Multi-allelic *APRR2* Gene is Associated with Fruit Pigment Accumulation in Melon and Watermelon

Elad Oren[1,3], Galil Tzuri[1], Lea Vexler[1,3], Asaf Dafna[1,3], Ayala Meir[1], Uzi Saar[1], Adi Faigenboim[2], Merav Kenigswald[1,3], Vitaly Portnoy[1], Arthur A Schaffer[2], Amnon Levi[4], Edward S. Buckler[5,6], Nurit Katzir[1], Joseph Burger[1], Yaakov Tadmor[1] and Amit Gur[1]*

[1]Plant Science Institute, Agricultural Research Organization, Newe Ya'ar Research Center, P.O. Box 1021, Ramat Yishay 3009500, Israel. [2]Plant Science Institute, Agricultural Research Organization, The Volcani Center, P.O. Box 15159, Rishon LeZiyyon 7507101, Israel. [3]The Robert H. Smith Institute of Plant Sciences and Genetics in Agriculture, Faculty of Agriculture, The Hebrew University of Jerusalem, Rehovot, Israel. [4]United States Department of Agriculture-Agricultural Research Service, US Vegetable Laboratory, 2700 Savannah Highway, Charleston, SC 29414, USA. [5]Plant Breeding and Genetics Section, Cornell University, Ithaca, NY 14853, USA. [6]United States Department of Agriculture-Agricultural Research Service, Robert W. Holley Center for Agriculture and Health, Ithaca, NY 14853, USA.

*Correspondence should be addressed to Amit Gur (e-mail: amitgur@volcani.agri.gov.il, Tel: +972-52-3486760)

Abstract

Color and pigment content are important aspects of fruit quality and consumer acceptance of cucurbit crops. Here, we describe the independent mapping and cloning of a common causative *APRR2* gene regulating pigment accumulation in melon and watermelon. We initially show that the *APRR2* transcription factor is causative for the qualitative difference between dark and light green rind in both crops. Further analyses establish the link between sequence or expression level variations in the *CmAPRR2* gene and pigments content in the rind and flesh of mature melon fruits. GWAS of young fruit rind color in a panel composed of 177 diverse melon accessions did not result in any significant association, leading to an earlier assumption that multiple genes are involved in shaping the overall phenotypic variation at this trait. Through resequencing of 25 representative accessions and allelism tests between light rind accessions, we show that multiple independent SNPs in the *CmAPRR2* gene are causative for the light rind phenotype. The multi-haplotypic nature of this gene explain the lack of detection power obtained through GBS-based GWAS and confirm the pivotal role of this gene in shaping fruit color variation in melon. This study demonstrates the power of combining bi- and multi-allelic designs with deep sequencing, to resolve lack of power due to high haplotypic diversity and low allele frequencies. Due to its central role and broad effect on pigment accumulation in fruits, the *APRR2* gene is an attractive target for carotenoids bio-fortification of cucurbit crops.

Key words: *APRR2*, BSA-Seq, carotenoids, chlorophyll, melon, fruit quality, GWAS, QTL, RNA-Seq, watermelon.

Introduction

Flesh and rind pigmentation are key components affecting the nutritional value and consumer preference of the major cucurbits crops, melon and watermelon. Both crops exhibit extreme diversity in fruit traits, including size, shape, color, texture, aroma and sugar content (Burger *et al.*, 2006*b*; Wehner, 2008). Regulation of rind color in cucurbits initiates early in fruit development and is expressed as green color intensity at the young fruit stage, reflecting chlorophyll concentrations (Tadmor *et al.*, 2010). Most watermelons remain green at maturity with chlorophyll being their main rind pigment, and therefore rind color variation in watermelon is mostly expressed as green pigment intensity in uniform or striped patterns (Gusmini and Wehner, 2005). Conversely, melon rind color transforms during development, leading to extensive variation in mature fruit pigment profiles that include different combinations of carotenoids, flavonoids and chlorophylls (Burger *et al.*, 2010; Tadmor *et al.*, 2010). The genetic basis of this variation is only partly resolved. Several external fruit color QTLs have been mapped in populations derived from a cross between *Piel de Sapo* line and *PI16375* (Monforte *et al.*, 2004). It has been previously reported that the mature yellow rind color of yellow casaba melon accessions (*C. melo*, var inodorous) is caused by the accumulation of naringenin chalcone, a yellow flavonoid pigment (Tadmor *et al.*, 2010). A Kelch domain-containing F-box protein coding gene (*CmKFB*) on chromosome 10 was identified as causative for the naringenin chalcone accumulation in melon fruit rind (Feder *et al.*, 2015). While it is logical to assume that young and mature fruit color intensity are correlated, and that common genetic factors may be involved, thus far, such genes have not been reported in melon.

Three major flesh color categories are defined in melon: green, white and orange, with β-carotene and chlorophyll being the predominant pigments of the orange and green phenotypes, respectively (Burger *et al.*, 2010). The major locus qualitatively differentiating between orange and non-orange flesh is green flesh (*gf)*, located on chromosome 9 (Cuevas *et al.*, 2009). *gf* was recently shown to be the *CmOr* gene, which governs carotenoids accumulation and orange flesh color (Tzuri *et al.*, 2015). A second qualitative flesh color locus, white flesh *(wf)*, which is associated with the difference between white and green flesh, has been previously described and mapped to chromosome 8 (Clayberg, 1992; Monforte *et al.*, 2004; Cuevas *et al.*, 2009). Another layer of quantitative variation in flesh pigment content and color intensity exists within

81   those color classes as defined by several QTL mapping studies (Monforte *et al.*, 2004;

82   Cuevas *et al.*, 2008, 2009; Paris *et al.*, 2008; Harel-Beja *et al.*, 2010; Diaz *et al.*, 2011).

83   These include the recent fine-mapping to a candidate causative gene level of flesh

84   carotenoids QTL using a recombinant inbred lines (RILs) population (Galpaz *et al.*,

85   2018). Thus far, however, causative genes governing this quantitative variation have

86   not been shown.

87   In recent years, a few transcription factors involved in regulation and synchronization

88   of chlorophyll and carotenoids accumulation were identified in plants. Among these,

89   Golden2-like (GLK2) transcription factors, which regulate chloroplast development

90   (Chen *et al.*, 2016). Allelic or expression variation in the GLK2 gene were shown to be

91   associated with levels of chlorophyll and carotenoids in Arabidopsis, tomato and pepper

92   (Waters *et al.*, 2008, 2009; powell *et al.*, 2012; Brand *et al.*, 2014). A related but distinct

93   transcription factor, the *ARABIDOPSIS PSEUDO-RESPONSE REGULATOR2-LIKE*

94   gene (*APRR2*) with phenotypic effects comparable to those of GLK2, was identified

95   and shown to regulate pigment accumulation in tomato and pepper (Pan *et al.*, 2013)

96   and over expression of the *APRR2* gene in tomato increased the number of plastids and

97   the color intensity. Recently, the *APRR2* gene was also shown to be causative of the

98   white immature rind color (*w*) in cucumber (*Cucumis sativus*) (Liu *et al.*, 2016), a close

99   relative of melon (*Cucumis melo*). The white rind phenotype of immature cucumbers

100   in this study was shown to be associated with reduced chloroplast number and

101   chlorophyll content.

102   In the current study, we used bi-parental populations to map and identify the *APRR2*

103   gene as a common causative regulator of pigment accumulation in both melon and

104   watermelon. We show that the effect of this transcription factor on pigment

105   accumulation is initially observed in the rind of young fruits (chlorophylls) and extends

106   to rind and flesh of mature melon fruits (chlorophylls and carotenoids). Through further

107   analysis of wider genetic variation in melon, we revealed a unique multi-allelic pattern

108   that inhibited our ability to detect a significant signal through GBS-based GWAS. By

109   zooming in on this allelic series, we confirmed the central role of this gene in shaping

110   the color variation of young fruit rind across melon diversity.

111

112   **Materials and methods**

Plant materials and field trials

The germplasm used in this study included four sets: (1) TAD×DUL RILs (F_7) – bi-parental segregating population derived from the cross of the dark rind line, 'Dulce' (DUL; *C. melo* var. *reticulatus*) with the light rind line, 'Tam Dew' (TAD; *C. melo* var. *inodorous*) (Tzuri *et al.*, 2015). One hundred and sixty-four F_7 recombinant inbred lines were developed through single-seed-descent. All RILs, F_1 and the parental lines were grown in a randomized block design (RCBD) in an open field at Newe Ya'ar Research Center, in the spring–summer seasons of 2016 and 2017. Each line was represented by two replicates of five plants per plot. (2) NA×DUL F3 and $F_{3:4}$ – An F3 population from this cross was grown in two repetitions in a greenhouse at Beit Elazari, Israel in 2013 as previously described (Ríos *et al.*, 2017). This population is derived from the cross between the light rind line 'Noy-Amid' (NA; *C. melo* var. *inodorous*) and the common dark rind parent, DUL. One hundred and fourteen $F_{3:4}$ families from F_3 genotyped plants alongside the parental lines and their F_1 were grown in RCBD in an open-field trial at Newe Ya'ar Research Center in the spring–summer season of 2017 in two replicates of six plants per plot. (3) Melo180 GWAS panel - a Newe-Ya'ar melon collection used in this study comprised 177 diverse accessions that represent the two melon subspecies (ssp. *agrestis* and ssp. *melo*) and 11 taxonomic groups. Each accession was represented by three plots of five plants each in a randomized block design (RCBD) in the open field at Newe-Ya'ar in summer 2015 (Gur *et al.*, 2017). (4) NY0016 x EMB $F_{2:3}$ - for mapping the light rind trait in watermelon, the light rind inbred accession NY0016 was crossed with the canary yellow accession Early Moon Beam (EMB) to produce 87 $F_{2:3}$ families (Branham *et al.*, 2017). During the summer of 2016 and 2017, ten plants per $F_{2:3}$ family and two plots of ten plants from the parents and F_1 were sown in the open field at Newe-Ya'ar. All the populations used in this study were grown under standard horticultural conditions open fields at Newe Ya'ar Research Center, northern Israel (32°43′05.4″N 35°10′47.7″E), soil type was grumusol, and the plants were drip-irrigated and drip-fertilized.

Fruit color phenotyping

In the melon populations, fruit images were taken with a digital camera on developing fruits of all accessions throughout the season, from anthesis to harvest. Rind color of young fruits was scored in the field at 10 to 15 days after anthesis and confirmed based on fruit images from the same developmental stage. Mature rind and flesh color were

5

147   measured on ripe fruits, which were harvested based on abscission in climacteric fruits,

148   or days after anthesis, rind color and TSS in non-climacteric fruits. Five mature fruits

149   per plot were photographed externally, then cut along the longitudinal section and

150   scanned for internal imaging, using a standard document scanner (Canon, Lide120) as

151   described previously (Gur *et al.*, 2017). Scanned images were analyzed using the

152   Tomato Analyzer software (Rodríguez *et al.*, 2010) for color (L, A, B, Chroma and

153   Hue) and morphological features. Rind and flesh tissues were sampled into 50 ml tubes

154   from at least three fruits per plot, immediately frozen in liquid nitrogen and then stored

155   at -80^{0}C for further analyses. For the watermelon mapping experiment (NY0016×EMB

156    $F_{2:3}$), ten F_3 individuals per $F_{2:3}$ family were harvested at maturity (~70 days post

157   sowing), imaged and phenotyped for rind color as above.

