Journal of Experimental Botany
https://doi.org/10.1093/jxb/erae355 Advance Access Publication 22 August 2024

RESEARCH PAPER

Irregular green netting of eggplant fruit peel: a domestication trait controlled by *SmGLK2* with potential for fruit colour diversification

Andrea Arrones[1], Silvia Manrique[1], Joaquin Gomis-Cebolla[1], Virginia Baraja-Fonseca[1], Mariola Plazas[1], Jaime Prohens[1], Ezio Portis[2], Lorenzo Barchi[2], Giovanni Giuliano[3], Pietro Gramazio[1,*], and Santiago Vilanova[1,*]

[1] Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Camino de Vera 14, 46022 Valencia, Spain
[2] Dipartimento di Scienze Agrarie, Forestali e Alimentari (DISAFA), Plant Genetics and Breeding, University of Turin, Grugliasco, Italy
[3] Agenzia Nazionale Per Le Nuove Tecnologie, L'energia e Lo Sviluppo Economico Sostenibile (ENEA), Casaccia Research Centre, Rome, Italy

* Correspondence: piegra@upv.es or sanvina@upv.es

Received 26 January 2024; Editorial decision 14 August 2024; Accepted 21 August 2024

Editor: John Lunn, MPI of Molecular Plant Physiology, Germany

Abstract

The distribution of chlorophylls in eggplant (*Solanum melongena*) peel exhibits either a uniform pattern or an irregular green netting pattern. The latter, manifested as a gradient of dark green netting that is intensified in the proximal part of the fruit on a pale green background, is common in wild relatives and some eggplant landraces. Despite the selection of uniform chlorophylls during domestication, the netting pattern contributes to a greater diversity of fruit colours. Here, we used over 2300 individuals from different populations, including a multi-parent advanced generation inter-cross population for candidate genomic region identification, an F_2 population for bulked segregant analysis by sequencing, and advanced backcrosses for edges-to-core fine-mapping, to identify *SmGLK2* gene as responsible for the irregular netting in eggplant fruits. We also analysed the gene sequence of 178 *S. melongena* accessions and 22 wild relative species for tracing the evolutionary changes that the gene has undergone during domestication. Three different mutations were identified leading to the absence of netting. The main causative indel induces a premature stop codon disrupting the protein conformation and function, which was confirmed by western blot analysis and confocal microscopy observations. *SmGLK2* has a major role in regulating chlorophyll biosynthesis in eggplant fruit peel.

Keywords: Domestication trait, eggplant, fruit colour diversification, fruit green netting, irregular chlorophyll pattern, *Solanum melongena*, *SmGLK2*.

Abbreviations: AB, advanced backcross; BSA-seq, bulked segregant analysis by sequencing; DAS1, *Solanum dasyphyllum*; FN, fruit green netting (presence); fn, fruit green netting absence; GLK, *GOLDEN2-LIKE*; IL, introgression line; INC1, *Solanum incanum*; INS1, *Solanum insanum*; MAGIC, multi-parent advanced generation inter-cross; MEL, *Solanum melongena*; NLS: nuclear localization signal.

Introduction

Variation in the presence or absence of chlorophylls in eggplant (*Solanum melongena* L.) fruit peel contributes to the wide range of fruit colours in this species (Arrones *et al.*, 2022). Furthermore, the distribution of chlorophylls can be uniform or irregular, the latter being referred to as netting, reticulation, or variegation (Tigchelaar, 1968; Daunay *et al.*, 2004). The presence of the fruit green netting (FN) trait is commonly present in eggplant wild species and landraces. Cultivated eggplant differs morphologically and physiologically from its wild ancestors. Indeed, wild eggplants produce small, bitter, green-netted fruits, while domesticated eggplants are characterized by desirable agricultural traits that were directly and indirectly selected and fixed during domestication (Frary and Doğanlar, 2003; Fuller, 2007). These include big and non-bitter fruits, with a diversification of immature fruit peel colour, resulting from the combination of chlorophyll and anthocyanin pigments (Taher *et al.*, 2017; Page *et al.*, 2019; Page and Chapman, 2021). While numerous studies have focused solely on the presence or absence of chlorophylls in the eggplant fruit peel, it is noteworthy that reticulated fruits often exhibit a gradient in netting density in a pale green background. This gradient often leads to a deep green colour in the proximal portion of the fruit, and gradually diminishes towards the distal end. Thus, FN contributes to further increase the colour diversity of eggplant, resulting in an irregular green or dark purple colour depending on the concurrent absence or presence of anthocyanins.

The gradient of expression of this trait is reminiscent of the green shoulder trait described in tomato, which is controlled by a single dominant gene known as the *Uniform ripening* (*U*) locus (Powell *et al.*, 2012; Nguyen *et al.*, 2014). Tomato green shoulder has been counter-selected by breeders, to improve fruit colour uniformity and facilitate maturity stage determination and harvesting (Powell *et al.*, 2012). Since tomato and eggplant display many significant similarities in their domestication syndromes (Chapman, 2019), a similar selection mechanism may have occurred in eggplant since immature fruits contain higher levels of chlorophyll in the peel.

In this study, we identified *GOLDEN2-LIKE2* (*SmGLK2*) as the candidate gene controlling eggplant FN by using different experimental and germplasm populations, combining wild and cultivated genepools: (i) identification of a candidate genomic region in a multi-parent advanced generation intercross (MAGIC) population; (ii) validation through bulked segregant analysis by sequencing (BSA-seq) in a contrasting F_2 population; (iii) narrowing down the genomic region by fine-mapping in two advanced backcross (AB) sets; (iv) search for causative allelic variants in the candidate gene sequence by the resequencing of accessions contrasting for the presence (FN) or absence (fn) of netting in the fruit peel; (v) tracing the changes that the gene has undergone through the domestication process in a set of accessions from a diverse germplasm collection; (vi) evaluation of the effect of the main mutation identified at the protein level by western blotting analysis; and (vii) observation of chlorophyll autofluorescence emission from FN and fn fruits by confocal microscopy. The identification of *SmGLK2* gene as responsible for the eggplant FN trait is useful to understand the genetic basis of fruit coloration in eggplant as well as for eggplant breeding.

Materials and methods

Plant materials

MAGIC population

A total of 420 individuals from the eggplant S3MEGGIC population developed by Mangino *et al.* (2022) were used to identify candidate genomic regions associated with the FN trait. This population was previously used for identifying *SmAPRR2* on chromosome 8 as the causative gene for uniform chlorophyll pigmentation in the eggplant fruit peel (Arrones *et al.*, 2022). The S3MEGGIC population was obtained by inter-crossing seven *Solanum melongena* accession and one wild *S. incanum* INC1 accession (Plazas *et al.*, 2016). Only the wild INC1 founder displayed the FN trait, while none of the seven *S. melongena* displayed this trait.

F_2 population

The white-fruited accession *S. melongena* MEL1 and the green netting-fruited *S. insanum* INS1 were used as parents to develop an F_2 population of 120 individuals segregating for the FN trait (Plazas *et al.*, 2016). This population was used for the validation of the candidate genomic region. These parents were selected for the easier visualization of the FN in the absence of anthocyanins (Supplementary Fig. S1).

Advanced backcross individuals and selfings

AB individuals from two programs of development of interspecific introgression line (IL) populations were used to narrow down the candidate genomic region responsible for the FN trait. One IL population was developed by crossing one white *S. melongena* accession, MEL1, and one accession of the wild *S. dasyphyllum*, DAS1, while the other IL population was developed from the interspecific cross between one *S. melongena*, MEL5, which presents fruit anthocyanin (Mangino *et al.*, 2020), and the wild *S. insanum*, INS1 (Plazas *et al.*, 2016, 2020). The two wild parents (DAS1 and INS1) presented FN phenotype while none of the *S. melongena* (MEL1 and MEL5) had chlorophyll pigmentation in the fruit. For this study, AB individuals from the advanced stages of the IL population development displaying the FN phenotype were selected. These AB individuals were used to confirm the FN candidate genomic region by locating the introgressed wild fragment with the single primer enrichment technology (SPET) high-throughput genotyping platform (Barchi *et al.*, 2019a). The AB individuals with the shortest introgressed fragment were selfed to find recombinants in their progeny and further narrow down the candidate region.

Germplasm core collection

Genomic sequences of 178 *S. melongena* accessions and 22 wild relative species available from the eggplant germplasm core collection established in the framework of the G2P-SOL project (http://www.g2p-sol.eu/G2P-SOL-gateway.html) were interrogated for the proposed candidate gene to find causative variants of the fn phenotype. This core collection includes accessions used for developing the first eggplant pan-genome (Barchi *et al.*, 2021). In addition, 22 wild relative species including *S. insanum* (12), *S. incanum* (4), *S. linneanum* (3), *S. macrocarpon* (1), *S. humile* (1),

and *S. tettense* (1), all with FN phenotype, were selected to confirm the putative candidate gene structure (Supplementary Table S1).

