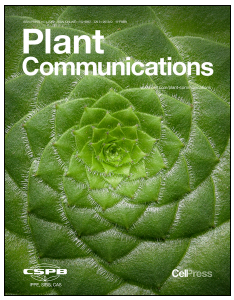
# Journal Pre-proof




Alternative Splicing in the *Anthocyanin fruit* Gene Encoding an R2R3 MYB Transcription Factor Affects Anthocyanin Biosynthesis in Tomato Fruits

Sara Colanero, Andrea Tagliani, Pierdomenico Perata, Silvia Gonzali

PII: S2590-3462(19)30006-9

DOI: https://doi.org/10.1016/j.xplc.2019.100006

Reference: XPLC 100006

To appear in: *PLANT COMMUNICATIONS*

Received Date: 8 August 2019

Revised Date: 16 September 2019

Accepted Date: 23 October 2019

Please cite this article as: Colanero, S., Tagliani, A., Perata, P., Gonzali, S., Alternative Splicing in the *Anthocyanin fruit* Gene Encoding an R2R3 MYB Transcription Factor Affects Anthocyanin Biosynthesis in Tomato Fruits, *PLANT COMMUNICATIONS* (2019), doi: https://doi.org/10.1016/j.xplc.2019.100006.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019

Journal Pre-proof

# Alternative Splicing in the *Anthocyanin fruit* Gene Encoding an R2R3 MYB Transcription Factor Affects Anthocyanin Biosynthesis in Tomato Fruits

**Sara Colanero, Andrea Tagliani, Pierdomenico Perata* and Silvia Gonzali**

PlantLab, Institute of Life Sciences, Scuola Superiore Sant'Anna, Pisa, Italy
*Correspondence: Pierdomenico Perata (p.perata@santannapisa.it) and Silvia Gonzali (s.gonzali@santannapisa.it)

**Running title: Alternative splicing in the *Anthocyanin fruit* gene leads to a non-functional R2R3 MYB transcription factor in tomato fruits**

**Short summary:** The *Anthocyanin fruit* tomato line accumulates anthocyanins in fruit peel through the introgression of *R2R3 MYB* genes from *Solanum chilense*. A comparative functional analysis of these genes revealed important differences for one of them, *SlAN2like*, between wild type and *Aft* plants. Remarkably, splicing mutations in the wild type allele determine a loss-of-function of the protein, explaining why domesticated tomato do not synthesize anthocyanins in its fruits.

Journal Pre-proof

Journal Pre-proof

**ABSTRACT**

Tomato (*Solanum lycopersicum* L.) fruits are typically red at ripening, with high levels of carotenoids and a low content in flavonoids. Considerable work has been done to enrich the spectrum of their health-beneficial phytochemicals, and interspecific crosses with wild species have successfully led to purple anthocyanin-colored fruits. The *Aft* (*Anthocyanin fruit*) tomato accession inherited from *Solanum chilense* the ability to accumulate anthocyanins in fruit peel through the introgression of loci controlling anthocyanin pigmentation, including four R2R3 MYB transcription factor encoding genes. Here, we carried out a comparative functional analysis of these transcription factors in wild type and *Aft* plants, testing their ability to take part in the transcriptional complexes that regulate the biosynthetic pathway and their efficiency in inducing anthocyanin pigmentation. Significant differences emerged for SlAN2like, both in the expression level and protein functionality, with splicing mutations determining a complete loss-of-function of the wild type protein. This transcription factor thus appears to play a key role in the anthocyanin fruit pigmentation. Our data provide new clues to the long-awaited genetic bases of the *Aft* phenotype and also contribute to clarify why domesticated tomato fruits display a homogeneous red coloration without the typical purple streaks observed in wild tomato species.

**Keywords:** *Solanum lycopersicum* L., tomato, *Anthocyanin fruit*, *Aft*, anthocyanin, R2R3 MYB transcription factors, MBW complex, purple pigmentation, *Solanum chilense* [(Dunal) Reiche]

Journal Pre-proof

Journal Pre-proof

## INTRODUCTION

Tomato (*Solanum lycopersicum* L.) is the most consumed vegetable worldwide, and 14% of global vegetable production (FAO, 2010). It belongs to the *Solanaceae* family, and is the only domesticated species within the fourteen of the tomato clade (*Solanum* genus, section *Lycopersicon*) (Bedinger et al., 2011). It was first cultivated in the pre-Columbian era in Central-South America, where it originated. In the 16th century it was introduced to Europe as an ornamental plant, and only two centuries later did its cultivation for human consumption gradually spread (Peralta and Spooner, 2007). The domestication of tomato experienced repeated bottlenecks, which strongly reduced its genetic diversity. Today it represents only 5% of the genetic diversity in the wild relative *Solanum* species (Bai and Lindhout, 2007), which thus constitute an invaluable reserve of genetic variability. Although there may be reproductive barriers (Bedinger et al., 2011), interspecific crosses can be carried out to improve tomato performance with new genes and allelic variants.

Human selection has progressively changed many of the original traits of tomato plants, also producing a wide variation in fruit size, morphology and color. *S. lycopersicum*, as well as *S. pimpinellifolium*, *S. galapagense* and *S. cheesmaniae*, bear orange/red fruits, with carotenoids as major pigments. The other *Solanum* species produce green fruits, which under favorable conditions display purple pigmentation on the peel (Bedinger et al., 2011). The purple color is conferred by the accumulation of anthocyanins, polyphenolic secondary metabolites belonging to the class of flavonoids (Liu et al., 2018).

Cultivated tomato cannot produce purple fruits: flavonoid biosynthesis is interrupted with the accumulation of intermediate compounds (mainly naringenin chalcone and the flavonol glycosides rutin and kaempferol-3-*O*-rutinoside) (Bovy et al., 2002), probably due to an inefficient activation of the pathway (Povero et al., 2011). However, due to their increasingly recognized health-promoting effects (Martin et al., 2011; Liu et al., 2017), considerable work has been done in recent years to enrich tomatoes with anthocyanins (Gonzali et a., 2009; Martì et al., 2016). Along with transgenic approaches (Butelli et al., 2008), biodiversity has been exploited with positive results (Mes et al., 2008; Gonzali et al., 2009).

The *Aft* (*Anthocyanin fruit*) line, selected in a segregant progeny from a cross between *S. lycopersicum* and *S. chilense* [(Dunal) Reiche] (Georgiev, 1972), is one of the genotypes mostly commonly used in tomato breeding to obtain purple peel fruits (Mes et al., 2008; Gonzali et al., 2009; Myers, 2012). In *Aft*, anthocyanin-spotted fruits are produced upon intense light exposure (Figure 1A). The phenotype is associated with a genomic region mapped on the distal part of the long arm of chromosome 10 (Mes et al., 2008; Sapir et al., 2008) (Figure 1B), introgressed from *S. chilense*. In this genomic region there therefore needs to be a major locus controlling fruit anthocyanin pigmentation. Interestingly, a major QTL responsible for most of the phenotypic variations in fruit anthocyanin content is already known to be in chromosome 10 of eggplant, and both the flower and tuber skin color of potato have been associated with genes mapped on chromosome 10 (Doganlar et al., 2002). Genetic mapping studies in pepper have identified a major region in chromosome 10 containing genes related to the accumulation of anthocyanins in the fruit (Wang et al., 2018). The association of fruit anthocyanin pigmentation with chromosome 10 observed in *Aft* tomato thus appears to be strongly conserved in domesticated *Solanaceae*.

