the plant journal

SEB Society for Experimental Biology

The Plant Journal (2018) **94,** 169–191 doi: 10.1111/tpj.13838

RESOURCE

Deciphering genetic factors that determine melon fruit-quality traits using RNA-Seq-based high-resolution QTL and eQTL mapping

Navot Galpaz[1,†], Itay Gonda[1,2], Doron Shem-Tov[3,4], Omer Barad[3], Galil Tzuri[1], Shery Lev[1,5], Zhangjun Fei[2,6], Yimin Xu[2], Linyong Mao[2,‡], Chen Jiao[2], Rotem Harel-Beja[1], Adi Doron-Faigenboim[7], Oren Tzfadia[8], Einat Bar[1], Ayala Meir[1], Uzi Sa'ar[1], Aaron Fait[9], Eran Halperin[4], Merav Kenigswald[1,5,10], Elazar Fallik[10], Nadia Lombardi[2,11], Guy Kol[3], Gil Ronen[3], Yosef Burger[1], Amit Gur[1], Ya'akov Tadmor[1], Vitaly Portnoy[1], Arthur A. Schaffer[7], Efraim Lewinsohn[1], James J. Giovannoni[2,6] and Nurit Katzir[1,*]

[1]*Department of Vegetable and Field Crops, Newe Ya'ar Research Center, Agricultural Research Organization, Ramat Yishay, Israel,*
[2]*Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, New York, USA,*
[3]*NRGENE, Park HaMada Ness Ziona, Israel,*
[4]*Department of Molecular Microbiology and Biotechnology, Tel-Aviv University, Tel-Aviv, Israel,*
[5]*Institute of Life Science, Hebrew University of Jerusalem, Jerusalem, Israel,*
[6]*USDA-ARS Robert W. Holley Center for Agriculture and Health, Ithaca, New York, USA,*
[7]*Department of Vegetable and Field Crops, Volcani Center, Agricultural Research Organization, Rishon LeZion, Israel,*
[8]*VIB Department of Plant Systems Biology, Ghent University, Gent, Belgium,*
[9]*The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Beer-Sheva, Israel,*
[10]*Department of Postharvest Science of Fresh Produce, Volcani Center, Agricultural Research Organization, Rishon LeZion, Israel, and,*
[11]*Department of Agricultural Sciences, University of Naples, Portici, Italy*

Received 7 September 2017; revised 19 December 2017; accepted 8 January 2018; published online 31 January 2018.
*For correspondence (e-mail katzirn@agri.gov.il).
[†]Present address:Northern Agriculture R&D, Kiryat Shmona, Israel.
[‡]Present address:Department of Biochemistry and Molecular Biology, Howard University, Washington, DC, USA.
N.G., I.G., D.S-T. and O.B. contributed equally to this work.

SUMMARY

Combined quantitative trait loci (QTL) and expression-QTL (eQTL) mapping analysis was performed to identify genetic factors affecting melon (*Cucumis melo*) fruit quality, by linking genotypic, metabolic and transcriptomic data from a melon recombinant inbred line (RIL) population. RNA sequencing (RNA-Seq) of fruit from 96 RILs yielded a highly saturated collection of > 58 000 single-nucleotide polymorphisms, identifying 6636 recombination events that separated the genome into 3663 genomic bins. Bin-based QTL analysis of 79 RILs and 129 fruit-quality traits affecting taste, aroma and color resulted in the mapping of 241 QTL. Thiol acyltransferase (*CmThAT1*) gene was identified within the QTL interval of its product, *S*-methyl-thioacetate, a key component of melon fruit aroma. Metabolic activity of CmThAT1-encoded protein was validated *in bacteria* and *in vitro*. QTL analysis of flesh color intensity identified a candidate *white-flesh* gene (*CmPPR1*), one of two major loci determining fruit flesh color in melon. *CmPPR1* encodes a member of the pentatricopeptide protein family, involved in processing of RNA in plastids, where carotenoid and chlorophyll pigments accumulate. Network analysis of > 12 000 eQTL mapped for > 8000 differentially expressed fruit genes supported the role of *CmPPR1* in determining the expression level of plastid targeted genes. We highlight the potential of RNA-Seq-based QTL analysis of small to moderate size, advanced RIL populations for precise marker-assisted breeding and gene discovery. We provide the following resources: a RIL population genotyped with a unique set of SNP markers, confined genomic segments that harbor QTL governing 129 traits and a saturated set of melon eQTLs.

Keywords: next-generation sequencing, transcriptome, *Cucumis melo*, network analysis, pentatricopeptide repeat-containing protein, thiol acyltransferase.

INTRODUCTION

Numerous important agricultural traits are governed by quantitative trait loci (QTL). QTL analysis is a powerful tool for the detection of genomic regions that control a given trait, and can be further used for gene discovery and marker-assisted selection (Holland, 2007; Hao and Lin, 2010). In the past, QTL studies were based on a limited number of markers with constrained mapping resolution, resulting in the detection of QTL covering large genomic intervals harboring many candidate genes (Paran *et al.*, 1997; Galpaz and Reymond, 2010; Harel-Beja *et al.*, 2010; Diaz *et al.*, 2011). As a result, expensive, laborious and time-consuming fine mapping and positional cloning strategies were frequently required to narrow down the boundaries of target loci and pinpoint the underlying genes (Cohen *et al.*, 2014). Moreover, direct application of QTL mapping to breeding programs was hindered because of the inherent linkage drag of deleterious alleles resulting from low-resolution mapping (Varshney *et al.*, 2014).

The introduction of next-generation sequencing (NGS) technologies has allowed the facile development of hundreds of thousands of single-nucleotide polymorphism (SNP) markers that differentiate between the parents of a given mapping population. The large-scale data generated by NGS, combined with powerful computational tools, enabled a major technological leap from low-resolution to high-resolution QTL mapping. This facilitated the gap-closing in gene-identification and breeding that is of special importance in non-model species (Xiao *et al.*, 2013; Unamba *et al.*, 2015; Bhat *et al.*, 2016). Resequencing and genotyping-by-sequencing (GBS; Elshire *et al.*, 2011) technologies have recently been employed for high-resolution QTL mapping in crop plants, such as rice and soybean (Huang *et al.*, 2009; Gao *et al.*, 2013; Spindel *et al.*, 2013; Xu *et al.*, 2013). Genetic linkage analyses of genome-wide expression patterns, obtained by microarray studies or NGS technologies, have also been applied to studies of genetic variation in gene expression. In these studies, mRNA transcript abundances were treated as quantitative traits that were mapped as gene-expression QTL (eQTL). Gene regulatory networks have also been constructed to further study the coordinated functionality of genes controlled by common eQTL (Jansen and Nap, 2001; Brem *et al.*, 2002; Schadt *et al.*, 2003; Kliebenstein *et al.*, 2006; Keurentjes *et al.*, 2007; Kliebenstein, 2009; Cubillos *et al.*, 2012; Tohge and Fernie, 2012; Zhang and Kim, 2014).

Recombinant inbred line (RIL) populations are powerful genetic resources for dissecting the genetic architecture of quantitative traits. RILs are developed by several generations of self-crosses of single-seed descents derived from a cross between two and more parental lines with significant genotypic and phenotypic variation. Each RIL is, in practice, homozygous for a unique mosaic of genomic intervals obtained from the parental genotypes, as a result of independent recombination events that occurred in each cycle of the self-crossing (Burr and Burr, 1991; Page and Grossniklaus, 2002).

Melon (*Cucumis melo* L.) is an important crop, with an estimated annual yield of more than 32 million tons worldwide (http://faostat3.fao.org). It is a diploid ($2n = 24$) plant with a sequenced genome and an estimated size of 450 Mb (Garcia-Mas *et al.*, 2012). Melon varieties exhibit extensive variation in fruit-quality traits: levels of soluble sugars, organic acids and aroma volatiles, rind and flesh color, firmness, fruit shape and size (Burger *et al.*, 2009; Nuñez-Palenius *et al.*, 2008). This wide polymorphic variation makes melon an important model crop for studying the genetics of fruit-quality traits. However, only a few genes affecting fruit-quality traits in melon have been identified, among them are genes affecting ethylene synthesis (Ayub *et al.*, 1996), aroma (Ayub *et al.*, 1996; Gonda *et al.*, 2016) and fruit acidity (Cohen *et al.*, 2014). Genes determining fruit aroma and color are of particular interest, due to their profound effects on the fruit palatability and consumer preferences. Aromatic melon varieties owe their aroma to volatile esters, sulfur-containing aroma compounds, sesquiterpenes, norisoprenoids, and short-chain alcohols and aldehydes (Gonda *et al.*, 2016; Jordan *et al.*, 2001; Schieberle *et al.*, 1990; Wyllie *et al.*, 1994). While certain genetic, biochemical and molecular processes that dictate the production of a number of these volatiles have been reported (El-Sharkawy *et al.*, 2005; Gonda *et al.*, 2010, 2013; Ibdah *et al.*, 2006; Lucchetta *et al.*, 2007; Manríquez *et al.*, 2006; Portnoy *et al.*, 2008), the biosynthesis of others remains largely unknown. Among the little studied but important volatile compounds are the thioesters (*S*-esters) that are present in melon and in other fruit and food products (Landaud *et al.*, 2008; Liu *et al.*, 2004; Tapp *et al.*, 2008; Vandendriessche *et al.*, 2013). Specifically, the thioester *S*-methyl thioacetate has been shown to contribute sulfuric-tropical fruit notes to certain melon cultivars (Wang and Lin, 2014). Methanethiol, which originates from L-methionine through the activity of methionine-γ-lyase enzyme (Gonda *et al.*, 2013), is the proposed substrate for the production of *S*-methyl thioesters. The *in vitro* enzymatic production of thioesters from thiols and acyl-coenzyme A (CoA) by thiol acyltransferase (ThAT) has been demonstrated in cell-free extracts of strawberry fruit (Noichinda *et al.*, 1999) and the fungus *Geotrichum candidum* (Helinck *et al.*, 2000). This enzymatic reaction catalyzes the production of thioesters in a manner similar to the production of *O*-esters, catalyzed by alcohol acyltransferases (AATs; Figure S1). AAT activities have been demonstrated in melon (El-Sharkawy *et al.*, 2005) and other fruit (Aharoni *et al.*, 2000), but ThAT activities have not been reported in

melon fruit. Moreover, genes encoding specific enzymes that esterify thiol substrates to generate thioesters have not yet been characterized in plants.

Flesh (mesocarp) color is another major attribute of fruit quality and a hallmark of fruit ripening in melon. Two major loci have been described as affecting flesh color in melon (Clayberg, 1992; Hughes, 1948; Iman *et al.*, 1972): *green-flesh* (*gf*), located on chromosome 9 (Perin *et al.*, 2002), recently shown to be *CmOr*, an ortholog of the cauliflower *BoOr* gene that governs carotenoid accumulation and orange flesh color (Tzuri *et al.*, 2015); and *white-flesh* (*wf*), localized to an approximately 5 Mbp region on chromosome 8 (Cuevas *et al.*, 2009; Diaz *et al.*, 2015; Monforte *et al.*, 2004). The two genes (designated here following Perin *et al.*, 2002) were suggested to interact epistatically, where combinations of their alleles determine orange, green and white flesh color. However, the gene underlying *wf* remains to be identified. Additional minor QTL and genes affecting flesh color in different genetic backgrounds have also been described in melon (Cuevas *et al.*, 2009; Galpaz *et al.*, 2013; Monforte *et al.*, 2004).

Over the past two decades, several melon genetic maps were developed with QTL analyses performed (http://www.icugi.org). We developed the 414 × Dul RIL population from a cross between PI 414723 (414) (*C. melo* var. *momordica*) and 'Dulce' (Dul) (*C. melo* var. *reticulatus*) (Harel-Beja *et al.*, 2010). PI 414723 is an Indian cooking melon with a non-sweet, pale orange fruit flesh and an unpleasant aroma due to the accumulation of high levels of sulfur-derived volatiles (Gonda *et al.*, 2013). 'Dulce' is an American cantaloupe-type with a sweet, dark-orange fruit flesh and a pleasant fruity aroma (Danin-Poleg *et al.*, 2002). The wide phenotypic variation between the two genotypes makes the 414 × Dul population an excellent tool for the study of the genetic basis of melon fruit-quality traits. A genetic map and QTL analysis of this population were described previously, based on 668 DNA markers (Harel-Beja *et al.*, 2010). This facilitated the cloning of the *CmPH* gene, that determines melon fruit acidity (Cohen *et al.*, 2014). Inter-metabolite association patterns and patterns of association of these metabolites with RNA-Seq-based gene-expression levels were previously explored in this population (Freilich *et al.*, 2015). A melon consensus linkage map, merging the data from eight populations, including the 414 × Dul, was constructed with 1592 markers (Diaz *et al.*, 2011). This map was recently anchored to the melon genome with 1850 markers (Diaz *et al.*, 2015).

Here, we applied a high-throughput RNA-Seq analysis of the 414 × Dul population for the performance of high-resolution QTL mapping. Several causative genes were mapped at up to a single-gene resolution, and two undescribed genes that affect fruit aroma and flesh color were identified and functionally or genetically validated: (i) *CmThAT1*, a gene encoding an enzyme catalyzing the formation of *S*-methyl-thioacetate from methanethiol and acetyl-CoA, and (ii) *CmPPR1*, involved in RNA editing in plastids and affecting flesh color.

RESULTS

RNA-Seq analysis and SNP calling

A systems approach was applied to link genotype, metabolite and gene-expression data of melon, using the above-described 414 × Dul RIL population at F_{6-8} (Harel-Beja *et al.*, 2010). RNA-Seq was performed on the flesh tissue of mature fruit (mesocarp) of 96 RILs and their parental lines, using the Illumina GAII platform (NCBI SRA repository accession SRP052934). Between 5 and 15 million 100-bp reads were obtained for each line (Freilich *et al.*, 2015). The transcriptome analysis provided both expression levels of genes transcribed in ripe melon fruit and a comprehensive set of SNPs. The reads were aligned to the melon reference genome (Garcia-Mas *et al.*, 2012), and 82 140 SNPs were detected. SNP filtration was performed: SNPs were phased, based on the parental information, and the selection of reliable SNPs in the population was based on Mendelian segregation criteria (> 10 lines with each of the homozygous SNP options and < 20 lines with the heterozygous SNP status). SNPs with genotype patterns that were inconsistent with their neighboring SNP patterns were filtered out. This resulted in a total of 58 328 SNPs (Data S1) that were used for genotyping each of the lines in the 414 × Dul RIL population; 79 RILs with low heterozygosity levels were selected for QTL analysis (Experimental Procedures; Figure 1; Data S1). Of the 27 427 predicted melon genes (Garcia-Mas *et al.*, 2012), about 16 000 genes were expressed in the fruit samples (Data S2; Freilich *et al.*, 2015).

High-resolution dissection of the genetic structure of the 414 × Dul RIL population

Genotyping of the 414 × Dul RILs, using RNA-Seq, allowed high-resolution dissection of the genetic structure of this population. Genetic blocks (bins) were defined by the transitions between PI 414723 and 'Dulce' genotypes. Unique bins (3663) were identified, resulting from 6636 recombination events that occurred in the 79 RILs during the development of the population (Figure 1a; Data S1 and Data S2). The number of recombination events per RIL ranged from 40 to 133, with a mean of 84. The average bin size was 72.6 kb (median of 5.5 kb), 55% of the bins were shorter than 10 kb, and the maximal bin size was 8563 kb (Data S3). The observed recombination-to-bin ratio of 1.8 suggested the existence of recombination hot spots in the RIL population. Thus, 2521 of the recombination events occurred in at least two RILs in the population, and 22 recombination events occurred in 10 or more RILs (Data S3). Genetic distances were calculated between each of the

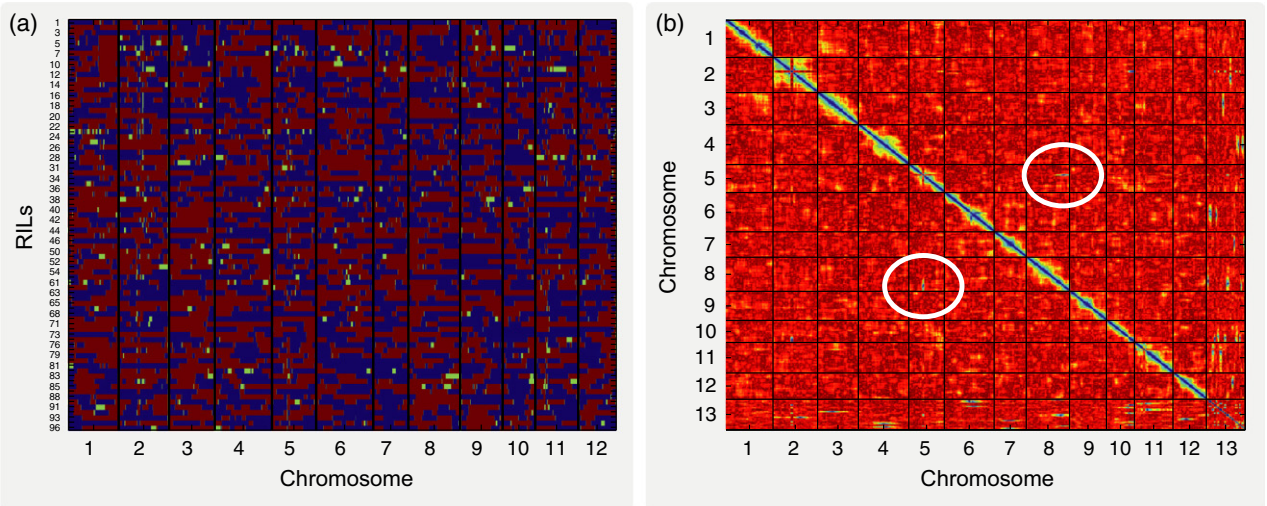


Figure 1. Recombination bin map (a) and heat map of genotyping bin distance matrix (b) of the 414 × Dul recombinant inbred line (RIL) population.
(a) Graphic representation of genotyping of 96 RILs from the 414 × Dul population that were analyzed by RNA-Seq. Color coding of the bins (genomic blocks): red – PI 414723; blue – 'Dulce'; green – heterozygous/unknown.
(b) Distance was calculated as the weighted sum over all of the genotyping differences. Blue – small distance; red – large distance. Bins are ordered along the chromosomes based on the physical map (*X*- and *Y*-axes). Adjacent bins with a small genotyping distance indicate agreement between the physical and genetic maps, while distant bins with a large genotyping distance indicate misassembly of the physical map (based on version 3.5; Garcia-Mas *et al.*, 2012). White circles mark the misassembly of a 64-gene fragment that should be relocated from chromosome 5 (scaffold 22) to chromosome 8 (adjacent to upper end of scaffold 2).

bin pairs to construct a genetic map with a total distance of 2047 cM (Figure 1b; Data S3), and mean density of 28.5 SNP cM^{-1} and 1.8 bin cM^{-1}. The total recombination rate, calculated as the total genetic map size divided by the estimated genome size, was 4.5 cM Mb^{-1}. The obtained genetic map was used to anchor unassembled scaffolds or to identify small regions of misassembly in version 3.5 of the melon genome assembly (Garcia-Mas *et al.*, 2012). Genomic regions with misassembled bins were detected in cases where the matching bin with minimal genetic distance was located on a different chromosome. Relocation of selected bins was also used to assign the chromosomal location of selected scaffolds with unknown chromosomal locations (virtual chromosome "0"; Garcia-Mas *et al.*, 2012; Argyris *et al.*, 2015), and is presented in Figure 1b and Data S4. This was in agreement with the recently published version 3.5.1 of the melon genome assembly (Argyris *et al.*, 2015), although small differences still exist. The bins were subsequently used to calculate recombination rates. The position of the change in slope in each of the chromosomes demonstrated reduced recombination rates, and localized the proposed centromeric region in each of the chromosomes (Figure S2), in accordance with previous reports (Argyris *et al.*, 2015; Garcia-Mas *et al.*, 2012).