158

159   *Carotenoids and chlorophyll quantification*

160   Carotenoids were extracted from 0.5 mg ground tissue samples in a mixture of

161   hexane:acetone:ethanol (2:1:1, v/v/v) as described previously (Tadmor *et al.*, 2005),

162   and separated using a Waters 2695 HPLC apparatus equipped with a Waters 996 PDA

163   detector (Milford, MA). Carotenoids were identified by their characteristic absorption

164   spectra, distinctive retention time and comparison to authentic standards. Quantification

165   was performed by integrating the peak areas with standard curves of authentic standards

166   with the Waters millennium chromatography software. Lutein and β-carotene were

167   relatively quantified at 450 nm and 270 nm respectively, by integrating their peak areas

168   and calculating their percentage from total integrated peak areas. Tissues for

169   chlorophyll determination were sampled as explained for carotenoid analysis.

170   Chlorophyll extraction was performed in dimmed light to avoid possible

171   photodegradation of chlorophyll. Chlorophyll was extracted by adding 5 mL of

172   dimethyl sulfoxide (DMSO) to 0.5 g, vortexing and incubating in the dark at room

173   temperature for 24 h. The extract was analyzed for absorbance in the wavelengths of

174   663 and 645nm using a Cary50Bio spectrophotometer (Varian). Chlorophyll

175   concentration was calculated as described by Tadmor *et al*., (2010).

176

177   *Genotyping*

178   DNA isolations were performed using the GenElute^{TM} Plant Genomic Miniprep Kit

179   (Sigma-Aldrich, St. Louis, MO). DNA quality and quantification were determined

180   using a Nanodrop ND-1000 (Nanodrop Technologies, Wilmington, DE)

181 Spectrophotometer, electrophoresis on agarose gel (1.0%) and Qubit® dsDNA BR

182 Assay Kit (Life Technologies, Eugene, OR).

183 *GBS analysis, SNP calling and map construction*:

184 (1) TAD×DUL RILs – DNA from 164 F_7 individuals was processed by Novogene

185 (Novogene Bioinformatics Institute, Beijing, China) for GBS analysis. 0.3~0.6 μg of

186 genomic DNA from each sample was digested with *ApeKI* restriction enzyme, based on

187 the *in silico* evaluation results, and the obtained fragments were ligated with two

188 barcoded adapters at each end of the digested fragment. Followed by several rounds of

189 PCR amplification, all the samples were pooled and size-selected for the required

190 fragments to complete the library construction. Samples where diluted to 1 ng/μl and

191 the insert size was assessed using the Agilent® 2100 bioanalyzer; qPCR was

192 performed to detect the effective concentration of each library. Libraries with a

193 concentration higher than 2 nM were sequenced on an Illumina HiSeq 2000/2500

194 platform as 144 bp, paired-end reads and mapped to the *C. melo* reference genome

195 DHL92 v3.5.1 (Garcia-Mas *et al.*, 2012; available at

196 https://melonomics.net/fles/Genome/Melon_genome_v3.5.1/). Over 570 million

197 reads were produced covering nearly 21% of the genome across more than 35 million

198 tags at an average read depth of 9 reads per site. SNP calling was carried out using

199 Broad Institute's genome analysis toolkit (GATK) (McKenna *et al.*, 2010) resulting

200 in 1,205,528 raw SNPs. Sites with a depth of less than three reads per site or more

201 than 50 percent missing data were filtered out using TASSEL v5.2.43 (Bradbury *et*

202 *al.*, 2007). Data was then imputed using full-sib families LD algorithm (Swarts *et al.*,

203 2014) followed by the removal of individuals with excess heterozygosity. The

204 genotypic dataset was phased to ABH format consisting of 89,343 SNPs across 146

205 lines. Binning was performed using SNPbinner (Gonda *et al.*, 2018) with a minimum

206 ratio between crosspoints set at 0.001 and minimum bin size of 1000 bp. Bin statistics

207 and genetic distance were calculated using in-house script developed in python, based

208 on the Kosambi mapping function (Kosambi, 1943). The final set included 2,853

209 recombination bins across 146 lines. Evaluation of genotypic data quality was done

210 by accurately mapping flesh color to a 55Kb interval spanning the previously

211 published *CmOr* gene (Melo3C05449) (Tzuri *et al.*, 2015).

212 (2) NA×DUL $F_{3:4}$ – DNA from 140 F_3 individuals was processed NRGene LTD (Nes

213 Ziyyona, Israel) for Restriction-site-Associated DNA sequencing (RAD-seq) (Ríos *et*

7

214   *al.*, 2017), SNP calling was performed following similar methods and the initial

215   marker set included 43,975 SNPs across 140 individuals with an average depth of 16

216   reads per site. Further filtering and imputation were performed as described above and

217   the final set for binning was composed of 19,015 SNPs across 134 individuals.

218   Binning and genetic map construction were carried out using the same parameters as

219   those used for the TAD×DUL population, yielding 1,321 bins across 134 individuals.

220   (3) GWAS180 - Genotyping of this diversity panel was performed using GBS, as

221   described by Gur *et al.*, (2017). The final SNP set included 23,931 informative SNPs

222   (at MAF>5%) across 177 accessions.

223   (4) The watermelon mapping population, NY0016×EMB, was genotyped by GBS.

224   Library construction, sequencing, and SNP calling were performed at the Genomic

225   Diversity Facility at Cornell University (Ithaca, NY) as described by Branham *et al.*,

226   (2017). Sequences in this project were aligned to the Charleston Gray genome, version

227   1 (available at ftp://www.icugi.org/pub/genome/watermelon/WCG/v1/).

228   *Bulk Segregant Analysis by sequencing (BSA-Seq) of the watermelon population*

229   DNA samples from 35 F_2 plants (from NY0016×EMB cross) homozygote for the rind

230   color trait (based on F_3 family's phenotypes) were prepared into two bulks (light rind: 19

231   F_2 samples and dark rind: 16 F_2 samples). These samples, as well as DNA samples of the

232   parental lines (EMB and NY0016) were used for whole-genome resequencing performed

233   at the DNA Services Center at the University of Illinois, Urbana-Champaign. Six shotgun

234   genomic libraries were prepared with the Hyper Library construction kit from Kapa

235   Biosystems (Roche) with no PCR amplification. The libraries were quantitated by qPCR

236   and sequenced on one lane for 151 cycles from each end of the fragments on a HiSeq

237   4000 using a HiSeq 4000 sequencing kit version1. Fastq files were generated and

238   demultiplexed with the bcl2fastq v2.17.1.14 Conversion Software (Illumina). Average

239   output per library was 44 million reads of 150 bp. All raw reads were mapped to the

240   Charleston Gray reference genome using the Burrows-Wheeler Aligner (BWA),

241   producing analysis ready BAM files for variant discovery with Broad Institute's Genome

242   Analysis Toolkit (GATK). Homozygous SNPs between the two parents were extracted

243   from the vcf file that was further filtered to total depth>20 reads per site. The read depth

244   information for the homozygous SNPs in the 'light' and 'dark' pools was obtained to

245   calculate the SNP-index (Takagi *et al.*, 2013). For each site we then calculated for each

246   bulk the ratio of the number of 'reference' reads to the total number of reads, which

247   represented the SNP index of that site. The difference between the SNP-index of two

248 pools was calculated as ΔSNP-index. Sliding window method was used to perform the

249 whole-genome scan and identify the trait locus confidence interval on chr9.

250 *Whole Genome re-Sequencing (WGS) of 25 representative diverse melon accessions*

251 DNA of the 25 core accessions was shipped to the Genomic Diversity Facility at Cornell

252 University (Ithaca, NY) for whole genome resequencing to an estimated 30X depth.

253 *Validation of rare alleles at the APRR2 genes* – the causative variants at the different

254 *APRR2* alleles in melon and watermelon, which were discovered based on NGS of

255 genomic DNA, were confirmed on the parental lines and relevant segregants through

256 Sanger sequencing of genomic DNA in melon and cDNA from mRNA that was extracted

257 from fruits in watermelon.

258

259 *qRT-PCR analysis*

260 Rind samples were peeled from fruits harvested throughout development, from

261 anthesis to maturity, and immediately frozen in liquid nitrogen. Three fruits were

262 sampled from each genotype at each developmental stage. 100-150 mg of frozen rind

263 tissue per sample was used for RNA extraction using a Plant/Fungi Total RNA

264 Purification Kit (NORGEN Biotek Corp., Canada). First-strand cDNA was

265 synthesized using a cDNA Reverse Transcription Kit (Applied Biosystems, USA).

266 The 10-μl qPCR volume included 1 μl of cDNA template, 0.2 μl of each primer (10

267 μM), 5 μl of Fast SYBR Green Master Mix (Applied Biosystems, USA), and RNase-

268 free water to a final volume of 10 μl. qRT-PCR, with an annealing temperature of

269 60°C, was performed in triplicate on a 96-well plate in the Step-One Plus Real-Time

270 PCR system (Applied Biosystems, USA). The melon *Cyclophilin A* gene

271 (Melo3C013375) was used as a control to normalize the qRT-PCR values across

272 different samples. Primers are listed in **Supplemental Table 5**.

273

274 *Data analysis*

275 *Trait mapping*

276 Genome-wide linkage analysis for young fruit rind color was performed in TASSEL

277 using a generalized linear model (GLM) in the bi-parental populations and confirmed

278 using single-marker analysis in the JMP V13.1 software package (SAS institute, Cary,

279 NC, USA). GWAS at the melon diversity collection was performed by a mixed linear

280 model (MLM) analysis in TASSEL, using both the population structure (Q matrix) and

281 relatedness (kinship (k) matrix) as covariates to control for population structure. Multiple

9

282   comparisons correction to significance thresholds were performed using the FDR

283   approach (Benjamini and Hochberg, 1995). All further statistical analyses (correlations

284   and analyses of variance) were performed using the JMP V13.1 software package.

285   *Population structure, Kinship and LD analysis*

286   Relatedness between the melon accessions in the diverse collection was estimated in

287   TASSEL software v5.2.43 using the pairwise kinship matrix (k matrix) through the

288   Centered IBS method. Linkage disequilibrium (LD) between intra-chromosomal pairs of

289   sites was done on chromosome 4 using the full matrix option in TASSEL.