The genetic nature of the *Aft* trait is still an open issue. Several studies have proposed putative candidates among the four R2R3 MYB encoding genes (*Solyc10g086250* = *SlMYB75* = *SlAN2*, *Solyc10g086260* =

Journal Pre-proof

*anthocyanin 1* = *SlANT1*, *Solyc10g086270* = *SlMYB28* = *SlANT1like*, *Solyc10g086290* = *SlMYB114* = *SlAN2like*) identified in this chromosome region (Figure 1B) (Sapir et al., 2008; Schreiber et al., 2011; Kiferle et al., 2015; Cao et al., 2017; Qiu et al., 2019). R2R3 MYB proteins are transcription factors (TFs) which are involved in the regulation of many aspects of cell identity and fate, including the control of secondary metabolism (Stracke et al., 2001; Liu et al., 2015). They can participate with subgroup IIIf bHLH factors and WDR proteins in the MYB-bHLH-WDR (MBW) transcriptional complexes that regulate the anthocyanin biosynthetic pathway (Xu et al., 2015; Liu et al., 2018), and their expression patterns may impact on the pigmentation patterns of a plant.

Activation of the anthocyanin synthesis is a consequence of a transcriptional regulatory cascade (Albert et al., 2014; Montefiori et al., 2015) (Figure 1C). The first MBW complex is composed of an R2R3 MYB protein, developmentally or environmentally regulated, and the constitutively expressed WDR and bHLH1 factors. This complex transcriptionally activates an inducible bHLH2-encoding gene, thus producing a second complex composed of the same MYB and WDR proteins as well as the new bHLH2 partner. Thanks to the MYB DNA-binding domains, the second MBW complex finally activates the transcription of “late biosynthetic genes” (LBGs). This produces the enzymes involved in the steps of the flavonoid pathway that lead to anthocyanins and are differently regulated from “early biosynthetic genes”, which encode the enzymes that act in earlier reactions of the pathway (Liu et al., 2018). The second complex also induces other positive regulators, including the same bHLH2 factor (“reinforcement mechanism”), and repressor MYB proteins, in a feedback loop finely titrating the accumulation of anthocyanins (Albert et al., 2014).

In this work we carried out a functional characterization of the *Aft* R2R3 MYB TFs, which contribute, individually or in combination, to the pigmentation of the fruit, compared with the wild type (WT) protein variants. We found some key differences in transcript levels and protein activities for one of these MYB factors, which thus appeared to be primarily involved in the *Aft* phenotype. We believe that our identification of splicing mutations in the WT allele of its gene finally contributes to the understanding of the lack of anthocyanin pigmentation in cultivated tomato.

## RESULTS AND DISCUSSION

### Structural and functional analyses of the R2R3 MYB proteins encoded by the genes located in the introgressed genomic region of *Aft*

The *R2R3 MYB* genes identified in the long arm of chromosome 10, where *Aft* was mapped (Sapir et al., 2008), encode proteins which are phylogenetically correlated with many other plant MYBs involved in anthocyanin synthesis (Figure 1D). They also show very similar sequences: the R2/R3 MYB domains, which specify DNA binding (Lin-Wang et al., 2010), are highly conserved, whereas the C-terminal regions, which influence the strength of the promoter activation (Heppel et al., 2013), are more variable (Supplemental Figure 1). In *Aft*, the four *R2R3 MYB* genes show sequence polymorphisms compared to the WT counterparts, which produce amino acid variants in the relative polypeptides (Supplemental Figure 2).

Anthocyanins are synthesized in tomato vegetative tissues upon different environmental stimuli, such as cold or intense light, with the R2R3 MYB TF SlAN2 representing the key MYB activator of the pathway (Kiferle et

al., 2015). Similarly to other dicots, a ternary MBW complex constitutes the key transcriptional regulator of the structural LBGs of the biosynthetic pathway, and SlAN2, as well as the bHLH factors SlJAF13 (bHLH1) (Nukumizu et al., 2013; Montefiori et al., 2015) and SlAN1 (bHLH2) (Qiu et al., 2016; Colanero et al., 2018; Gao et al., 2018), and the WDR protein SlAN11 (Gao et al., 2018), have been proven to interact with each other and to be essential for the synthesis of anthocyanins. SlAN2 shows the conserved [DE]Lx2[RK] x3Lx6Lx3R motif containing the bHLH-binding site (Zimmermann et al., 2004) in the R3 domain, the amino acidic signature [A/S/G]NDV and the KPRPR[ST]F motif typical of dicot R2R3 MYBs involved in anthocyanin synthesis (Stracke et al., 2001; Lin-Wang et al., 2010; Heppel et al., 2013) (Supplemental Figure 2A). All these features are also present in the other three WT and four *Aft* R2R3 MYB factors (Supplemental Figures 2A-2D). On the basis of their strict sequence similarities, all these MYB TFs should therefore be able to activate the synthesis of anthocyanins. For SlANT1 this has already been demonstrated (Mathews et al., 2003; Schreiber et al., 2011; Kiferle et al., 2015).

To directly compare all these TFs, either from the WT or from *Aft*, in the activation of the anthocyanin pathway, we tested them in a transactivation assay in tomato protoplasts. We used a reporter *luciferase* gene driven by the promoter of *Dihydroflavonol 4-reductase* (*SlDFR*), a marker LBG (Kiferle et al., 2015). Each MYB protein was expressed starting from its genomic sequence and in combination with the bHLH2 factor SlAN1. An ectopic WDR protein was not included in the test as SlAN11 is constitutively expressed (Gao et al., 2018). Whereas all the four *Aft* MYB TFs were able to strongly transactivate the reporter gene, only three WT MYBs activated it, with SlAN2like being ineffective (Figure 2A). This incapacity was also verified in combination with the bHLH1 factor SlJAF13 (Figure 2B). The WT SlAN2like protein (hereafter SlAN2like$^{WT}$) thus behaved very differently from the *Aft* SlAN2like (hereafter SlAN2like$^{Aft}$), with only the latter being active with both bHLHs (Figure 2B).

To obtain an *in vivo* confirmation of the different functionality of the two SlAN2like variants, we agro-infiltrated tobacco leaves with vectors expressing *SlAN2like*$^{WT}$or *SlAN2like*$^{Aft}$. Again, whereas SlAN2like$^{Aft}$ induced ectopic anthocyanin synthesis both by interacting with or without its partner SlAN1 (likely engaging a tobacco bHLH factor), SlAN2like$^{WT}$ was non-functional (Figures 2C and 2D).

The fruits of *S. chilense*, the wild progenitor of *Aft* (Georgiev, 1972), show anthocyanin pigmentation when exposed to light (Figure 2E). Therefore, if SlAN2like$^{Aft}$ is involved in the *Aft* phenotype, the corresponding protein of *S. chilense*, ScAN2like, whose sequence differs from SlAN2like$^{WT}$ in relation to many amino acid variants already found in SlAN2like$^{Aft}$ (Supplemental Figure 3), should be functional. In fact, when expressed in tomato protoplasts, ScAN2like activated the *SlDFR* promoter similarly to SlAN2like$^{Aft}$ (Figure 2F).

Both *in vitro* and *in vivo* analyses thus indicated that, among the R2R3 MYB factors encoded by the genes identified in chromosome 10, the WT TF SlAN2like was non-functional (Figures 2A and 2B), unlike its *Aft* and *S. chilense* orthologous proteins (Figure 2F).

**SlAN2like$^{Aft}$ plays a primary role in *Aft* fruit pigmentation**

The previous results prompted us to focus on the possible role of SlAN2like in the pigmentation of the fruit. We then grew *Aft* and WT plants under light conditions that induce anthocyanin production. In *Aft,*

Journal Pre-proof

anthocyanins were synthesized from the green stage in the part of the fruit peel developed directly under light, corresponding to the stem-end of the epicarp, whereas the stylar-end remained green (Figures 3A and 3B). By contrast, anthocyanins were not produced in WT fruits, not even in the stem-end of the epicarp, developed directly under light (Figure 3B).