High-resolution QTL mapping of fruit-quality traits in the 414 × Dul RIL population

To explore phenotypic variation in the 414 × Dul RIL population, 129 fruit traits, predominately levels of defined metabolites, were scored from the same mature fruit samples used for the RNA-Seq analysis (Data S5; Freilich *et al.*, 2015). Most of the traits phenotyped (2006, 2007 and/or 2010) belonged to the major categories affecting fruit quality: taste (sugars and acidity); color (carotenoids); and aroma volatile compounds (Data S5).

The collected phenotypic and genotypic data were used for QTL analysis. In each of the 3663 bins, the RILs were divided into two groups: homozygous 'Dulce' and homozygous PI 414723, and correlation to phenotype was determined by *t*-test. Using a false discovery rate (FDR) cut-off of < 10% (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2005; Francesconi and Lehner, 2013; Josephs *et al.*, 2015; Kinnersley *et al.*, 2015), a total of 241 QTL were identified, explaining the phenotypic variation of 86 out of 129 fruit-quality traits, most of which were levels of defined metabolic QTL (Data S6). When a less stringent statistical criterion was applied (FDR < 30%), 138 QTL and eight traits not identified at FDR < 10% were detected (Data S7).

The QTL peak was defined as the bin with the highest r^2 value, and the QTL interval was defined as the genomic segment surrounding the peak with r^2 above 90% of the peak value. The average length of the QTL intervals of the 241 fruit-quality QTL was 602 kb, and the average number of predicted genes in the QTL interval was 41. The median QTL interval was 274.2 kb and 20 genes. Moreover, 25% of the QTL intervals harbored only 1–10 genes, and 67% of the QTL intervals contained less than 30 genes (Data S6). Eighteen QTL were responsible for skewing the mapping resolution down, with large QTL intervals containing

between 100 and 486 genes, 159 genes on average (Data S6).

Validation of the high-resolution mapping using known genes

To validate the precision of the high-density-marker QTL mapping, we examined four QTL with known causative genes: (i) fruit acidity, determined by the *PH* gene encoding a PILS-type membrane transporter (*CmPH*; Cohen *et al.*, 2014), (ii) fruit length, affected by the *a* gene (*CmACS7*), encoding ACC synthase, an ethylene-biosynthesis enzyme shown to determine flower sexual type and to have pleiotropic effects on several traits, including fruit length and size (Boualem *et al.*, 2008; Monforte *et al.*, 2014), (iii) levels of the volatile methanethiol, derived from L-methionine by the methionine-γ-lyase, encoded by the gene *CmMGL* (Gonda *et al.*, 2013), and (iv) levels of the alcohols and the volatile esters derived from these alcohols by AATs, encoded mainly by *CmAAT1* (El-Sharkawy *et al.*, 2005).

Figure 2 and Data S6 depict the QTL peaks (determined by the bin with the highest LOD score for each of the QTL) and their known causative genes: (i) a major QTL for fruit pH, explaining 88% of the total phenotypic variance of the trait (LOD = 40.3, r^2 = .88), was precisely located to the *CmPH* gene (MELO3C025264), (ii) a QTL for fruit length was located only 2 genes distal to *CmACS7* (MELO3C015444, LOD = 17.8, r^2 = .60); notably, QTL for fruit weight, total soluble solids (TSS) and ethylene contents were also mapped to the same peak, suggesting a pleiotropic effect of *CmACS7* on these traits (Figure S3); (iii) one of the QTL for methanethiol levels was located within a peak of 9 genes that included *CmMGL* (MELO3C013774, LOD = 3.2), and (iv) more than 30 volatile ester compounds from various metabolic pathways were analyzed for QTL that affect their levels in mature fruits. A major QTL located within a cluster of *AAT* genes that encode for the enzymes that synthesize these esters from alcohols was identified in a subgroup of alcohols and esters. Figure 2d depicts examples of QTL of alcohols and esters mapped precisely to the location of *CmAAT1* and *CmAAT2* (MELO3C024771 and MELO3C024766, respectively, LOD = 4.02–5.84, r^2 = .19–.26 for the first five compounds). As expected for a complex trait in such a variable group of compounds, additional QTL had major impact in other subgroups (e.g. Figure 2d, compounds 6–8). The four examples of QTL that were mapped precisely or adjacent to known causative genes demonstrated the power and reliability of the high-resolution obtained here.

Identification and functional analysis of a ThAT gene

S-methyl thioesters are a unique group of volatile esters formed by the condensation of methanethiol and an acyl-CoA (Figure 3). Although thioesters are important compounds in many food products, their enzymatic production is poorly understood and the genes involved are only partially known. The first step in the pathway to *S*-methyl thioesters in melon fruit, namely the production of methanethiol from L-methionine (Figure 3a), was previously shown to be catalyzed by the enzyme methionine-γ-lyase, encoded by *CmMGL* (Gonda *et al.*, 2013). This was further supported here by the precise mapping of *CmMGL* to one of the QTL of methanethiol (Figure 2c). Deciphering of the predicted second part of the pathway, i.e. the generation of *S*-methyl thioacetate from methanethiol and acetyl-CoA, was made possible here due to the high-resolution QTL analysis.

Quantitative trait loci mapping of *S*-methyl thioacetate levels yielded a single, significant QTL on chromosome 1 (LOD = 10.38, r^2 = .4). A putative acetyl-CoA acetyltransferase (MELO3C024190, subsequently termed *CmThAT1*) was located 10 genes proximal to the QTL peak, within a QTL interval comprising 58 genes and 8 bins (Figure 3b; Data S6). According to the Pfam database (http://pfam.xfam.org/), CmThAT1 belongs to the thiolase-like superfamily, with both C- and N-terminal thiolase domains. Analysis by NCBI conserved domain tool (https://www.ncbi.nlm.nih.gov/cdd) classified CmThAT1 as a thiolase belonging to the condensing enzyme superfamily, members of which are involved in both synthesis and degradation processes. MELO3C026065, a single homolog of CmThAT1, was identified in the melon genome, located on chromosome 12, with 78% amino acid identity to CmThAT1. MELO3C026065 was termed CmAACT1, following the observations below. AtAACT1 and AtAACT2 (respectively, 77% and 86% AA identity) are the ortholog enzymes from *Arabidopsis thaliana*, and their function as biosynthetic thiolases was previously validated (Jin *et al.*, 2012). A multiple sequence alignment of CmThAT1 and CmAACT1 was preformed, concurrent with AtAACT1, AtAACT2, HaAACT (a sunflower AACT with validated biochemical function; Dyer *et al.*, 2009) and HsAACTc (a human cytosolic enzyme, the crystal structure of which has been determined; Kursula *et al.*, 2005; Figure S4). All three catalytic residues identified in the human HsAACTc, as well as the NEAF motif that is highly conserved in thiolase sequences, were conserved in CmThAT1 and all other plant AACTs (Figure S4).

To unravel the function of CmThAT1, we first demonstrated that fruit cell-free extracts supported *S*-methyl thioacetate biosynthesis, from methanethiol and acetyl-CoA, at a much higher rate than its spontaneous non-enzymatic production (Figure S5). The next stage was the evaluation of volatiles produced by *Escherichia coli* expressing the coding region of *CmThAT1* and co-expressing *CmMGL* (in order to generate the putative substrate, methanethiol). The bacteria co-expressing these two genes produced significant amounts of *S*-methyl thioacetate, while control bacteria expressing *CmMGL* alone accumulated

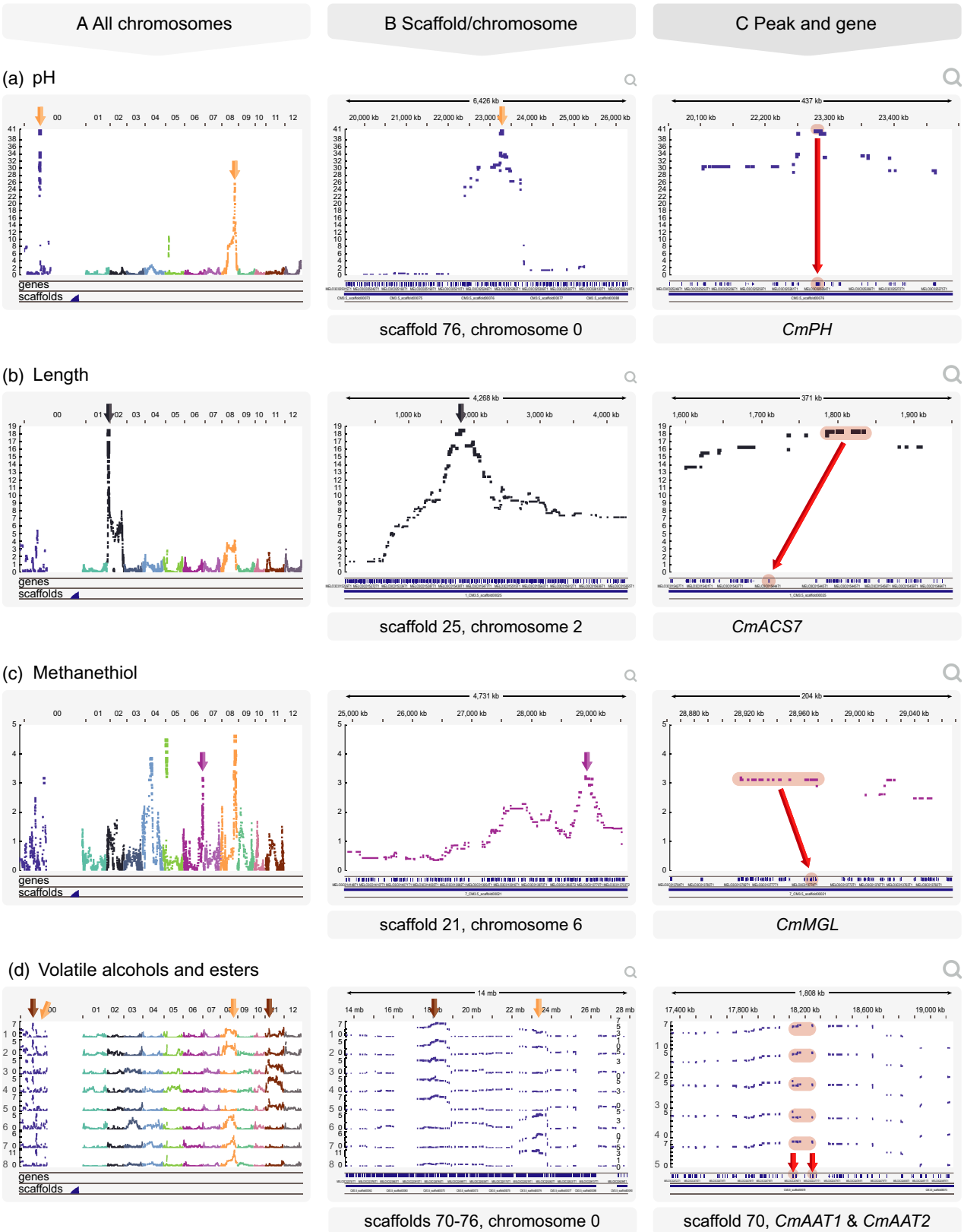


Figure 2. Integrative genomics viewer (IGV; Robinson *et al.*, 2011) snapshots of genomic location of quantitative trait loci (QTL) for selected traits relative to the known genes underlying them (red arrows).
(a) Fruit acidity, controlled by *CmPH* (MELO3C025264). The left side orange arrow indicates the peak on scaffold 76 (chromosome "0"), belonging to the truncated peak on chromosome 8.
(b) Fruit length, controlled by *CmACS7* (MELO3C015444).
(c) Levels of methanethiol, affected by *CmMGL* (MELO3C013774).
(d) Levels of volatile alcohols and esters, affected by *CmAAT1* & *CmAAT2* (MELO3C024771 & MELO3C024766, respectively). The left side brown arrow indicates the peak on scaffold 70 (chromosome "0"), belonging to the truncated peak on chromosome 11. The left side orange arrow indicates the peak on scaffold 76 (chromosome "0"), belonging to the truncated peak on chromosome 8. (1) 2-(methylthio) ethanol; (2) 3-(methylthio) propanol; (3) ethyl 2-methylbutanoate; (4) ethyl tiglate; (5) ethyl 3-(methylthio) propanoate; (6) benzyl alcohol; (7) 2-(methylthio) ethyl acetate; (8) benzyl acetate.
A Whole genome view (all chromosomes); B zoom in to specific chromosome and scaffold; C zoom in to QTL peak and causative gene. *X*-axis: the 12 melon chromosomes, in addition to the artificial chromosome "0" (i.e. all scaffolds not linked to chromosomes in version 3.5 of the genome; Garcia-Mas *et al.*, 2012). *Y*-axis: –log (Pval).

only traces of *S*-methyl thioacetate (Figure 3c). In addition, the synthesis of lower levels of *S*-methyl propanethioate was noted (Figure 3d). Finally, the recombinant CmThAT1 was purified from bacteria and tested for *in vitro* production of thioesters. CmThAT1 catalyzed the enzymatic production of *S*-methyl thioacetate from methanethiol and acetyl-CoA (Figure 3e), and of *S*-methyl propanethioate from methanethiol and propanoyl-CoA, *in vitro* (Figure 3f). The recombinant enzyme lacked propanol acetyltransferase enzymatic activity (Figure S6), indicating a certain degree of substrate specificity of the enzyme. We did not detect any thioester-synthesis activity for CmAACT1 in similar co-expression and *in vitro* assays.

The expression levels of *CmThAT1* in the RIL population were correlated with the levels of *S*-methyl thioacetate ($r = .66$, Spearman correlation; Freilich *et al.*, 2015). The coding regions of *CmThAT1*, isolated from PI 414723 and 'Dulce', were identical, yet polymorphism was detected in the 5′ and 3′ UTRs (Figure S7). This polymorphism was associated with the expression levels of both *CmThAT1* and *S*-methyl thioacetate. In the 39 lines homozygous for the 'Dulce' allele, mean values for gene expression (1122 RPKM) and *S*-methyl thioacetate [37.1 ng g^{-1} fresh weight (FW)] were 15- and 10-fold higher compared with the 52 lines homozygous for the PI 414723 allele (74 RPKM and 3.9 ng g^{-1} FW). Intermediate expression levels were recorded in the five heterozygous RILs (Data S2, Data S5 and Data S6). In compliance, a single and significant *cis*-acting eQTL (LOD = 36.8; $r^2 = .87$) of *CmThAT1* was localized to a bin of *CmThAT1.* This supports the view that *cis* polymorphic elements explain the variability of *CmThAT1* expression levels.

Identification and linkage analyses of a *CmPPR1* gene, likely affecting carotenoid accumulation and flesh color

Both parental lines of the 414 × Dul population have orange flesh color and, as expected, carry the $CmOr^{orange}$ allele that governs carotenoids accumulation in melon (Tzuri *et al.*, 2015). However, the parents differ in pigmentation intensity, from pale orange in PI 414723 to deep orange in 'Dulce' (Harel-Beja *et al.*, 2010). Accordingly, orange color intensity and carotenoid content segregated among the RILs of the 414 × Dul population (Data S5). QTL analysis of carotenoids measured by high-performance liquid chromatography (HPLC) or other color evaluation methods (visual or spectrophotometer) resulted in the identification of 49 QTL (Figure 4; Data S6). The most prominent QTL of eight of the 12 color-related evaluations were co-localized to a QTL peak of a single gene, MELO3C003069, on chromosome 8 (LOD = 3.7–15.2, $r^2 = .18–.55$; Data S6). This same QTL, in the remaining four traits, was either second in rank or less precise. Moreover, the peak of three QTL of apocarotenoid volatiles was located to the same genomic region. What appears as an additional QTL for carotenoids and apocarotenoids, localized to chromosome 5, with a peak of a single gene (MELO3C014263) in seven of the 12 color-related traits, is in fact contiguous to MELO3C003069 (Figure 4; Data S6). According to our genetic mapping, MELO3C014263 should be located adjacent to MELO3C003069, being one of 64 genes that were relocated from chromosome 5 to chromosome 8 (Figure 4; Data S4). The relocation of 38 of these genes was confirmed by the new version of the melon genome (Argyris *et al.*, 2015), whereas the relocations of the 26 genes most proximate to MELO3C003069 are not reported there. The order of ortholog genes around MELO3C003069, including the relocation of all 64 genes, was further supported by the precise synteny of genes in the genomes of melon (*C. melo*), cucumber (*Cucumis sativus*) and watermelon (*Citrullus lanatus*; Figure S8).