Cultivation conditions

Seeds from the studied individuals were germinated in Petri dishes, following the protocol developed by Ranil *et al.* (2015). They were subsequently transferred to seedling trays in a climatic chamber under a photoperiod and temperature regime of 16 h light (25 °C, 100–112 µmol m^{-2} s^{-1}) and 8 h dark (18 °C). After acclimatization, plantlets were grown either in a pollinator-free benched glasshouse or an open field plot at the UPV campus in Valencia, Spain (GPS coordinates: latitude, 39° 28′ 55″ N; longitude, 0° 20′ 11″ W; 7 m above sea level). Plants were spaced 1.2 m between rows and 1.0 m within the row, fertirrigated using a drip irrigation system, and trained with vertical strings (greenhouse) or bamboo canes (open field). Pruning was done manually to regulate vegetative growth and flowering. Phytosanitary treatments were performed when necessary. To accelerate fruit setting, inflorescences of plants grown in the greenhouse were vibrated with a mechanical vibrator.

F_2 population development for BSA-seq

The complete F_2 population obtained by the interspecific cross between MEL1 and INS1 parents was phenotyped for the FN trait and a chi-square (χ^2) test was performed to assess the goodness-of-fit to a 3:1 segregation model. This population was used for BSA-seq. As the recommended size of each pool population is considered to be around 0.25, 30 individuals were selected from each of the FN and fn pools.

Young leaves were harvested from both parents and 60 F_2 individuals selected by their different FN phenotypes. Genomic DNA was extracted using the silica matrix (SILEX) protocol (Vilanova *et al.*, 2020) and checked for quality and integrity by agarose electrophoresis and NanoDrop (Thermo Fisher Scientific, Waltham, MA, USA) ratios (260/280 and 260/230), while its concentration was estimated with Qubit 2.0 Fluorometer (Thermo Fisher Scientific). Pools were made by mixing an equal amount of DNA from each selected F_2 individual (20 ng μl^{-1}) and shipped to Novogene (Novogene Europe, Cambridge, UK) where genomic libraries [paired end (PE) 150, insert 350 bp] were constructed and sequenced.

All raw reads were trimmed using the fastq-mcf tool from the Ea-utils package (Aronesty, 2013) with 'q 30 -l 50' parameters, and the overall quality was checked using FastQC v.0.12.1 (Andrews *et al.*, 2010). Clean reads were mapped against the 67/3 v.3 eggplant reference genome (Barchi *et al.*, 2019b) using BWA-MEM v.0.7.17–r1188 with default parameters (Li and Durbin, 2009). The ΔSNP index was estimated using QTL-seq software v.2.2.3 (Takagi *et al.*, 2013) from BAM files with 'n1 30 -n2 30 -D 90 -d 5' options. The calculated ΔSNP indexes were imported into R v.4.3.1 for the final plot. The raw data of the BSA-seq analysis are available at NCBI SRA (BioProject ID PRJNA1114430).

Fine-mapping of the advanced backcross introgressed fragments

Parental ILs (MEL1, DAS1, MEL5, and INS1) and the selected AB individuals were high-throughput genotyped using the 5k eggplant SPET platform (Barchi *et al.*, 2019a). The single nucleotide polymorphisms (SNPs) identified were filtered using TASSEL software to retain the most reliable ones (minor allele frequency >0.01, missing data <20%, and maximum marker heterozygosity <80%) and the graphical visualization of the genotypes was performed by using Flapjack software (Milne *et al.*, 2010). The raw data of the genotyping of the AB individuals are available at NCBI SRA (BioProject ID PRJNA1117349).

To further narrow down the candidate region, AB individuals with the shortest wild introgressed fragment were selfed and the progeny was genotyped by high resolution melting (HRM) on a LightCycler 480 Real-Time PCR (Roche Diagnostics, Meylan, France) to identify recombinants with shorter introgressions. To design primers pairs spanning the region and detect recombination breakpoints with higher precision, the genome of the four IL parents was resequenced (PE 150, insert 300 bp, 24 Gb) at the Beijing Genomics Institute (BGI Genomics, Hong Kong, China). Trimming, quality check, mapping, and SNP calling were performed against the 67/3 v.3 eggplant reference genome (Barchi *et al.*, 2019b) as in Gramazio *et al.* (2019). Integrative Genomics Viewer (IGV) tool v.2.15.2 was used for the visual exploration and detection of variants among parental genome sequences (Robinson *et al.*, 2023). Primers were designed to detect differential indel variants, which are listed in Supplementary Table S2. The informative recombinants were grown in a pollinator-free benched glasshouse and phenotyped for FN.

Candidate genes for fruit green netting and allelic variants

The genomic region narrowed down by fine-mapping was explored for candidate genes, retrieved from the functional annotation of the 67/3 v.3 reference genome (Barchi *et al.*, 2019b). Simultaneously, the whole genome resequencing data of the eight founders of the S3MEGGIC populations (Gramazio *et al.*, 2019) and the four parents of the two IL populations were interrogated for 'high' impact allelic variants predicted by SnpEff (Cingolani *et al.*, 2012) that contrasted between FN and fn. The nucleotide sequence for the best candidate gene, *SmGLK2*, was retrieved by a BLASTx search (e-value cut-off of 1e-5) against the HQ-1315 and GUIQIE-1 eggplant reference genomes (Wei *et al.*, 2020; Li *et al.*, 2021). Additionally, transcriptomic data from *S. melongena* 67/3 (SRR3884608), *S. incanum* INC1 (SRR2289250), and *S. insanum* MM0686 accession (SRR8736646) were downloaded from the NCBI SRA database. Transcripts were mapped against the *SmGLK2* gene sequence of the 67/3 v.3 reference genome using RNA STAR (Dobin and Gingeras, 2015) and checked using the IGV tool. The gene sequence of the tomato orthologue (*SlGLK2*, Solyc10g008160.3) was also analysed to elucidate the *SmGLK2* gene structure. A conservative domain analysis was performed by assessing the NCBI conserved domain server (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). The detection of nuclear localization signals (NLS) was performed through cNLS Mapper (https://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi).

Allelic variants of the *SmGLK2* gene were also interrogated in whole-genome resequencing data of 178 *S. melongena* accessions and 22 wild relative species from the eggplant G2P-SOL germplasm core collection (Gramazio *et al.*, 2019; Barchi *et al.*, 2021 and unpublished data) for which FN phenotyping was available. For this purpose, raw reads were trimmed by SOAPnuke software (Chen *et al.*, 2018) with filter parameters '-l 20 -q 0.5 -n 0.03 -A 0.28' and aligned against de 67/3 v.3 eggplant reference genome (Barchi *et al.*, 2019b) using the BWA-MEM program v.0.7.17–r1188 (Li and Durbin, 2009). Subsequently, the clean reads mapped in the genomic region of *SmGLK2* were extracted with samtools (Danecek *et al.*, 2021) and *de novo* assembled using Megahit v.1.2.9 with default parameters (Li *et al.*, 2015). The raw data of the assembled *SmGLK2* from the G2P-SOL core collection is available at NCBI SRA (BioProject ID PRJNA977872). A multiple sequence alignment of the assembled sequences was then performed using the MAFFT program v.7 (Katoh and Toh, 2008) and the results were visualized in the Jalview alignment editor v.2.11.2.6 (Waterhouse *et al.*, 2009). The resulting alignments were used to identify the candidate variants in the *SmGLK2* gene of the G2P-SOL core collection.

To gain insight into the relationship of the different *GLK*s (i.e. *GLK1* and *GLK2*) across different species, a phylogenetic analysis was also performed using the predicted protein sequences from Arabidopsis, eggplant, tomato, and pepper. Protein alignments and dendrogram construction were performed using MEGA v.11.0.10 software (Kumar *et al.*, 2018) via the maximum likelihood method with default settings, and graphical

representation was displayed and edited using the iTOL v.4 software (Letunic and Bork, 2019).

Protein extraction and immunoblot analysis

Fruit peel tissues from eggplant accessions MEL1 and INS1 were harvested and ground with liquid nitrogen. Total proteins were extracted with extraction buffer [50 mM Tris–HCl (pH 8), 10 mM $MgCl_2$, 5% glycerol, 0.1% Triton X-100, 14 mM β-mercaptoethanol]. Samples were centrifuged twice at 16 200 *g* for 15 min to remove cell debris and recover the cleanest supernatant containing proteins. Fourteen milligrams of protein was denatured at 95 °C for 5 min in reducing loading buffer and separated by 12% SDS-PAGE. Proteins were transferred (30 V, 1 h, 4 °C) to polyvinylidene fluoride membranes (Immobilon-E PVDF Membrane) and blocked overnight at 4 °C with 5% non-fat dry milk (Sveltesse®, Nestlé) and gentle shacking. To detect the gene product of *SmGLK2*, membranes were incubated with an anti-sera peptide-based polyclonal rabbit GLK2-antibody (1:3,000) and a peroxidase-conjugated anti-rabbit antibody (1:10 000) (Merck, Darmstadt, Germany) for 1 h at room temperature in TBST [20 mM Tris–HCl (pH 7.6), 20 mM NaCl, 0.1% (v/v) Tween-20] supplemented with 2% non-fat dry milk. The anti-sera peptide-based polyclonal rabbit GLK2-antibodies were raised against the SmGLK2 N-terminal sequence (synthetic peptide VSPPLSYTNENENY, 5–18 aa; and NMKSKSKEAKKSSG, 70–83 aa) (Davids Biotechnologie GmbH, Regensburg, Germany). Peroxidase activity was developed in ECL Plus western blotting detection reagents (GE Healthcare, Little Chalfont, UK) and the chemiluminescence signal was captured with the LAS-1000 imaging system (Fujifilm, Tokyo, Japan).