A qPCR analysis carried out in *Aft* skin at the mature green stage showed differences between the top and the bottom halves of the fruit. In the peel directly exposed to light (top epicarp), we observed the expression of several genes involved in the anthocyanin pathway, including R2R3 MYBs *SlAN2* and *SlAN2like*, bHLH *SlJAF13* and *SlAN1*, and WDR *SlAN11* (Figure 3C). Both *SlAN2* and *SlAN2like* responded to light intensity, showing a higher expression in the top than in the bottom half of the fruit; however, *SlAN2like* was much more expressed than *SlAN2* (Figure 3C). The expression of the other two MYBs, *SlANT1* and *SlANT1like*, was barely detectable (Figure 3C). *SlAN1*, along with the LBGs, *SlDFR* and *SlANS*, only showed high expression levels in the top peel (Figures 3C). The same was found for *SlMYB-ATV* (Figure 3C), encoding an R3 MYB repressor of the pathway which has been recently characterized (Cao et al., 2017; Colanero et al., 2018).

In WT fruit peel, the transcript analysis showed a low expression of all four MYBs in both the stem- and stylar-end of the fruit (Figure 3C). However, in the part of the fruit developed under light, *SlAN2* was more expressed in WT than in *Aft* fruit, whereas *SlAN2like* was much less expressed (Figure 3C). *SlJAF13* and *SlAN11* were expressed in both halves of the fruit at similar levels, confirming their constitutive expression, whereas very few transcripts were measured for *SlAN1*, *SlMYB-ATV*, *SlDFR* and *SlANS* in all the fruit (Figure 3C).

Transcript qPCR data clearly indicated that the actors of the activation mechanism were present in *Aft* peel under light, with *SlAN2like*$^{Aft}$ as the major R2R3 MYB expressed gene. The interaction among this light-induced MYB activator and the bHLH1 and WDR factors, SlJAF13 and SlAN11, thus produced the first MBW complex, hierarchically activating the transcription of the inducible bHLH2 gene, *SlAN1*. The SlAN1 protein then participated with SlAN2like$^{Aft}$ and SlAN11 in the second MBW complex, inducing the LBGs and the anthocyanin accumulation. In WT fruits, based on transcript analyses, the absence of anthocyanins was due to a scarce activation of the LBGs, which, in turn, could be attributed to the failure of the assembly of the MBW complexes, particularly the second one, which could have been formed only at negligible concentrations, given the very low expression levels of *SlAN1* (Figure 3C).

On the whole, the qPCR analysis indicated that: i) in fruit peel at the mature green stage, *SlANT1* and *SlANT1like* levels appeared insignificant in both *Aft* and WT fruits; ii) whereas in WT, *SlAN2* was the most expressed *MYB* gene, in *Aft SlAN2like* was the main *MYB*, while its expression was very low in WT; and iii) the expression of the *MYB* gene *SlAN2* in WT fruit peel was not sufficient to trigger anthocyanin synthesis. The high expression of *SlAN2like* that we observed in *Aft* fruit confirmed findings in other tomato lines expressing the *Aft* gene (Cao et al., 2017; Qiu et al., 2019).

It is known from other species, particularly Arabidopsis (Nesi et al., 2000; Zhang et al., 2003; Ramsay and Glover, 2005), that R2R3 MYBs determine the pathway specificity of the MBW complexes, whereas bHLH and WDR factors can control different aspects of cell identity participating in different MBW complexes. It is thus possible that the global level of the R2R3 MYBs promoting anthocyanins and expressed in fruit peel

need to reach a certain threshold to be able to recruit enough bHLH and WDR partners to produce sufficient MBW complexes to activate the anthocyanin pathway. If this holds true, in *Aft* fruits, under appropriate light conditions, the SlAN2like$^{Aft}$ level may become high enough to activate anthocyanin synthesis. By contrast in WT fruits, *SlAN2like* is poorly expressed and, most importantly, SlAN2like proteins are not functional, while SlAN2 cannot reach an adequate level to activate a significant transcription of *SlAN1*. Overexpression of *SlAN2* (as well as *SlANT1*) in tomato WT plants can lead to purple fruit pigmentation (Kiferle et al., 2015). The level of expression of this R2R3 MYB TF in fruit peel is therefore crucial to activate the anthocyanin pathway. As a consequence of the insufficient transcription of *SlAN2* and the inefficiency of SlAN2like, all the genes that are under the transcriptional control of the second MBW complex cannot be properly expressed in WT fruit peel. This is the case for the LBGs and for the same *bHLH2* gene, *SlAN1*, whose weak induction makes its final protein level insufficient for the sufficient activation of the pathway. In line with this, the repressor R3 MYB protein, SlMYB-ATV, whose transcription is stimulated through a feedback mechanism by the same MBW complex activating *SlAN1* and the *LBGs* (Colanero et al., 2018), was much less expressed in WT than in the *Aft* fruit peel (Figure 3C).

To understand whether the differences in the expression levels of *SlAN2like* in *Aft* and WT fruits depended on the different activation of the gene, we cloned the promoter regions. The sequence amplified in WT plants overlapped with the one deposited in the SOL Genomics database. The promoter of the *Aft* gene was instead cloned thanks to the data available with the recent publication of the reference genome of *S. chilense* (Stam et al., 2019). In fact, the *Aft* promoter was very similar to the region upstream of the *ScAN2like* gene, although, as with the coding sequence (cds), a few polymorphisms between them were found (Supplemental Figure 4).

R2R3 MYB TFs are often prone to auto-activation which has also been found to characterize MYB proteins involved in anthocyanin synthesis (Brendolise et al., 2017). We found MYB cis regulatory elements in both WT and *Aft* promoters of *SlAN2like* (Supplemental Figure 4). Therefore, to test whether they could be transactivated by the MYB proteins produced by their respective genes and involved in the MBW complex, we expressed the promoters of *SlAN2like* fused to the *luciferase* gene in protoplasts transfected with the components of the MBW complex. We found similar basal expression levels of the two reporter genes and no activation for either the WT or the *Aft* promoters (Figure 3D). On the other hand the promoter of *SlDFR*, included in the test as a control, was transactivated by the complex including the SlAN2like$^{Aft}$ MYB protein (Figure 3D), as already observed (Figures 2A, 2B). This result indicated that the higher expression of *SlAN2like*$^{Aft}$ in *Aft* fruit peel was not due to auto-activation by the MBW complex which induces anthocyanin synthesis. Finally, to test whether the two promoters showed different activation states in the respective fruits, we also expressed them fused to the *luciferase* reporter gene in protoplasts isolated from the fruit peel sampled from mature green WT (Figure 3E) and *Aft* fruits (Figure 3F). Again, no significant differences were measured between the basal activities of the promoter of *SlAN2like*$^{WT}$ and of *SlAN2like*$^{Aft}$ either in WT or in *Aft* fruit protoplasts (Figures 3E, 3F). Moreover, in both of them a very low basal activation of the *SlAN2like* promoters was observed compared to the activity of the promoter of *SlDFR* transactivated by the MBW complex including the SlAN2like$^{Aft}$ protein and included in the test as a control (Figures 3E, 3F).

**Splicing mutations affect the *SlAN2like* transcripts produced in tomato fruit peel**

To understand why SlAN2like$^{WT}$ is non-functional, we amplified the *SlAN2like* transcripts from the top epicarp of WT and *Aft* fruits (Figure 4A). Whereas the cds of *SlAN2like$^{Aft}$* was well-aligned with the tomato *SlAN2like* cds bioinformatically predicted (*Solyc10g086290.1.1*) (Supplemental Figures 5A, 5B), in WT fruit peel we amplified two slightly different shorter sequences, lacking one or two nucleotide strings, at the end of the first or second exon (Supplemental Figures 5C-5E). As the genomic sequence of *SlAN2like$^{WT}$* from our plants was identical to the reference sequence (*Solyc10g086290.1*), an alternative processing of the pre-mRNA should have occurred leading to these transcript arrangements. To verify whether this was linked to the variety Ailsa Craig that we used as WT, we also cloned the *SlAN2like* transcript from another variety, Heinz 1706, the one used as a reference for the tomato genome (Tomato Genome Consortium, 2012). Again, from the fruit peel at the mature green stage, we obtained a shorter sequence than the expected one, lacking the same nucleotide string at the end of the second exon already identified in one of the transcripts of Ailsa Craig fruits (Supplemental Figure 6A).