Adjacent to MELO3C003069 is MELO3C003070, and together they are homologs of two parts of a single Arabidopsis gene (AT5G02830, with 51% amino acid identity based on Blastx), annotated as a pentatricopeptide repeat-containing protein (PPR). The PPR family of proteins is associated with post-transcriptional RNA editing and processing of genes encoded in chloroplasts and mitochondria (Barkan and Small, 2014). The full-length sequencing of cDNA clones obtained from both parental lines of the population using primers from MELO3C003069 (5′) and MELO3C003070 (3′) (Table S1) further confirmed MELO3C003069–70 to be a single gene, designated CmPPR1 gene. The transcribed protein of the CmPPR1 gene comprised

850 amino acids (accession numbers KX228499–502), similar to other PPR genes (e.g. AT5G02830). The full-length sequences were analyzed using the recently developed Plant PPR tool (http://www.plantppr.com; Cheng *et al.*, 2016) that linked the gene to the P-class PPR proteins and suggested the structure of a composite gene comprising 11 P-class motifs (Figure S9). A chloroplast transit peptide (cTP) of CmPPR1 was predicted by ChloroP1.1 (http://www.cbs.dtu.dk/services/ChloroP(/and chloroplast localization by LocTree3 (https://rostlab.org/services/loctree3/), similar to the prediction of its Arabidopsis ortholog by SUBA3 database (http://suba3.plantenergy.uwa.edu.au).

Five polymorphic SNPs in the exons of MELO3C003069, three of which were non-synonymous, distinguished the parental lines of the population ('Dulce' allele: A/C/G/C/A; PI 414723 allele: T/G/T/A/T; Data S1 ss08_21171832–ss08_21173492 and Data S8). To highlight the contribution of these polymorphisms to the genetic variation of fruit color within the species, we surveyed a *C. melo* core collection of 148 cultivars representing the major groups of the species, capturing the full spectrum of fruit color diversity in melon (orange, green and white; Figure 5; Data S8). A single non-synonymous SNP (C/G^{358} with amino acids Val/Leu^{120}, respectively), located in the first exon, explained 26% of the green and white color variation among the non-orange genotypes (Figure 5a). Another SNP (G/T^{441}, Asp/Glu^{147}) was associated with pale orange genotypes. G/T^{441} showed low minor allele frequency, and only six of the 148 cultivars, including PI 414723, carried this allele. The third non-synonymous SNP was located in the third exon (A/T^{802}, Ser/Thr^{268}) with no additional effect. All three non-synonymous mutations were located in two of the P-class motifs (Figure S9). Additional SNPs were identified by sequencing *CmPPR1* of other lines (e.g. 'Piel de Sapo' and PI 161375; Data S8).

CmOr (*gf*; Tzuri *et al.*, 2015) was previously described to be epistatic to *wf*, proposed here to be *CmPPR1*. To assess the relationships between the two genes, we genotyped *CmOR* in the same core collection of 148 cultivars described above (Figure 5a; Data S8). A joint, 3-SNP haplotype ($CmOR^{orange/non-orange}$_*CmPPR1*_ C/G^{358}; G/T^{441}) explained 85% of the color variation in an analysis of variance across the tested core collection. The *CmPPR1* SNP C/G^{358} genotype provided significant (but not complete) partitioning between green and white color in lines homozygous for the $CmOR^{non\text{-}orange}$ allele in the diversity panel (Figure 5a). SNP G/T^{441} further improved the statistical model by partitioning between orange and pale orange lines (Figure 5a). A model describing flesh color determination based on the 3-SNP haplotype of the two genes is presented in Figure 5b.

Supportive information for the location of the *wf* locus was recently obtained by a QTL analysis of an F_3 population (NA × Dul), derived from a cross between the Noy-Amid cultivar (white fleshed, *inodorous* type) and the Dulce cultivar (orange fleshed, parent also of the 414 × Dul RIL population). F_3 plants (131), segregating for orange, green and white colors, were evaluated for flesh color intensity in each color category. The same 131 plants were genotyped using 76 988 SNPs identified by RAD-Seq (Miller *et al.*, 2007). QTL analysis was performed separately for each color category. One of the QTL for white flesh color was identified in chromosome 8 at the location of *CmPPR1* (Figure S10). For orange and green flesh color, *CmOR* was the major QTL (Mendelian, LOD = 47; $r^2 = .87$) and obscured the *wf* QTL. While the information is preliminary (Figure S10), the similarity of the peak strongly supports the location of the QTL in the 414 × Dul population.

The *CmPPR1* gene is located within the 2–5 Mb region of the previously described *wf* locus (Table S2; Cuevas *et al.*, 2009; Harel-Beja *et al.*, 2010; Tzuri *et al.*, 2015), and we therefore suggest that *CmPPR1* corresponds to the melon *wf* locus. Here, the QTL interval was significantly reduced and its peak was mapped to a single gene (MELO3C003069) in most of the color-related QTL (Data S6).

Major effects of *CmPPR1* on the expression of chloroplast-targeted genes revealed by eQTL analysis

An advantage of RNA-Seq fine mapping is that the RNA-Seq data also provided quantitative information on the

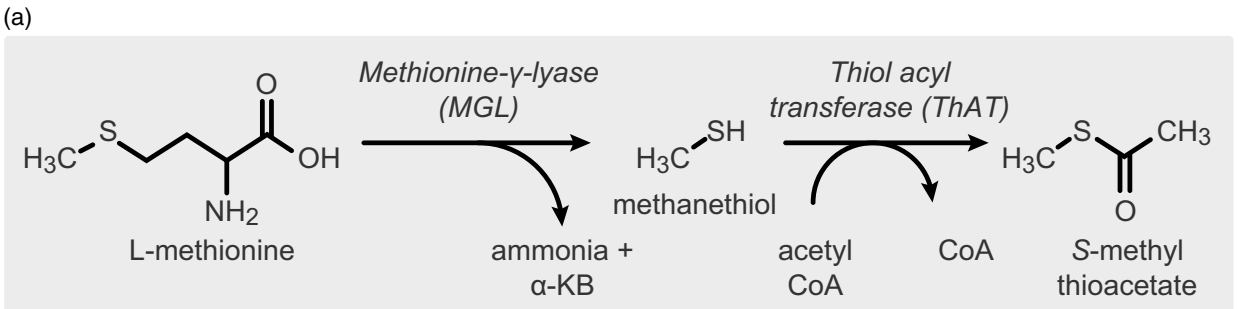
Figure 3. Metabolic pathway and quantitative trait loci (QTL) analysis of *S*-methyl thioacetate and the identification of the thiol acyltransferase (ThAT) gene involved in *S*-methyl thioacetate formation.

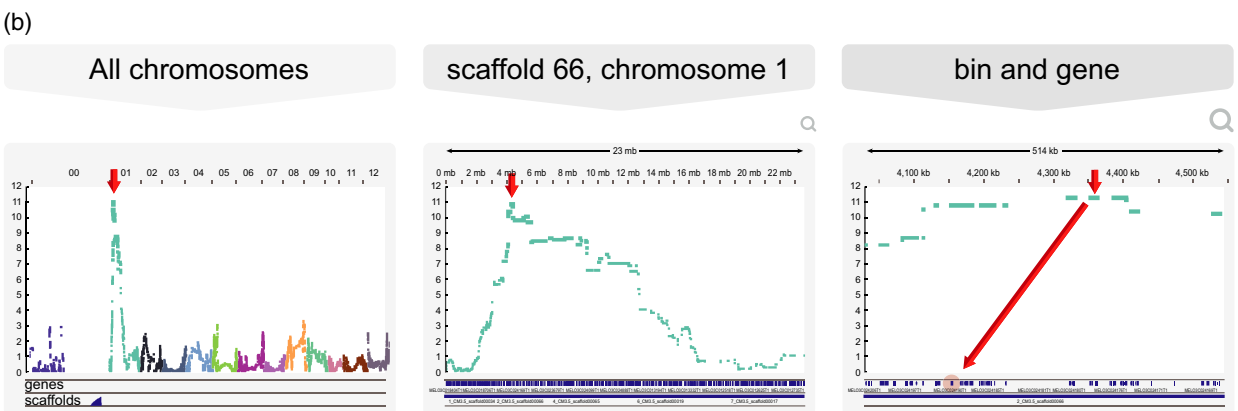
(a) Proposed biosynthetic pathway for *S*-methyl thioacetate from L-methionine. The pathway includes cleavage of L-methionine by *CmMGL* followed by proposed *S*-acylation. α-KB, α-ketobutyrate; CoA, coenzyme A.

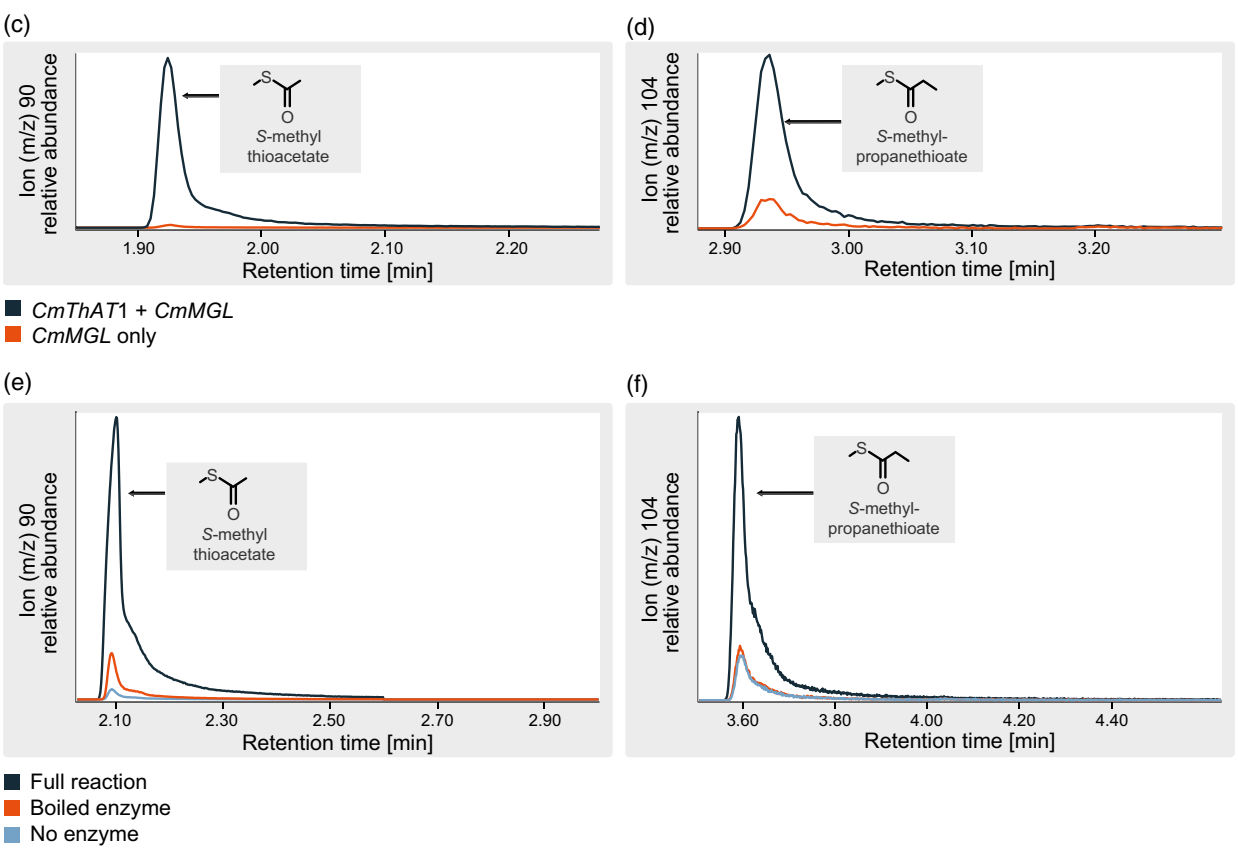
(b) Identification of the *CmThAT1* gene (MELO3C024190). QTL analysis of the volatile *S*-methyl thioacetate: whole genome view (left panel); zoom in on chromosome 1 (middle panel) and zoom in on the QTL peak (right panel) and *CmThAT1* location. *Y*-axis: −log (Pval).

(c and d) Solid-phase microextraction–gas chromatography–mass spectrometry (SPME–GC–MS) analysis of CmThAT1 enzymatic products in the headspace of *Escherichia coli* harboring *CmThAT1* and *CmMGL* (the latter was used for methanethiol production). The *in vivo* production of the thioesters was determined after induction with IPTG and overnight incubation at 37°C. (c) Single ion chromatograms of m/z = 90 depicting *S*-methyl thioacetate. (d) Single ion chromatograms of m/z = 104 depicting *S*-methyl propanethioate.

(e and f) *In vitro* functional analyses of the recombinant CmThAT1. Enzymatic assays were performed with His-tagged purified CmThAT1 enzyme, 1 mM methanethiol and 0.2 mM acyl-CoA incubated overnight at 30°C and analyzed by SPME–GC–MS. (e) Single ion chromatogram of m/z = 90 depicting *S*-methyl thioacetate production from methanethiol and acetyl CoA. (f) Single ion chromatogram of m/z = 104 depicting *S*-methyl propanethioate production from methanethiol and propanoyl CoA. Compounds were identified based on their retention index and mass-spectrum based on literature values and verified with an authentic standard of S-methyl thioacetate. Chromatograms are typical for at least four replicates based on at least two independent experiments. m/z, mass to charge ratio.







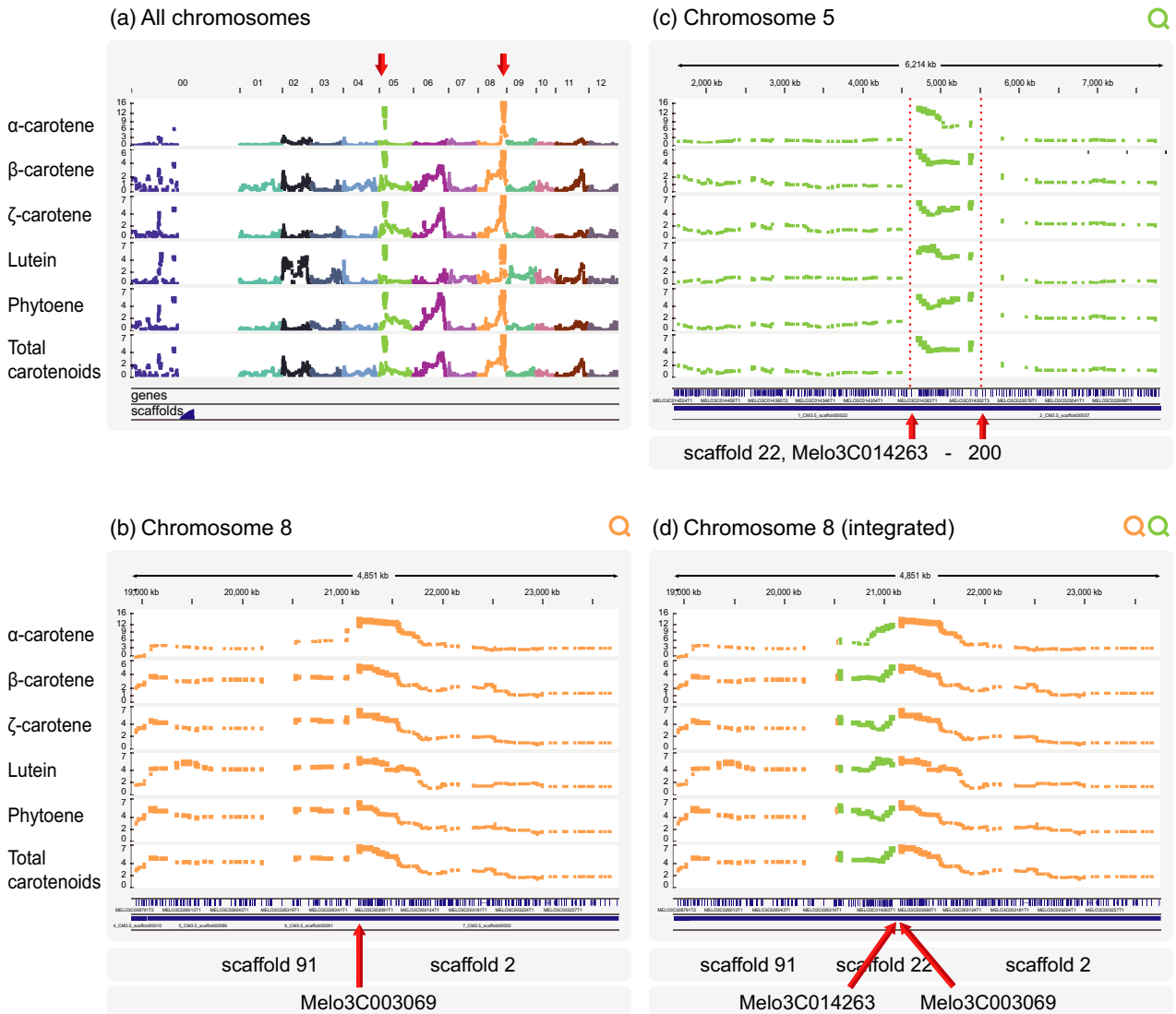
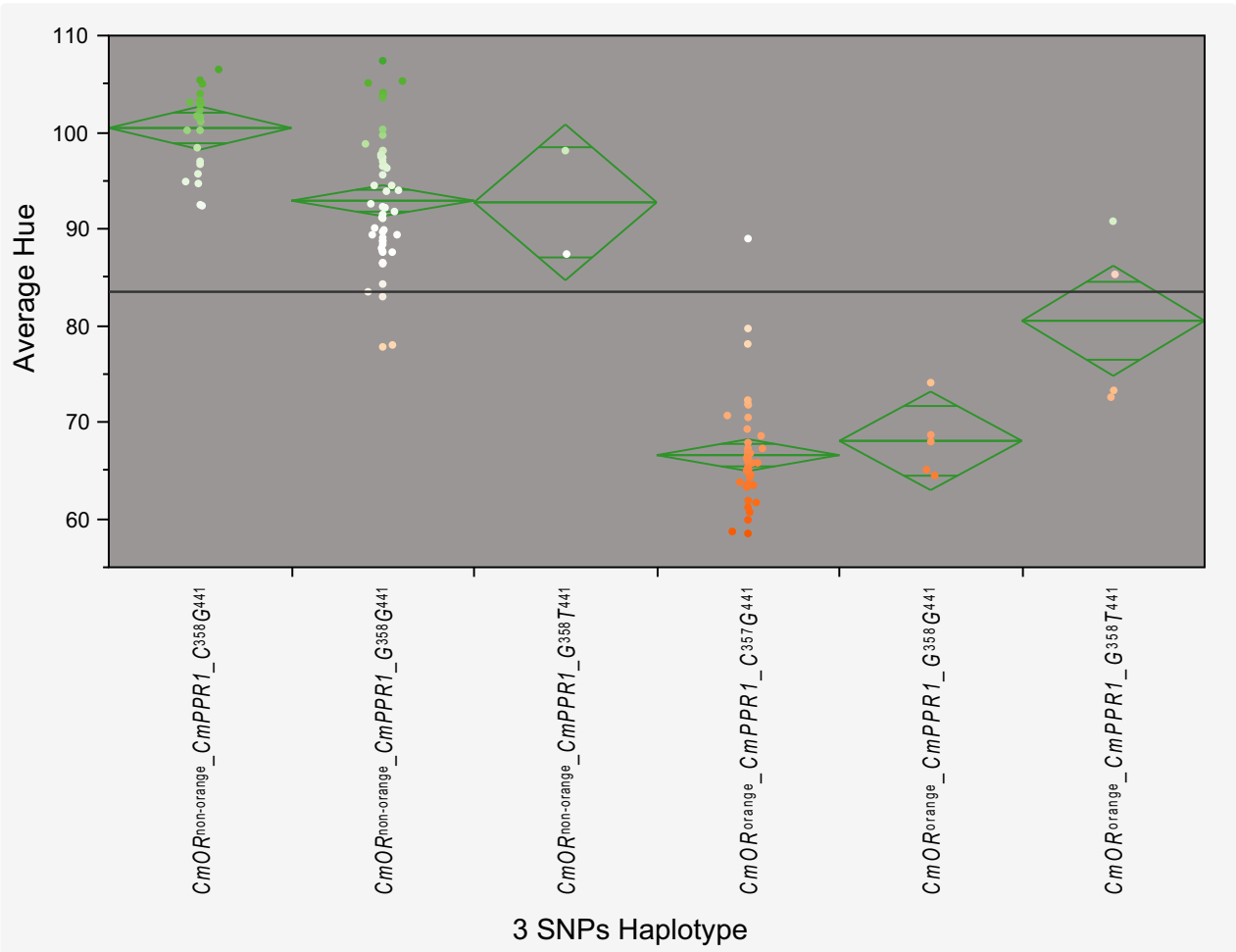