290   *Sequence analyses*

291   Sequence alignments and comparison of *APRR2* alleles were performed using the BioEdit

292   software package (Hall, 1999) and the integrative genomics viewer (IGV) package

293   (Robinson *et al.*, 2011). Comparative analysis of haplotype diversity across 2,200 genes

294   on melon chromosome 4 was performed following these steps: 1) A VCF file containing

295   ~4,000,000 high quality SNPs across the core set of 25 melon lines (MAF>0.1 and less

296   than 10% missing data per SNP) was created based on alignments to the melon genome

297   version 3.5.1. 2) The corresponding gene annotations file was used to create a subset of

298   exonic SNPs on all annotated genes on chromosome 4. 3) The number of exonic-SNPs

299   haplotypes per gene was calculated.

300

Results

302   *GWAS of young fruit rind color in melon*

303   Most melons can be visually classified into two distinct young fruit (~10 days post

304   anthesis) rind colors; dark or light green, reflecting qualitative variation in chlorophyll

305   content. Light immature rind color was previously reported to display a recessive

306   single-gene inheritance in a bi-parental segregating population (Burger *et al.*, 2006*a*).

307   In the current study, young fruit rind color was visually scored on a previously

308   described diverse melon collection composed of 177 accessions (Gur *et al.*, 2017). The

309   collection was genotyped genome-wide with 23,931 informative, GBS-derived SNP

310   markers, and was shown to be an effective resource for mapping simple traits in melon

311   (Gur *et al.*, 2017). Here, we used a subset composed of 120 accessions with a clearly

312   defined dark or light rind phenotype (Example in **Figure 1a)** for genome-wide

313   association analysis. Accessions with prominent non-uniform rind color (stripes or

10

314   dots) were excluded from this analysis. We also excluded *Charentais* lines, as their

315   dominant grayish light rind is exceptional and phenotypically distinct from the common

316   light rind in other melon types. While the dark and light phenotypes were distributed

317   uniformly across the genetic variation and were represented in balanced proportions

318   across this set (41% and 59%, respectively, **Figure 1b**), a genome-wide population-

319   structure-corrected analysis did not result in any significant marker-trait association.

320   This result has led to the assumption that while this highly heritable trait may show

321   simple inheritance in a specific bi-allelic cross, it is possibly more complex and

322   explained by multiple loci across a multi-allelic diverse collection.

323

324   *Mapping and cloning of the young fruit light rind gene in melon*

325   In order to further dissect this trait using a simpler genetic design, we analyzed two

326   segregating bi-parental populations: the first is composed of 164 RILs (F_7) from a cross

327   between a light rind honeydew parent (Tam-Dew; TAD) and a dark rind *reticulatus*

328   parent (Dulce; DUL, **Figure1b**). The second population is composed of 114 $F_{3:4}$

329   families derived from a cross of DUL with another light rind accession, a yellow casaba

330   *inodorous* melon (Noy-Amid; NA, **Figure 1b**). These segregating populations were

331   visually phenotyped for young fruit rind color over two seasons and a consistent single

332   gene (Mendelian) ratio was observed in dark:light phenotypes. The populations were

333   then genotyped through GBS and 89,343 (TAD×DUL RILs) and 43,975 (NA×DUL

334    $F_{3:4}$) informative SNP markers were identified and used for mapping. Whole-genome

335   linkage analysis using the four datasets (two populations over two growing seasons)

336   resulted in the identification of a single highly significant consistent trait locus on

337   chromosome 4 (**Figure 1c**). The common confidence interval for this trait locus spans

338   a 290 Kb region (Chr4: 640-930 Kb, **Figure 1d**) on the melon reference genome

339   (Garcia-Mas *et al.*, 2012; http://cucurbitgenomics.org/organism/3), as confirmed also

340   through substitution mapping using recombinants within this interval in the TAD×DUL

341   RILs population (**Figure 2a**). Annotation of the genomic sequence at this interval

342   revealed 33 putative genes (**Sup. Table 1**) including a strong candidate,

343   Melo3C003375, which is annotated as an *ARABIDOPSIS PSEUDO-RESPONSE*

344   *REGULATOR2-LIKE* (*APRR2)* gene, the melon homolog of a recently reported

345   causative gene of the recessive white rind (*w*) mutation in cucumber (Liu *et al.*, 2016).

11

346   We then compared the *CmAPRR2* gene (Melo3C003375) sequence between the

347   parental lines of the mapping populations. Genomic and mRNA sequencing revealed

348   multiple polymorphisms, including two different exonic polymorphisms causing

349   independent stop codons in each of the light rind parents compared to the common dark

350   parent (DUL). G to T substitution in exon 8 in TAD compared to DUL lead to a

351   premature stop-codon and a predicted aberrant protein of 292 amino-acids (AA)

352   compared to the normal 527 AA protein of DUL (**Figure 2b-c, Sup Figure 1**). A 13-

353   bp insertion in exon 9 of NA result in a frame-shift leading to a different premature

354   stop-codon in this line and a predicted protein of 430 AA (**Figure 2c, Sup Figure 1**).

355   Furthermore, we crossed TAD and NA with each other and with DUL (as a reference

356   testcross) and phenotyped the F_1s for young fruit rind color. Both testcrosses with DUL

357   resulted, as expected, in a dark rind in the F_1. However, the F_1 of TAD×NA had a clear

358   light rind and confirmed the allelic nature of these recessive phenotypes (**Figure 3a**).

359   This further corroborates that these independent predicted causative mutations in the

360   *CmAPRR2* gene are indeed allelic.

361

362   *Expression pattern of the CmAPRR2 gene in melon fruit*

363   Fruits from the light (TAD) and dark (DUL) parental lines were sampled during

364   development from anthesis to maturity and mRNA levels of the *CmAPRR2* gene were

365   analyzed by qRT-PCR. We show here that *CmAPRR2* has higher expression level in

366   fruit comparted to leaves (**Figure 2d**), as shown also in cucumber (Liu *et al.*, 2016) and

367   pepper (Brand *et al.*, 2014), and in agreement with the Melonet-DB gene expression

368   atlas (Yano *et al.*, 2018). In accordance with these studies, we also show that the

369   *CmAPRR2* peak expression in fruit rind occur around 15 days post anthesis (DPA),

370   before the initiation of ripening and color change. Comparison between the parental

371   lines of the mapping population also showed significantly lower levels of *CmAPRR2*

372   expression in light rind fruits throughout fruit development. Both light and dark lines

373   have reduced *CmAPRR2* expression levels at the mature fruit stage and were not

374   significantly different from each other at that stage (**Figure 2d**).

375

376   *Multiple independent causative mutations in the CmAPRR2 gene across melon diversity*

377   In light of the conflict between the clear identification of a single causative gene through

378   linkage mapping in two different bi-parental crosses on one-hand, and the absence of

379   any significant genome-wide signal from the GWAS analysis on the other hand, we re-

380   sequenced and compared the genomic sequence of the *CmAPRR2* gene across a core

381   panel of 25 diverse melon lines. This core panel was selected to represent the different

382   groups and overall diversity in our collection as described previously (**Figure 3b**) (Gur

383   *et al.*, 2017). Nineteen lines from the panel, showing a clear dark or light young fruit

384   rind phenotype, were used for the sequence comparison. Seventeen SNPs and InDels

385   within exons in the *CmAPRR2* gene were identified across this panel (**Figure 2b**). Eight

386   of these SNPs were either synonymous or did not show a distinct allelic state between

387   dark and light accessions (in gray). The remaining nine polymorphisms are

388   independently inherited (not in LD with each other) and display low frequency alleles

389   (0.05-0.15) which are unique to the light rind accessions (**Figure 2e**). Four of these

390   polymorphisms (2, 6, 13 and 14) are SNPs that change a single amino acid (in black).

391   Three are InDels (4, 7 and 12) that cause frame-shifts leading to major modification in

392   predicted protein sequence (in red), and the remaining two (9 and 11) are the causative

393   polymorphisms described above, leading to premature stop codons as in TAD and NA.

394   These five major polymorphisms (4, 7, 9, 11 and 12) explain the light rind phenotype

395   in eight of the eleven light rind accessions in the core panel. The non-synonymous

396   polymorphisms 2, 6 and 13 are potentially causative of the light rind phenotype in

397   BAHC and QME. The only light rind accession that could not be explained by non-

398   synonymous variation within the *CmAPRR2* coding sequence is SAS. However, the low

399   mRNA expression of *CmAPRR2* in young fruit (5 DPA) rinds of SAS, which was

400   similar to the expression level in TAD, and significantly lower compared to DUL

401   (**Figure 2d**), suggest expression level variation as a possible causative element for the

402   light rind phenotype in this line. To test whether all these 'light' accessions are indeed

403   allelic and caused by different mutations in the *CmAPRR2* gene, we performed allelism

404   tests where all the 'light' accessions (n=11) were intercrossed and the resulting 55 F_1s

405   were phenotypically evaluated for young fruit rind color. As a reference, these 'light'

406   accessions were crossed with two 'dark' testers (DUL and Ananas Yoqne'am; AY).

407   **Figure 3c** shows that all 55 'light'×'light' F_1 hybrids displayed light immature fruit rinds,

408   while all 22 'light'×'dark' testcrosses displayed dark rinds. These results confirm the

409   allelism between the 11 'light' lines, including the light rind phenotype of SAS.

13

410 Combined interpretation of the sequence variation and allelism tests across this

411 representative core panel indicate that most of the young-fruit rind color variation can

412 be explained by multiple independent polymorphisms related to the *CmAPRR2* gene.

413 Extrapolation of the results from this core set suggests that each of these causative

414 variants is most likely also present at low allele frequency, across the wider diversity

415 panel, leading to the non-significant associations observed in our GBS-based GWAS

416 experiment. It is worth noting that these independent mutations are not in LD with each

417 other, leading to the high haplotype diversity in this gene. In a comparative analysis of

418 haplotype diversity based on exonic SNPs across 2,200 genes on melon chromosome

419 4, we found that *CmAPRR2* is indeed the second most diverse gene, irrespective of

420 number of SNPs and transcript length (**Sup Figure 2** and methods). Analysis of the LD

421 pattern in the genomic region surrounding the *CmAPRR2* locus, confirmed the low LD

422 between SNPs in this region (**Sup Figure 3**). The fact that these allelic polymorphisms

423 are not in LD with each other allow us aggregate them into a theoretical unified

424 functional polymorphism, resulting in increased frequency of aberrant *CmAPRR2*

425 allele (0.52, **Figure 2e** right column). This analysis, in turn, produces a significant

426 association between the *CmAPRR2* gene and the light rind trait (**Figure 2f**).