The SlAN2like$^{WT}$ polypeptide bioinformatically predicted from the reference gene sequence (Solyc10g086290, Supplemental Figures 1 and 2) derives from a mature mRNA assembled by using the splicing sites which are indeed recognized by the splicing machinery in the sequence of the primary transcript of *SlAN2like$^{Aft}$* (Figure 4B; Supplemental Figures 5B, 5C). These splicing sites can be considered as “canonical”, since they produce a mature mRNA translated into a functional protein. By contrast, in the WT pre-mRNA, alternative 5’ splicing sites in the first and second introns are recognized by the spliceosome, thus producing shorter transcripts (Figures 4C, 4D; Supplemental Figures 5D, 5E). Interestingly, in these transcripts the loss of various nucleotides led to a frameshift, which produced an early stop codon at the beginning of the third exon (Supplemental Figures 5D, 5E). The corresponding proteins should thus present a premature truncation resulting in a much lower size than the theoretical one, with the loss of most of the residues downstream of the R2 domain (Figures 4C, 4D; Supplemental Figure 7A).

By directly comparing the WT and *Aft SlAN2like* transcript variants in protoplasts, we confirmed that the WT proteins derived from the fruit peel transcripts were non-functional, whereas the *Aft* activated the *SlDFR* promoter (Figure 4E), similarly to its corresponding genomic sequence (Figures 2A, 2B). We also found transactivation of the reporter gene by expressing a synthetic cds corresponding to the version of *SlAN2like$^{WT}$* produced through the canonical splicing (Figure 4E). The “correctly spliced” version of this MYB TF was able to transactivate the *SlDFR* promoter. Its efficiency was lower than that of SlAN2like$^{Aft}$ (Figure 4E), probably due to the presence of polymorphisms in the C-terminal region (Supplemental Figure 7B), which is part of the activation domain of the TF. The alternative splicing leading to the fruit peel transcripts of *SlAN2like$^{WT}$* thus prevented the translation of a functional protein.

By examining the structure of the truncated SlAN2like$^{WT}$ protein, it seems evident that the absence of the R3 domain, containing the bHLH-binding signature, prevents it from forming MBW complexes. In fact, a split-luciferase complementation assay carried out in protoplasts showed that SlAN2like$^{WT}$ did not interact with the bHLH factor SlAN1, unlike SlAN2like$^{Aft}$, which showed a clear interaction with the bHLH partner (Figure 4F). On the contrary, the WDR protein SlAN11 did not bind either SlAN2like$^{WT}$ or SlAN2like$^{Aft}$ (Figure 4F), confirming previous data indicating that WDR proteins can only bind bHLH factors and not MYBs (An et al., 2012; Montefiori et al., 2015; Gao et al., 2018). A bimolecular fluorescence complementation assay confirmed the interaction between SlAN2like$^{Aft}$ and SlAN1, taking place in the nucleus (Figure 4G), the

Journal Pre-proof

cellular compartment where transcription occurs. Protein-protein interaction assays thus demonstrated that SlAN2like$^{WT}$ is unable to associate with bHLH partners, and thus to participate in the MBW complexes which induce anthocyanin synthesis.

Non-canonical splicing events are becoming more and more frequently identified in plants, and over 60% of intron-containing genes are estimated to be prone to alternative splicing (AS) (Syed et al., 2012). AS can affect transcript levels and stability. Aberrant transcripts, containing premature termination codons, may induce a nonsense-mediated decay (NMD) leading to degradation of the same mRNAs (Syed et al., 2012; Sibley et al., 2016). If a similar mechanism affected *SlAN2like*$^{WT}$ transcripts, it would explain why we found very low expression levels in WT fruits (Figure 3C). Interestingly, tomato fruits of the Heinz 1706 variety not only produced aberrant *SlAN2like* transcripts such as Ailsa Craig, but also displayed similar low expression levels in fruit peel compared to *Aft* (Supplemental Figure 6B). Also in this variety, therefore, the red color of the fruit peel is associated with an alternative splicing of *SlAN2like* and a low expression of this gene. Heinz fruits also showed a low transcription of the other *R2R3 MYB* gene, *SlAN2* (Supplemental Figure 6B). These data suggest that what observed was not peculiar to the variety chosen in the study, but may be a general feature of domesticated tomato. The AS of *SlAN2like*$^{WT}$, preventing the translation of a functional MYB TF, thus would impede WT tomatoes from responding to excess light and to synthesize anthocyanins.

There is increasing evidence that AS represent a way of further regulating gene expression and at the same time increasing the protein-coding capacity of a genome. It thus contributes to the adaptation of plants to the environment (Syed et al., 2012). Wild tomato species mainly come from the Andean regions of South America (Chetelat et al., 2009), environments where high altitudes are common and there is ultraviolet radiation-enriched light. In these conditions, it is plausible that the capacity to synthesize protective anthocyanins not only in vegetative tissues but also in fruit peel is common and, in fact, most of the wild species still found in these areas (e.g. *S. chilense*, *S. peruvianum*, *S. lycopersicoides*) show green/purple fruits. With gradual diffusion in low altitude areas, also as a consequence of domestication and cultivation, it is possible that such a characteristic was progressively lost or counter-selected, perhaps because of a more appealing uniform red color. Interestingly, the *AN2like* cds is very conserved in *S. lycopersicum*, its more direct ancestor species, *S. pimpinellifolium*, and other more distant wild species, such as *S. chilense*, *S. pennellii* and *S. lycopersicoides* (Supplemental Figure 8), and all of these cds can be translated into functional proteins. The intronic regions of the gene appear more variable, and *S. lycopersicum* and *S. pimpinellifolium*, which both bear red fruits, are considerably more interrelated in terms of their intronic sequences than the other green/purple fruited species (Supplemental Figure 8). Transcriptomic data related to the expression levels of the gene *AN2like* in wild species are not available. However an RNASeq experiment carried out in *S. pimpinellifolium* indicated a level of expression of *AN2like* in the fruit which is not so different from that of *S. lycopersicum* (Supplemental Figure 9). It is tempting to speculate that intronic mutations in the *SlAN2like* gene could lead to the production of those cis elements that force the spliceosome to recognize the non-canonical splicing sites with a consequent reduction in splicing fidelity. Further studies are needed however to verify how much this process has spread among tomato varieties and when it originated.

## METHODS

Journal Pre-proof

**Plant material and growth conditions**

Seeds of *S. lycopersicum* cv. Ailsa Craig (LA2838A), representing WT tomato, *Aft*/*Aft* (LA1996), cv Heinz 1706 (LA 4345) and *S. chilense* (LA1930) were provided by the Tomato Genetic Resource Center (https://tgrc.ucdavis.edu/). Accession LA1930 was chosen, as the line of *S. chilense* that was originally crossed with *S. lycopersicum* (Georgiev, 1972) is not known. Seeds were germinated in rock-wool plugs (Grodan, https://www.grodan.com/) soaked in a nutritive solution (Kiferle et al., 2013). Two-week-old seedlings were transplanted in pots containing a 70:30 soil (Hawita Flor, https://www.hawita-gruppe.de/en/) / expanded clay mixture, and placed in a growth chamber with 12h daylight, 300 µmol photons $m^{-2}$ $s^{-1}$, 28℃/21℃ day/night temperature, 50% relative humid ity. Fruits were sampled at the mature green stage, divided into two halves, and the peel was removed from the top and bottom parts, frozen in liquid nitrogen and stored at -80℃ until use.