Figure 4. IGV snapshots (Robinson *et al.*, 2011) of carotenoid quantitative trait loci (QTL).
(a) All chromosomes.
(b) The most significant QTL was localized to chromosome 8, scaffold 2 (Melo3C003069, *CmPPR1*). Peak shape is incomplete due to misplacement of a group of genes to scaffold 22, chromosome 5, in version 3.5 of the genome (Argyris *et al.*, 2015; Garcia-Mas *et al.*, 2012).
(c) The group of 64 genes (Melo3C014263- 200) from scaffold 22, chromosome 5 that should be relocated to chromosome 8 (in opposite orientation).
(d) The reconstructed peak on chromosome 8, with Melo3C003069 next to Melo3C014263.
Y-axis: −log (Pval).

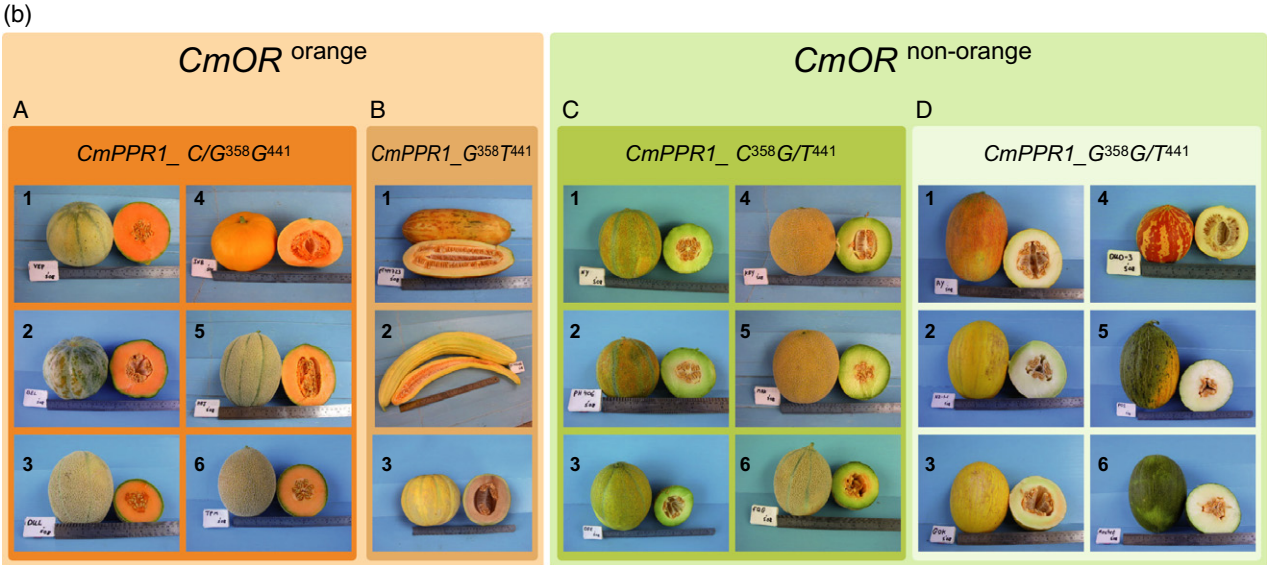
expression levels of about 16 000 fruit-expressed genes across the RIL population (Data S2). This enabled high-resolution mapping of 12 703 eQTL for the 8405 genes that were differentially expressed among the lines of the 414 × Dul RIL population (Data S9). Thus, the number of eQTL of a given target gene ranged between 1 and 6, and the

Figure 5. Genetic interactions between the *CmOr* and *CmPPR1* alleles in a core collection representing the full spectrum of fruit color diversity in melon (orange, green and white).
(a) One-way analysis of variance for the joint effect of a 3-SNP haplotype (*CmOR_CmPPR1*_C/G^{358}, G/T^{441}) across the tested core collection of 148 genotypes.
(b) A model presenting the effect of the combined *CmOR_CmPPR1* haplotype in the core collection (see a).
A Deep orange flesh color: *CmOR*^{orange} *_CmPPR1_* C/G^{358}G^{441} – (1) 'Vedrantais', (2) 'Bellegarde', (3) 'Dulce', (4) 'Indian Best', (5) 'Hale's Best Jumbo', (6) 'Top Mark'.
B Pale orange/cream flesh color: *CmOR*^{orange}*_CmPPR1_*G^{358}T^{441} – (1) PI 414723, (2) PI 435288, (3) CUM64/1979.
C Green flesh color: *CmOR*$^{non\text{-}orange}$*_CmPPR1_* C^{358}G/T^{441} – (1) 'Noy Yizre'el', (2) PH406, (3) 'Ogen', (4) 'Krimka', (5) 'Magyar Kines', (6) 'Fordhook Gem'.
D White flesh color: *CmOR*$^{non\text{-}orange}$*_CmPPR1_*G^{358}G/T^{441} – (1) 'Ananas Yoqne'am', (2) 'Noy Zahov', (3) 'Gold King', (4) 'Dudaim 3', (5) 'Piel de Sapo', (6) 'Rochet'.

(a)



(b)



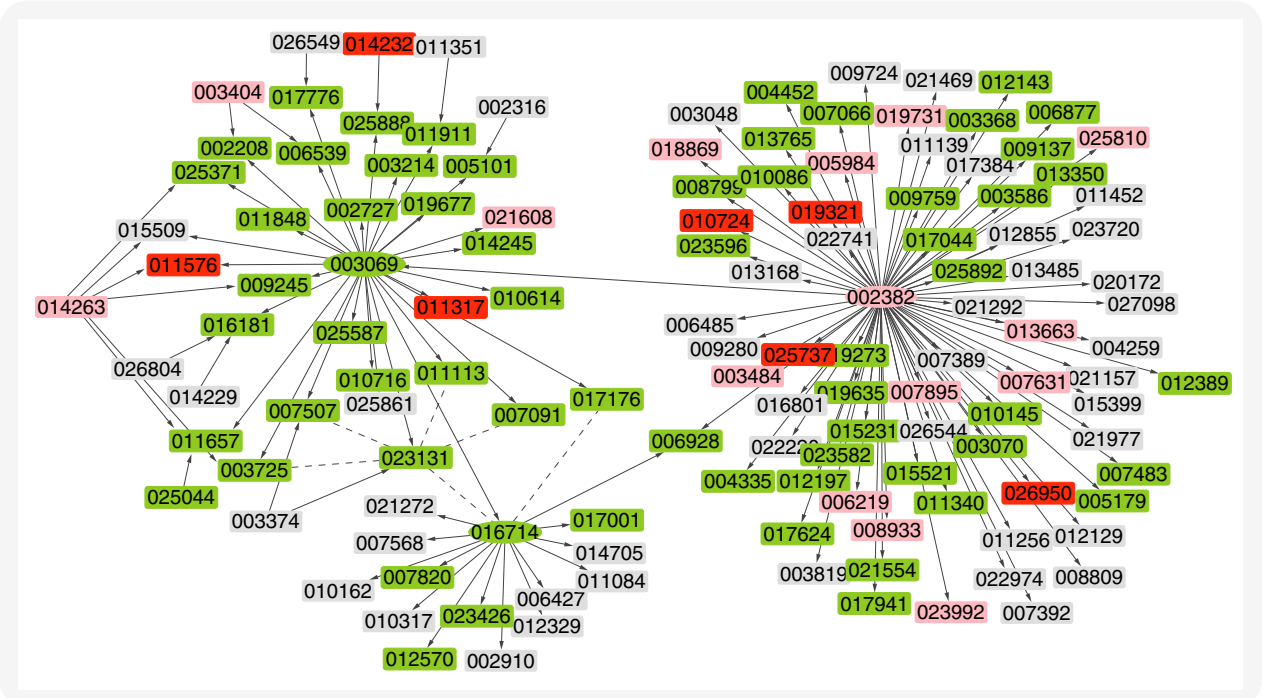


Figure 6. A regulatory trans-expression-quantitative trait loci (eQTL) network developed using Cytoscape.
The complete network included 4485 *trans*-eQTL with a confidence interval peak of a single candidate effector gene. Depicted is a subnetwork produced by extraction of the first and second neighbors of the *CmPPR1* gene (MELO3C03069). Each node represents a gene, and a directional arrow represents a putative regulatory relationship. Broken lines connect the genes of the co-expression network that was overlaid on the eQTL subnetwork. Based on GO terms and annotation, plastid genes are highlighted in green; transcription factors (TFs) in red; other regulatory genes in pink; and all the rest in gray.

average length of the QTL interval of the 6930 eQTL, mapped to a single scaffold, was 162 kb. Of the 12 703 eQTL, 4326 (34%) were co-located (up to 100 genes) with the candidate effector gene and classified as *cis*-eQTL (Data S9). Similar *cis*- to *trans*-eQTL ratios have been reported for other plants, for example, rice and Arabidopsis (37% and 41%, respectively; Keurentjes *et al.*, 2007; Wang *et al.*, 2014).

A regulatory network of the eQTL was developed using the network-drawing software Cytoscape (Shannon *et al.*, 2003). The analysis included only *trans*-eQTL (4485) selected from the eQTL with a confidence interval peak of a single candidate effector gene (6214 of the total 12 703; FDR < 10%).

Gene ontology (GO) terms were assigned to the genes (Ashburner *et al.*, 2000). Of the 3127 melon plastid targeted genes annotated in the database, 820 were represented in the network. Of the 1538 melon transcription factors (TFs) listed in the Plant Transcription Factor Database (http://planttfdb.cbi.pku.edu.cn/), 241 were represented in the network.

The differential expression of >1000 of the fruit-expressed genes represented in the network was attributed to only 22 of the eQTLs, each affecting ≥30 genes, and we refer to these 22 as major impact genes (MIGs). One of these MIGs was the *CmPPR1* gene described above (MELO3C003069 part of the gene), that was found to affect mainly the expression of genes targeted to the chloroplast.

Figure 6 depicts a subset of the complete genetic network of the eQTL associated with the *CmPPR1* gene up to the second order (produced by extraction of the first and second neighbors of the gene). According to this subnetwork, MELO3C003069 governs the expression of 30 genes, 27 (90%) of which are chloroplast-targeted genes (Data S10). One of the affected genes, MELO3C16714, further affects a group of 15 genes, including five chloroplast-targeted genes (Figure 6).

The analysis revealed that the expression of MELO3C003069 is governed (LOD = 6.5, r^2 = .30) by a single gene, MELO3C002382, a putative serine/threonine-protein kinase (best similarity to At1 g54610). This gene governs the expression of 78 genes, including 33 chloroplast-targeted genes. As expected, the expression of MELO3C003070 (shown above to be part of the *CmPPR1* gene), was also found to be governed by MELO3C002382 (LOD = 5.3, r^2 = .25), strengthening the accuracy of the eQTL analysis and the combination of the two into a single gene.

Gene ontology term analysis of this subnetwork (132 genes) detected significant enrichment (using Plant MetGenMap with FDR < 0.05) in chloroplast (61/132

genes), photosynthesis (32/132), plastid organization (29/154), chloroplast organization (19/132) and binding (99/132). Data S10 depicts all the significantly enriched GO terms of the 132 genes. Additional regulatory genes were manually picked on the basis of the NCBI annotation.

We calculated the correlation between RNA-Seq-based gene expression levels of genes encoded by the chloroplast genome (Data S2) and MELO3C003069 plus the 30 chloroplast-targeted genes (Figure 6), the expression of which was governed by MELO3C003069 (Figure S11). The 30 genes were clustered into three main groups (Figure S11): the largest (green cluster) was negatively correlated with most of the chloroplast encoded genes and positively correlated with four genes, including *petN*, *petL* and *petG*. These genes were previously shown to have a role in assembly, stability and dimerization of the cytochrome b6f complex in tobacco (Schwenkert *et al.*, 2007). A small cluster (red cluster, four genes, including MELO3C003069) showed a negative correlation with the above-mentioned chloroplast encoded genes, and the third cluster (blue cluster) did not show a distinct pattern of correlations. The correlation pattern suggests a common regulation and possibly a common functional pathway of most of the 30 genes of the subnetwork.

Co-expression patterns can highlight relevant genes that are potentially active in the same pathway but have independent functions. A co-expression network was therefore constructed based on gene-expression data (Data S2), using CoExpNetViz (Tzfadia *et al.*, 2016). The co-expression network was overlaid on the eQTL subnetwork (Figure 6) using Module Analysis via Topology of Interactions and Similarity Sets (MATISSE; Ulitsky and Shamir, 2009) to search for modules common to both networks. These modules comprise cellular components whose interactions might be attributed to a specific biological function. We detected a 'mini hub' around MELO3C023131 (encoding for a magnesium-chelatase subunit H ortholog) that interlinks seven co-expressed genes with eQTL connections (Figure 6). All seven genes were: (i) directly connected to the *CmPPR1* gene (MELO3C003069), (ii) chloroplast-targeted genes, and (iii) significantly correlated with carotenoid and apocarotenoid levels in the RIL population (Table S3; Freilich *et al.*, 2015).

DISCUSSION

RNA-Seq-based analysis of a relatively small-size RIL population allows for high-resolution QTL and eQTL mapping

Our aim here was to elucidate genetic factors that determine melon fruit-quality traits. To this end we employed RNA-Seq-based QTL and eQTL mapping. This enabled the achievement of a high-resolution QTL mapping of well over 100 fruit-quality traits obtained with a comprehensive set of SNP markers, and the quantification of gene-expression levels for the mapping of over 12 000 eQTLs and the construction of a gene-expression regulatory network. Thus, RNA-Seq proved itself to be an excellent alternative to the more frequently used DNA sequencing-based QTL analysis methods, including GBS, genome resequencing and DNA-Seq (Chen *et al.*, 2014; Gao *et al.*, 2013; Huang *et al.*, 2009; Spindel *et al.*, 2013; Xu *et al.*, 2013).

Previous studies applied the 'genetical genomics' approach, combining microarray expression data with available QTL maps for marker density enrichment and the detection of candidate genes in QTL intervals harboring a high number of genes (Drost *et al.*, 2015; Graham *et al.*, 2014; Hansen *et al.*, 2008; Joosen *et al.*, 2009; Jordan *et al.*, 2007; Kliebenstein *et al.*, 2006). The direct measurements of full-length transcripts, compared with indirect measurements derived from the hybridization of transcripts to pre-designed probes, confer substantial advantages for RNA-Seq over microarray for transcriptome profiling. These include improved sensitivity and accuracy for extremely low or high transcript abundance, the obtaining of complete expressed gene sequences, alternative splicing sites, no prerequisite for a reference genome and minimal technical biases introduced during hybridization of microarrays (Sîrbu *et al.*, 2012; Zhao *et al.*, 2014).