427

428 *Mapping and cloning of the light rind color gene in watermelon*

429 In parallel with melon, we also studied the genetics of rind color in watermelon. The

430 main difference is that in watermelon, due to its non-climacteric fruit ripening,

431 chlorophylls are the main rind pigments also during fruit maturity and therefore light

432 and dark green were visually scored on mature fruits. Our light rind source in this study

433 was an heirloom accession named NY0016 (Tadmor *et al.*, 2005) that was crossed on

434 different lines, varying in their rind color, to produce F_1s and F_2s. All the F_1 hybrids

435 had dark rind fruits (irrespective of the stripes pattern in the 'dark' parents), proving the

436 recessive nature of the light rind phenotype of NY0016. All F_2 populations were

437 phenotyped for rind color and a consistent 3:1 Mendelian ratio was observed for dark

438 and light rinds, respectively (**Figure 4a**, **Sup Table 2**). The cross between the light rind

439 accession (NY0016) and a striped (dark) parent (Early Moon Beam; EMB) was selected

440 for linkage mapping and this population was advanced to F_3 to perform $F_{2:3}$ analysis

441 (**Figure 4b**). GBS of the F_2 population (N= 87) resulted in a final high-quality set of

14

442   3,160 filtered SNPs (Branham *et al.*, 2017) that was used for genetic mapping of the

443   light rind phenotype. Seven to ten fruits per F_3 family were visually scored and each

444   family was classified into a defined category: fixed for light rind, fixed for dark rind or

445   segregating for rind color (**Sup. Figure 4**). The observed 1:2:1 frequencies of light,

446   segregating and dark across the F_3 families supported a single gene inheritance for this

447   trait (**Figure 4b**). Whole-genome linkage analysis resulted in the identification of a

448   single significant trait locus on chromosome 9 (R^2=0.62, P=2.9x10^{-18}, **Figure 4c**). The

449   confidence interval for this locus spanned 1.7 Mb with 80 predicted genes on the

450   watermelon   reference   genome   (Charleston   Gray,

451   http://cucurbitgenomics.org/organism/4). In order to narrow down this genomic

452   interval, we performed mapping-by-sequencing of DNA bulks (BSA-seq) from 35

453   selected F_2 segregants that were homozygous for light or dark rind (on the basis of fixed

454   F_3 phenotypes). The ~30X whole-genome re-sequencing resulted in the identification

455   of 400,000 high quality SNPs differentiating between parental lines. Comparison of

456   allele frequencies between the light and dark bulks across the 400,000 SNPs (ΔSNP-

457   index analysis) confirmed the trait locus on chromosome 9 and allowed us to narrow

458   the genomic confidence interval to a 900Kb region with 30 predicted genes (**Figure 4d,**

459   **Sup Table 3**). Review of the list of genes within the confidence interval revealed a

460   strong candidate, CICG09G012330, the watermelon homolog *ClAPRR2* gene, highly

461   similar to the causative melon (Melo3C003375) and cucumber (Csa3G904140.3 (Liu

462   *et al.*, 2016)) genes. Comparison of the genomic sequence of CICG09G012330 between

463   the mapping population parental lines revealed several SNPs, none of them within

464   exons. The only putative causative SNP at that point was at intron 6/exon 7 junction

465   (**Figure 4e**). Parents and segregants mRNA sequence comparison revealed an

466   alternative splicing in the intron6/exon7 junction, leading to a 16 bp deletion at the

467   mRNA of the light rind parent and corresponding segregants carrying the 'light' allele.

468   This 16 bp deletion, which created a frame shift leading to a premature stop codon and

469   a predicted aberrant protein, is most likely causative for the light rind phenotype of

470   NY0016 (**Figure 4f**, **g**).

471

472   *Allelic variation in the CmAPRR2 gene is associated with mature fruit rind and flesh*

473   *pigmentation in melon*

15

474   Earlier analyses of rinds from TAD, DUL and selected F_3 families from their cross,

475   demonstrated the correlation between young fruit chlorophyll content and mature fruit

476   carotenoids (**Sup. Figure 5**). To test whether the *CmAPRR2* gene is also associated

477   with mature fruit rind pigmentation in melon, we analyzed mature fruits from the

478   TAD×DUL RILs population for rind color and carotenoids content. We harvested 10

479   mature fruits per line, external images of the fruits were taken for color scoring and the

480   rinds were sampled for carotenoids profiling. Color intensity variation at the young

481   (green) stage in this population is qualitative and could be visually classified into two

482   distinct classes (dark or light green) on a single-fruit basis. However, at the mature stage

483   any effect of the *CmAPRR2* gene on rind color is visually of a quantitative nature

484   (**Figure 5a**), as shown also in tomato (Pan *et al.*, 2013). Rind netting, which segregates

485   in this population, further masked rind color and complicated visual scoring and

486   carotenoids quantification. For the parental lines, TAD, with the light green rind at

487   young fruit stage, has a cream-yellowish rind at maturity and DUL, with the dark green

488   rind at young stage, has an orange rind masked by dense rind netting (**Figure 2d**).

489   Visual observation of standardized external images of selected mildly-netted mature

490   fruits, representing both alleles in the *CmAPRR2* gene, suggested a possible effect of

491   this gene on rind color intensity, such that on average the 'dark' allele is associated with

492   deeper orange color (**Figure 5a**). We confirmed this effect through analysis of

493   carotenoids content in rinds of 50 selected mildly netted segregants (25 RILs carrying

494   each allele in the *CmAPRR2* gene). The 'light' allele was significantly associated with

495   more than 10 fold reduction in total carotenoids in the rind, which is consistent with the

496   reduced chlorophyll levels observed in young fruits of this group (**Figure 5b, Sup**

497   **Figure 5**). Allelic variation in this gene is explaining 37% of the variation in lutein

498   content ($P=2\times10^{-6}$), 34% of the variation in β-carotene content ($P=8\times10^{-6}$), and 37% of

499   total carotenoids ($P=3\times10^{-6}$) in fruit rind in this population.

500   To examine whether the effect of the *CmAPRR2* gene extends also to mature fruit flesh

501   color, we phenotyped the TAD×DUL RILs for flesh color intensity using longitudinal

502   fruit section scanning and quantitative image analysis (n=145 lines x 10 fruits per line).

503   This population is segregating for the main flesh color gene in melon, *CmOr*,

504   discriminating between orange and non-orange flesh (Tzuri *et al.*, 2015) and it is

505   therefore composed of orange flesh lines (48%) and green flesh lines (49%), the

506   remaining 3% of the lines are segregating due to residual heterozygosity (**Figure 5c**).

507　In order to test the association of the *CmAPRR2* gene with flesh color, we analyzed the

508　orange and green fruits separately. A significant association between the *CmAPRR2*

509　allelic segregation and color intensity was found in the orange-flesh group, which

510　accumulated β-carotene as the main flesh pigment ($R^2=0.25$, $P=4.7 \times 10^{-5}$, **Figure 5d-e**)

511　and as expected, the 'dark' allele was associated with stronger pigmentation and higher

512　predicted β-carotene content ($R^2=0.62$, $P=0.0053$, see Materials and methods and **Sup.**

513　**Figure 6**). Significant association of this gene with flesh color was found also in the

514　green flesh group, which accumulated chlorophyll as the main flesh pigment ($R^2=0.31$,

515　$P=2 \times 10^{-6}$) as the 'dark' allele was significantly associated with higher green Chroma

516　(**Figure 5f-g**) reflecting higher chlorophyll content.

517

518　*Expression level of the CmAPRR2 gene is associated with mature flesh color in melon*

519　In order to test whether the expression level of the *CmAPRR2* gene is associated with

520　mature fruit pigmentation, we analyzed RNA-seq data and mature fruit flesh

521　carotenoids on a different RILs population derived from a cross between DUL and an

522　Indian phut snapmelon (Momordica group), PI414723 (hereafter called 414) (Galpaz *et*

523　*al.*, 2018). While 414 has a spotted rind (and not a clear light or dark phenotype), we

524　assume based on testcrosses with some of the light rind accessions that, as DUL, it also

525　carry a 'dark' allele of the *CmAPRR2*. This assumption is supported by the fact that it

526　does not show any of the predicted 'light' non-synonymous polymorphisms found at the

527　*CmAPRR2* gene (**Figure 2 b, e**). DUL and 414 are genetically and phenotypically

528　distant and differ in their mature fruit flesh color and carotenoids content (**Figure 6a**).

529　DUL has dark orange flesh while 414 has light (salmon-colored) orange flesh, and

530　accordingly the RILs population segregates for these traits (Harel-Beja *et al.*, 2010;

531　Galpaz *et al.*, 2018). RNA-Seq was previously performed on mature fruit flesh of 96

532　RILs from this cross (Freilich *et al.*, 2015) and now allowed us to execute a genome-

533　wide eQTL analysis for the *CmAPRR2* gene (Melo3C003375). A single, highly

534　significant, cis-eQTL was mapped to chromosome 4 and defined by 270 Kb interval

535　flanking this locus (Galpaz *et al.*, 2018**, Figure 6b-c**). This result confirmed the

536　heritable variation in *CmAPRR2* expression level in this population and that the

537　expression of this gene is mostly regulated by cis-acting sequence variants. We then

538　tested the correlation between *CmAPRR2* expression level and flesh β-carotene content

539　across the 96 RILs and found a significant positive correlation ($R=0.38$, $P=0.0008$,

17

540   **Figure 6d**). Since both parental lines of this population carry the 'dark' allele based on

541   the coding sequence of the *CmAPRR2* gene, this result provides a quantitative support

542   for the possible relationship between expression level of this gene and pigment

543   accumulation in melon. It is important to note that this population is segregating at

544   additional QTLs that affect carotenoids content in mature fruit flesh, including a major

545   QTL on chromosome 8, which was recently mapped to a candidate gene level (Diaz *et*

546   *al.*, 2011; Galpaz *et al.*, 2018). This variation further masked the specific effect of

547   *CmAPRR2* expression level on flesh carotenoids content. We also assume that the

548   observed correlation is an underestimation, as gene expression in this experiment was

549   measured on mature fruits whereas the peak of expression of the *CmAPRR2* gene is

550   much earlier before fruit ripening (~15 DPA).

551

552   *Expression of CmAPRR2 in melon is correlated with plastid-development related genes*

553   To characterize co-expression patterns associated with the *CmAPRR2* gene, we

554   calculated the correlations between the expression of Melo3C003375 and all annotated

555   melon genes (n=27,557), using RNA-Seq data from mature fruits of the 414×DUL RILs

556   population (n=96) (Freilich *et al.*, 2015). Fourteen thousand genes expressed in mature

557   fruit flesh were used for this correlation analysis. Gene ontology (GO) enrichment

558   analysis was performed on 200 genes that had the strongest correlations with

559   Melo3C003375 (R>0.43, FDR adjusted P<0.001). The four most significant functional

560   groups that were enriched are related to photosynthesis, light reaction and plastid

561   organization and the fifteen most enriched components are related to plastids and

562   chloroplasts (**Sup Table 4**). These results support the predicted involvement of the

563   *CmAPRR2* gene in the regulation of chloroplasts and chromoplasts development.