RNA extraction, sequencing, and expression analysis

Samples from different INS1 tissues were collected and frozen immediately in liquid nitrogen (all samples contained at least five different individuals) and kept at −80 °C for later processing. Samples were ground in liquid nitrogen using a mortar and pestle, and 100 mg of each was used for RNA extraction using an RNAqueous Total RNA Isolation kit (Thermo Fisher Scientific) following the manufacturer's instructions. Purity and quantity of RNA was assessed in a NanoDrop (Thermo Fisher Scientific), while integrity was checked by 1.5% agarose electrophoresis. For sequencing of the *GLK2* transcripts, 0.5 µg of total RNA was treated with 1 U of RNase-free DNaseI (Thermo Fisher Scientific) following the manufacturer's protocol and then retro-transcribed using the UltraScript cDNA Synthesis Kit (PCRBio) using oligo(dT) for priming. The *GLK2* transcripts were amplified using LaTaq (Takara, Japan) and sequenced by Sanger sequencing (DNA Sequencing Service, IBMCP-UPV, Valencia, Spain).

For RT-qPCR analysis, 1 µg of total RNA was treated with 1 U of RNase-free DNaseI (Thermo Fisher Scientific) following manufacturer's protocol. Treated RNA was retro-transcribed using the UltraScript cDNA Synthesis Kit (PCRBio) using oligo(dT) for priming. A control without RT (RT−) was included. For qPCR, the obtained cDNAs were diluted 1:5 and 2 µl of such dilution was used as template for qPCR. The expression of *GLK1* and *GLK2* was measured by qPCR in a LightCycler 480 System (Roche) using HOT FIREPol EvaGreen HRM Mix (Solis BioDyne). The program used, as suggested by Solis BioDyne was: 12 min incubation at 95 °C, followed by 40 cycles of 15 s at 95 °C, 20 s at 60 °C, and 15 s at 72 °C. At the end of this program, a melting curve program was added following the manufacturer's recommendations to assess the presence of a single amplification product for each primer pair. All samples contained three technical replicates. Triplicates for the RT− and a non-template control for each gene were also included giving no amplification signal. The RT-qPCR primers showed an amplification efficiency close to 100% so, the $2^{-\Delta\Delta C_T}$ method (Livak and Schmittgen, 2001) was used for the calculation of the transcript expression changes. The *PP2AA2* (SMEL010g337680) and *AC97* (SMEL011g378080) genes were used as reference genes for normalization, and the seedling sample was used as calibrator for the rest of the tissues. To assess significant differences between groups, Student's *t*-test was performed using R software. All primers used are listed in Supplementary Table S3.

Confocal microscopy and chlorophyll content measurement

To determine differences in chloroplast number and structure between the netting and the uniform distribution of chlorophylls in the fruit peel, three replicates from FN and fn tissues were analysed. Exploiting chlorophyll autofluorescence and emission spectra, confocal microscopy was used for monitoring chloroplast appearance and pigment content. Following the confocal laser scanning microscopy protocol developed by D'Andrea *et al.* (2014), thin sections of pericarp were cut with a vegetable peeler or a single-edge razor blade. Samples were mounted on a microscope slide with a drop of water with the exocarp facing the coverslip. Samples were imaged in an AxioObserver 780 (Zeiss) confocal microscope using a 488 nm argon laser and a ×40 water-immersion objective, and chlorophyll emission was recorded between 634 and 723 nm. For each sample, a stack of 10 nm was taken, taking one optical section every 1 µm, as well as a lambda scan between 500 nm and 700 nm with a bandwidth of 8 nm and a step size of 5 nm to verify that no other pigments were emitting in this range.

Chlorophyll content was measured in five replicates of each FN and fn fruit. Samples of 100 mg were extracted with 1 ml of ice-cold 80% acetone (v/v) by mixing for 24 h in darkness. After centrifugation at 13 800 *g* for 15 min at 4 °C, the supernatant was diluted 10-fold with 80% acetone, and the absorbance was measured at 663, 646, and 470 nm. Chlorophyll *a* and *b* concentrations were then calculated according to Lichtenthaler and Wellburn (1983) and expressed in mg g^{-1} fresh weight. To assess significant differences between groups, Student's *t*-test was performed using R software.

Results

BSA-seq for the fruit green netting trait

Through a genome-wide association study (GWAS) in the S3MEGGIC population (Mangino *et al.*, 2022), Arrones *et al.* (2022) identified a significant association peak for chlorophyll content in the eggplant fruit peel on chromosome 8, which co-localized with the *SmAPRR2* gene. They also identified a minor but significant peak on chromosome 4 in the genomic region between 3.23 and 6.35 Mb. However, no clear candidate gene was identified since the SnpEff software did not predict high-effect variants contrasting for the S3MEGGIC founders with and without chlorophylls in the fruit peel. The phenotyping of this population for fruit chlorophyll pigmentation was performed using a binary classification (presence/absence) without discriminating between uniform distribution or netting. This fact, together with previous studies relating this genomic region with the FN in eggplant (Tigchelaar, 1968; Doganlar *et al.*, 2002; Frary *et al.*, 2014), led us to hypothesize a possible relationship of this peak with the FN trait.

To confirm the candidate genomic region for the FN, an F_2 population was developed by crossing MEL1 (fn) and INS1 (FN) parents. The FN trait displayed classical patterns of Mendelian inheritance of a single dominant gene

in the F_1 and F_2 populations. Among the 120 F_2 individuals, 74.17% (89 out of 120) were FN while 25.83% (31 out of 120) were fn (Fig. 1A). Chi-square (χ^2) analysis was in agreement with the 3:1 segregation with a value of 0.044 (d.f. = 1.0, $P < 0.05$). BSA-seq was performed using 30 individuals from each of the FN and fn pools from the F_2 segregating population. DNA pool sequencing resulted in an average of 117 million 150 bp raw reads yielded per bulk with a mean coverage of 33×. About 9 436 870 high-quality SNPs differentiating between parental lines were identified. Of these, 2 192 129 SNPs were selected by QTL-seq software to compute the ΔSNP index with a 99% confidence interval identifying a unique significant region of 2.76 Mb on chromosome 4, between 3 742 532 bp and 6 497 820 bp, which was consistent with the results obtained in the S3MEGGIC population (Fig. 1A; Supplementary Table S4).

Fine-mapping of the candidate genomic region

Since two AB populations segregating for the FN trait were available, both were used for further inspecting the candidate genomic region identified by GWAS and BSA-seq analyses. Specifically, five AB individuals from the MEL1×DAS1 set and 20 AB individuals from the MEL5×INS1 were selected, all showing the FN phenotype. All of the 25 AB individuals presented a wild DAS1 or INS1 introgression in a *S. melongena* background (MEL1 and MEL5, respectively) at the beginning of chromosome 4. The size and the physical location of the wild introgressed fragments varied according to the recombination and the selection made in the previous generation. The genotyping of these AB individuals by the eggplant SPET platform allowed us to narrow down the candidate region to 0.80 Mb, between 5 202 058 bp and 5 996 586 bp (Fig. 1B; Supplementary Table S5). There were approximately 48 putative candidate genes in this narrowed region (Supplementary Table S6).

To further delimit this region, 1761 individuals were generated by selfing the AB individuals with the shortest wild introgressed fragment and screened following the edges-to-core approach using the primers listed in Supplementary Table S2. As a result, 105 recombinant individuals were identified (Supplementary Table S7) and based on their genotype and phenotype, the region of interest was reduced to 0.36 Mb, between 5 412 659 bp and 5 774 878 bp of chromosome 4 (Fig. 1C). The number of putative candidate genes in this region was reduced to 14 genes (Fig. 1D and highlighted genes are shown in Supplementary Table S6).

Identification of a responsible candidate gene

The 14 candidate genes annotated according to the 67/3 v.3 reference genome (Barchi *et al.*, 2019b) were interrogated for 'high' impact allelic variants following SnpEff analysis, but none of the genes carrying such variants was consistent with the contrasting phenotypes. However, we decided to further interrogate the *SmGLK2* gene (SMEL004g203570, 5 457 658–5 461 306 bp), which was within the confidence interval (Fig. 1D), since it had previously been described as a positive regulator of chloroplast development and pigment accumulation in other *Solanaceae* species (Powell *et al.*, 2012; Brand *et al*, 2014; Nguyen *et al.*, 2014).