**Plasmid construction**

The genes *Solyc10g086250 (SlAN2)*, *Solyc10g086260 (SlANT1)*, *Solyc10g086270 (SlANT1like)*, *Solyc10g086290 (SlAN2like)*, *Solyc09g065100* (*SlAN1*), *Solyc08g081140* (*SlJAF13*) and *Solyc03g097340* (*SlAN11*) (SOL Genomics Network, https://sgn.cornell.edu) were amplified by PCR starting from WT and/or *Aft* genomic DNAs using the "Phusion High-Fidelity DNA Polymerase" (Thermo Fisher Scientific, https://corporate.thermofisher.com) and the oligonucleotide primers reported in Supplemental Table 1. The promoters of *SlAN2like*$^{WT}$ and *SlAN2like*$^{Aft}$ were amplified by PCR as above described and using the primers reported in Supplemental Table 1. The *S. chilense AN2like* was amplified from DNA extracted from dry seeds. The cds of WT (Ailsa Craig and Heinz 1706) or *Aft SlAN2like* was amplified from RNA extracted from fruit peel using the "Spectrum Plant Total RNA Kit" (Sigma–Aldrich, https://www.sigmaaldrich.com), treated with DNase and reverse-transcribed with the SuperScript IV Reverse Transcriptase (Thermo Fisher Scientific). The "synthetic" *SlAN2like*$^{WT}$ cds was purchased from GeneArt Gene Synthesis (Thermo Fisher Scientific). The amplified sequences were cloned into pENTR/D-TOPO vector (Thermo Fisher Scientific) and the entry clones were recombined with different destination vectors, as described below, via Invitrogen™ Gateway™ recombination cloning technology (Thermo Fisher Scientific). Multiple sequence alignments were performed using ClustalW (www.genome.jp/tools-bin/clustalw) and DNAMAN sequence analysis softwares.

**Phylogenetic Analysis**

The analysis was performed on the Phylogeny.fr platform (Dereeper et al., 2008). R2R3 MYB protein sequences were aligned with MUSCLE (v3.8.31) configured for highest accuracy (MUSCLE with default settings). Ambiguous regions (i.e. containing gaps and/or poorly aligned) were removed with Gblocks (v0.91b) using the following parameters: minimum length of a block after gap cleaning: 10; no gap positions were allowed in the final alignment; all segments with contiguous nonconserved positions bigger than 8 were rejected; minimum number of sequences for a flank position: 85%. The phylogenetic tree was reconstructed using the maximum likelihood method implemented in the PhyML program (v3.1/3.0 aLRT). The WAG substitution model was selected assuming an estimated proportion of invariant sites (of 0.145) and 4

Journal Pre-proof

gamma-distributed rate categories to account for rate heterogeneity across sites. The gamma shape parameter was estimated directly from the data (gamma=1.135). Reliability for internal branch was assessed using the aLRT test (SH-Like). Graphical representation and edition of the phylogenetic tree were performed with TreeDyn (v198.3).

**Anthocyanin quantification**

Anthocyanins were extracted and quantified as described in Colanero et al., 2018, and finally expressed as microgram petunidin-3-(*p*-coumaroyl rutinoside)-5-glucoside gram$^{-1}$ fresh weight (Kiferle et al., 2015).

**Tomato protoplast isolation**

Leaf protoplasts were isolated following the protocol in Shi et al. (2012) from 3-week-old tomato plants, cv. Micro-Tom, grown as reported above. Fruit peel protoplasts were isolated from mature green WT and *Aft* fruits with the same protocol. Polyethylene glycol-mediated protoplast transformation was carried out as in Yoo et al. (2007).

**Transactivation assays**

Transactivation assays by dual-luciferase system were performed exploiting the *Renilla reniformis* (Renilla) and *Photinus pyralis* (Firefly) luciferase (Luc) enzymes. The effector constructs *35S:SlAN2*, *35S:SlANT1*, *35S:SlANT1like*, *35S:SlAN2like*, *35S:ScAN2like*, *35S:SlJAF13* and *35S:SlAN1*, with R2R3 MYB genomic sequences and bHLH cds, as well as the promoter *SlDFR:FireflyLuc* and *SlAN2like:FireflyLuc* reporter constructs were produced as reported in Colanero et al. (2018). A *35S:RenillaLuc* vector was used to normalize luminescence values detected in protoplasts (Weits et al., 2014). Effector and reporter plasmids were co-transfected in protoplasts and luminescence relative levels were measured as described in Kiferle et al. (2015). In each assay data were expressed as relative luciferase activity (RLU) (FireflyLuc/RenillaLuc). Each experiment was repeated three times with similar results.

**Agro-infiltration assay**

Transient expression assay was performed using *Nicotiana benthamiana* plants placed in a growth chamber with 16 h daylight, 100 µmol photons m$^{-2}$ s$^{-1}$, 23℃/20℃ day/night temperature. Overexpression v ectors were generated by recombining the entry clones containing the genomic sequences of *SlAN2like*$^{WT}$, *SlAN2like*$^{Aft}$ and the cds of *SlAN1* with the Gateway™ compatible binary vector pK7WG2 (Karimi et al., 2002). *Agrobacterium tumefaciens* GV3101 (MP90) strains hosting the different constructs were infiltrated in Nicotiana leaves following the protocol of Li (2011). Each leaf was infiltrated in four different points with different constructs, as shown in Figure 2C. Non-recombined pK7WG2 vectors were used as negative controls. Three different leaves in three tobacco plants were analyzed as biological replicates for each combination of plasmids. Anthocyanins were quantified in single portions sampled from leaves in relation to the infiltrated areas at four days after infiltration. The experiment was repeated twice with similar results.

**Split-luciferase complementation assay**

The Gateway™ compatible bait vector pDuEx-Dn6 and prey vector pDuEx-Ac6 (Fujikawa and Kato, 2007), containing the C-terminal half and the N-terminal half of the Renilla *luciferase* gene, respectively, were used for the recombination of *SlAN2like*$^{WT}$, *SlAN2like*$^{Aft}$, *SlAN1* and *SlJAF13* entry clones. Leaf protoplasts were transfected with mixtures of two different recombined bait and prey vectors. As the control, the NLuc-half protein was expressed in combination with each of the two CLuc-SlAN2like$^{WT}$ or CLuc-SlAN2like$^{Aft}$ fusion proteins and the CLuc-half protein was expressed in combination with each of the two SlAN1-NLuc or SlAN11-NLuc fusion proteins. Luciferase activity was analyzed as described (Colanero et al., 2018). Data were expressed as relative luciferase activity (RLU) (RenillaLuc/protein content). The experiment was repeated twice with similar results.

**RNA isolation, cDNA synthesis, and real-time PCR analysis**

Total RNA, extracted from fruit peel as described above, was subjected to DNase treatment and then reverse transcribed into cDNA using the “Maxima First Strand cDNA Synthesis Kit for RT-qPCR, with dsDNase” (Thermo Fisher Scientific). Quantitative RT-PCR was performed with an ABI Prism 7300 Sequence Detection System (Thermo Fisher Scientific) using the “PowerUp™ SYBR® Green Master Mix” (Thermo Fisher Scientific) and the primers listed in Supplemental Table 2. *Elongation Factor 1-alpha* (*SlEF1A*) (Kiferle et al., 2015) and *Abscisic stress ripening gene1* (*SlASR1*) (Bovy et al., 2002) were used as reference genes. Expression levels relative to the geometric averaging of the reference genes were quantified for each target gene.

**Bimolecular Fluorescence Complementation (BiFC) assay**

The Gateway™ compatible destination vectors used were pDH51-GW-YFPN and pDH51-GW-YFPC (Zhong et al., 2008), enabling the fusion of the N-terminus or C-terminus of the yellow fluorescent protein (YFP) moieties, respectively, to the C-terminus of the protein of interest. Control vectors were pDH51-YFPC and pDH51-YFPN (Zhong et al., 2008). Protoplasts were isolated as described, transformed with one microgram DNA for each plasmid, and incubated in the dark at 25℃ for 16 h before subsequent analysis. Fluorescence of YFP was analyzed with a ZEISS LSM 880 with Airyscan microscope, using YFP, TRITC and 4’6-diamidino-2-phenylindole filters.