Moreover, RNA-Seq-based QTL mapping is particularly advantageous for large and complex genomes, where genomic DNA-based genotyping is prone to misalignment, due to high frequencies of repetitive sequences (Zhong *et al.*, 2011). An additional advantage of RNA-Seq-based markers in terms of precision is their localization, by definition, within genes. Here, SNP/gene ratio was 3.7 (for the genes expressed in the fruit), and most SNPs were within the coding sequences of the genes. This ratio highlights the dual utility of RNA-Seq, providing a saturated set of markers and comprehensive information on allelic variation in the population. This has clear implications for understanding the functional consequences of the genetic variation.

Mapping resolution, expressed by the size of the genomic interval of a given QTL, is determined by the number of recombination breakpoints in the studied population and marker density. Here, although only 79 F_6–F_8 generation RILs were used, high-resolution was achieved based on 3633 unique bins, resulting from 6636 recombination events that occurred during the development of the 414 × Dul RIL population. The number of recombinations in our study (6636) was close to the theoretical rough estimate (7584; calculated as 12 chromosomes × 4 early generation × 2 recombinations/generation × 79 lines). Mapping resolution in the present RNA-Seq study was substantially higher compared with an earlier work (Harel-Beja *et al.*, 2010) using the same population, due to an increase in

marker-number that was two orders of magnitude higher (58 324 and 668, respectively) and a marker/gene ratio (2.12 and 0.02, calculated per all melon genes).

Advanced RIL populations are an excellent tool for QTL detection, due to: (i) a high level of homozygosity, which results in high numbers of informative markers, (ii) high levels of recombination breakpoints, obtained by a series of self-crosses, (iii) immortality, allowing the use of one-time genotyping for unlimited phenotypic data acquired in diverse experiments (Keurentjes *et al.*, 2007). The power of advanced RIL populations in high-resolution QTL mapping has been recently demonstrated in soybean, where an F_8 RIL population, consisting of 246 lines genotyped by whole genome resequencing, was utilized to define 3509 bins. In that work, a major QTL for nematode resistance was mapped to a 29.7-kb interval harboring only three genes (Xu *et al.*, 2013).

In populations with a different structure, such as introgression lines (IL) and derived backcross inbred lines (BIL), a substantial enhancement of mapping resolution was achieved by increasing the number of lines and marker number. Using the BIL population of tomato (*Solanum lycopersicum* × *Solanum pennellii*), 446 BIL genotyped by SNP-chip generated 663 bins (Ofner *et al.*, 2016), and 545 lines genotyped by DNA sequencing generated 1049 bins (Fulop *et al.*, 2016). For comparison, the original 76 IL population comprised 106 bins (Eshed and Zamir, 1995). Although the melon RIL population here was several times smaller than the tomato BIL population, a similar mapping resolution was achieved. This may be explained by population structure, namely the higher number of recombination breakpoints in the RIL population.

The production of thioesters in melon fruit is mediated by the ThAT gene *CmThAT1*

Thioesters are important contributors to the aroma of melon fruit as well as other crops, yet often impart undesirable hedonic effects when in excess. Despite their importance, the biosynthetic pathways that lead to thioester formation are poorly understood. While evidence describing the enzymatic production of thioesters in microorganisms (Helinck *et al.*, 2000), strawberry fruit (Noichinda *et al.*, 1999) and mung-bean roots (Riov and Jaffe, 1972) is available, the genes encoding those enzymes are unknown. Here, high-resolution QTL mapping enabled the discovery of *CmThAT1*, encoding an enzyme that supports this catalysis. As evident from the *in vitro* experiments, CmThAT1 is a ThAT enzyme that produces *S*-methyl thioacetate and *S*-methyl propanethioate from methanethiol and acyl-CoA substrates (Figure 3). In addition, our co-expression experiments utilizing *E. coli* expressing both *CmThAT1* and *CmMGL* (for methanethiol production) highlighted the effectiveness of CmThAT1 in the production of straight-chain thioesters. While the production of *O*-esters (Figure S1) is mediated by AATs in melon and other fruits (Aharoni *et al.*, 2000; El-Sharkawy *et al.*, 2005; Yahyaoui *et al.*, 2002), the production of thioesters (*S*-esters) is mediated by this ThAT. Although catalyzing similar reactions, CmThAT1 differs from AAT enzymes that are members of the BAHD enzyme family. *CmThAT1* does not display the typical conserved BAHD motifs (Ma *et al.*, 2005) and belongs to the thiolase superfamily. Enzymes of the thiolase superfamily may be divided into two main categories: enzymes that break thioesteric bonds (EC 2.3.1.16); and enzymes that form thioesteric bonds (EC 2.3.1.9; Peretó *et al.*, 2005). CmThAT1 shares 77% and 86% amino acid identity with *A. thaliana* acetoacetyl thiolase 1 and 2 (AtAACT1 and AtAACT2), respectively. These function as biosynthetic thiolases that condense two acetyl-CoA molecules to form acetoacetyl-CoA, the precursor of 3-hydroxy-3-methyl-glutaryl-CoA, the key initial precursor of the cytosolic mevalonate pathway for terpene biosynthesis (Jin *et al.*, 2012). These authors further demonstrated that AtAACT2, but not AtAACT1, is crucial to plant development by altering sterol biosynthesis via the mevalonate pathway in Arabidopsis (Jin *et al.*, 2012). Here, we report a function for CmThAT1, a thiolase-like enzyme, that catalyzes the formation of volatile thioesters, analogous to the function of AACT enzymes in acetoacetyl-CoA production in other plants (Dyer *et al.*, 2009; Jin *et al.*, 2012). While CmThAT1 may have additional functions in acetoacetyl-CoA synthesis, its function in volatile thioester biosynthesis is clear and supported by genetic, co-expression and *in vitro* analyses (Figure 3). Still, the biochemical and *in vivo* function of CmAACT1 remains to be determined.

A *CmPPR1* gene potentially underlying the *wf* locus and affecting plastid-nucleus communication

The major QTL governing the levels of all carotenoids, as well as several of their apocarotenoid derivatives, in the 414 × Dul population, was mapped to the distal end of chromosome 8, with a peak comprising MELO3C003069. This QTL was located within the wider chromosomal segment of the previously described *wf* locus (2–5 Mb; Monforte *et al.*, 2004; Cuevas *et al.*, 2009; Harel-Beja *et al.*, 2010; Diaz *et al.*, 2015), one of the two major loci affecting flesh color in melon (Cuevas *et al.*, 2009; Diaz *et al.*, 2011; Harel-Beja *et al.*, 2010; Monforte *et al.*, 2004). Orange flesh color, dominant to both green and white, is governed by *CmOR*, located on chromosome 9 at the site of the *gf* trait (Tzuri *et al.*, 2015), while *wf* is associated with white or green (Monforte *et al.*, 2004; Tzuri *et al.*, 2015) or with the intensity of orange (Cuevas *et al.*, 2009; Harel-Beja *et al.*, 2010).

The role of *CmPPR1* (MELO3C003069–70) as the causative gene underlying the *wf* locus is supported by the QTL analysis described here, by the combined *CmOR_CmPPR1* haplotype in the core collection that explains approximately

85% of the color variation of the species and by the preliminary QTL analysis of the NA × Dul population. A single non-synonymous mutation was associated with the difference between white-flesh and green-flesh melons (C/G^{358}), while a rare mutation found in the pale orange parental line (PI 414723) was associated with the difference between pale orange and deep orange melons (G/T^{441}; Figure 5a). The various mutations identified in *CmPPR1* comply with the polymorphism in each of the crosses used in previous studies (Data S8; Perin *et al.*, 2002; Monforte *et al.*, 2004; Cuevas *et al.*, 2008, 2009).

The *CmPPR1* gene encodes a melon ortholog of a P-class PPR protein from Arabidopsis (At5 g02830). It belongs to one of the largest families of nuclear-encoded proteins that are targeted to plastids and mitochondria, and involved in post-transcriptional modifications of genes expressed in these organelles (Barkan and Small, 2014; Cheng *et al.*, 2016). Both *CmPPR1* and At5 g02830 are predicted to be targeted to the plastids. Members of this family bind organellar transcripts and influence their expression by altering their sequence, turnover, processing or translation (Barkan and Small, 2014). Studies of PPR mutants, with distinct molecular and physiological phenotypes, elucidated their involvement in plastid function, including their role in photosynthetic defects and changes in leaf pigmentation (Barkan *et al.*, 1994; Chateigner-Boutin *et al.*, 2008; Liu *et al.*, 2010). To the best of our knowledge, a phenotype associated with At5 g02830 has not yet been described. Earlier studies suggested little redundancy between members of the family due to the specific recognition between PPR proteins and single or several organellar transcripts.

The combined eQTL and network analyses revealed that *CmPPR1* affects multiple gene expression (Figure 6). Most of the affected genes are chloroplast-targeted, expressed in the nucleus and transferred to the plastids as proteins, which represent > 95% of the genes that function in the chloroplasts. It is therefore not expected that all 30 affected genes will be edited or affected directly by *CmPPR1*, assumed to edit RNA in the chloroplast, but rather perhaps through a signal mediated by *CmPPR1*. This is possibly part of the plastid-to-nucleus communication pathway, known as retrograde signaling (Barajas-López *et al.*, 2013; Chan *et al.*, 2010; Mochizuki *et al.*, 2001; Woodson *et al.*, 2013). One of the first genes shown to be involved in this pathway, Genomes Uncoupled (*GUN1*) of *A. thaliana*, is also a pentatricopeptide protein (Koussevitzky *et al.*, 2007). Additional *GUN* mutants (#2–#6) that continue to transcribe nuclear-encoded photosynthetic genes, irrespective of photosynthesis inhibition by various factors, were previously described (Koussevitzky *et al.*, 2007; Mochizuki *et al.*, 2001; Strand *et al.*, 2003). Little is known of specific fruit retrograde signals that coordinate the expression of chloroplast-targeted genes, at plastid biogenesis, including the transition from chloroplasts to chromoplasts (Pesaresi *et al.*, 2014). Related to our findings, a tetratricopeptide repeat protein (TPR) was recently shown to affect carotenoid biosynthesis and chloroplast development in Monkeyflowers, and was proposed to act through the chromoplast-to-nucleus retrograde signaling (Stanley *et al.*, 2017). The TPR family contains motifs that are related to the PPR (Small and Peeters, 2000).

In melon, during the first stages of fruit development, the light green mesocarp contains chloroplasts, as evidenced by the presence of chlorophylls (Chayut *et al.*, 2015). Later in development, there are three possible scenarios: an increase in the number of chloroplasts (green flesh melons), a disintegration of the chloroplasts (white flesh melons) and a transition of chloroplast-to-chromoplast (orange flesh melons). The postulated involvement of *CmPPR* in plastid-nucleus communication may explain the three different flesh color options. White melon genotypes (and other cucurbit species) may be considered the primitive or progenitor genotypes, considering that the bulk of *subsp. agrestis* genotypes are white fleshed, while orange flesh characterizes the more modern genotypes of *subsp. melo*, var. cantaloupensis and var. reticulatus (Pitrat *et al.*, 2000). These white genotypes are here suggested to carry the $CmPPR^{white}$ allele that fails to maintain plastid biogenesis. The $CmPPR^{green}$ mutation could enable a coordinated expression of genes involved in chloroplast biogenesis, and these indeed are found in the *subsp. melo* group. The orange mutation in *CmOr* may be considered to be more recently evolved, and enabled the transition of chloroplast to chromoplast and the accumulation of carotenoids in the orange flesh melons (Chayut *et al.*, 2015, 2017; Tzuri *et al.*, 2015). $CmOr^{orange}$ is epistatic to both white and green, $CmPPR^{white}$ and $CmPPR^{green}$, respectively. However, the $CmPPR^{light\ orange}$ allele possibly disrupts the communication only partially, resulting in light orange melons. The model presented here is in accordance with the combined effect of *CmOR-CmPPR1* haplotype observed in the core collection (Figure 5).

Several of the 30 genes affected by *CmPPR1* (Figure 6) are orthologous of Arabidopsis genes associated with retrograde signaling. These include two GUN genes, magnesium-chelatase subunit H (GUN5, MELO3C23231; Mochizuki *et al.*, 2001) and an Mg-protoporphyrin IX binding protein (*GUN4*, MELO3C011657; Strand *et al.*, 2003). A gene involved in chlorophyll synthesis is *CmGGH* (MELO3C017176), an ortholog of the geranylgeranyl reductase of *Nicotiana tabacum* (*Chl P* gene; Tanaka *et al.*, 1999) that catalyzes the reduction of geranylgeranyl diphosphate to phytyl diphosphate, the precursor of chlorophyll. *Chl P* has been shown to affect both chlorophyll and total carotenoid levels in tobacco (Tanaka *et al.*, 1999), in accordance with the *wf* phenotype observed in melon fruit. The expression of *CmGGH* was significantly correlated with

carotenoid and apocarotenoid contents in the 414 × Dul population (Table S3). Moreover, in a recent microarray-based differential expression study, *CmGGH* was shown to have a distinctly higher expression in three melon genotypes with orange or green/orange fruit flesh, compared with the white-fleshed genotype (Saladié *et al.*, 2015). Among the genes are also two TFs (CONSTANS-like zinc finger proteins that contain a CCT domain) involved in light signal transduction (MELO3C021608 and MELO3C011317).

The combined *CmOr_CmPPR1* haplotype presented in our model explains approximately 85% of the flesh color variability in our core collection. Additional QTL detected here (Figure 4; Data S6) and in previous studies (Cuevas *et al.*, 2009; Diaz *et al.*, 2011, 2015; Harel-Beja *et al.*, 2010), as well as color-related mutants (Galpaz *et al.*, 2013), are likely to explain much of the remainder. This is in agreement with Monforte *et al.* (2004) and Cuevas *et al.* (2009), where a model of at least three genes was required to explain flesh color distribution in each of their study populations. Notably, a QTL localized to the lower part of chromosome 6 is second in the hierarchy of color-related QTL in our study, as well as in previous studies (Cuevas *et al.*, 2009; Diaz *et al.*, 2015; Harel-Beja *et al.*, 2010).

The findings presented above with regard to the *wf* gene are primarily supported by genetic analyses, and demonstrate the power of RNA-Seq-based QTL and eQTL analyses in linking candidate causative genes of unknown function with phenotypes of interest. Further functional studies, beyond the scope of this manuscript, are needed to confirm our genetic findings.

EXPERIMENTAL PROCEDURES

Plant material

A RIL population (414 × Dul, F_{6-8}), developed from a cross between PI 414723 (*C. melo* var. *momordica*) and 'Dulce' (*C. melo* var. *reticulatus*) was used in this work (Danin-Poleg *et al.*, 2002; Harel-Beja *et al.*, 2010). Ninety-nine lines of the 414 × Dul population, F_1 and their parental lines were grown in a completely randomized design in an open field at Newe Ya'ar Research Center, in the spring–summer seasons of 2006, 2007 and 2010, under standard conditions as described previously (Freilich *et al.*, 2015; Harel-Beja *et al.*, 2010). Each line was represented by 10–12 plants. Flowers were manually pollinated and tagged at anthesis, and one-two fruits were allowed to develop per plant. The fruits were sampled at ripening, when the abscission layer was fully developed. Fruit sampling in 2006 and 2007 was as described in Harel-Beja *et al.* (2010). In 2010, flesh tissues (mesocarp) were sampled from five fruits per line for transcriptomic and metabolomic analyses (see further on). Tissues were immediately frozen in liquid nitrogen and stored at −80°C. In addition, flesh tissues (mesocarp) were sampled from five fruits of 10 selected lines (five high- and five low-sucrose-accumulating lines in 2010). The RIL population is available to the community.

A core collection of 148 melon cultivars (Data S8) was grown in an open field at Newe Ya'ar Research Center under standard conditions as above. Each cultivar was planted in three replications of five plants per plot. Forty-nine of the cultivars were previously described by Tzuri *et al.* (2015). Genotyping was as described in Data S8.

A population (NA × Dul) developed from a cross between Noy-Amid cultivar (white fleshed, *C. melo* var. *inodorous*, NA, Yellow Canary type, white fleshed) and 'Dulce' (see above, Dul, orange fleshed) was also used. One-hundred and forty F_3 plants (selected by ethylene emission) were grown in two repetitions in a greenhouse at Beit Elazari (Israel, 2013). The plants segregated for orange, green cream and white flesh color. An evaluation of color intensity was performed for each of these color categories. RAD-Seq-based (Miller *et al.*, 2007) QTL analysis was performed as described below.

Evaluation of fruit traits

Data S5 presents the phenotyping data for 129 traits analyzed in the 79 RILs selected for the QTL analysis. The following evaluation procedures were performed (aroma volatiles and ethylene emission were analyzed in 2010 only).

Carotenoids and tocopherols. These were extracted from 0.5 g FW of frozen fruit flesh tissue finely ground in a hexane:acetone:ethanol (50:25:25, v/v) mixture as described previously (Freilich *et al.*, 2015; Harel-Beja *et al.*, 2010). They were analyzed, identified and quantified using a Waters (Milford, MA, USA) 2695 HPLC apparatus equipped with a Waters 996 PDA detector (total carotenoid content, α-carotene, β-carotene, ζ-carotene, phytoene and lutein) and Waters 2475 fluorescence detector (total tocopherol content, α-tocopherol, γ-tocopherol and δ-tocopherol) as described previously (Tadmor *et al.*, 2000, 2005).

Soluble sugars. Glucose, fructose and sucrose were extracted from 1 g FW of fresh fruit flesh tissue in hot 80% ethanol and analyzed by HPLC using an Aminex® Fast Carbohydrate Column (100 × 7.8 mm; BioRad Laboratories, Hercules, CA, USA) as described in Harel-Beja *et al.* (2010).