Although no high-effect variants were predicted by SnpEff in the coding sequence of *SmGLK2*, some intronic variants were shared by the FN S3MEGGIC founders and IL parents (INC1, DAS1, and INS1), compared with fn ones. Studying the gene structure in the other available eggplant reference genomes, HQ-1315 and GUIQIE-1 (Wei *et al.*, 2020; Li *et al.*, 2021), it was observed that the *SmGLK2* gene model was split into two different genes (Smechr0400347-8 and EGP26827-8) between the second and third exon (Fig. 2A). Furthermore, slight differences between *SmGLK2* gene structure compared with its tomato orthologue were identified (*SlGLK2*, Solyc10g008160.3). The *SlGLK2* 5′-untranslated region (UTR) was divided into three different exons, the latter two coinciding with the two first exons of the *SmGLK2*. In addition, one extra exon was observed between the eggplant second and third exons (Fig. 2A). Then, different transcriptomic data were analysed to reveal whether a gene mis-annotation existed in the eggplant reference genomes. When analysing the 67/3, INC1, and MM0686 transcriptomes against the *SmGLK2* gene sequence of the 67/3 v.3 reference genome (Barchi *et al.*, 2019b), some transcripts also aligned with the intronic regions between the second and third exon (Fig. 2B), suggesting an incorrect annotation in the available eggplant reference genomes, all of them developed from fn accessions. We therefore propose the presence of an extra exon between the second and third exon for the 67/3 v.3 reference genome and the merging of the two annotated genes in the HQ-1315 and GUIQIE-1 eggplant reference genomes into a single one (Wei *et al.*, 2020; Li *et al.*, 2021; Fig. 2C).

To investigate the *GLK*s relationship in relevant species in the *Solanaceae* family, a phylogenetic analysis was performed revealing orthology among GLK2-like proteins within the family (Fig. 3; Supplementary Dataset S1).

Identification of causative allelic variants

In the suggested additional exon, a frameshift mutation was identified when comparing FN accessions against the 67/3 v.3 eggplant reference genome (Barchi *et al.*, 2019b), which was derived from an fn accession. This mutation consisted of a small insertion A(T)G in the 5 459 673 bp position (Fig. 2C). The absence of this insertion leads to a premature stop codon (Fig. 2D). Analysing the resequencing data, all FN parents (the wild INC1, DAS1, and INS1) presented the T insertion, while the fn parents were identical to the 67/3 reference (Supplementary Dataset S2). The absence or presence of this insertion was also

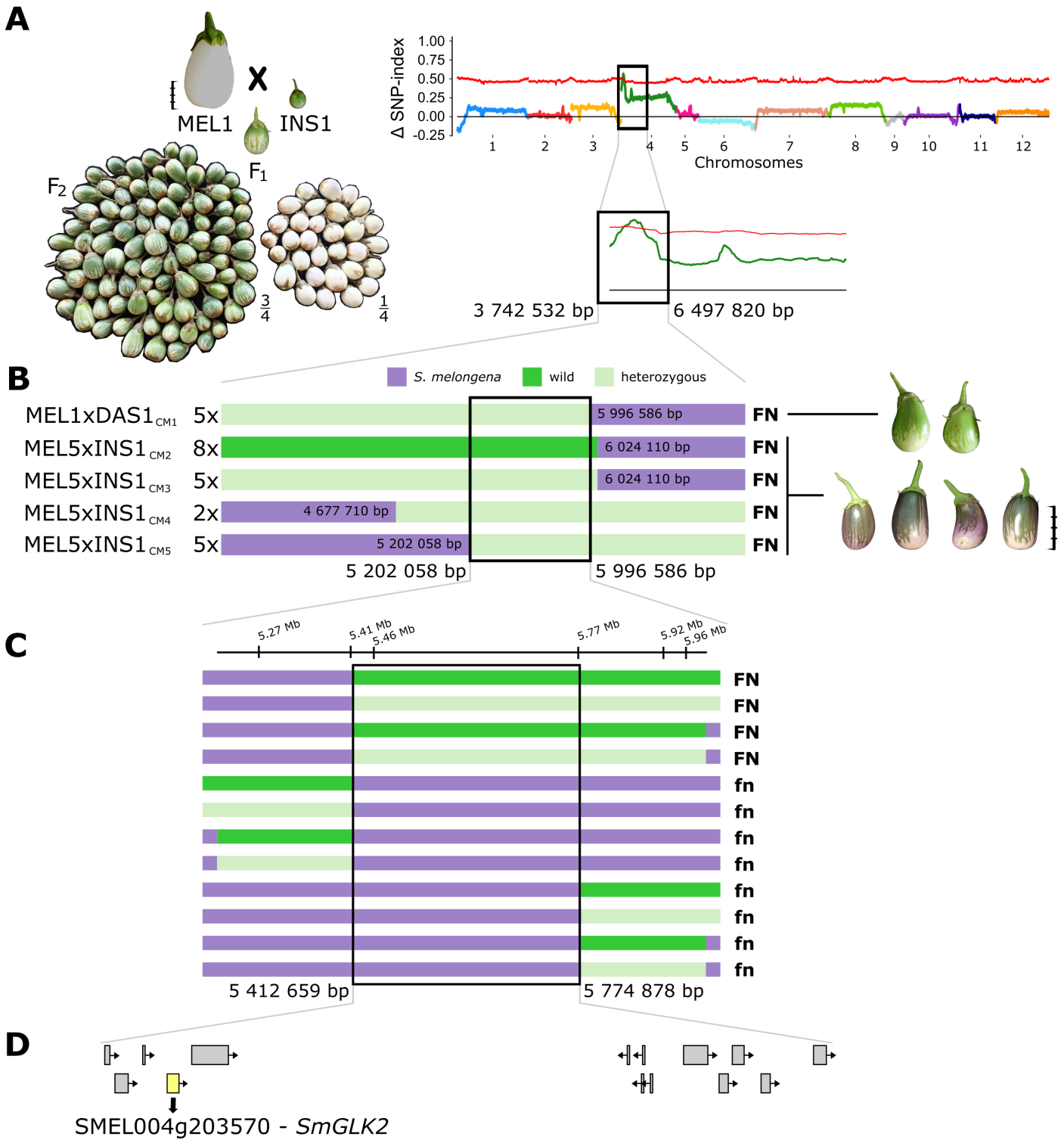

Fig. 1. Process followed to narrow down the candidate genomic region for the fruit green netting (FN) trait. (A) On the left, segregation of the F_2 population developed by the inter-cross of a white-fruited *S. melongena* MEL1 accession and the green netting-fruited *S. insanum* INS1 accession. Scale bar: 5 cm. On the right, BSA-seq results from the green netting-fruited and non-netting-fruited pools. The association significance threshold (ΔSNP index) is represented with a red line. The black box delimits the significant candidate region. (B) Representation of the 25 selected advanced backcross (AB) individuals from two IL populations used for coarse mapping (CM), MEL1×DAS1 (five accessions) and MEL5×INS1 (20 accessions), covering the candidate region on chromosome 4. In purple the *S. melongena* background (coming from MEL1 and MEL5), in dark green the wild introgressions coming from *S. dasyphyllum* (DAS1) and *S. insanum* (INS1), and in light green the heterozygotes. On the right, the phenotype of the fruits obtained from

the selected AB at a scale based on the real fruit size, all of them presenting the fruit green netting (FN). Scale bar: 5 cm. (C) Representation of a set of the recombinant individuals generated by selfing. The genomic indel positions used for the edges-to-core fine-mapping are specified at the top. The phenotype of the different individuals is indicated on the right. (D) The 14 candidate genes annotated under the fine-mapped region according to the 67/3 v.3 reference genome (Barchi *et al.*, 2019b). The candidate *SmGLK2* gene (SMEL004g203570) is indicated in yellow.

confirmed *in silico* in the 67/3 (fn), INC1 (FN), and MM0686 (FN) transcripts (Fig. 2B), and *in vivo* in INS1 (FN) transcripts (Supplementary Fig. S2).

To further study the *SmGLK2* allelic variants responsible for the loss of the FN trait over the course of domestication, a diverse germplasm collection was analysed. The aim was to trace the changes that the gene had undergone from wild to cultivated species. Therefore, we selected 178 *S. melongena* accessions (76 FN and 102 fn) and 22 wild relative species (all of which showed the FN phenotype) from the eggplant G2P-SOL germplasm core collection (Supplementary Table S1). All 76 *S. melongena* FN accessions presented the T insertion. However, only 97 out of 102 fn accessions showed the absence of the T insertion, similar to 67/3. The remaining five accessions exhibited a different haplotype, with two new frameshift mutations: a 68 bp deletion on exon 1 from 5 457 779 to 5 457 847 bp and a 41 bp deletion on exon 5 from 5 461 264 to 5 461 305 bp. The 22 wild relative species analysed also presented the T insertion, indicating a complete and fully functional *GLK2* gene.

Differences in GLK2 protein structure

Analysis of the full-length amino acid sequence of *SmGLK2* in the NCBI conserved domain server, confirmed that the premature stop codon generated by the absence of the T insertion was translated into a truncated protein in fn accessions. Precisely, protein sequence was reduced from 313 aa to 209 aa, directly affecting the golden2-like transcription factor domain at the C-terminal region with a reduction in 148 aa (from a 224 aa to a 76 aa golden2-like transcription factor domain), although without affecting the NLS since it is located in the N-terminal part of the protein (Fig. 2E). Western blot analysis with a polyclonal rabbit GLK2 antibody showed a differential molecular mass banding pattern of the target protein for INS1 (FN) and MEL1 (fn) samples. A difference in band migration was observed, with the apparent molecular masses of full-length and truncated versions of GLK2 being of approximately 30.5 kDa for FN and 28.75 kDa for fn, respectively (Fig. 2E). RT-qPCR analysis of RNA extracted from INS1 seedlings, young leaves, flowers, and fruits revealed that only *GLK2* is expressed in fruits (Fig. 2E; Supplementary Fig. S3).

