*. *PLoS ONE, 8*(12), e83444. https://doi.org/10.1371/journal.pone.0083444

Liu, Z., Ren, Z., Zhang, J., Chaung, C. C., Kandaswamy, E., Zhou, T., & Zou, L. (2018). Role of ROS and nutritional antioxidants in human diseases. *Frontiers in Physiology, 9*, 477. https://doi.org/10.3389/fphys.2018.00477

Nan, H., & Goe, L. (2019). Genome-wide analysis of WRKY gene and their responses to hormone and mechanical stress in carrot. *Frontiers in Genetics, 10*, 363. https://doi.org/10.3389/fgene.2019.00363

Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., & Salazar, G. A. (2022). InterPro. *Nucleic Acids Research, 51*(1), 412–427. https://doi.org/10.1093/nar/gkac993

Przybylska, S. (2020). A bioactive carotenoid offering multiple health benefits: A review. *International Journal of Food Science & Technology, 55*(1), 11–32.

Reddy, A. S., Marquez, Y., Kalyna, M., & Barta, A. (2013). Complexity of the alternative splicing landscape in plants. *The Plant Cell, 25*(10), 3657–3683.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution, 4*, 406–425.

Seymour, G. B., Chapman, N. H., Chew, B. L., & Rose, J. K. (2013). Regulation of ripening and opportunities for control in tomato and other fruits. *Plant Biotechnology Journal, 11*, 269–278. https://doi.org/10.1111/j.1467-7652.2012.00738.x

Shi, Y., Wang, R., Lou, Z., Jin, L., Liu, P., Chen, Q., Li, F., Wei, C., Wu, M., Wei, P., Xie, H., Qu, L., Lin, F., & Yang, J. (2014). Molecular cloning and functional characterization of the lycopene ε-cyclase gene via virus-induced gene silencing and its expression pattern in *Nicotiana tabacum*. *International Journal of Molecular Sciences, 15*, 14766–14785.

Staiger, D., & Brown, J. W. (2013). Alternative splicing at the intersection of biological timing, development, and stress responses. *The Plant Cell, 25*(10), 3640–3656.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., & Nei, M. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution, 28*, 2731–2739.

Van der Hoeven, R. S., Ronning, C., Jim, G., Gregory, M., & Steven, K. (2002). Deductions about the number, organization, and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. *The Plant Cell, 14*(7), 1441–1456.

Vogel, J. T., Tieman, D. M., Sims, C. A., Odabasi, A. Z., Clark, D. G., & Klee, H. J. (2010). Carotenoid content impacts flavor acceptability in tomato (*Solanum lycopersicum*). *Journal of the Science of Food and Agriculture, 90*, 2233–2240.

Wenxue, C., Peng, W., Limei, Y., Zhiyuan, F., Yangyong, Z., Mu, Z., Honghao, L., Yong, W., & Jialei, J. (2021). Carotenoid biosynthetic genes in cabbage: Genome-wide identification, evolution, and expression analysis. *Genes, 12*(12), 2027. https://doi.org/10.3390/genes12122027

Xiangpeng, L., Peipei, W., Chen, W., Xudong, Z., Xiaopeng, L., Hongyan, L., Qian, M., Ao, L., Zhongjie, L., & Jinggui, F. (2017). Genome-wide identification and characterization of genes involved in carotenoid metabolism in three stages of grapevine fruit development. *Scientific Reports, 7*, 4216. https://doi.org/10.1038/s41598-017-04004-0

Yoo, H. J., Park, W. J., Lee, G. M., Oh, C. S., Yeam, I., Won, D. G., Kim, C. K., & Lee, J. M. (2017). Inferring the genetic determinants of fruit colors in tomato by carotenoid profiling. *Molecules, 22*, 764.

Zhao, T., Xia, H., Liu, J., & Ma, F. (2014). The gene family of dehydration responsive element-binding transcription factors in grape (*Vitis vinifera*): Genome-wide identification and analysis, expression profiles, and involvement in abiotic stress resistance. *Molecular Biology Reports, 41*, 1577–1585. https://doi.org/10.1007/s11033-013-3004-6