564

565   **Discussion**

566   *Color variation in immature fruit rind in melon*

567   External fruit color is an important attribute in melons as it is a key factor defining

568   consumers' preference. Melon rind color transforms during fruit development and

569   ripening, mainly by shifting from the green rind of immature fruits, where chlorophyll

570   is the main pigment, to variable rind colors composed of different combinations of

18

571 chlorophylls, carotenoids and flavonoids (Tadmor *et al.*, 2010). Inheritance of external

572 color of immature fruit has been previously described in two different studies: The

573 white color of immature fruits was reported by Kubicki, (1962) to be dominant to green

574 immature fruits and controlled by a single gene named *Wi* (Dogimont, 2011). Burger

575 *et al.*, (2006*a*) described a recessive gene for light immature exterior color in a cross

576 between an American muskmelon (*Reticulatus* group) and an American honeydew-type

577 melon, (*Inodorous* group). Recently, a major gene for external color of immature fruit

578 was mapped in a cross between "Védrantais", a *Charentais* line from the *Cantalupensis*

579 group, and "Piel de Sapo", from the *Inodorous* group (Pereira *et al.*, 2018). The

580 dominant light rind from the "Védrantais" parent, that most likely correspond to the *Wi*

581 gene, was mapped to ~1.6 Mb interval on chromosome 7 (Pereira *et al.*, 2018). These

582 results confirm our observation that the dominant light grayish rind of *Charentais*

583 accessions is phenotypically and genetically distinct and controlled by a different gene

584 from the one we identified in the current study, which correspond to the recessive gene

585 described by Burger *et al.*, (2006*a*).

586 *APRR2-like transcription factors are key regulators of fruit pigmentation*

587 The results of the current study support the pivotal role of *APRR2-like* genes in

588 regulation of pigmentation in fruits. Pan *et al.* (2013) showed that over expression of

589 an *APRR2* gene in tomato resulted in increased chlorophyll content in immature fruits

590 and higher carotenoids level in ripe tomatoes. Both effects resulted from an increase in

591 plastid number. They also provided evidence for association between null mutation in

592 an *APRR2* gene and external fruit color intensity in green peppers. These results are

593 complementary to reports on the role of a related transcription factors group, GLKs,

594 which were shown to be associated with levels of chlorophyll and carotenoids in

595 Arabidopsis, tomato and pepper (Waters *et al.*, 2008, 2009; powell *et al.*, 2012; Brand

596 *et al.*, 2014). Recently, Liu *et al.*, (2016) reported that an *APRR2* gene is causative for

597 the white rind (*w*) mutation in cucumber, expressed as reduced chloroplast density and

598 chlorophyll content in young cucumber fruits. In all these crop plant species (tomato,

599 pepper and cucumber), there is also a correlation between expression levels of either

600 *APRR2* or *GLK* genes and pigments intensity. In the current study, we performed high-

601 resolution NGS-based mapping in segregating populations and found that null

602 mutations in the *CmAPRR2* and *ClAPRR2* genes are associated with light rind color in

603 melon and watermelon, respectively. We also showed that expression of the *CmAPRR2*

604   gene is correlated with pigment intensity in melon (**Figures 2d, 6d**) and that, as in

605   pepper (Brand *et al.*, 2014) and cucumber (Liu *et al.*, 2016), these transcription factors

606   show their strongest expression in fruit and reach their peak expression around 10-20

607   DPA and before fruit ripening. Our results expand the extent of experimental data that

608   demonstrate the conserved function of *APRR2*-like genes in regulating fruit

609   pigmentation as shown by the comparable expression profiles and analogous

610   phenotypes associated with variation in these genes.

611   *CmAPRR2 is associated with pigment accumulation across fruit developmental stages*

612   *and tissues*

613   Variation in ripe fruit color is substantially wider compared to the variation in the

614   immature stage. While in the immature stage color variation mostly reflect chlorophyll

615   concentrations, in the mature stage biosynthetic pathways of additional pigments (i.e.

616   carotenoids, flavonoids) are involved, leading to extended complexity of the genetic

617   architecture. This complexity is also expressed by the independent genetic control of

618   flesh and rind colors in melon as best demonstrated by the ability to combine different

619   rind and flesh colors through classical breeding. Rind and flesh color QTLs were

620   mapped in multiple melon populations (Monforte *et al.*, 2004; Cuevas *et al.*, 2008,

621   2009; Harel-Beja *et al.*, 2010; Galpaz *et al.*, 2018; Pereira *et al.*, 2018) and a few

622   causative color genes were identified (Feder *et al.*, 2015; Tzuri *et al.*, 2015). However,

623   so far, a common regulator that affects pigmentation throughout the different

624   developmental stages and fruit tissues (rind and flesh) has not been described in melon.

625   In the current study, we showed that the *CmAPRR2* gene is such a key regulator,

626   associated with pigments concentrations during the course of fruit development and

627   across fruit tissues (**Figures 1, 5**). Furthermore, the mapping population in this study

628   (TAD×DUL RILs), which independently segregated for both *CmOr* and *CmAPRR2*

629   genes, allowed us to show that the *CmAPRR2* effect is also independent of the type of

630   pigment accumulated in the flesh, as it was associated with variation in both chlorophyll

631   and carotenoids concentrations (**Figure 5c-g**). We also showed here, using a different

632   segregating population (414×DUL RILs) that was previously subjected to mature fruit

633   RNA-Seq and carotenoid analysis (Freilich *et al.*, 2015; Galpaz *et al.*, 2018), that the

634   cis-regulated *CmAPRR2* expression variation is correlated with flesh β-carotene content

635   in mature fruits (**Figure 6**). Since both parental lines of this population carry a predicted

636   'dark' allele based on the coding sequence of the *CmAPRR2* gene, we assume that this

637   experiment provided another piece of evidence for the relationship between expression

638   level of *CmAPRR2* and flesh pigments content. The proposed involvement of this

639   transcription factor in regulation of plastid development in fruits (Pan *et al.*, 2013; Liu

640   *et al.*, 2016) support the broad effect of *CmAPRR2* that was described here.

641   *Multi-allelic nature of the CmAPRR2 gene in melon*

642   The genetic architecture that describes a phenotypic trait is strongly dependent on the

643   type and structure of the germplasm used for the study. For quantitative polygenic traits,

644   QTL segregation and detection will not necessarily overlap across different mapping

645   populations. The same can apply to simple traits, where independent mutations in

646   different genes, which are involved in a common biological process, lead to the same

647   discrete phenotype. While bi-parental populations will draw only part of the picture in

648   such cases, diverse collections or multi-parental segregating populations are more

649   effective in comprehensively characterizing this architecture. In the current study we

650   tried to genetically characterize the light immature rind phenotype in melon using a

651   diverse collection, assuming it is under a simple genetic control as previously described

652   (Burger *et al.*, 2006*a*). In GWAS, lack of detection power can result from low

653   heritability, low frequency of the phenotype under investigation, strong confounding

654   effect of population structure or insufficient markers density (Korte and Farlow, 2013).

655   While none of these factors seemed to apply in our case (Gur *et al.*, 2017 and **Figure**

656   **1**), we did not obtain any significant GWA signal, which led to the intuitive assumption

657   that multiple genes are associated with the light immature rind phenotype in our

658   collection. The identification of two independent allelic nonsense mutations in the

659   *CmAPRR2* gene, through linkage analyses (**Figures 1**, **2**), indicated that we might be

660   looking at a different scenario. Through complementary resequencing of a diverse core

661   panel and comprehensive allelism testing (**Figures 2, 3**) we were able to demonstrate

662   that this trait is a unique case of simple genetic architecture. On the functional level, it

663   seems to be controlled by a single gene that segregates in a Mendelian manner in bi-

664   parental crosses, but the multi-allelic pattern at the *CmAPRR2* gene drove reduced

665   power through GWAS, which masked this simplicity and created the observed

666   contradiction between the different mapping strategies. These results provide a thought-

667   provoking example for another possible inherent complexity that can arise in GWAS -

668   independent low-frequency causative variants within a common gene. In the current

669   scenario, even whole-genome deep resequencing of the GWAS panel, which would

670   target each of the variants in the *CmAPRR2* gene, would not necessarily resolve the lack

671   of detection power, as each of these independent variants remain at low frequency.

672   Genetic mapping studies are rapidly shifting towards sequencing-based genotyping and

673   in most cases marker density is no longer a bottleneck in GWAS (Yano *et al.*, 2016;

674   Misra *et al.*, 2017). A key challenge remains in prioritizing GWAS signals and

675   improving weak signals obtained from low frequency causative variants (Lee and Lee,

676   2018). The availability of whole-genome assemblies and corresponding protein-coding

677   gene annotations, alongside additional layers of information, such as expression

678   profiles from RNA-Seq experiments, now facilitate the integration of multiple data

679   layers to improve GWAS results (Shim *et al.*, 2017; Lee and Lee, 2018; Schaefer *et al.*,

680   2018). Our example of the *CmAPRR2* gene suggests that adding functional annotation

681   prediction to GWAS SNPs and treating predicted genes as integral functional units

682   could potentially be used as an informative layer that can boost the signal of causative

683   weak associations.

684   In summary, we have identified the *CmAPRR2* gene as a common regulator of fruit

685   pigmentation in melon and watermelon. The conserved and broad effect of this gene

686   across species, fruit tissues, developmental stages and different types of pigment

687   accumulated, suggest its potential as a useful target for carotenoids bio-fortification of

688   cucurbits and other fruits.

689

690   **Supplementary data**

691   **Supplementary Figure 1**: cDNA sequence comparison between the mapping population

692   parental lines DUL, TAD and NA.

693   **Supplementary Figure 2:** Number of annotated transcript haplotypes vs number of SNPs

694   per bp (a) and transcript length (b) across 2,200 genes on chromosome 4. The *CmAPRR2*

695   gene is highlighted in red.

696   **Supplementary Figure 3**: *CmAPRR2* gene is located in a low LD region. a) LD (R^2) heat

697   map by physical position in a 400Kb window surrounding the *CmAPRR2* gene. b) Sliding-

698   window trend line of pairwise SNP LD across 1M bp interval surrounding the *CmAPRR2*

699   gene on chr4.

22

700 **Supplementary Figure 4:** Segregation and scoring of rind color of 87 F_3 families from
701 the NY0016×EMB cross. a) Examples of the three phenotypic rind color classes in the
702 population. b) Segregation of the color classes across F_3 families.

703 **Supplementary Figure 5**: Analysis of rind chlorophyll and carotenoids during fruit
704 development on TAD, DUL and selected F_3 families from their cross. a) Analysis of rind
705 chlorophyll content during fruit development in TAD and DUL. b) Comparison of rind
706 chlorophyll content between light and dark F_3 families at 20 DAA. c) Comparison of rind
707 total carotenoids content between light and dark F_3 families at maturity (40 DAA). d)
708 Correlation between chlorophyll content at 20 DAA and carotenoids at 40 DAA across
709 the selected F3 families and parental lines.