Cytological observations

Confocal microscopy was used to determine differences in chloroplast number and chlorophyll accumulation in FN and fn fruits. Exocarp from areas displaying netting (FN) and uniform green coloration (fn) exhibited important differences (Fig. 4A). Since there is a single peak in the fluorescence emission spectra (Fig. 4B), it is confirmed that the fluorescence observed comes exclusively from chlorophyll. In the FN sample coming from the proximal part of a fruit with dark green netting, imaging of chlorophyll autofluorescence indicated higher chlorophyll accumulation due to the presence of bigger and brighter spots. Instead, in the fn sample, chlorophyll autofluorescence was dim and the spots showed irregular shapes suggesting that the levels of accumulation of chlorophyll in the chloroplasts was low. In addition, significant differences in chlorophyll content were observed between FN and fn samples, with significantly lower chlorophyll *a* and *b* content in fn samples (Fig. 4C). Overall, our results indicate that the presence of a full-length GLK2 protein leads to the accumulation of higher amounts of chlorophylls, which results in the dark-green netted pattern in eggplant fruits.

Discussion

Eggplant domestication resulted in a diversification of fruit colour at the physiologically unripe stage, characteristic of commercial maturity. However, in recent times, modern breeding has resulted in a predominance of purple and dark purple colours, especially in western markets (Taher *et al.*, 2017; Page *et al.*, 2019; Page and Chapman, 2021). Consumers demand new vegetable products with different and unusual aesthetic characteristics, broadening the consumers' options and fostering vegetable consumption (Di Gioia *et al.*, 2020). In this study, we have successfully harnessed different plant materials, combining germplasm and classical bi-parental (F_2) populations with advanced bi- (ILs) and multi-parent (MAGIC) populations, for the identification of the genes underlying FN in eggplant. This is of interest as it increases the diversity of the eggplant fruit colour palette.

Research on the genetics in eggplants of FN, also known as green reticulation or variegation, has been limited. Daunay *et al.* (2004) associated FN with a trait under monogenic dominant control following an F_2 proportion of 3:1. We confirmed the dominance hypothesis by the development of the F_1 hybrid and F_2 population from an FN×fn cross. In the published studies on this trait, *S. linneanum* was used as a parental donor since FN is commonly present in this wild species (Doganlar *et al.*, 2002; Daunay *et al.*, 2004; Frary *et al.*, 2014). Although a few commercial varieties display the FN trait, it has been counter-selected during domestication because the green colour was erroneously related to immaturity (Page *et al.*, 2019). As occurs

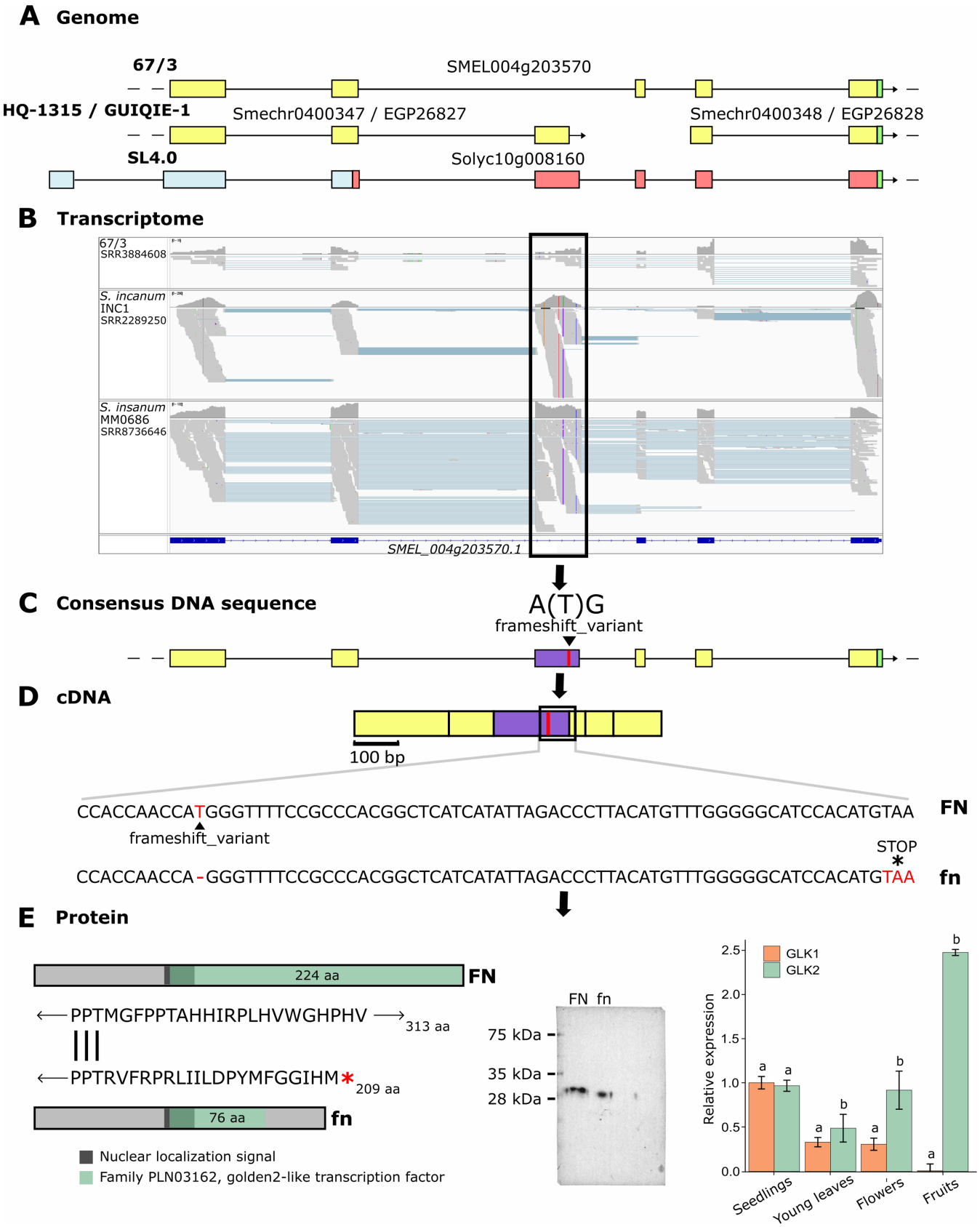

Fig. 2. From the genomic sequence of the *SmGLK2* gene to protein structure. (A) *SmGLK2* annotated gene structure in 67/3, HQ-1315, and GUIQIE-1 eggplant reference genomes and SL4.0 tomato genome [5′-untranslated region (UTR) in blue, eggplant annotated exons in yellow, tomato annotated exons in red, and 3′-UTR in green]. (B) The *S. melongena* 67/3 (SRR3884608), *S. incanum* INC1 (SRR2289250), and *S. insanum* MM0686 accession

(SRR8736646) transcripts aligned against the *SmGLK2* gene sequence from the 67/3 v.3 eggplant reference genome and visualized in the IGV tool (Robinson *et al.*, 2023). The candidate structural variation identified is indicated with a purple line. (C) The suggested *SmGLK2* gene structural annotation as a consensus of previous information with the extra proposed exon in purple. The identified structural variation in the gene sequence is indicated with a black arrowhead and a red line. (D) Comparison of mRNA sequences for presence (FN) or absence (fn) of the fruit green netting trait. The premature stop codon downstream of the indel is indicated with an asterisk. (E) From left to right: comparison of the protein structure and sequence around the indel site for FN and fn with the nuclear localization signal (NLS) indicated in dark grey and the golden2-like transcription factor domain in green; western blot showing the difference in apparent molecular mass of the GLK2 protein from FN and fn tissues; and differences in the expression level of *GLK1* and *GLK2* from different tissues of *S. insanum* INS1. *PP2AA2* gene was used as a reference gene for normalization. Columns and error bars represent means and standard deviations of three replicates. Different letters above the columns indicate statistical significance ($P<0.05$) according to Student's *t*-test.

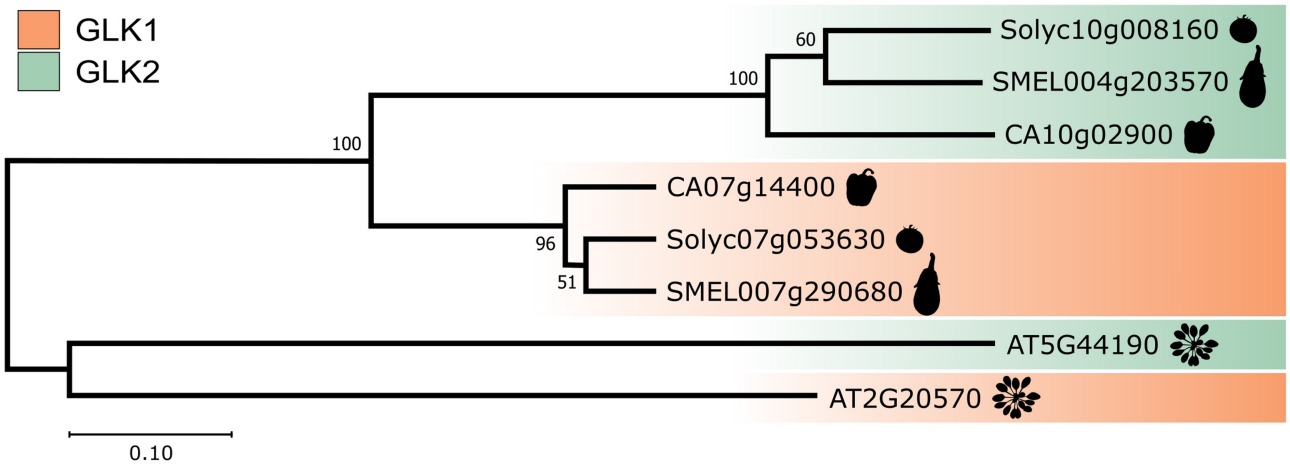

Fig. 3. Maximum likelihood dendrogram of GLK proteins from Arabidopsis and *Solanaceae* species (eggplant, tomato, and pepper). In orange, GLK1-like proteins and in green, GLK2-like proteins. AT, *Arabidopsis thaliana*; SMEL, *Solanum melongena* (eggplant); Solyc, *S. lycopersicum* (tomato); CA, *Capsicum annuum* (pepper). Scale bar refers to a 0.10 distance and gene/IDs correspond to those from Solgenomics.net.