Ethylene emission. This was measured using a modification of a previously described method (Sitrit *et al.*, 1986). Each fruit was enclosed in a Cooky® bag (24 × 38 cm; Schestowitz, Tirat HaCarmel, Israel) equipped with a rubber septum at the bottom and a valve at the top. After sealing, the air was evacuated through the valve using a vacuum pump (MZ 2C, VACUUBRAND GMBH + CO KG, Wertheim, Germany). After incubation for 30 min at room temperature, ethylene was measured in a 1-mL air sample withdrawn from the bag through the rubber septum with a hypodermic syringe and injected into a gas chromatograph (HP 5890 Series II PLUS GC with FID; Hewlett-Packard, Palo Alto, CA, USA) equipped with an SS-packed HAYESEP Q column (80/100, 6′ × 1/8″; Restek, http://www.restek.com/) to measure the ethylene peak area. Ethylene production rate (ppm kg^{-1} FW per hour) was calculated by converting the peak area based on the injection of 1 mL of 1 ppm ethylene in N_2 standard. Ethylene emission of five fruits for each RIL was measured, and the average production rate was calculated (Data S5).

Volatile analyses. These were performed using frozen, uniformly finely ground fruit samples (3 g) that were placed in a 20-mL glass vial containing 1 g solid NaCl. To each vial, 5 mL of a 20% (w/v) NaCl solution and 0.3 μg of 2-heptanone, which was used as an internal standard, were added. The vial was then sealed and stored at 4°C for up to 1 week until analysis. Solid-

phase microextraction (SPME) sampling was conducted according to Davidovich-Rikanati *et al.* (2007) with slight modifications. Each sample was preheated to 50°C, agitated for 5 min at 500 rpm, and then a 65-μm fused silica SPME fiber coated with polydimethylsiloxane/divinylbenzene/carboxen (PDMS/DVB/CAR; Supelco, http://www.sigmaaldrich.com/analytical-chromatography.html) was inserted into the vial and exposed to the sample headspace. After 25 min, the SPME fiber was introduced into the injector port of a gas chromatography–mass spectroscopy (GC–MS) apparatus (Agilent Technologies, http://www.agilent.com/home) equipped with an RXI-5 SIL MS (30 m × 0.25 mm, 0.25 μm) fused-silica capillary column (Restek, http://www.restek.com/). Thermal desorption from the fiber was allowed for 10 min. The injector temperature was 250°C, set for splitless injection. The oven was set to 50°C for 1 min (45°C for 5 min for *in vitro* assays), and the temperature was subsequently increased to 180°C at a rate of 5°C min^{-1}, and to 280°C at 20°C min^{-1}. Helium was used as a carrier gas in a constant pressure mode. The detector temperature was 280°C. The mass range was recorded from 41 to 250 m/z, with electron energy of 70 eV. A mixture of straight-chain alkanes (C7–C23) was injected into the column under the aforementioned conditions for determination of retention times. Volatile compounds were analyzed and quantified in a GC–MS apparatus as described previously (Davidovich-Rikanati *et al.*, 2007; Gonda *et al.*, 2010).

Internal flesh color of 148 genotypes. For the diversity core collection, 15 representative mature fruits per cultivar (five fruits per plot in three replications) were phenotyped for flesh color using image analysis (Data S8). Internal sides of all cut fruit were scanned using a document scanner (CanonScan Lide120). Images were further analyzed using the Tomato Analyzer program (Brewer *et al.*, 2006) to obtain multiple color parameters, including hue index, which is a main property defining the color. L* a* b* color indexes derived from image analysis were confirmed to be highly correlated to parallel measurements taken on fruit flesh using a digital Chroma meter (Minolta CR-400). Hue index was calculated from a* to b* using the equation: hue = arctan(b*/a*). Statistical analysis of the color data (Figure 5a) was performed using the Fit Y by X function in JMP V.8 software package (SAS Institute, Cary, NC, USA).

Additional fruit traits. Fruit length, width, weight and softness (firmness) were measured (Harel-Beja *et al.*, 2010). Fruit pH was measured by pH meter in juice squeezed from the fruit flesh.

Sequence analysis of genes related to flesh color traits

Genomic and complementary DNA sequencing was conducted using an Applied Biosystems 3130xl Genetic Analyzer (Foster City, CA, USA) according to the manufacturer's instructions. The following primers were used: (i) MELO3C003069-F1 and MELO3C003069-R1 were used to detect SNP^{358} in MELO3C003069–70; MELO3C003069-start and MELO3C003070-stop were used to sequence the full-length cDNA of MELO3C003069–70 (primer sequences presented in Table S1), (ii) OR-F4 and OR-R15 were applied to detect the SNPs in *CmOr* (Tzuri *et al.*, 2015). Resequencing data of *CmPPR1* (MELO3C003069–70) of 'Dulce', PI 414723, 'Piel De Sapo' and PI 161375 were deposited in the NCBI repository under accession numbers KX228499–KX228502.

Cloning of *CmThAT1* and its heterologous expression in *Escherichia coli*

Specific primers (*CmThAT1*-F and *CmThAT1*-R; Table S1) were designed to clone *CmThAT1* (MELO3C024190) from 'Dulce' cultivar into the pET21a vector (Novagen, EMD Biosciences, an affiliate of Merck KGaA, Darmstadt, Germany) without any tag at its terminal ends. cDNA from ripe fruit flesh tissues was used as the template (see above). The recombinant plasmid was transformed into *E. coli* BL21 (DE3) using ampicillin selection. New electrocompetent cells were prepared from bacteria carrying the plasmid with the *CmThAT1* insert. The competent cells were transformed again with *CmMGL* in pBK-CMV (extracted from cDNA library clone no. DV633932), using kanamycin selection. Bacteria harboring the two clones were incubated overnight with shaking in 3 mL LB medium containing 30 μg mL^{-1} ampicillin and 50 μg mL^{-1} kanamycin at 37°C, and then 0.5 mL was transferred to 2.5 mL fresh LB medium in 20-mL SPME glass vials and incubated for another hour at 37°C with shaking. Then, isopropylthio-β-galactoside was added to a final concentration of 0.5 mM, and bacteria were incubated for another night at room temperature with shaking (225 rpm). Finally, the headspace of the bacteria was analyzed by SPME–GC–MS as described above.

CmThAT1 *in vitro* enzymatic assays

For *in vitro* analysis, primers were designed to insert *CmThAT1* with a hexahistidine tag in its C-terminal domain into the pET21a vector. The recombinant plasmid was transformed and expressed in BL21(DE3)pLys *Escherichia coli* (Promega, www.promega.com; Gonda *et al.*, 2013). CmThAT1 was purified on an Ni–NTA column using a buffer containing 250 mM imidazole for elution, according to the procedure described there for purification of CmMGL (Gonda *et al.*, 2013). The purified enzyme was assayed for ThAT activity as described below.

Enzymatic assays, with a minimum of four repetitions per experiment, were performed by mixing 200 μL of purified and desalted CmThAT1 (about 3 μg protein) or melon cell-free extract (Gonda *et al.*, 2013) with 50 mM Bis-Tris pH 7.0 buffer containing 0.2 mM acyl-CoA (or propanoyl-CoA), 1 mM sodium methanethiolate and 1 mM dithiothreitol in a total volume of 400 μL in a 2-mL glass vial. The reactions were incubated overnight at 30°C and analyzed by SPME–GC–MS as described above. Control assays were performed at times with no CmThAT1, boiled CmThAT1 no acyl-CoA or no sodium methanethiolate. Authentic standards of thioacetate were run at times to confirm the identification of the product.

RNA isolation

Flesh tissue was harvested at the mature stage from the parental lines, F_1 and RILs. Total RNA was extracted from a pool of five fruit per line, as described previously (Portnoy *et al.*, 2011). Total RNA was also extracted from the flesh of 10 RILs that were sampled 10 and 20 days after pollination (DAP). In brief, frozen, uniformly ground fruit samples (10 g) were mixed by vortexing in a 50-mL tube with 10 mL of extraction buffer and incubated at 65°C for 5 min. Then 30% (w/v) sodium lauroylsarcosine was added to a final concentration of 2% (v/v), and the mixture was vortexed and incubated at 65°C for 2–3 min. An equal volume of phenol was added to the solution, vortexed, and centrifuged at 5000 ***g*** for 5 min. Following three rounds of chloroform–isoamyl alcohol (24:1, v/v) extractions, nucleic acids were precipitated with 1/10 volume of 3 M sodium acetate (pH 5.3) and 2 volumes of 95% (v/v) ethanol. The resulting nucleic acid pellet was dissolved in 10 mL of 2 M LiCl at 4°C overnight. Total RNA was precipitated by centrifugation at 15 000 ***g*** for 10 min at 4°C and dissolved in 0.5 mL diethyl pyrocarbonate (DEPC) water. After reprecipitation, the pellet was dissolved in 100 μL DEPC water. The quality of the RNA was analyzed by ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA) and by electrophoresis on a

formaldehyde agarose gel. Total RNA was used for the preparation of Illumina RNA-Seq libraries. First-strand cDNA synthesis was performed using the Verso cDNA kit (Thermo Fisher Scientific, Grand Island, NY, USA).

RNA-Seq analysis: library preparation, sequencing and data analysis

A total of 112 strand-specific RNA-Seq libraries were constructed from fruit samples collected in 2010, using the protocol described in Zhong *et al.* (2011). These include libraries of 96 RILs (mature fruits); 10 RILs (mixed fruit of 10 and 20 DAP), F1 and parental lines (NCBI SRA repository, accession identifier SRP052934). The libraries were barcoded, and 8–20 libraries were pooled and sequenced on one lane of the Illumina GAII and HiSeq 2000 systems, respectively. RNA-Seq reads were first aligned to the ribosomal RNA database (Quast *et al.*, 2013) using Bowtie, allowing two mismatches (Langmead *et al.*, 2009), to remove any possible rRNA sequence contamination (Data S11). The resulting filtered reads were aligned to the melon genome (Garcia-Mas *et al.*, 2012) using TopHat (Trapnell *et al.*, 2009), allowing one segment mismatch (Data S11).

Following alignments, raw counts for each melon gene were normalized to RPKM (Mortazavi *et al.*, 2008). RPKM values for 27 427 gene accessions across the RIL population are available for the mature fruit (Freilich *et al.*, 2015). RPKM values of 10 and 20 DAP are available for 10 of the RILs. Expression values of chloroplast genome encoded genes are presented under a second tab.

SNP calling

To identify SNPs between RILs, RNA-Seq reads of the RILs and parental lines were aligned to the melon reference genome (Garcia-Mas *et al.*, 2012) using the Burrows–Wheeler Alignment tool (Li and Durbin, 2009) and only uniquely mapped reads were kept. Identification of 82 140 SNPs was based on the mpileup files generated by SAMtools (Li *et al.*, 2009). The identified SNPs were supported by at least three reads and had an allele frequency of at least 0.8. Next, SNPs were phased based on the parental information and searched for reliable SNPs in the population based on Mendelian segregation criteria (> 10 lines with each of the alternative homozygous SNP options and < 20 lines with heterozygous SNP status). SNPs with a genotyping pattern that was inconsistent with their neighbor's SNP pattern were filtered out; 58 328 reliable SNPs were identified and used for genotyping.

Genetic map construction and identification of miss-assembled scaffolds

The determination of the genotype of the SNPs in each line was done using a three states [Hom-A, Het (also capturing not defined regions) and Hom-B] Hidden Markov model with the following transition and emission probabilities.

1) Transition probability between adjacent SNPs, capturing Poisson event determined by the expected recombination rate, that was set to be:
   a. transition from Hom to Het states: 2*PROB(1-PROB)
   b. transition from Hom-A to Hom-B (or vice versa): PROB^2
   c. staying at the same Hom state: (1-PROB)^2
   d. transition from Het to Hom state: PROB*(1-PROB)
   e. staying at the Het state: 1-2*PROB*(1-PROB)

   where PROB=1 – exp (-SNPs_distance/(23*10^6))
2) Emission probability:
   a. Hom state:
      i. missing information = 0.9
      ii. only phased SNP exist = 0.08
      iii. unphased SNP exist = 0.02
   b. Het state:
      i. missing information = 0.9
      ii. only one allele exist = 0.05
      iii. both alleles exist = –0.05

In a RIL population a small heterozygous block between distinct homozygous blocks marks an uncertainty genotype region in which the genotype cannot be determined due to missing information and lack of imputation ability. High heterozygosity levels and frequent recombination-like events were detected in 17 lines (> 500 heterozygous bins), and they were excluded from further analysis. The remaining 79 lines generated overall 3663 distinct genotyping patterns (bins) in the melon chromosomes 1–12, corresponding to 6636 recombination events that occurred in the RIL population. Genetic map of 2050 cM (note that this map includes the errors in the physical map in chr5, chr2, etc.), which clearly makes our observation an overestimation of the genetic map size (1880 cM estimated size after correction).

Based on the obtained genotypic patterns of each bin, several regions of misassembly in the draft reference genome were identified (Garcia-Mas *et al.*, 2012), and the location of several unmapped scaffolds (grouped earlier to the virtual chromosome "0") were updated. The genetic distance between each pair of bins was calculated (weighted sum over all of the differences in genotyping status). Corrected locations of bins were discovered in cases where the matching bin with minimal genotyping distance was located on a different chromosome (Data S3).

QTL and eQTL analyses

Phenotypic data collected in 2006, 2007 and 2010 (Data S5) were analyzed for each year independently.

Genotypic patterns (bins) were used for QTL analysis. In each pattern, the RILs were divided into two groups: Homozygous–' and Homozygous–PI 414723, based on the genotyping, and correlation to phenotype or gene expression was measured by *P*-value (using *t*-test) and explained the variance ($r^2 = 1 - RSS_0/RSS_1$, where $RSS_0 = sum_{i:over\ all\ samples}[phenotype_i - average_{all\ samples}(phenotype)]^2$ and $RSS_1 = sum_{i:over\ all\ samples\ with\ DL\ genotype}[phenotype - average_{DL\ samples}(phenotype)]^2 + sum_{i:over\ all\ samples\ with\ PI\ genotype}[phenotype_i - average_{PI\ samples}(phenotype)]^{2)}$ and the LOD score ($LOD = n/2\log_{10}[RSS_0/RSS_1]$). The resulting raw *P*-values were adjusted for multiple testing using FDR (Benjamini and Hochberg, 1995). Statistically significant QTL were identified if they had FDR < 10%, and QTL intervals were defined as the region surrounding the QTL peak with $r^2 > .9$ (max r^2); QTL and eQTL were excluded from the analysis of mapping resolution when their QTL intervals spanned more than one scaffold. As 18.3% of the melon scaffolds in version 3.5 of the melon genome were not mapped to their chromosomal location and were assigned to virtual chromosome "0" (Garcia-Mas *et al.*, 2012), the integration of those scaffolds into the genome could disrupt the determination of QTL intervals in cases of QTL with more than one scaffold. In addition, to avoid mistakes, QTL on both the unmapped scaffolds and their predicted neighboring scaffolds (Data S4) were kept, such that the final number of QTL/eQTL is probably lower.

QTL analysis of NA × Dul F3 population

In the NA × Dul F3 population, restriction-site-associated DNA sequencing (RAD-Seq; Miller *et al.*, 2007) and QTL analyses (similarly to described above) were performed by NRGene LTD (Nes Ziyyona, Israel) using 131 individual F3 plants representing the whole scale of ethylene emission.

Network analysis

The eQTL networks were plotted using Cytoscape 2.8.2 (http://www.cytoscape.org/; Shannon *et al.*, 2003). GO terms of the genes in the network were selected with the help of the Melonomics database (https://melonomics.net/) and the Plant MetGenMap package (http://bioinfo.bti.cornell.edu/cgi-bin/MetGenMAP/home.cgi).

A whole-genome co-expression network was constructed from the RNA-Seq dataset (Data S2, mature fruit data; Freilich *et al.*, 2015), using CoExpNetViz (Tzfadia *et al.*, 2016). The network was filtered to hold correlations with *r*-value cut-offs corresponding to the top 95 percentile and bottom 5 percentile of all correlations measured (which correspond to *r*-value > .8 and *r*-value < −.2, respectively, Pearson correlation coefficient).

We searched for 'functional modules' (i.e. groups of cellular components whose interactions can be attributed to a specific biological function) with MATISSE (Ulitsky and Shamir, 2009). Functional modules were represented as groups of genes forming subnetworks (modules) in both the co-expression and eQTL networks.

ACCESSION NUMBERS

Illumina sequencing data can be found in the NCBI SRA repository under accession identifier SRP052934. The allelic value and location of 58 328 SNPs can be found in the Cucurbit Genomics Database (http://www.icugi.org/cgi-bin/ICuGI/index.cgi).

Resequencing data of *CmThAT1* (MELO3C024190) of 'Dulce' and PI 414723 can be found in the NCBI repository under accession numbers KX017232–KX017233. Resequencing data of *CmPPR1* (MELO3C003069–70) of 'Dulce', PI 414723, 'Piel De Sapo' and PI 161375 can be found in the NCBI repository under accession numbers KX228499–KX228502.

ACKNOWLEDGEMENTS

We are most thankful to Tamar Zakai for critically reviewing the manuscript. This research was supported by Research Grant Award No. IS–4223–09C from BARD, The United States–Israel Binational Agricultural Research and Development Fund, Grant Award ICOB 2012 from NSF-BSF Joint Funding Program and partially by the Chief Scientist of the Ministry of Agriculture of Israel (Grant No. 261-1049-14).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Enzymatic production of thioesters (*S*-esters) and *O*-esters.

Figure S2. Correlation between genetic and physical maps.

Figure S3. IGV snapshots (Robinson *et al.*, 2011) of QTL analysis of fruit length and weight, TSS, total sugars and ethylene emission.

Figure S4. Multiple amino acid sequence alignment of melon thiolases and other functionally characterized acetoacetyl-coenzyme A thiolases (AACTs).

Figure S5. Methanethiol acetyltransferase enzymatic activity in melon soluble protein extracts.