710 **Supplementary Figure 6:** Prediction of flesh β-carotene based on fruit-section image
711 analyses in the TAD×DUL RILs. a) Analysis of flesh color (Chroma) on fruit sections
712 scans of ~700 orange fruits in the TAD×DUL RILs. b) Calibration curve for the relation
713 between flesh Chroma and β-carotene concentration as measured across 73 diverse orange
714 flesh accessions from our GWAS panel. Logarithmic equation representing the best fit is
715 shown. c) Calculation of predicted β-carotene in the TAD×DUL RILs based on Chroma
716 values from (a) that were placed in the equation described in (b). Analysis of predicted β-
717 carotene concentrations across the TAD×DUL RILs. Comparison between the 'light' and
718 'dark' alleles in the *APRR2* gene.

719 **Supplementary Table 1:** Melon young fruit light rind QTL interval: annotations and
720 positions of genes.

721 **Supplementary Table 2:** Segregation of light rind in four F_2 watermelon populations.

722 **Supplementary Table 3:** Light rind QTL interval in watermelon: annotations and
723 positions of genes.

724 **Supplementary Table 4:** Gene Ontology (GO) enrichment analysis for 200 genes
725 correlated with expression of the *APRR2* gene (Melo3C003375) in melon fruit.

726 **Supplementary Table 5:** list of primers used for RT-qPCR.

727

728 **Acknowledgments**

23

736

737   References

738   **Benjamini Y, Hochberg Y**. 1995. Controlling the false discovery rate: a practical and

739   powerful approach to multiple testing. Journal of the Royal Statistical Society B **57**,

740   289–300.

741   **Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES**.

742   2007. TASSEL: Software for association mapping of complex traits in diverse

743   samples. Bioinformatics **23**, 2633–2635.

744   **Brand A, Borovsky Y, Hill T, Rahman KAA, Bellalou A, Van Deynze A, Paran I**.

745   2014. CaGLK2 regulates natural variation of chlorophyll content and fruit color in

746   pepper fruit. Theoretical and Applied Genetics **127**, 2139–2148.

747   **Branham S, Vexler L, Meir A, Tzuri G, Frieman Z, Levi A, Wechter P, Tadmor**

748   **Y, Gur A**. 2017. Genetic mapping of a major codominant QTL associated with β -

749   carotene accumulation in watermelon. Molecular Breeding.

750   **Burger Y, Bhasteker D, Sa'ar U, Katzir N, Paris HS**. 2006*a*. A Recessive Gene for

751   Light Immature Exterior Color of Melon. Genetics, Cucurbit Report, Cooperative **18**,

752   17–18.

753   **Burger Y, Paris HS, Cohen R, Katzir N, Tadmor Y, Lewinsohn E, Schaffer AA**.

754   2010. Genetic Diversity of Cucumis Melo. Horticultural Reviews. Wiley-Blackwell,

755   165–198.

756   **Burger Y, Sa'ar U, Paris HS, Lewinsohn E, Katzir N, Tadmor Y, Schaffer AA**.

757   2006*b*. Genetic variability for valuable fruit quality traits in Cucumis melo. Israel

758   Journal of Plant Sciences **54**, 233–242.

759   **Chen M, Ji M, Wen B, Liu L, Li S, Chen X, Gao D, Li L**. 2016. GOLDEN 2-LIKE

24

760   Transcription Factors of Plants. Frontiers in Plant Science **7**, 1–5.

761   **Clayberg CD**. 1992. Interaction and linkage test of flesh colour genes in Cucumis
762   melo L. Cucurbit Genetics Cooperative **15**, 53.

763   **Cuevas HE, Staub JE, Simon PW, Zalapa JE**. 2009. A consensus linkage map
764   identifies genomic regions controlling fruit maturity and beta-carotene-associated
765   flesh color in melon (Cucumis melo L.). Theoretical and Applied Genetics **119**, 741–
766   756.

767   **Cuevas HE, Staub JE, Simon P, Zalapa JE, McCreight JD**. 2008. Mapping of
768   genetic loci that regulate quantity of beta-carotene in fruit of US Western Shipping
769   melon (Cucumis melo L.). TAG. Theoretical and applied genetics. Theoretische und
770   angewandte Genetik **117**, 1345–1359.

771   **Diaz A, Fergany M, Formisano G, et al.** 2011. A consensus linkage map for
772   molecular markers and quantitative trait loci associated with economically important
773   traits in melon (Cucumis melo L.). BMC plant biology **11**, 111.

774   **Dogimont C**. 2011. 2011 Gene List for Melon. Cucurbit Genetics Cooperative Report
775   33-34 **133**, 104–133.

776   **Feder A, Burger J, Gao S, et al.** 2015. A Kelch domain-containing F-box coding
777   gene negatively regulates flavonoid accumulation in Cucumis melo L. Plant
778   Physiology **169**, 1714–1726.

779   **Freilich S, Lev S, Gonda I, et al.** 2015. Systems approach for exploring the intricate
780   associations between sweetness, color and aroma in melon fruits. BMC Plant Biology
781   **15**.

782   **Galpaz N, Gonda I, Shem-Tov D, et al.** 2018. Deciphering Genetic Factors that
783   Determine Melon Fruit-Quality Traits Using RNA-Seq-Based High-Resolution QTL
784   and eQTL Mapping. The Plant Journal **94**, 169–191.

785   **Garcia-Mas J, Benjak a., Sanseverino W, et al.** 2012. The genome of melon
786   (Cucumis melo L.). Proceedings of the National Academy of Sciences **109**, 11872–
787   11877.

788   **Gonda I, Ashrafi H, Lyon DA, et al.** 2018. Sequencing-Based Bin Map Construction
789   of a Tomato Mapping Population, Facilitating High-Resolution Quantitative Trait

790   Loci Detection. The Plant Genome.

791   **Gur A, Tzuri G, Meir A, Sa'ar U, Portnoy V, Katzir N, Schaffer AA, Li L,**
792   **Burger J, Tadmor Y**. 2017. Genome-Wide Linkage-Disequilibrium Mapping to the
793   Candidate Gene Level in Melon (Cucumis melo). Scientific Reports **7**, 9770.

794   **Gusmini G, Wehner TC**. 2005. Genes Determining Rind Pattern Inheritance in
795   Watermelon : A Review. Hort Science **40**, 1928–1930.

796   **Hall TA**. 1999. BioEdit: a user-friendly biological sequence alignment editor and
797   analysis program for Windows 95/98/NT. Nucleic acids symposium series. [London]:
798   Information Retrieval Ltd., c1979-c2000., 95–98.

799   **Harel-Beja R, Tzuri G, Portnoy V, *et al.*** 2010. A genetic map of melon highly
800   enriched with fruit quality QTLs and EST markers, including sugar and carotenoid
801   metabolism genes. TAG. Theoretical and applied genetics. Theoretische und
802   angewandte Genetik **121**, 511–533.

803   **Korte A, Farlow ashley**. 2013. The advantages and limitations of trait analysis with
804   GWAS : a review. Plant methods **9**, 29.

805   **Kosambi DD**. 1943. The estimation of map distances from recombination values.
806   Annals of Eugenics **12**, 172–175.

807   **Kubicki B**. 1962. Inheritance of some characters in muskmelons (Cucumis melo).
808   Genet. Pol. **3**, 265–274.

809   **Lee T, Lee I**. 2018. AraGWAB: Network-based boosting of genome-wide association
810   studies in Arabidopsis thaliana. Scientific Reports **8**, 1–6.

811   **Liu H, Jiao J, Liang X, Liu J, Meng H, Chen S, Li Y, Cheng Z**. 2016. Map-based
812   cloning, identification and characterization of the w gene controlling white immature
813   fruit color in cucumber (Cucumis sativus L.). Theoretical and Applied Genetics **129**,
814   1247–1256.

815   **McKenna A, Hanna M, Banks E, *et al.*** 2010. The Genome Analysis Toolkit: a
816   MapReduce framework for analyzing next-generation DNA sequencing data.
817   Genome research **20**, 1297–1303.

818   **Misra G, Badoni S, Anacleto R, Graner A, Alexandrov N, Sreenivasulu N**. 2017.

819 Whole genome sequencing-based association study to unravel genetic architecture of

820 cooked grain width and length traits in rice. Scientific Reports **7**, 12478.

821 **Monforte AJ, Oliver M, Gonzalo MJ, Alvarez JM, Dolcet-Sanjuan R, Arus P**.

822 2004. Identification of quantitative trait loci involved in fruit quality traits in melon (

823 Cucumis melo L.). Theoretical and Applied Genetics **108**, 750–758.

824 **Pan Y, Bradley G, Pyke K,** *et al.* 2013. Network Inference Analysis Identifies an

825 APRR2-Like Gene Linked to Pigment Accumulation in Tomato and Pepper Fruits.

826 Plant Physiology **161**, 1476–1485.

827 **Paris MK, Zalapa JE, Mccreight JD, Staub JE**. 2008. Genetic dissection of fruit

828 quality components in melon (Cucumis melo L.) using a RIL population derived from

829 exotic × elite US Western Shipping germplasm. Molecular Breeding **22**, 405–419.

830 **Pereira L, Ruggieri V, Pérez S, Alexiou KG, Fernández M, Jahrmann T, Pujol**

831 **M**. 2018. QTL mapping of melon fruit quality traits using a high-density GBS-based

832 genetic map. BMC Plant Biology **18**, 1–17.

833 **powell A l. t., Nguyen C v., Hill T,** *et al.* 2012. Uniform ripening Encodes a Golden.

834 Science **336**, 1711–1715.

835 **Ríos P, Argyris J, Vegas J,** *et al.* 2017. ETHQV6.3 is involved in melon climacteric

836 fruit ripening and is encoded by a NAC domain transcription factor. Plant Journal **91**,

837 671–683.

838 **Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G,**

839 **Mesirov JP**. 2011. Integrative genomics viewer. Nature Biotechnology **29**, 24.

840 **Rodríguez GR, Moyseenko JB, Robbins MD, Morejón NH, Francis DM, van der**

841 **Knaap E**. 2010. Tomato Analyzer: a useful software application to collect accurate

842 and detailed morphological and colorimetric data from two-dimensional objects.

843 JoVE.

844 **Schaefer R, Michno J-M, Jeffers J, Hoekenga OA, Dilkes BP, Baxter IR, Myers**

845 **C**. 2018. Integrating co-expression networks with GWAS to prioritize causal genes in

846 maize. The Plant Cell, doi:10.1105/tpc.18.00299.