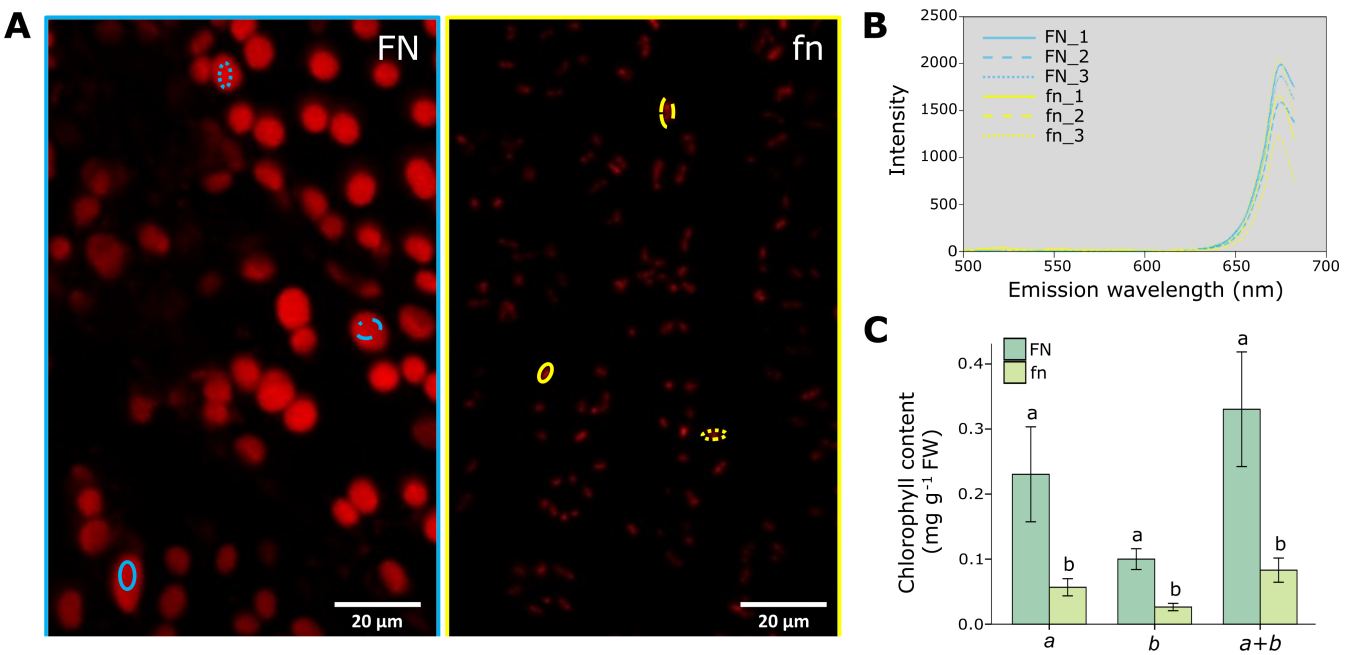

Fig. 4. Evaluation of chloroplast number and structure and chlorophyll content. (A) Confocal images of exocarp layer showing presence (FN) or absence (fn) of the fruit green netting trait. The FN sample comes from the proximal part of a fruit showing a dark green netting (*SmGLK2* expression), while the fn sample comes from the proximal part of a fruit with uniform distribution of chlorophylls. (B) Fluorescence emission spectra of FN and fn samples ($n=3$) after excitation at 488 nm. Fluorescence intensity is represented relative to the total fluorescence of the sample. (C) Concentrations of chlorophyll *a*, chlorophyll *b*, and total chlorophyll content in FN and fn samples ($n=5$). Different letters above the columns indicate statistical significance ($P<0.05$) according to Student's *t*-test.

with other domestication traits, fn is recessive to the dominant allele of the wild species (Doganlar *et al.*, 2002; Frary *et al.*, 2014). This supports the paradigm that domestication usually involves the loss of gene function or regulation (Lester and Hasan, 1991).

In our previous study, through a GWAS in the S3MEGGIC population, we identified significant associations for the presence of chlorophyll pigmentation in the eggplant fruit peel (Arrones *et al.*, 2022). The Manhattan plot revealed one major peak on chromosome 8, which resulted in the identification of a gene similar to *ARABIDOPSIS PSEUDO RESPONSE REGULATOR2* (*APRR2*, SMEL008g315370). The *SmAPRR2* gene was suggested as the best candidate gene for uniform fruit chlorophyll pigmentation. The GWAS also revealed a minor peak, although significant, at the beginning of chromosome 4. This region was rejected as SnpEff did not predict contrasting high-effect variants among the S3MEGGIC founders. Since previous studies identified a QTL related to FN on chromosome 4 explaining 67–78% of the variation (Doganlar *et al.*, 2002; Daunay *et al.*, 2004; Frary *et al.*, 2014), we decided to further investigate this genomic region. Moreover, the minor peak could have been obtained by phenotyping imprecisions as there was no discrimination between netting or uniform distribution of chlorophylls in the Arrones *et al.* (2022) study. These results together with a BSA-seq of an F_2 population segregating for the trait and the fine-mapping of two ABs populations, allowed us to confirm the candidate genomic region for FN and to narrow it down to 0.36 Mb (Fig. 1). None of the 14 genes in this region presented variants consistent with the FN and fn phenotypes. However, one of these genes annotated as similar to *GOLDEN2-LIKE2* (*SmGLK2*, SMEL004g203570) was studied as the best candidate. This gene belongs to the widely conserved GARP family of MYB transcription factors whose functions are reported to be involvement in chloroplast development in different species such as *Zea mays* (maize), *Oryza sativa* (rice), *Physcomitrella patens* (moss), and Arabidopsis (Hall *et al.*, 1998; Rossini *et al.*, 2001; Fitter *et al.*, 2002; Yasumura *et al.*, 2005; Waters *et al.*, 2009). Mechanism of action and roles in regulating chloroplast development and photosynthesis have been defined (Zubo *et al.*, 2018; Li *et al.*, 2020; Yeh *et al.*, 2022).

Through different complementary analyses, we identified an erroneous annotation of the *SmGLK2* in the available reference genomes, all of them developed from accessions with fn phenotype (Barchi *et al.*, 2019b; Wei *et al.*, 2020; Li *et al.*, 2021). An improved annotation of the *SmGLK2* gene was achieved thanks to the availability of resequencing and transcriptomic data (Fig. 2). As a result, we found an extra coding exon where we identified a small insertion. The complete sequence of the *SmGLK2* gene is given by the presence of this insertion, while its absence results in a frameshift mutation and a premature stop codon in the mRNA sequence. This was the reason for the fn phenotypes in our populations. This could also explain the HQ-1315 and GUIQIE-1 eggplant reference genomes annotation of *SmGLK2* as two different genes (Wei *et al.*, 2020; Li *et al.*, 2021). Furthermore, we verified the effects of this mutation at the protein level by western blotting analysis. The presence of the small insertion resulted in a complete and fully functional *SmGLK2* gene product, while its absence resulted in a truncated golden2-like transcription factor domain leading to a protein of smaller molecular mass (Fig. 2).

Reverse genetics approaches, such as overexpression of the candidate genes or knock-out by CRISPR/Cas are the optimal gene validation techniques. However, due to the recalcitrance of eggplant to genetic transformation (García-Fortea *et al.*, 2020; Mir *et al.*, 2021), there is a need for alternative indirect validation methods in this crop until new, genotype-independent genetic transformation protocols become available. Validation through diverse experimental populations and the screening of large germplasm collections can provide a good alternative (Arrones *et al.*, 2022). For this reason, a set of *S. melongena* accessions and wild species, including *S. insanum*, *S. incanum*, *S. linneanum*, *S. macrocarpon*, *S. humile*, and *S. tettense*, from the eggplant G2P-SOL germplasm core collection were analysed. This methodology also allows for tracing the evolutionary changes that the gene has undergone. The complete *SmGLK2* gene sequence was confirmed in the wild relative species, all of them with FN phenotype. The small insertion identified was validated in the *S. melongena* accessions by correlating genotypic and phenotypic data. Besides, two additional, rare frameshift mutations responsible for the fn phenotype were identified. This indicates that the lack of the FN trait has arisen and been selected independently several times during eggplant domestication and diversification.