Figure S6. Methanethiol acetyltransferase activity of the recombinant CmThAT1 enzymes.

Figure S7. Sequences and expression polymorphisms in 5′ and 3′ untranslated regions (UTRs) of *CmThAT1* in the parental lines and RILs of the 414 × Dul population.

Figure S8. Synteny of genes around *CmPPR1* in melon, cucumber and watermelon.

Figure S9. Analysis of the *CmPPR1* gene using the PPR gene tool.

Figure S10. QTL analysis of white flesh color in the NA × Dul F3 population.

Figure S11. A heatmap of the correlation of expression of 30 genes governed by the eQTL of MELO3C003069 versus chloroplast encoded genes.

Table S1 List of primers, their sequences and their uses

Table S2 Mapping the *wf* locus in the current and previous studies

Table S3 Correlation of expression between genes and metabolites (α-carotene, α-ionone, β-carotene and β-ionone)

Data S1. Allelic value and location of SNPs of the RILs of the 414 × Dul population.

Data S2. Level of genes' expression across RILs of the 414 × Dul population.

Data S3. Bin information.

Data S4. Suggested genomic rearrangements according to the genetic map.

Data S5. Metabolic and phenotypic data of fruit traits.

Data S6. QTL analysis of fruit traits (FDR < 10%).

Data S7. QTL analysis of fruit traits (FDR < 30%).

Data S8. Genotyping and phenotyping of a core collection for flesh color and candidate genes.

Data S9. eQTL of genes expressed in fruit (FDR < 10%).

Data S10. GO terms and annotation of the 132 genes comprising the subnetwork of eQTL associated with the *CmPPR1* gene.

Data S11. Statistics of melon RNA-Seq libraries.

REFERENCES

Aharoni, A., Keizer, L.C.P., Bouwmeester, H.J. *et al.* (2000) Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. *Plant Cell*, **12**, 647–662. https://doi.org/10.1105/tpc.12.5.647

Argyris, J., Ruiz-Herrera, A., Madriz-Masis, P., Sanseverino, W., Morata, J., Pujol, M., Ramos-Onsins, S. and Garcia-Mas, J. (2015) Use of targeted SNP selection for an improved anchoring of the melon (*Cucumis melo* L.) scaffold genome assembly. *BMC Genom.*, **16**, https://doi.org/10.1186/s12864-014-1196-3.

Ashburner, M., Ball, C.A., Blake, J.A. *et al.* (2000) Gene Ontology: Tool for the unification of biology. *Nat. Genet.*, **25**, 25–29. https://doi.org/10.1038/75556

Ayub, R., Guis, M., Amor, M.B., Gillot, L., Roustan, J.-P., Latche, A., Bouzayen, M. and Pech, J.-C. (1996) Expression of ACC oxidase antisense gene inhibits ripening of cantaloupe melon fruits. *Nat. Biotechnol.*, **14**, 862–866. https://doi.org/10.1038/nbt0796-862

Barajas-López, J.D.D., Blanco, N.E. and Strand, Å. (2013) Plastid-to-nucleus communication, signals controlling the running of the plant cell. *Biochim. Biophys. Acta*, **1833**, 425–437. https://doi.org/10.1016/j.bbamcr.2012.06.020

Barkan, A. and Small, I. (2014) Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.*, **65**, 415–442. https://doi.org/10.1146/annurev-arplant-050213-040159

Barkan, A., Walker, M., Nolasco, M. and Johnson, D. (1994) A nuclear mutation in maize blocks the processing and translation of several chloroplast mRNAs and provides evidence for the differential translation of alternative mRNA forms. *EMBO J.*, **13**, 3170–3181.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B*, **57**, 289–300.

Benjamini, Y. and Yekutieli, D. (2005) False discovery rate–adjusted multiple confidence intervals for selected parameters. *J. Am. Stat. Assoc.*, **100**, 71–81. https://doi.org/10.1198/016214504000001907

Bhat, J.A., Ali, S., Salgotra, R.K. *et al.* (2016) Genomic selection in the era of next generation sequencing for complex traits in plant breeding. *Front. Genet.*, **7**, 221.

Boualem, A., Fergany, M., Fernandez, R. *et al.* (2008) A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. *Science*, **321**, 836–838. https://doi.org/10.1126/science.1159023

Brem, R.B., Yvert, G., Clinton, R. and Kruglyak, L. (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science*, **296**, 752–755. https://doi.org/10.1126/science.1069516

Brewer, M.T., Lang, L., Fujimura, K., Dujmovic, N., Gray, S. and Van Der Knaap, E. (2006) Development of a controlled vocabulary and software application to analyze fruit shape variation in tomato and other plant species. *Plant Physiol.*, **141**, 15–25. https://doi.org/10.1104/pp.106.077867

Burger, Y., Paris, H., Cohen, R., Katzir, N., Tadmor, Y., Lewinsohn, E. and Schaffer, A.A. (2009) Genetic diversity of *Cucumis melo*. *Hortic. Rev.*, **36**, 165–198.

Burr, B. and Burr, F.A. (1991) Recombinant inbreds for molecular mapping in maize: Theoretical and practical considerations. *Trends Genet.*, **7**, 55–60.

Chan, A.P., Crabtree, J., Zhao, Q., Lorenzi, H., Orvis, J., Puiu, D., Melake-Berhan, A., Jones, K.M., Redman, J. and Chen, G. (2010) Draft genome sequence of the oilseed species Ricinus communis. *Nat. Biotechnol.*, **28**, https://doi.org/10.1038/nbt.1674.

Chateigner-Boutin, A.-L., Ramos-Vega, M., Guevara-García, A. *et al.* (2008) CLB19, a pentatricopeptide repeat protein required for editing of rpoA and clpP chloroplast transcripts. *Plant J.*, **56**, 590–602. https://doi.org/10.1111/j.1365-313X.2008.03634.x

Chayut, N., Yuan, H., Ohali, S. *et al.* (2015) A bulk segregant transcriptome analysis reveals metabolic and cellular processes associated with Orange allelic variation and fruit β-carotene accumulation in melon fruit. *BMC Plant Biol.*, **15**, 274. https://doi.org/10.1186/s12870-015-0661-8

Chayut, N., Yuan, H., Ohali, S. *et al.* (2017) Distinct mechanisms of the ORANGE protein in controlling carotenoid flux. *Plant Physiol.*, **173**, 376–389. https://doi.org/10.1104/pp.16.01256

Chen, Z., Wang, B., Dong, X., Liu, H., Ren, L., Chen, J., Hauck, A., Song, W. and Lai, J. (2014) An ultra-high density bin-map for rapid QTL mapping for tassel and ear architecture in a large F(2) maize population. *BMC Genom.*, **15**, 433. https://doi.org/10.1186/1471-2164-15-433.

Cheng, S., Gutmann, B., Zhong, X. *et al.* (2016) Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *Plant J.*, **85**, 532–547. https://doi.org/10.1111/tpj.13121

Clayberg, C.D. (1992) Interaction and linkage test of flesh colour genes in *Cucumis melo* L. *Cucurbit Genet. Coop.*, **15**, 53.

Cohen, S., Itkin, M., Yeselson, Y. *et al.* (2014) The PH gene determines fruit acidity and contributes to the evolution of sweet melons. *Nat. Commun.*, **5**, https://doi.org/10.1038/ncomms5026.

Cubillos, F.A., Coustham, V. and Loudet, O. (2012) Lessons from eQTL mapping studies: Non-coding regions and their role behind natural phenotypic variation in plants. *Curr. Opin. Plant Biol.*, **15**, 192–198. https://doi.org/10.1016/j.pbi.2012.01.005

Cuevas, H.E., Staub, J.E., Simon, P.W., Zalapa, J.E. and Mccreight, J.D. (2008) Mapping of genetic loci that regulate quantity of beta-carotene in fruit of US Western Shipping melon *Cucumis melo* L.). *Theor. Appl. Genet.*, **117**, 1345–1359. https://doi.org/10.1007/s00122-008-0868-2

Cuevas, H.E., Staub, J.E., Simon, P.W. and Zalapa, J.E. (2009) A consensus linkage map identifies genomic regions controlling fruit maturity and beta-carotene-associated flesh color in melon (*Cucumis melo* L.). *Theor. Appl. Genet.*, **119**, 741–756. https://doi.org/10.1007/s00122-009-1085-3

Danin-Poleg, Y., Tadmor, Y., Tzuri, G., Reis, N., Hirschberg, J. and Katzir, N. (2002) Construction of a genetic map of melon with molecular markers and horticultural traits, and localization of genes associated with ZYMV resistance. *Euphytica*, **125**, 373–384. https://doi.org/10.1023/A:1016021926815

Davidovich-Rikanati, R., Sitrit, Y., Tadmor, Y. *et al.* (2007) Enrichment of tomato flavor by diversion of the early plastidial terpenoid pathway. *Nat. Biotechnol.*, **25**, 899–901. https://doi.org/10.1038/nbt1312

Diaz, A., Fergany, M., Formisano, G. *et al.* (2011) A consensus linkage map for molecular markers and Quantitative Trait Loci associated with economically important traits in melon (*Cucumis melo* L.). *BMC Plant Biol.*, **11**, 111. https://doi.org/10.1186/1471-2229-11-111.

Diaz, A., Forment, J., Argyris, J., Fukino, N., Tzuri, G., Harel-Beja, R., Katzir, N., Garcia-Mas, J. and Monforte, A. (2015) Anchoring the consensus ICuGI genetic map to the melon (*Cucumis melo* L.) genome. *Mol. Breed.*, **35**, 1–7.

Drost, D., Puranik, S., Novaes, E., Novaes, C., Dervinis, C., Gailing, O. and Kirst, M. (2015) Genetical genomics of Populus leaf shape variation. *BMC Plant Biol.*, **15**, 166. https://doi.org/10.1186/s12870-015-0557-7.

Dyer, J.H., Maina, A., Gomez, I.D., Cadet, M., Oeljeklaus, S. and Schiedel, A.C. (2009) Cloning, expression and purification of an acetoacetyl CoA thiolase from sunflower cotyledon. *Int. J. Biol. Sci.*, **5**, 736–744. https://doi.org/10.7150/ijbs.5.736

El-Sharkawy, I., Manríquez, D., Flores, F., Regad, F., Bouzayen, M., Latché, A. and Pech, J.-C. (2005) Functional characterization of a melon alcohol acyl-transferase gene family involved in the biosynthesis of ester volatiles. Identification of the crucial role of a threonine residue for enzyme activity. *Plant Mol. Biol.*, **59**, 345–362. https://doi.org/10.1007/s11103-005-8884-y

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. and Mitchell, S.E. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, **6**, e19379. https://doi.org/10.1371/journal.pone.0019379.

Eshed, Y. and Zamir, D. (1995) An introgression line population of Lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated Qtl. *Genetics*, **141**, 1147–1162.

Francesconi, M. and Lehner, B. (2013) The effects of genetic variation on gene expression dynamics during development. *Nature*, **505**, 208.

Freilich, S., Lev, S., Gonda, I. *et al.* (2015) Systems approach for exploring the intricate associations between sweetness, color and aroma in melon fruits. *BMC Plant Biol.*, **15**, 71. https://doi.org/10.1186/s12870-015-0449-x.

Fulop, D., Ranjan, A., Ofner, I. *et al.* (2016) A new advanced backcross tomato population enables high resolution leaf QTL mapping and gene identification. *G3: Genes|Genomes|Genetics*, **6**, 3169–3184.

Galpaz, N. and Reymond, M. (2010) Natural variation in Arabidopsis thaliana revealed a genetic network controlling germination under salt stress. *PLoS ONE*, **5**, e15198. https://doi.org/10.1371/journal.pone.0015198.

Galpaz, N., Burger, Y., Lavee, T. *et al.* (2013) Genetic and chemical characterization of an EMS induced mutation in *Cucumis melo* CRTISO gene. *Arch. Biochem. Biophys.*, **539**, 117–125. https://doi.org/10.1016/j.abb.2013.08.006

Gao, Z.-Y., Zhao, S.-C., He, W.-M. *et al.* (2013) Dissecting yield-associated loci in super hybrid rice by resequencing recombinant inbred lines and improving parental genome sequences. *Proc. Natl Acad. Sci. USA*, **110**, 14 492–14 497. https://doi.org/10.1073/pnas.1306579110

Garcia-Mas, J., Benjak, A., Sanseverino, W. *et al.* (2012) The genome of melon (*Cucumis melo* L.). *Proc. Natl Acad. Sci. USA*, **109**, 11 872–11 877. https://doi.org/10.1073/pnas.1205415109

Gonda, I., Bar, E., Portnoy, V. *et al.* (2010) Branched-chain and aromatic amino acid catabolism into aroma volatiles in *Cucumis melo* L. fruit. *J. Exp. Bot.*, **61**, 1111–1123. https://doi.org/10.1093/jxb/erp390

Gonda, I., Lev, S., Bar, E. *et al.* (2013) Catabolism of L-methionine in the formation of sulfur and other volatiles in melon (*Cucumis melo* L.) fruit. *Plant J.*, **74**, 458–472. https://doi.org/10.1111/tpj.12149

Gonda, I., Burger, Y., Schaffer, A.A., Ibdah, M., Tadmor, Y., Katzir, N., Fait, A. and Lewinsohn, E. (2016) Biosynthesis and perception of melon aroma. In *Biotechnology in Flavor Production (Havkin-Frenkel, D.* (Dudai, N., ed). Oxford, UK: Wiley-Blackwell, pp. 281–305. https://doi.org/10.1002/9781118354056

Graham, N.S., Hammond, J.P., Lysenko, A. *et al.* (2014) Genetical and comparative genomics of *Brassica* under altered Ca supply identifies *Arabidopsis* Ca-transporter orthologs. *Plant Cell*, **26**, 2818–2830. https://doi.org/10.1105/tpc.114.128603

Hansen, B.G., Halkier, B.A. and Kliebenstein, D.J. (2008) Identifying the molecular basis of QTLs: eQTLs add a new dimension. *Trends Plant Sci.*, **13**, 72–77. https://doi.org/10.1016/j.tplants.2007.11.008

Hao, W. and Lin, H.-X. (2010) Toward understanding genetic mechanisms of complex traits in rice. *J. Genet. Genomics*, **37**, 653–666. https://doi.org/10.1016/S1673-8527(09)60084-9

Harel-Beja, R., Tzuri, G., Portnoy, V. *et al.* (2010) A genetic map of melon highly enriched with fruit quality QTLs and EST markers, including sugar and carotenoid metabolism genes. *Theor. Appl. Genet.*, **121**, 511–533. https://doi.org/10.1007/s00122-010-1327-4

Helinck, S., Spinnler, H.E., Parayre, S., Dame-Cahagne, M. and Bonnarme, P. (2000) Enzymatic versus spontaneous S-methyl thioester synthesis in *Geotrichum candidum*. *FEMS Microbiol. Lett.*, **193**, 237–241. https://doi.org/10.1111/j.1574-6968.2000.tb09430.x

Holland, J.B. (2007) Genetic architecture of complex traits in plants. *Curr. Opin. Plant Biol.*, **10**, 156–161. https://doi.org/10.1016/j.pbi.2007.01.003

Huang, X., Feng, Q., Qian, Q. *et al.* (2009) High-throughput genotyping by whole-genome resequencing. *Genome Res.*, **19**, 1068–1076. https://doi.org/10.1101/gr.089516.108

Hughes, M. (1948) The inheritance of two characters of *Cucumis melo* and their interrelationship. *Proc. Am. Soc. Hortic. Sci.*, **59**, 399–402.

Ibdah, M., Azulay, Y., Portnoy, V. *et al.* (2006) Functional characterization of CmCCD1, a carotenoid cleavage dioxygenase from melon. *Phytochemistry*, **67**, 1579–1589. https://doi.org/10.1016/j.phytochem.2006.02.009

Iman, M.K.L., Abo-Bakr, M.A. and Hanna, H.Y. (1972) Inheritance of some economic characters in crosses between sweet melon and snake cucumber. I. Inheritance of qualitative characters. *Assiut. J. Agric. Sci.*, **3**, 363–380.

Jansen, R.C. and Nap, J.-P. (2001) Genetical genomics: The added value from segregation. *Trends Genet.*, **17**, 388–391. https://doi.org/10.1016/S0168-9525(01)02310-1

Jin, H., Song, Z. and Nikolau, B.J. (2012) Reverse genetic characterization of two paralogous acetoacetyl CoA thiolase genes in Arabidopsis reveals their importance in plant growth and development. *Plant J.*, **70**, 1015–1032. https://doi.org/10.1111/j.1365-313X.2012.04942.x

Joosen, R.V.L., Ligterink, W., Hilhorst, H.W.M. and Keurentjes, J.J.B. (2009) Advances in genetical genomics of plants. *Curr. Genomics*, **10**, 540–549. https://doi.org/10.2174/138920209789503914

Jordan, M.J., Shaw, P.E. and Goodner, K.L. (2001) Volatile components in aqueous essence and fresh fruit of *Cucumis melo* cv. athena (muskmelon) by GC-MS and GC-O. *J. Agric. Food Chem.*, **49**, 5929–5933. https://doi.org/10.1021/jf010954o

Jordan, M.C., Somers, D.J. and Banks, T.W. (2007) Identifying regions of the wheat genome controlling seed development by mapping expression quantitative trait loci. *Plant Biotech. J.*, **5**, 442–453. https://doi.org/10.1111/j.1467-7652.2007.00253.x

Josephs, E.B., Lee, Y.W., Stinchcombe, J.R. and Wright, S.I. (2015) Association mapping reveals the role of purifying selection in the maintenance of genomic variation in gene expression. *Proc. Natl Acad. Sci. USA*, **112**, 15 390–15 395. https://doi.org/10.1073/pnas.1503027112

Keurentjes, J.J.B., Fu, J., Terpstra, I.R., Garcia, J.M., Van Den Ackerveken, G., Snoek, L.B., Peeters, A.J.M., Vreugdenhil, D., Koornneef, M. and Jansen, R.C. (2007) Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. *Proc. Natl Acad. Sci. USA*, **104**, 1708–1713. https://doi.org/10.1073/pnas.0610429104

Kinnersley, B., Labussière, M., Holroyd, A. *et al.* (2015) Genome-wide association study identifies multiple susceptibility loci for glioma. *Nauret. Commun.*, **6**, 8559. https://doi.org/10.1038/ncomms9559.