847 **Shim JE, Bang C, Yang S, Lee T, Hwang S, Kim CY, Singh-Blom UM, Marcotte**

848 **EM, Lee I**. 2017. GWAB: A web server for the network-based boosting of human

849   genome-wide association data. Nucleic Acids Research **45**, W154–W161.

850   **Swarts K, Li H, Romero Navarro JA,** *et al.* 2014. Novel Methods to Optimize
851   Genotypic Imputation for Low-Coverage, Next-Generation Sequence Data in Crop
852   Plants. The Plant Genome **7**.

853   **Tadmor Y, Burger J, Yaakov I,** *et al.* 2010. Genetics of flavonoid, carotenoid, and
854   chlorophyll pigments in melon fruit rinds. Journal of Agricultural and Food Chemistry
855   **58**, 10722–10728.

856   **Tadmor Y, Hirschberg J, Lewinsohn E**. 2005. Comparative fruit colouration in
857   watermelon and tomato. **38**, 837–841.

858   **Takagi H, Abe A, Yoshida K,** *et al.* 2013. QTL-seq: Rapid mapping of quantitative
859   trait loci in rice by whole genome resequencing of DNA from two bulked populations.
860   Plant Journal **74**, 174–183.

861   **Tzuri G, Zhou X, Chayut N,** *et al.* 2015. A "golden" SNP in CmOr governs fruit
862   flesh color of melon (Cucumis melo). The Plant Journal **82**, 267–279.

863   **Waters MT, Moylan EC, Langdale JA**. 2008. GLK transcription factors regulate
864   chloroplast development in a cell-autonomous manner. Plant Journal **56**, 432–444.

865   **Waters MT, Wang P, Korkaric M, Capper RG, Saunders NJ, Langdale JA**.
866   2009. GLK Transcription Factors Coordinate Expression of the Photosynthetic
867   Apparatus in Arabidopsis. The Plant Cell **21**, 1109–1128.

868   **Wehner TC**. 2008. Watermelon. Vegetables I. Springer, 381–418.

869   **Yano R, Nonaka S, Ezura H**. 2018. Melonet-DB, a Grand RNA-Seq Gene
870   Expression Atlas in Melon (Cucumis melo L.). Plant & cell physiology **59**, e4.

871   **Yano K, Yamamoto E, Aya K,** *et al.* 2016. Genome-wide association study using
872   whole-genome sequencing rapidly identifies new genes influencing agronomic traits
873   in rice. Nature Genetics **48**, 927.

874

875

876

Figure legends

Figure 1: Characterization of variation and genetic mapping of young fruit rind color in melon. a) Example of accessions with light and dark rinds. b) Genetic PCA plot of 120 light and dark accessions from the diverse collection (Gur *et al.*, 2017). Dot color correspond to light and dark immature rind. Parental lines of the mapping populations are shown to the right of the plot (TAD: Tam Dew, NA: Noy-Amid). The pie chart on the left bottom corner summarizes the frequencies of young fruit rind colors. c) Manhattan plots for mapping of young fruit rind color across two populations over two seasons. d) Zoom in on chromosome 4. The 290Kb confidence interval is highlighted.

Figure 2: Fine mapping and candidate gene characterization. a) Substitution mapping at the TAD×DUL RILs population. Nine recombinant RILs at the QTL interval are shown. Marker bins physical positions (bp) are indicated on the top of each column. Light and dark colors correspond to parental alleles (TAD and DUL, respectively). RILs young fruit rind color phenotypes are shown on the right column. Trait mapping interval is bounded with thick vertical lines. Position of the candidate gene Melo3C003375 is shown. b) Melo3C003375 gene structure and exonic sequence variants. Black boxes represent exons. Light-gray arrows represent synonymous SNPs. Dark-gray arrows are non-synonymous SNPs that did not show a distinct allelic state between dark and light accessions (15, 16). Black arrows are SNPs causing a single amino-acid change, and red arrows are polymorphisms (SNPs or InDels) causing major change in protein (frame-shift, stop codon). c) Predicted protein size of the dark (DUL; Dulce) and light rind parents (TAD; Tam-Dew, NA; Noy-Amid). d) Expression pattern of the *CmAPRR2* gene through fruit development, and comparison between parental lines. DPA: days post anthesis. e) Table of non-synonymous allelic variants in Melo3C003375 that distinguish between light and dark phenotypes across 19 diverse melon lines. Accessions are colored by their horticultural group. f) Association tests of independent causative variants and combined 'functional variant' with rind color, across the core set.

Figure 3: Allelism tests for light rind accessions. a) Allelism test for the mapping populations light rind parental lines (TAD and NA). b) Genetic PCA plot with 25 selected founders highlighted by rind color. In gray are lines with non-distinct rind color. c) half-diallele allelism tests across 11 light rind accessions. Two dark lines were used as reference testers.

909 **Figure 4:** Mapping and cloning of the light rind color gene in watermelon. a) Three
910 testcrosses and F_2 segregation for rind color. OFT: Orange Flesh Tender sweet, EMB:
911 Early Moon Beam, DL: Dixie Lee. b) Parents of the mapping population and rind color
912 frequency distribution in the $F_{2:3}$ population. c) Manhattan plot of whole-genome linkage
913 analysis in the $F_{2:3}$ population using 3,160 GBS-derived SNPs. d) BSA-Seq results across
914 35 fixed $F_{2:3}$ families. Association significance (ΔSNP-index analysis) is expressed using
915 blue-to-red color scale. Position of the watermelon *APRR2* gene (CICG09G012330) is
916 shown. e) *ClAPRR2* (CICG09G012330) annotated gene structure. Exons are represented
917 as black boxes. Parental genomic sequence alignment and SNP (C/G) at the intron 6 exon
918 7 junction. f) Comparison of mRNA sequence of parental lines (EMB and NY0016) and
919 three segregants from each rind color group. Stop codon downstream to the 16 bp indel is
920 shown. g) Alignment of parental lines translated protein sequence around the InDel site.

921 **Figure 5:** Association of allelic variation in the *CmAPRR2* gene with mature fruit rind and
922 flesh pigmentation in TAD×DUL RILs. a) External images of fruits from RILs with dark
923 and light genotype in the *CmAPRR2* gene. b) Analysis of carotenoids in mature fruit rinds.
924 c) Segregation of orange and green flesh controlled by the *CmOr* gene, across 166 RILs.
925 d) Representative scans of orange fruits segregating for dark and light alleles at the
926 *CmAPRR2* gene. e) Analysis of *CmAPRR2* allelic effect on flesh color of mature orange
927 (*CmOR/CmOR*) fruits. f) Representative scans of green fruits segregating for dark and
928 light alleles in the *CmAPRR2* gene. g) Analysis of *CmAPRR2* allelic effect on flesh color
929 of mature green (*cmor/cmor*) fruits.