All these results together provide strong evidence of *SmGLK2* gene as the gene controlling the FN trait. These results are also supported by previous studies performed in other *Solanaceae* species. In pepper, *CaGLK2* has also been described as a major gene controlling chlorophyll content and chloroplast compartment size in immature fruit (Brand *et al*, 2014). Three null mutations resulting in the appearance of premature stop codons and leading to truncated proteins were identified as responsible for a reduction in the chlorophyll content of pepper fruits. In tomato, *SlGLK2* has been related to the green shoulder trait, also referred to as the *Uniform ripening* (*U*) locus, which is very similar to FN in eggplant (Powell *et al.*, 2012). It has been demonstrated that the overexpression of *SlGLK2* increases chloroplast number and size, producing homogeneously dark-green fruits with enhanced nutritional quality (Powell *et al.*, 2012; Nguyen *et al.*, 2014). An A insertion causing a frameshift and a premature stop codon encoding a truncated protein were identified as responsible of the *u* phenotype (Powell *et al.*, 2012). Therefore, *GLK2* homologs have been shown to share evolutionary relationship and potentially similar roles in chloroplast development within the *Solanaceae* (eggplant, tomato, and pepper). It is noteworthy that the three *Solanaceae GLK2* genes are orthologues of each other, as are the three *GLK1* genes (Fig. 3). Thus, the most likely scenario

is that a duplication event arose early during the evolution of *Solanaceae*, and the specialization of these *GLK* genes occurred independently of each other. This is further supported by their differential expression in different plant organs. In eggplant, tomato, and pepper, it has been observed that *GLK1* expression is more important in vegetative tissues, whereas *GLK2* predominates in fruits (Brand *et al.*, 2014; Nguyen *et al.*, 2014). Although further analysis could be performed to reach the numerous conclusions and evidence provided in other species, there is no reason to doubt that *SmGLK2* is functionally equivalent to its orthologues.

Moreover, different studies proposed distinct roles for *SlGLK2* and *SlAPRR2* in regulating chloroplast development in tomato fruits (Pan *et al.*, 2013; Nadakuduti *et al.*, 2014). Ultrastructural analysis of chloroplasts in netted areas of tomato fruits showed more developed chloroplasts, with more grana in the stroma than those of non-netted tomatoes (Powell *et al.*, 2012; Nguyen *et al.*, 2014). Likewise, the results obtained by confocal imaging of chlorophyll autofluorescence and chlorophyll content measurements also suggest differences in chloroplast development and chlorophyll accumulation in eggplant fruits as a result of *SmGLK2* and *SmAPRR2* expression leading to two different phenotypic traits. This is coherent with our observation of higher chlorophyll contents in eggplant netted areas (Fig. 4). Overall, these results indicate that *SmGLK2* expression leads to more developed chloroplasts with higher chlorophyll accumulation. This may also be the reason why the green colour coming from the netting is persistent and remains throughout ripening, until turning yellow when over-ripe, while chlorophylls from uniform pigmentation fade rapidly and give rise to yellow pigments (Page and Chapman, 2021).

By using multiple *in silico* and *in vivo* methodologies we have found that *SmGLK2* gene is responsible for the FN trait, with a 100% association between disruptive mutations in this gene and fn phenotypes. The eggplant FN has important implications for fruit visual quality particularly because eggplant is commercialized when it is still physiologically immature, and the expression of the trait is very intense in the proximal part of the fruit. The rescue of this domestication trait for future eggplant breeding programmes could be of interest to widen the diversity of the fruit colour palette. In addition, due to a higher concentration of functional chloroplasts and chlorophyll accumulation, *SmGLK2* could also be a valuable target for enhancing fruit nutritional quality.

Supplementary data

The following supplementary data are available at *JXB* online.

Fig. S1. Appearance of eggplant fruits in presence (FN) or absence (fn) of the green netting.

Fig. S2. Alignment of *Solanum insanum* INS1 transcripts against the *SmGLK2* gene sequence.

Fig. S3. Expression of *GLK1* and *GLK2* in different tissues of *Solanum insanum* INS1, normalized with *ACT97* gene.

Table S1. List of *Solanum melongena* accessions and wild relative species from the eggplant G2P-SOL germplasm core collection.

Table S2. List of primers used for fine-mapping.

Table S3. List of primers used for RNA sequencing and expression analyses.

Table S4. BSA-seq results.

Table S5. Correspondence between the AB individuals and the NCBI SRA repository Bioproject ID PRJNA1117349.

Table S6. List of putative candidate genes.

Table S7. Genotype of the 1761 accessions used for the edges-to-core fine-mapping approach.

Dataset S1. MEGA alignment of GLKs protein sequences from *Arabidopsis thaliana* and *Solanaceae* species (eggplant, tomato, and pepper).

Dataset S2. Resequencing data of the four IL parents (MEL1, DAS1, MEL5, and INS1).

Author contributions

JP, PG, and SV conceived the idea and supervised the study. AA, MP, and PG performed the field trials. AA, VB-F, PG, and SV performed the analysis of the S3MEGGIC, F_2 and AB populations and the *GLK2* gene structure. EP, LB, and GG performed the analyses of the G2P-SOL core collection. AA and SM performed confocal and transcript and protein analysis. AA and PG prepared a first draft of the manuscript and the rest of the authors reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This work has been funded by grants PID2021-128148OB-I00 funded by MCIN/AEI/10.13039/501100011033/and by 'ERDF A way of making Europe', PDC2022-133513-I00 funded by MCIN/AEI/10.13039/501100011033/and European Union Next Generation EU/PRTR, CIPROM/2021/020 from Conselleria d'Innovació, Universitats, Ciència i Societat Digital (Generalitat Valenciana, Spain), and by European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 677379 (G2P-SOL project: Linking genetic resources, genomes and phenotypes of Solanaceous crops). AA is grateful to Spanish Ministerio de Ciencia, Innovación y Universidades for a predoctoral (FPU18/01742) contract. VB-F is grateful to Generalitat Valenciana for grant INVEST/2022/146, funded by European Union, Next Generation EU. SM would like to thank funding support from UPV and the Spanish Ministerio de Universidades under the program María Zambrano funded by the European Union Next Generation-EU. PG is grateful to Spanish Ministerio de Ciencia

e Innovación for a post-doctoral grant (RYC2021–031,999-I) funded by MCIN/AEI/10.13039/501,100,011,033 and the European Union through NextGenerationEU/PRTR.

Data availability

The data presented in the study are deposited in the NCBI SRA repository, BioProject IDs PRJNA392603 (MAGIC founders resequencing), PRJNA1114430 (BSA-seq), PRJNA1117349 (AB individuals SPET genotyping), and PRJNA977872 (GLK2 sequence of the G2P-SOL core collection accessions).

References

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Aronesty E. 2013. Comparison of sequencing utility programs. The Open Bioinformatics Journal **7**, doi: 10.2174/1875036201307010001.

Arrones A, Mangino G, Alonso D, Plazas M, Prohens J, Portis E, Barchi L, Giuliano G, Vilanova S, Gramazio P. 2022. Mutations in the SmAPRR2 transcription factor suppressing chlorophyll pigmentation in the eggplant fruit peel are key drivers of a diversified colour palette. Frontiers in Plant Science **13**, 1025951.

Barchi L, Acquadro A, Alonso D, *et al*. 2019a. Single Primer Enrichment Technology (SPET) for high-throughput genotyping in tomato and eggplant germplasm. Frontiers in Plant Science **10**, 1005.

Barchi L, Pietrella M, Venturini L, *et al*. 2019b. A chromosome-anchored eggplant genome sequence reveals key events in Solanaceae evolution. Scientific Reports **9**, 11769.

Barchi L, Rabanus-Wallace MT, Prohens J, Toppino L, Padmarasu S, Portis E, Rotino GL, Stein N, Lanteri S, Giuliano G. 2021. Improved genome assembly and pan-genome provide key insights into eggplant domestication and breeding. The Plant Journal **107**, 579–596.

Brand A, Borovsky Y, Hill T, Rahman KAA, Bellalou A, Van Deynze A, Paran I. 2014. CaGLK2 regulates natural variation of chlorophyll content and fruit color in pepper fruit. Theoretical and Applied Genetics **127**, 2139–2148.

Chapman MA. 2019. Introduction: The importance of eggplant. In: Chapman MA, ed. The eggplant genome. Cham: Springer, 1–10.

Chen Y, Chen Y, Shi C, *et al*. 2018. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. GigaScience **7**, gix120.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly **6**, 80–92.

D'Andrea L, Amenós M, Rodríguez-Concepción M. 2014. Confocal laser scanning microscopy detection of chlorophylls and carotenoids in chloroplasts and chromoplasts of tomato fruit. Methods in Molecular Biology **1153**, 227–232.

Danecek P, Bonfield JK, Liddle J, *et al*. 2021. Twelve years of SAMtools and BCFtools. GigaScience **10**, giab008.