Kliebenstein, D. (2009) Quantitative genomics: Analyzing intraspecific variation using global gene expression polymorphisms or eQTLs. *Annu. Rev. Plant Biol.*, **60**, 93–114. https://doi.org/10.1146/annurev.arplant.043008.092114

Kliebenstein, D., West, M., Van Leeuwen, H., Loudet, O., Doerge, R. and St Clair, D. (2006) Identification of QTLs controlling gene expression networks defined *a priori*. *BMC Bioinformatics*, **7**, 308. https://doi.org/10.1186/1471-2105-7-308.

Koussevitzky, S., Nott, A., Mockler, T.C., Hong, F., Sachetto-Martins, G., Surpin, M., Lim, J., Mittler, R. and Chory, J. (2007) Signals from chloroplasts converge to regulate nuclear gene expression. *Science*, **316**, 715–719. https://doi.org/10.1126/science. 1140516

Kursula, P., Sikkilä, H., Fukao, T., Kondo, N. and Wierenga, R.K. (2005) High resolution crystal structures of human cytosolic thiolase (CT): A comparison of the active sites of human CT, bacterial thiolase, and bacterial KAS I. *J. Mol. Biol.*, **347**, 189–201. https://doi.org/10.1016/j.jmb.2005.01.018

Landaud, S., Helinck, S. and Bonnarme, P. (2008) Formation of volatile sulfur compounds and metabolism of methionine and other sulfur compounds in fermented food. *Appl. Microbiol. Biotechnol.*, **77**, 1191–1205. https://doi.org/10.1007/s00253-007-1288-y

Langmead, B., Trapnell, C., Pop, M. and Salzberg, S. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25. https://doi.org/10.1186/gb-2009-10-3-r25.

Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Subgroup, G.P.D.P. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liu, S.Q., Holland, R. and Crow, V.L. (2004) Esters and their biosynthesis in fermented dairy products: A review. *Int. Dairy J.*, **14**, 923–945. https://doi.org/10.1016/j.idairyj.2004.02.010

Liu, X., Yu, F. and Rodermel, S. (2010) An arabidopsis pentatricopeptide repeat protein, SUPPRESSOR OF VARIEGATION7, is required for FtsH-mediated chloroplast biogenesis. *Plant Physiol.*, **154**, 1588–1601. https://doi.org/10.1104/pp.110.164111

Lucchetta, L., Manriquez, D., El-Sharkawy, I. *et al.* (2007) Biochemical and catalytic properties of three recombinant alcohol acyltransferases of melon. Sulfur-containing ester formation, regulatory role of CoA-SH in activity, and sequence elements conferring substrate preference. *J. Agric. Food Chem.*, **55**, 5213–5220. https://doi.org/10.1021/jf070210w

Ma, X., Koepke, J., Panjikar, S., Fritzsch, G. and Stöckigt, J. (2005) Crystal structure of vinorine synthase, the first representative of the BAHD superfamily. *J. Biol. Chem.*, **280**, 13 576–13 583. https://doi.org/10.1074/jbc.M414508200

Manríquez, D., El-Sharkawy, I., Flores, F., El-Yahyaoui, F., Regad, F., Bouzayen, M., Latché, A. and Pech, J.-C. (2006) Two highly divergent alcohol dehydrogenases of melon exhibit fruit ripening-specific expression and distinct biochemical characteristics. *Plant Mol. Biol.*, **61**, 675–685. https://doi.org/10.1007/s11103-006-0040-9

Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A. and Johnson, E.A. (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res.*, **17**, 240–248. https://doi.org/10.1101/gr.5681207

Mochizuki, N., Brusslan, J.A., Larkin, R., Nagatani, A. and Chory, J. (2001) *Arabidopsis genomes uncoupled 5 (GUN5)* mutant reveals the involvement of Mg-chelatase H subunit in plastid-to-nucleus signal transduction. *Proc. Natl Acad. Sci. USA*, **98**, 2053–2058. https://doi.org/10.1073/pnas.98.4.2053

Monforte, A.J., Oliver, M., Gonzalo, M.J., Alvarez, J.M., Dolcet-Sanjuan, R. and Arus, P. (2004) Identification of quantitative trait loci involved in fruit quality traits in melon (*Cucumis melo* L.). *Theor. Appl. Genet.*, **108**, 750–758.

Monforte, A.J., Diaz, A., Caño-Delgado, A. and Van Der Knaap, E. (2014) The genetic basis of fruit morphology in horticultural crops: Lessons from tomato and melon. *J. Exp. Bot.*, **65**, 4625–4637.

Mortazavi, A., Williams, B.A., Mccue, K., Schaeffer, L. and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Meth.*, **5**, 621–628. https://doi.org/10.1038/nmeth.1226

Noichinda, S., Ueda, Y., Imahori, Y. and Chachin, K. (1999) Thioester production and thioalcohol specificity of alcohol acetyltransferase in strawberry fruit. *Food Sci. Technol. Res.*, **5**, 99–103. https://doi.org/10.3136/fstr.5.99

Nuñez-Palenius, H.G., Gomez-Lim, M., Ochoa-Alejo, N., Grumet, R., Lester, G. and Cantliffe, D.J. (2008) Melon fruits: Genetic diversity, physiology, and biotechnology features. *Crit. Rev. Biotechnol.*, **28**, 13–55. https://doi.org/10.1080/07388550801891111

Ofner, I., Lashbrooke, J., Pleban, T., Aharoni, A. and Zamir, D. (2016) *Solanum pennellii* backcross inbred lines (BILs) link small genomic bins with tomato traits. *Plant J.*, **87**, 151–160. https://doi.org/10.1111/tpj.13194

Page, D.R. and Grossniklaus, U. (2002) The art and design of genetic screens: *Arabidopsis thaliana*. *Nat. Rev. Genet.*, **3**, 124–136. https://doi.org/10.1038/nrg730

Paran, I., Goldman, I. and Zamir, D. (1997) QTL analysis of morphological traits in a tomato recombinant inbred line population. *Genome*, **40**, 242–248. https://doi.org/10.1139/g97-034

Peretó, J., López-García, P. and Moreira, D. (2005) Phylogenetic analysis of eukaryotic thiolases suggests multiple proteobacterial origins. *J. Mol. Evol.*, **61**, 65–74. https://doi.org/10.1007/s00239-004-0280-8

Perin, C., Gomez-Jimenez, M., Hagen, L., Dogimont, C., Pech, J.C., Latche, A., Pitrat, M. and Lelievre, J.M. (2002) Molecular and genetic characterization of a non-climacteric phenotype in melon reveals two loci conferring altered ethylene response in fruit. *Plant Physiol.*, **129**, 300–309. https://doi.org/10.1104/pp.010613

Pesaresi, P., Mizzotti, C., Colombo, M. and Masiero, S. (2014) Genetic regulation and structural changes during tomato fruit development and ripening. *Front. Plant Sci.*, **5**, 124.

Pitrat, M., Hanelt, P. and Hammer, K. (2000) Some comments on infraspecific classification of cultivars of melon. In *VII Eucarpia Meeting on Cucurbit Genetics and Breeding* (Katzir, N. and Paris, H.S., eds). Israel: ISHS Acta Hortic, pp. 29–36.

Portnoy, V., Benyamini, Y., Bar, E. *et al.* (2008) The molecular and biochemical basis for varietal variation in sesquiterpene content in melon (*Cucumis melo* L.) rinds. *Plant Mol. Biol.*, **66**, 647–661. https://doi.org/10.1007/s11103-008-9296-6

Portnoy, V., Diber, A., Pollock, S. *et al.* (2011) Use of non-normalized, non-amplified cDNA for 454-based RNA sequencing of fleshy melon fruit. *Plant Genome*, **4**, 36–46. https://doi.org/10.3835/plantgenome2010.11.0026

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glöckner, F.O. (2013) The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.*, **41**, D590–D596.

Riov, J. and Jaffe, M.J. (1972) Evidence for enzymatic thioltransacetylation in plant extracts. *Phytochemistry*, **11**, 2437–2438. https://doi.org/10.1016/S0031-9422(00)88512-X

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P. (2011) Integrative genomics viewer. *Nat. Biotechnol.*, **29**, 24–26. https://doi.org/10.1038/nbt.1754

Saladié, M., Cañizares, J., Phillips, M.A., Rodriguez-Concepcion, M., Larrigaudière, C., Gibon, Y., Stitt, M., Lunn, J.E. and Garcia-Mas, J. (2015) Comparative transcriptional profiling analysis of developing melon (*Cucumis melo* L.) fruit from climacteric and non-climacteric varieties. *BMC Genom.*, **16**, 440. https://doi.org/10.1186/s12864-015-1649-3.

Schadt, E.E., Monks, S.A., Drake, T.A. *et al.* (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature*, **422**, 297–302. https://doi.org/10.1038/nature01434

Schieberle, P., Ofner, S. and Grosch, W. (1990) Evaluation of potent odorants in cucumbers (*Cucumis sativus*) and muskmelons (*Cucumis melo*) by aroma extract dilution analysis. *J. Food Sci.*, **55**, 193–195. https://doi.org/10.1111/j.1365-2621.1990.tb06050.x

Schwenkert, S., Legen, J., Takami, T., Shikanai, T., Herrmann, R.G. and Meurer, J. (2007) Role of the low-molecular-weight subunits PetL, PetG, and PetN in assembly, stability, and dimerization of the cytochrome *b(6)f* complex in tobacco. *Plant Physiol.*, **144**, 1924–1935. https://doi.org/10.1104/pp.107.100131

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504. https://doi.org/10.1101/gr.1239303

Sîrbu, A., Kerr, G., Crane, M. and Ruskin, H.J. (2012) RNA-Seq vs dual- and single-channel microarray data: Sensitivity analysis for differential expression and clustering. *PLoS ONE*, **7**, e50986. https://doi.org/10.1371/journal.pone.0050986.

Sitrit, Y., Riov, J. and Blumenfeld, A. (1986) Regulation of ethylene biosynthesis in avocado fruit during ripening. *Plant Physiol.*, **81**, 130–135. https://doi.org/10.1104/pp.81.1.130

Small, I.D. and Peeters, N. (2000) The PPR motif – a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.*, **25**, 45–47. https://doi.org/10.1016/S0968-0004(99)01520-0

Spindel, J., Wright, M., Chen, C., Cobb, J., Gage, J., Harrington, S., Lorieux, M., Ahmadi, N. and Mccouch, S. (2013) Bridging the genotyping gap: Using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. *Theor. Appl. Genet.*, **126**, 2699–2716. https://doi.org/10.1007/s00122-013-2166-x

Stanley, L.E., Ding, B., Sun, W., Mou, F., Hill, C., Chen, S. and Yuan, Y.-W. (2017) A tetratricopeptide repeat protein regulates carotenoid biosynthesis and chromoplast development in monkeyflowers (*Mimulus*). *bioRxiv*. https://doi.org/10.1101/171249.

Strand, A., Asami, T., Alonso, J., Ecker, J.R. and Chory, J. (2003) Chloroplast to nucleus communication triggered by accumulation of Mg-protoporphyrinIX. *Nature*, **421**, 79–83. https://doi.org/10.1038/nature01204

Tadmor, Y., Larkov, O., Meir, A., Minkoff, M., Lastochkin, E., Edelstein, M., Levin, S., Wong, J., Rocheford, T. and Lewinsohn, E. (2000) Reversed-phase high performance liquid chromatographic determination of vitamin E components in maize kernels. *Phytochem. Anal.*, **11**, 370–374. https://doi.org/10.1002/(ISSN)1099-1565

Tadmor, Y., Paris, H.S., Meir, A., Schaffer, A.A. and Lewinsohn, E. (2005) Dual role of the pigmentation gene B in affecting carotenoid and vitamin E content in squash (*Cucurbita pepo*) mesocarp. *J. Agric. Food Chem.*, **53**, 9759–9763. https://doi.org/10.1021/jf0520591

Tanaka, R., Oster, U., Kruse, E., Rüdiger, W. and Grimm, B. (1999) Reduced activity of geranylgeranyl reductase leads to loss of chlorophyll and tocopherol and to partially geranylgeranylated chlorophyll in transgenic tobacco plants expressing antisense RNA for geranylgeranyl reductase. *Plant Physiol.*, **120**, 695–704. https://doi.org/10.1104/pp.120.3.695

Tapp, E.J., Cummins, I., Brassington, D. and Edwards, R. (2008) Determination and isolation of a thioesterase from passion fruit (*Passiflora edulis* Sims) that hydrolyzes volatile thioesters. *J. Agric. Food Chem.*, **56**, 6623–6630. https://doi.org/10.1021/jf800793q

Tohge, T. and Fernie, A.R. (2012) Co-expression and co-responses: Within and beyond transcription. *Front. Plant Sci.*, **3**, 248. https://doi.org/10.3389/fpls.2012.00248.

Trapnell, C., Pachter, L. and Salzberg, S.L. (2009) TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**, 1105–1111. https://doi.org/10.1093/bioinformatics/btp120

Tzfadia, O., Diels, T., De Meyer, S., Vandepoele, K., Aharoni, A. and Van De Peer, Y. (2016) CoExpNetViz: Comparative co-expression networks construction and visualization tool. *Front. Plant Sci.*, **6**, https://doi.org/10.3389/fpls.2015.01194.

Tzuri, G., Zhou, X., Chayut, N. *et al.* (2015) A 'golden' SNP in CmOr governs the fruit flesh color of melon (*Cucumis melo*). *Plant J.*, **82**, 267–279. https://doi.org/10.1111/tpj.12814

Ulitsky, I. and Shamir, R. (2009) Identifying functional modules using expression profiles and confidence-scored protein interactions. *Bioinformatics*, **25**, 1158–1164. https://doi.org/10.1093/bioinformatics/btp118

Unamba, C.I.N., Nag, A. and Sharma, R.K. (2015) Next generation sequencing technologies: The doorway to the unexplored genomics of non-model plants. *Front. Plant Sci.*, **6**, 1074. https://doi.org/10.3389/fpls.2015.01074.

Vandendriessche, T., Vermeir, S., Mayayo Martinez, C., Hendrickx, Y., Lammertyn, J., Nicolaï, B.M. and Hertog, M.L.a.T.M. (2013) Effect of ripening and inter-cultivar differences on strawberry quality. *LWT - Food Sci. Technol.*, **52**, 62–70. https://doi.org/10.1016/j.lwt.2011.12.037

Varshney, R.K., Terauchi, R. and Mccouch, S.R. (2014) Harvesting the promising fruits of genomics: Applying genome sequencing technologies to crop breeding. *PLoS Biol.*, **12**, e1001883. https://doi.org/10.1371/journal.pbio.1001883.

Wang, Y. and Lin, J. (2014) Gas chromatography–olfactometry and gas chromatography–tandem mass spectrometry analysis of fresh cantaloupe (*Cucumis melo* L. var. cantalupensis Naudin) aroma. *Flavour Frag. J.*, **29**, 87–94. https://doi.org/10.1002/ffj.3183

Wang, J., Yu, H., Weng, X., Xie, W., Xu, C., Li, X., Xiao, J. and Zhang, Q. (2014) An expression quantitative trait loci-guided co-expression analysis for constructing regulatory network using a rice recombinant inbred line population. *J. Exp. Bot.*, **65**, 1069–1079. https://doi.org/10.1093/jxb/ert464

Woodson, J.D., Perez-Ruiz, J.M., Schmitz, R.J., Ecker, J.R. and Chory, J. (2013) Sigma factor-mediated plastid retrograde signals control nuclear gene expression. *Plant J.*, **73**, 1–13. https://doi.org/10.1111/tpj.12011

Wyllie, S.G., Leach, D.N., Wang, Y. and Shewfelt, R.L. (1994) Sulfur volatiles in *Cucumis melo* cv. Makdimon (muskmelon) aroma. Sensory evaluation by gas chromatography-olfactometry. In *Sulfur Compounds in Foods* (Mussinan, C.J. and Keelan, M.E., eds). Washington DC: American Chemical Society, pp. 36–48. https://doi.org/10.1021/symposium

Xiao, M., Zhang, Y., Chen, X. *et al.* (2013) Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *J. Biotech.*, **166**, 122–134. https://doi.org/10.1016/j.jbiotec.2013.04.004

Xu, X., Zeng, L., Tao, Y. *et al.* (2013) Pinpointing genes underlying the quantitative trait loci for root-knot nematode resistance in palaeopolyploid

soybean by whole genome resequencing. *Proc. Natl Acad. Sci. USA*, **110**, 13 469–13 474. https://doi.org/10.1073/pnas.1222368110

Yahyaoui, F.E.L., Wongs-Aree, C., Latche, A., Hackett, R., Grierson, D. and Pech, J.C. (2002) Molecular and biochemical characteristics of a gene encoding an alcohol acyl-transferase involved in the generation of aroma volatile esters during melon ripening. *Eur. J. Biochem.*, **269**, 2359–2366. https://doi.org/10.1046/j.1432-1033.2002.02892.x

Zhang, L. and Kim, S. (2014) Learning gene networks under SNP perturbations using eQTL datasets. *PLoS Comput. Biol.*, **10**, e1003420. https://doi.org/10.1371/journal.pcbi.1003420.

Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K. and Liu, X. (2014) Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS ONE*, **9**, e78644. https://doi.org/10.1371/journal.pone.0078644.

Zhong, S., Joung, J.-G., Zheng, Y., Chen, Y.-R., Liu, B., Shao, Y., Xiang, J.Z., Fei, Z. and Giovannoni, J.J. (2011) High-throughput illumina strand-specific RNA sequencing library preparation. *Cold Spring Harb. Protoc.*, 2011, 940–949. pdb.prot5652. doi: 10.1101/pdb.prot5652.