**Statistics**

Statistical analyses were performed with GraphPad Prism 6.01 (www.graphpad.com/scientific-software/prism/). Data were analyzed by one-way ANOVA, and differences were tested using the Tukey honest significant difference (HSD) multiple comparisons test.

Journal Pre-proof

## FIGURE LEGENDS

**Figure 1. Anthocyanin synthesis in *Aft* tomato is associated with four R2R3 MYB genes introgressed into chromosome 10.**

**(A)** *Aft* tomato fruit at mature green (left) and red ripening (right) stages.

**(B)** Location of the four R2R3 MYB encoding genes in the distal part of the long arm of chromosome 10 introgressed into *Aft* from *Solanum chilense*.

**(C)** Model describing the regulatory mechanism controlling anthocyanin synthesis in dicots. Wavy orange arrows represent inductive environmental or developmental stimuli that trigger anthocyanin production. Black arrows indicate activation. Red arrows indicate repression. Adapted from Albert et al. (2014) and Liu et al. (2018).

**(D)** Phylogenetic tree showing the relatedness of the tomato R2R3 MYB proteins under study with other plant R2R3 MYB factors involved in anthocyanin synthesis. Protein sequences were identified on the Sol Genomics Network and NCBI websites. The relative accession numbers were as follows: SlAN2 (Solyc10g086260.1.1), SlAN2_*Aft* (ACT36608.1), SlANT1 (Solyc10g086250.1.1), SlANT1_*Aft* (ABO26065.1), SlANT1like (Solyc10g086270.1.1), SlANT1like_*Aft* (MN242013), SlAN2like (Solyc10g086290.1.1), SlAN2like_*Aft* (MN242011), ScAN2like (MN242012), StAN1 (AAX53089.1), StAN2 (AAX53091.1), PhAN2 (ABO21074.1), PhDPL (HQ116169), PhPHZ (HQ116170), PhPH4 (BAP28594.1), PhODO1 (Q50EX6.1), NtAN2 (ACO52470.1), AmROS1 (ABB83826.1), AmROS2 (ABB83827.1), AtMYB75 (AAG42001.1), AtMYB113 (NM_105308), AtMYB114 (NM_105309), ZmC1 (AAA33482), ZmPl (AAA19819), MdMYB10 (ABB84753), ScANT1(ABO26065.1), ScAN2(ACT36604.1), CsRuby (NP_001275818.1), CaMYBA (BBJ25251.1), MdMYB1 (ADQ27443.1), VvMYB5b (NP_001267854.1).

**Figure 2. Functional analysis of the R2R3 MYB proteins from WT and *Aft* plants.**

**(A)** Transactivation of the *SlDFR* promoter driving firefly *luciferase* in protoplasts with effector plasmids containing the *MYB SlAN2*, *SlANT1*, *SlANT1like* or *SlAN2like* genomic sequences from WT or *Aft* plants, in combination with the effector plasmid containing the *bHLH* factor *SlAN1*. Data are expressed as relative luciferase activity (RLU) (FireflyLuc/RenillaLuc) with the value of the promoter basal level set to one and are means of four biological replicates ± SE. One-way ANOVA with Tukey's HSD post-hoc test was performed. ns means P>0.5, and "****" means P≤0.0001, respectively.

**(B)** Transactivation of the *SlDFR* promoter driving firefly *luciferase* in protoplasts with effector plasmids containing the *MYB SlAN2like* genomic sequence from WT or *Aft* plants, in combination with effector plasmids containing the *bHLH* factor *SlAN1* or *SlJAF13*. Data are expressed as RLU with the value of the promoter basal level set to one and are means of four biological replicates ± SE. One-way ANOVA with Tukey's HSD post-hoc test was performed. Different letters indicate significant differences at P≤0.05.

**(C)** Anthocyanin accumulation in tobacco leaves agro-infiltrated with effector plasmids containing the *SlAN2like* genomic sequence cloned in WT or *Aft* plants expressed with or without the effector plasmid containing the *bHLH* factor *SlAN1*. White dotted circles indicate the agro-infiltrated areas.

Journal Pre-proof

**(D)** Quantification of the anthocyanins produced in the areas of tobacco leaves agro-infiltrated with WT or *Aft* *SlAN2like* in combination with *SlAN1*. Anthocyanins are expressed in µg petunidin-3-(*p*-coumaroyl rutinoside)-5-glucoside $g^{-1}$ fresh weight (FW). Data are means of eight biological replicates ± SE. One-way ANOVA with Tukey's HSD post-hoc test was performed. Different letters indicate significant differences at P≤0.05.

**(E)** *Solanum chilense* mature fruits (picture reproduced with the permission of the author from https://giorgetta.ch/fl_solanaceae_solanum_chilense.htm).

**(F)** Transactivation of the *SlDFR* promoter driving the firefly *luciferase* gene in protoplasts transfected with effector plasmids containing *SlAN2like* genomic sequences from WT or *Aft* plants or *ScAN2like* genomic sequence from *S. chilense*, in combination with the effector plasmid containing the *bHLH* factor *SlAN1*. Data are expressed as RLU with the value of the promoter basal level set to one and are means of four biological replicates ± SE. One-way ANOVA with Tukey's HSD post-hoc test was performed. Different letters indicate significant differences at P≤0.05.

**Figure 3. *SlAN2like*$^{Aft}$ is the major R2R3 MYB factor promoting anthocyanin synthesis in *Aft* fruit peel.**

**(A)** Top half (stem-end) and bottom half (stylar-end) of *Aft* fruit developed under permissive light conditions and photographed at the mature green stage.

**(B)** Anthocyanin content measured in the peel sampled from top and bottom halves of WT and *Aft* fruits at the mature green stage. Anthocyanins are expressed in µg petunidin-3-(*p*-coumaroyl rutinoside)-5-glucoside $g^{-1}$ fresh weight (FW). Data are means of three biological replicates ± SE. One-way ANOVA with Tukey's HSD post-hoc test was performed. Different letters indicate significant differences at P≤0.05.

**(C)** qPCR analysis of regulatory *R2R3 MYB* (*SlAN2*, *SlAN2like*, *SlANT1*, *SlANT1like*), *bHLH* (*SlJAF13*, *SlAN1*), *WDR* (*SlAN11*) and *R3 MYB* (*SlMYB-ATV*) genes performed in the skin from top and bottom halves of WT and *Aft* fruits at the mature green stage. Data are means of eight biological replicates ± SE. One-way ANOVA with Tukey's HSD post-hoc test was performed. Different letters indicate significant differences at P≤0.05.

**(D)** Transactivation of the *SlAN2like* promoters from WT and *Aft* plants and of the *SlDFR* promoter, all of them driving the firefly *luciferase* gene, in leaf protoplasts. As positive control, transactivation of the *SlDFR* promoter in protoplasts transfected with the effector plasmids containing the *SlAN2like*$^{Aft}$ genomic sequence and the *bHLH* factor *SlAN1* is shown. Data are expressed as relative luciferase activity (RLU) (FireflyLuc/RenillaLuc) with the value of the promoter basal level set to one and are means of four biological replicates ± SE. One-way ANOVA with Tukey's HSD post-hoc test was performed. ns means P>0.5, "*" means P≤0.05, and "****" means P≤0.0001, respectively

**(E)** Transactivation of the *SlAN2like* promoters from WT and *Aft* plants and of *SlDFR* promoter, all of them driving the firefly *luciferase* gene, in fruit peel protoplasts isolated from WT fruits at the mature green stage. As positive control, transactivation of the *SlDFR* promoter in protoplasts transfected with the effector plasmids containing the *SlAN2like*$^{Aft}$ genomic sequence and the *bHLH* factor *SlAN1* is shown. Data are

expressed as RLU and are means of four biological replicates ± SE. One-way ANOVA with Tukey's HSD post-hoc test was performed. ns means P>0.5, "*" means P≤0.05, and "****" means P≤0.0001, respectively.