Daunay M-C, Aubert S, Frary A, *et al.* 2004. Eggplant (*Solanum melongena*) fruit colour: pigments, measurements and genetics. In: Voorrips RE, ed. XIIth Meeting on Genetics and Breeding of *Capsicum* and Eggplant, May 17–19, 2004, Noordwijkerhout, The Netherlands, 108–116.

Di Gioia F, Tzortzakis N, Rouphael Y, Kyriacou MC, Sampaio SL, C.F.R. Ferreira I, Petropoulos SA. 2020. Grown to be blue—antioxidant properties and health effects of colored vegetables. Part II: Leafy, fruit, and other vegetables. Antioxidants **9**, 97.

Dobin A, Gingeras TR. 2015. Mapping RNA-seq reads with STAR. Current Protocols in Bioinformatics **51**, 11.14.1–11.14.19.

Doganlar S, Frary A, Daunay MC, Lester RN, Tanksley SD. 2002. Conservation of gene function in the Solanaceae as revealed by comparative mapping of domestication traits in eggplant. Genetics **161**, 1713–1726.

Fitter DW, Martin DJ, Copley MJ, Scotland RW, Langdale JA. 2002. *GLK* gene pairs regulate chloroplast development in diverse plant species. The Plant Journal **31**, 713–727.

Frary A, Doğanlar S. 2003. Comparative genetics of crop plant domestication and evolution. Turkish Journal of Agriculture and Forestry **27**, 59–69.

Frary A, Frary A, Daunay MC, Huvenaars K, Mank R, Doğanlar S. 2014. QTL hotspots in eggplant (*Solanum melongena*) detected with a high resolution map and CIM analysis. Euphytica **197**, 211–228.

Fuller DQ. 2007. Contrasting patterns in crop domestication and domestication rates: recent archaeobotanical insights from the Old World. Annals of Botany **100**, 903–924.

García-Fortea E, Lluch-Ruiz A, Pineda-Chaza BJ, García-Pérez A, Bracho-Gil JP, Plazas M, Gramazio P, Vilanova S, Moreno V, Prohens J. 2020. A highly efficient organogenesis protocol based on zeatin riboside for *in vitro* regeneration of eggplant. BMC Plant Biology **20**, 6.

Gramazio P, Yan H, Hasing T, Vilanova S, Prohens J, Bombarely A. 2019. Whole-genome resequencing of seven eggplant (*Solanum melongena*) and one wild relative (*S. incanum*) accessions provides new insights and breeding tools for eggplant enhancement. Frontiers in Plant Science **10**, 1220.

Hall LN, Rossini L, Cribb L, Langdale JA. 1998. GOLDEN 2: a novel transcriptional regulator of cellular differentiation in the maize leaf. The Plant Cell **10**, 925–936.

Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. Briefings in Bioinformatics **9**, 286–298.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution **35**, 1547–1549.

Lester RN, Hasan SMZ. 1991. Origin and domestication of the brinjal eggplant, *Solanum melongena*, from *S. incanum*, in Africa and Asia. In: Hawkes JG, Lester RN, Nee M, Estrada N, eds. Solanaceae III: taxonomy, chemistry, evolution. London: The Royal Botanic Gardens, 369–387.

Letunic I, Bork P. 2019. Interactive Tree of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Research **47**, W256–W259.

Li D, Liu CM, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics **31**, 1674–1676.

Li D, Qian J, Li W, *et al*. 2021. A high-quality genome assembly of the eggplant provides insights into the molecular basis of disease resistance and chlorogenic acid synthesis. Molecular Ecology Resources **21**, 1274–1286.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics **25**, 1754–1760.

Li X, Wang P, Li J, Wei S, Yan Y, Yang J, Zhao M, Langdale JA, Zhou W. 2020. Maize GOLDEN2-LIKE genes enhance biomass and grain yields in rice by improving photosynthesis and reducing photoinhibition. Communications Biology **3**, 151–162.

Lichtenthaler HK, Wellburn AR. 1983. Determinations of total carotenoids and chlorophylls *a* and *b* of leaf extracts in different solvents. Biochemical Society Transactions **11**, 591–592.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $_2{}^{-\Delta\Delta C_T}$ method. Methods **25**, 402–408.

Mangino G, Arrones A, Plazas M, Pook T, Prohens J, Gramazio P, Vilanova S. 2022. Newly developed MAGIC population allows identification of strong associations and candidate genes for anthocyanin pigmentation in eggplant. Frontiers in Plant Science **13**, 847789.

Mangino G, Plazas M, Vilanova S, Prohens J, Gramazio P. 2020. Performance of a set of eggplant (*Solanum melongena*) lines with introgressions from its wild relative *S. incanum* under open field and screenhouse conditions and detection of QTLs. Agronomy **10**, 467.

Milne I, Shaw P, Stephen G, Bayer M, Cardle L, Thomas WTB, Flavell AJ, Marshall D. 2010. Flapjack-graphical genotype visualization. Bioinformatics **26**, 3133–3134.

Mir R, Calabuig-Serna A, Seguí-Simarro JM. 2021. Doubled haploids in eggplant. Biology **10**, 685.

Nadakuduti SS, Holdsworth WL, Klein CL, Barry CS. 2014. *KNOX* genes influence a gradient of fruit chloroplast development through regulation of *GOLDEN2-LIKE* expression in tomato. The Plant Journal **78**, 1022–1033.

Nguyen CV, Vrebalov JT, Gapper NE, Zheng Y, Zhong S, Fei Z, Giovannoni JJ. 2014. Tomato GOLDEN2-LIKE transcription factors reveal molecular gradients that function during fruit development and ripening. The Plant Cell **26**, 585–601.

Page AML, Chapman MA. 2021. Identifying genomic regions targeted during eggplant domestication using transcriptome data. The Journal of Heredity **112**, 519–525.

Page AML, Daunay M-CR, Aubriot X, *et al.* 2019. Domestication of eggplants: a phenotypic and genomic insight. In: Chapman MA, ed. The eggplant genome. Cham: Springer, 193–212.

Pan Y, Bradley G, Pyke K, *et al*. 2013. Network inference analysis identifies an *APRR2-like* gene linked to pigment accumulation in tomato and pepper fruits. Plant Physiology **161**, 1476–1485.

Plazas M, Gramazio P, Vilanova S, *et al.* 2020. Introgression breeding from crop wild relatives in eggplant landraces for adaptation to climate change. In: Maxted N, ed. Crop Wild Relatives. Birmingham: University of Birmingham, 32–36.

Plazas M, Vilanova S, Gramazio P, *et al*. 2016. Interspecific hybridization between eggplant and wild relatives from different genepools. Journal of the American Society for Horticultural Science **141**, 34–44.

Powell ALT, Nguyen CV, Hill T, *et al*. 2012. *Uniform ripening* encodes a *Golden 2-like* transcription factor regulating tomato fruit chloroplast development. Science **336**, 1711–1715.

Ranil RHG, Niran HML, Plazas M, Fonseka RM, Fonseka HH, Vilanova S, Andújar I, Gramazio P, Fita A, Prohens J. 2015. Improving seed germination of the eggplant rootstock *Solanum torvum* by testing multiple factors using an orthogonal array design. Scientia Horticulturae **193**, 174–181.

Robinson JT, Thorvaldsdóttir H, Turner D, Mesirov JP. 2023. igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). Bioinformatics **39**, btac830.

Rossini L, Cribb L, Martin DJ, Langdale JA. 2001. The maize *Golden2* gene defines a novel class of transcriptional regulators in plants. The Plant Cell **13**, 1231–1244.

Taher D, Solberg S, Prohens J, *et al*. 2017. World vegetable center eggplant collection: Origin, composition, seed dissemination and utilization in breeding. Frontiers in Plant Science **8**, 1484.

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

Tigchelaar EC, Janick J, Erickson HT. 1968. The genetics of anthocyanin coloration in eggplant (*Solanum melongena* L.). Genetics **60**, 475–491.

Vilanova S, Alonso D, Gramazio P, *et al*. 2020. SILEX: A fast and inexpensive high-quality DNA extraction method suitable for multiple sequencing platforms and recalcitrant plant species. Plant Methods **16**, 110.

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. Bioinformatics **25**, 1189–1191.

Waters MT, Wang P, Korkaric M, Capper RG, Saunders NJ, Langdale JA. 2009. GLK transcription factors coordinate expression of the photosynthetic apparatus in *Arabidopsis*. The Plant Cell **21**, 1109–1128.

Wei Q, Wang J, Wang W, Hu T, Hu H, Bao C. 2020. A high-quality chromosome-level genome assembly reveals genetics for important traits in eggplant. Horticultural Research **7**, 153.

Yasumura Y, Moylan EC, Langdale JA. 2005. A conserved transcription factor mediates nuclear control of organelle biogenesis in anciently diverged land plants. The Plant Cell **17**, 1894–1907.

Yeh SY, Lin HH, Chang YM, *et al*. 2022. Maize Golden2-like transcription factors boost rice chloroplast development, photosynthesis, and grain yield. Plant Physiology **188**, 442–459.

Zubo YO, Blakley IC, Franco-Zorrilla JM, Yamburenko MV, Solano R, Kieber JJ, Loraine AE, Schaller GE. 2018. Coordination of chloroplast development through the action of the GNC and GLK transcription factor families. Plant Physiology **178**, 130–147.