930 **Figure 6:** *CmAPRR2* (Melo3C00375) is differentially expressed, and correlated with β-
931 carotene content in the 414×DUL RILs. a) Population parents (DUL and 414). b)
932 Manhattan plot for whole-genome eQTL mapping of *CmAPRR2* (Melo3C003375)
933 expression (RPKM) in mature fruit flesh. c) Zoom in on cis-eQTL spanning
934 Melo3C003375 on chr.4. d) Correlation between Melo3C00375 expression (RPKM) and
935 β-carotene content (μg/g fresh weight) in the fruit flesh.
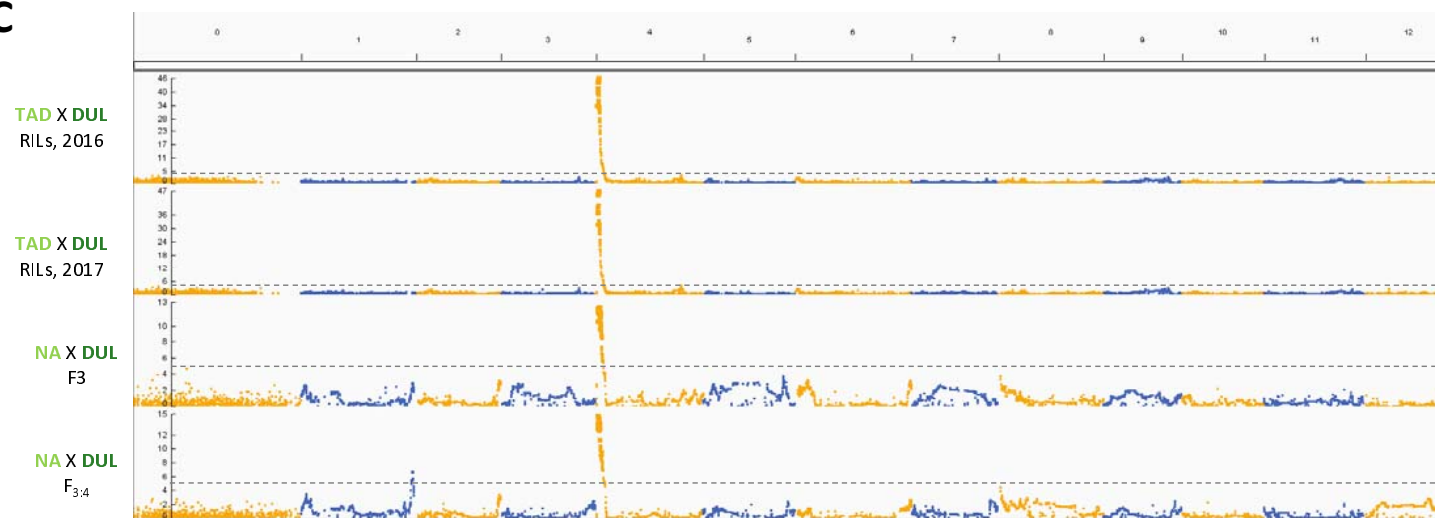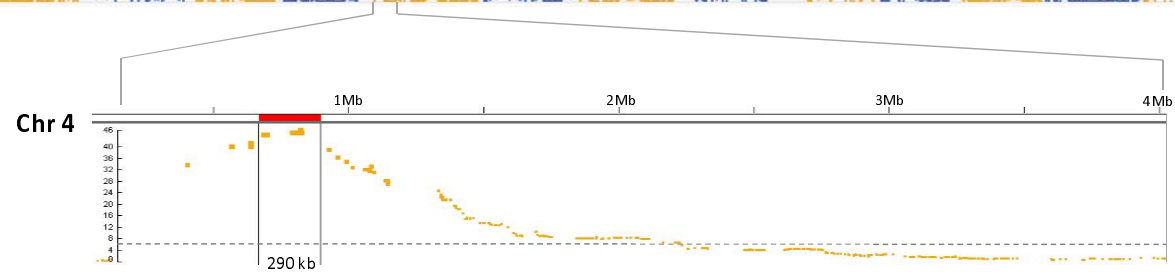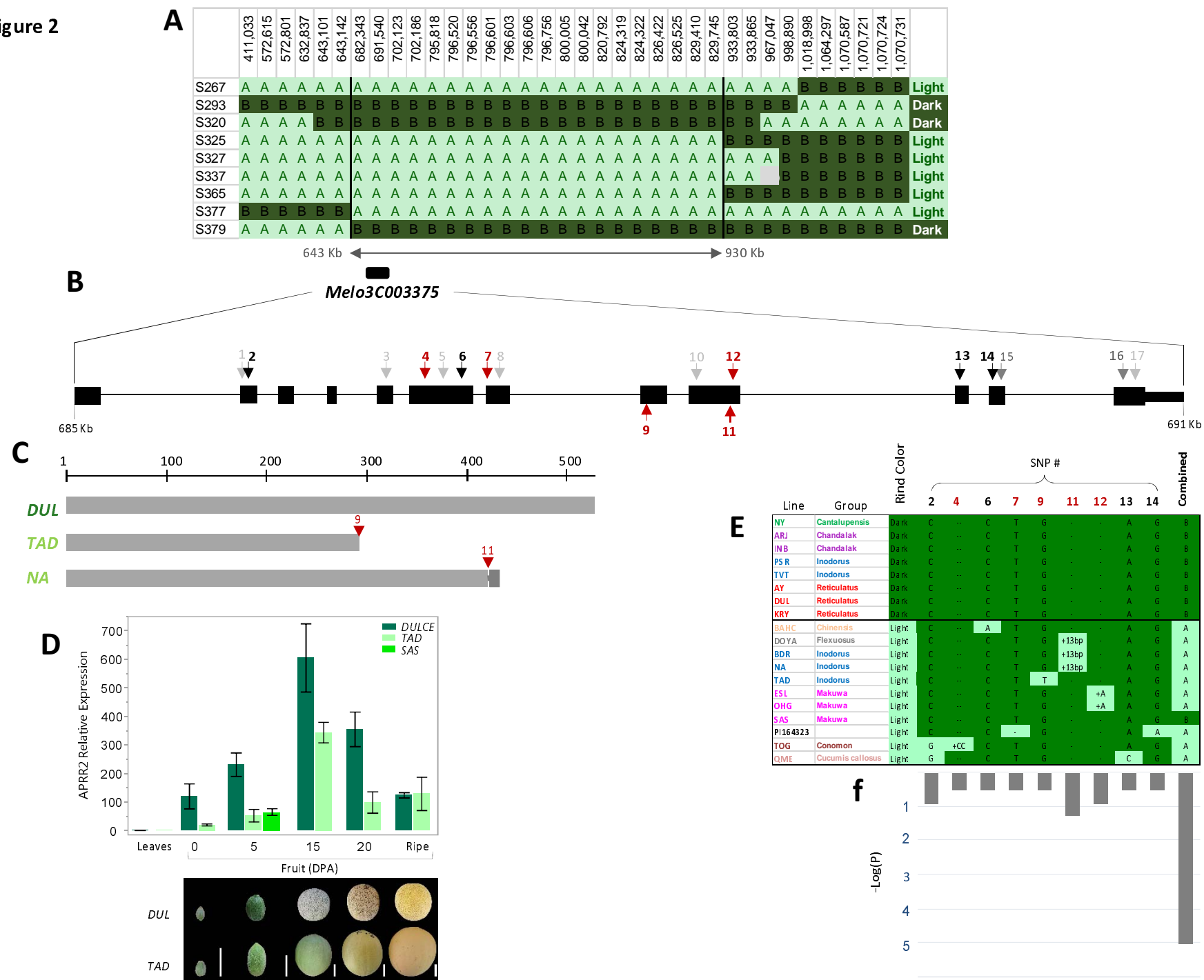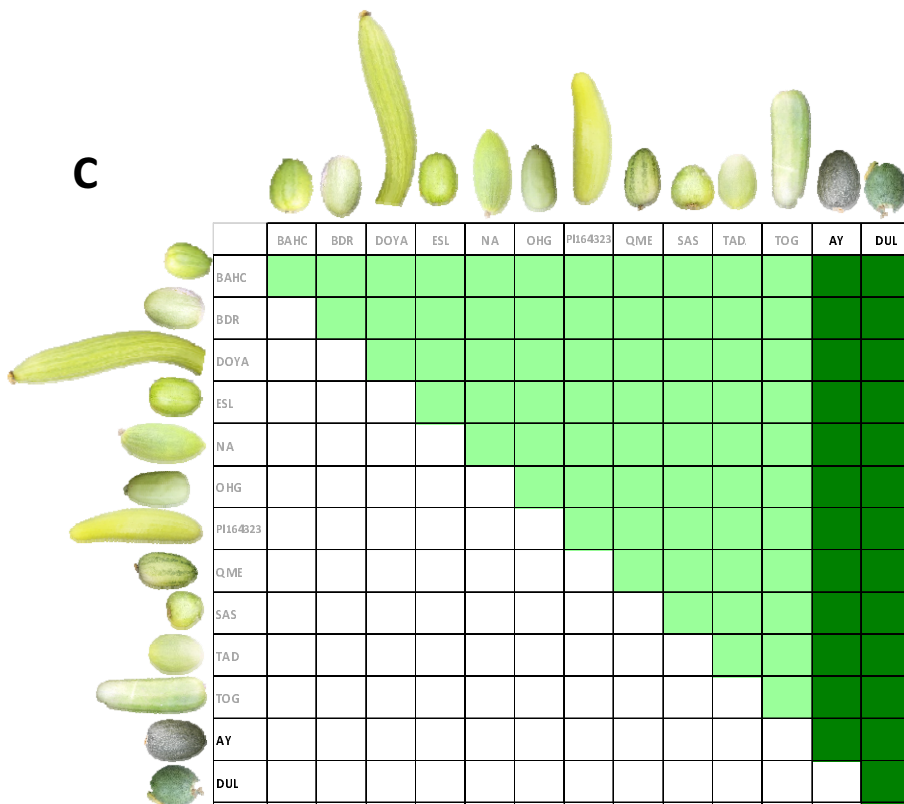
Figure 1
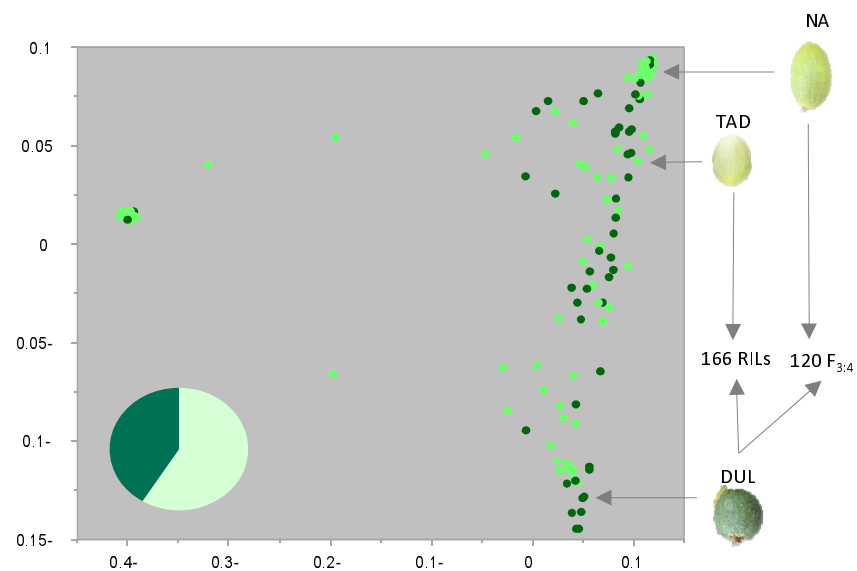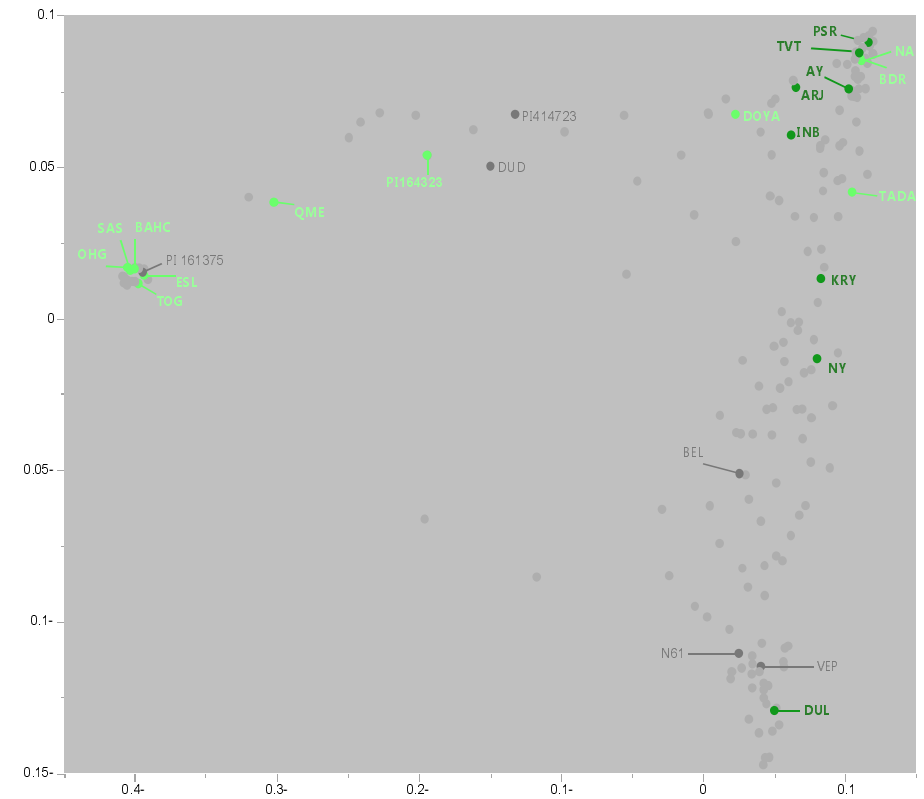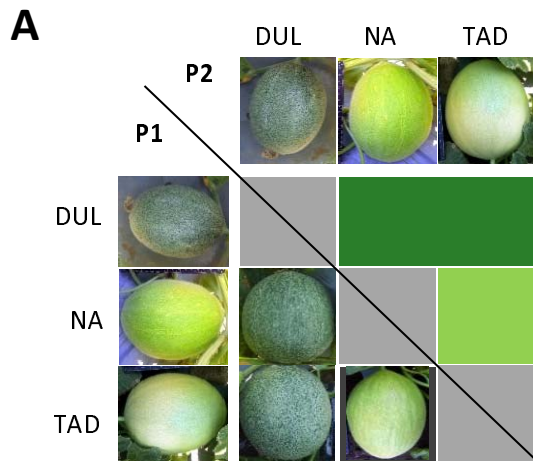
Figure 2

Figure 3

Figure 4

Figure 5

Figure 6