**(F)** Transactivation of the *SlAN2like* promoters from WT and *Aft* plants and of *SlDFR* promoter, all of them driving the firefly *luciferase* gene, in fruit peel protoplasts isolated from *Aft* fruits at the mature green stage. As positive control, transactivation of the *SlDFR* promoter in protoplasts transfected with the effector plasmids containing the *SlAN2like*$^{Aft}$ genomic sequence and the *bHLH* factor *SlAN1* is shown. Data are expressed as RLU and are means of four biological replicates ± SE. One-way ANOVA with Tukey's HSD post-hoc test was performed. "****" means P≤0.0001.

**Figure 4. Structural and functional analysis of the SlAN2like factors produced from the transcripts identified in WT and Aft fruit peel.**

**(A)** Agarose gel electrophoresis of the RT-PCR products showing the SlAN2like transcripts amplified from WT and Aft fruit peel cDNAs. The expected length of the WT SlAN2like cds (Solyc10g086290.1.1) is 798 bp.

**(B)** Schematic representation of intron-exon structure of the WT genomic sequence of SlAN2like with the positions of the "canonical" splicing sites (black arrows) which produce the theoretical transcript registered in the SOL Genomics Network database (Solyc10g086290.1.1 ) (above), and protein produced from its mature mRNA with major functional domains (below). Gene and protein sequences are shown at different scales.

**(C)** Schematic representation of intron-exon structure of the WT genomic sequence of SlAN2like with the positions of the "canonical" splicing sites (black arrows) and the alternative ones (red arrows) which produce the first shorter transcript identified in fruit peel (above), and protein produced from its mature mRNA with major functional domains (below). Gene and protein sequences are shown at different scales.

**(D)** Schematic representation of intron-exon structure of the WT genomic sequence of SlAN2like with the positions of the "canonical" splicing sites (black arrows) and the alternative ones (red arrows) which produce the second shorter transcript identified in fruit peel, and protein produced from its mature mRNA with major functional domains (below). Gene and protein sequences are shown at different scales.

**(E)** Transactivation of the *SlDFR* promoter driving the firefly *luciferase* gene in protoplasts with effector plasmids containing the *SlAN2like* transcripts cloned in WT and *Aft* fruit peel and the *SlAN2like* synthetic cds (corresponding to the theoretical transcript produced from the WT pre-mRNA using the "canonical" splicing sites used in the processing of the pre-mRNA of *SlAN2like*$^{Aft}$). MYB proteins were expressed in combination with SlAN1. Data are expressed as relative luciferase activity (RLU) (FireflyLuc/RenillaLuc) with the value of the promoter basal level set to one and are means of four biological replicates ± SE. One-way ANOVA with Tukey's HSD post-hoc test was performed. Different letters indicate significant differences at P≤0.05.

**(F)** Split-luciferase complementation assay in WT protoplasts expressing the fusion proteins NLuc-SlAN2like$^{WT}$ or NLuc-SlAN2like$^{Aft}$ with CLuc-SlAN1 or CLuc-SlAN11. Combinations of each construct with the empty vectors expressing the complementary half of the luciferase gene represent negative controls. Data are expressed as Relative Luciferase Activity (RLU) and are means of four biological replicates ± SE. One-way ANOVA with Tukey's HSD post-hoc test was performed. Each box was compared with the first one, and asterisks indicate significant differences at P≤0.0001.

Journal Pre-proof

**(G)** Bimolecular fluorescence complementation assay analyzing the interaction between SlAN2likeAft and SlAN1 in tomato protoplasts expressing the fusion proteins YFPN-SlN2likeAft and YFPC-SlAN1. As a control, YFPC-half protein was expressed in combination with YFPN-SlAN2likeAft fusion protein. Figure 4. Structural and functional analysis of the SlAN2like factors produced from the transcripts identified in WT and *Aft* fruit peel.

## ACCESSION NUMBERS

Genomic sequences of *AN2like* from *Aft* and *S. chilense* as well as genomic *ANT1like* sequence from *Aft* were deposited in the GenBank database with the following accession numbers: *SlAN2like*$^{Aft}$: MN242011, *ScAN2like*: MN242012, *SlANT1like*$^{Aft}$: MN242013.

## AUTHOR CONTRIBUTIONS

P.P. and S.G. conceived and designed the project. S.C. performed molecular cloning, transactivation assays, split-luciferase complementation and BiFC assays, anthocyanin measurements. S.G. performed gene expression analysis and sequence analyses. A.T. carried out microscope analysis. S.C., P.P. and S.G. wrote the manuscript.

## ACKNOWLEDGEMENTS

We would like to thank Dr. Remko Stam for his assistance in the analysis of the *ScAN2like* promoter sequence. The authors declare no competing interests.

## REFERENCES

Albert, N.W., Davies, K.M., Lewis, D.H., Zhang, H., Montefiori, M., Brendolise, C., Boase, M.R., Ngo, H., Jameson, P.E., and Schwinn, K.E. (2014). A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. Plant Cell **26**:962-980.

An, X.H., Tian, Y., Chen, K.Q., Wang, X.F., and Hao, Y.J. (2012). The apple WD40 protein MdTTG1 interacts with bHLH but not MYB proteins to regulate anthocyanin accumulation. J. Plant Physiol. **169**:710-717.

Bai, Y., and Lindhout, P. (2007). Domestication and breeding of tomatoes: what have we gained and what can we gain in the future? Ann. Bot. **100**: 1085-1094.

Bedinger, P.A., Chetelat, R.T., McClure, B., Moyle, L.C., Rose, J.K., Stack, S.M., van der Knaap, E., Baek, Y.S., Lopez-Casado, G., Covey, P.A. et al. (2011). Interspecific reproductive barriers in the tomato clade: opportunities to decipher mechanisms of reproductive isolation. Sex. Plant Reprod. **24**:171-187.

Bovy, A., de Vos, R., Kemper, M., Schijlen, E., Almenar Pertejo, M., Muir, S., Collins, G., Robinson, S., Verhoeyen, M., Hughes, S., et al. (2002). High-flavonol tomatoes resulting from the heterologous expression of the maize transcription factor genes LC and C1. Plant Cell **14**:2509-2526.

Journal Pre-proof

Brendolise, C., Espley, R.V., Lin-Wang, K., Laing, W., Peng, Y., McGhie, T., Dejnoprat, S., Tomes, S., Hellens, R.P., Allan, A.C. (2017). Multiple Copies of a Simple MYB-Binding Site Confers Trans-regulation by Specific Flavonoid-Related R2R3 MYBs in Diverse Species. Front. Plant Sci. **8**:1864.

Butelli, E., Titta, L., Giorgio, M., Mock, H.P., Matros, A., Peterek, S., Schijlen, E.G., Hall, R.D., Bovy, A.G., Luo, J., et al. (2008). Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nat. Biotechnol. **26**:1301-1308.

Cao, X., Qiu, Z., Wang, X., Van Giang, T., Liu, X., Wang, J., Wang, X., Gao, J., Guo, Y., Du, Y., et al. (2017). A putative R3 MYB repressor is the candidate gene underlying *atroviolacium*, a locus for anthocyanin pigmentation in tomato fruit. J. Exp. Bot. **68**:5745-5758.

Chetelat, R.T., Pertuzé, R.A., Faúndez, L., Graham, E.B., Jones, C.M. (2008). Distribution, ecology and reproductive biology of wild tomatoes and related nightshades from the Atacama Desert region of northern Chile. Euphytica **167**:77–93.

Colanero, S., Perata, P., and Gonzali, S. (2018). The *atroviolacea* Gene Encodes an R3-MYB Protein Repressing Anthocyanin Synthesis in Tomato Plants. Front. Plant Sci. **9**:830. doi: 10.3389/fpls.2018.00830.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M. et al. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. **36**:W465-469.

Doganlar, S., Frary, A., Daunay, M.C., Lester, R.N., and Tanksley, S.D. (2002).Conservation of gene function in the *Solanaceae* as revealed by comparative mapping of domestication traits in eggplant. Genetics **161**:1713-1726.

Fantini, E., Falcone, G., Frusciante, S., Giliberto, L., and Giuliano, G. (2013). Dissection of tomato lycopene biosynthesis through virus-induced gene silencing. Plant Physiol. **163**:986-998.

FAO. (2010). Plant genetic resource for food and agriculture. Rome, Food and Agriculture Organisation of the United Nations.

Fujikawa, Y., and Kato, N. (2007). Split luciferase complementation assay to study protein-protein interactions in Arabidopsis protoplasts. Plant J. **52**:185–195

Gao, Y., Liu, J., Chen, Y., Tang, H., Wang, Y., He, Y., Ou, Y., Sun, X., Wang, S., and Yao, Y. (2018). Tomato SlAN11 regulates flavonoid biosynthesis and seed dormancy by interaction with bHLH proteins but not with MYB proteins. Hortic. Res. **5**:27. doi: 10.1038/s41438-018-0032-3

Georgiev, C. (1972). Anthocyanin fruit (Af). Rep. Tomato Genet. Coop. **22**:10.

Gonzali, S., Mazzucato, A., and Perata, P. (2009). Purple as a tomato: towards high anthocyanin tomatoes. Trends Plant Sci. **14**:237-241.

Heppel, S.C., Jaffé, F.W., Takos, A.M., Schellmann, S., Rausch, T., Walker, A.R., and Bogs, J. (2013). Identification of key amino acids for the evolution of promoter target specificity of anthocyanin and proanthocyanidin regulating MYB factors. Plant Mol. Biol. **82**:457-471.

Journal Pre-proof

Karimi, M., Inzé, D., and Depicker, A. (2002). GATEWAY vectors for Agrobacterium-mediated plant transformation. Trends Plant Sci. **7**:193-195.

Kiferle, C., Fantini, E., Bassolino, L., Povero, G., Spelt, C., Buti, S., Giuliano, G., Quattrocchio, F., Koes, R., Perata, P., et al. (2015). Tomato R2R3-MYB Proteins SlANT1 and SlAN2: Same Protein Activity, Different Roles. PLoS One **10**:e0136365. doi: 10.1371/journal.pone.0136365

Kiferle, C., Gonzali, S., Holwerda, H.T., Ibaceta, R.R., and Perata, P. (2013). Tomato fruits: a good target for iodine biofortification. Front. Plant Sci. **4**:205. doi: 10.3389/fpls.2013.00205.

Li, X. (2011). Infiltration of *Nicotiana benthamiana* Protocol for Transient Expression via Agrobacterium. Bio-101:e95.

Lin-Wang, K., Bolitho, K., Grafton, K., Kortstee, A., Karunairetnam, S., McGhie, T.K., Espley, R.V., Hellens, R.P., and Allan, A.C. (2010). An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in *Rosaceae*. BMC Plant Biol. **10**:50.

Liu, C., Zhu, L., Fukuda, K., Ouyang, S., Chen, X., Wang, C., Zhang, C.J., Martin, B., Gu, C., Qin, L. et al. (2017). The flavonoid cyanidin blocks binding of the cytokine interleukin-17A to the IL-17RA subunit to alleviate inflammation in vivo. Sci. Signal. **10:**467.

Liu, J., Osbourn, A., and Ma, P. (2015). MYB Transcription Factors as Regulators of Phenylpropanoid Metabolism in Plants. Mol. Plant. **8**:689-708.

Liu, Y., Tikunov, Y., Schouten, R.E., Marcelis, L.F.M., Visser, R.G.F., and Bovy, A. (2018). Anthocyanin Biosynthesis and Degradation Mechanisms in Solanaceous Vegetables: A Review. Front. Chem**. 6**:52.

Martí, R., Roselló, S., and Cebolla-Cornejo, J. (2016). Tomato as a Source of Carotenoids and Polyphenols Targeted to Cancer Prevention. Cancers (Basel) **8**:6. doi: 10.3390/cancers8060058

Martin, C., Butelli, E., Petroni, K., and Tonelli, C. (2011). How can research on plants contribute to promoting human health? Plant Cell **23**:1685-1699.

Mathews, H., Clendennen, S.K., Caldwell, C.G., Liu, X.L., Connors, K., Matheis, N., Schuster, D.K., Menasco, D.J., Wagoner, W., Lightner, J., et al. (2003). Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. Plant Cell **15**:1689-1703.

Mes, P.J., Boches, P., and Myers, J.R. (2008). Characterization of Tomatoes Expressing Anthocyanin in the Fruit. J. Amer. Soc. Hort. Sci. **133**:262–269.

Montefiori, M., Brendolise, C., Dare, A.P., Lin-Wang, K., Davies, K.M., Hellens, R.P., and Allan, A.C. (2015). In the *Solanaceae*, a hierarchy of bHLHs confer distinct target specificity to the anthocyanin regulatory complex. J. Exp. Bot. **66**:1427-1436.

Myers, J. (2012). Breeding Tomatoes for Increased Flavonoids. (2012). Strengthening Community Seed Systems. In Proceedings of the 6$^{th}$ Organic Seed Growers Conference, Port Townsend, Washington, USA, 19-21 January, pp:50–51.

Nesi, N., Debeaujon, I., Jond, C., Pelletier, G., Caboche, M., and Lepiniec, L. (2000). The TT8 gene encodes a basic helix-loop-helix domain protein required for expression of DFR and BAN genes in Arabidopsis siliques. Plant Cell **12**:1863–1878.

Nukumizu, Y., Wada, T., and Tominaga-Wada, R. (2013). Tomato (*Solanum lycopersicum*) homologs of TRIPTYCHON (SlTRY) and GLABRA3 (SlGL3) are involved in anthocyanin accumulation. Plant Signal Behav. **8**:e24575.

Peralta, I.E., and Spooner, D.M. (2007). History, origin and early cultivation of tomato (*Solanaceae*). Genetic improvement of solanaceous crops, Vol. 2. Tomato, Science Publishers.

Povero, G., Gonzali, S., Bassolino, L., Mazzucato, A., and Perata, P. (2011). Transcriptional analysis in high-anthocyanin tomatoes reveals synergistic effect of *Aft* and *atv* genes. J. Plant Physiol. **168**:270-279.

Qiu, Z., Wang, H., Li, D., Yu, B., Hui, Q., Yan, S., Huang, Z., Cui, X., and Cao, B. (2019). Identification of Candidate HY5-Dependent and -Independent Regulators of Anthocyanin Biosynthesis in Tomato. Plant Cell Physiol. **60**:643-656.

Qiu, Z., Wang, X., Gao, J., Guo, Y., Huang, Z., and Du, Y. The Tomato Hoffman's Anthocyaninless Gene Encodes a bHLH Transcription Factor Involved in Anthocyanin Biosynthesis That Is Developmentally Regulated and Induced by Low Temperatures. (2016). PLoS One **11**:e0151067.

Ramsay, N. A., and Glover, B. J. (2005). MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. Trends Plant. Sci. **10**:63–70.

Sapir, M., Oren-Shamir, M., Ovadia, R., Reuveni, M., Evenor, D., Tadmor, Y., Nahon, S., Shlomo, H., Chen, L., Meir, A. et al. (2008). Molecular aspects of *Anthocyanin fruit* tomato in relation to *high pigment-1*. J. Hered. **99**:292-303.

Schreiber, G., Reuveni, M., Evenor, D., Oren-Shamir, M., Ovadia, R., Sapir-Mir, M., Bootbool-Man, A., Nahon, S., Shlomo, H., Chen, L., et al. (2011). ANTHOCYANIN1 from *Solanum chilense* is more efficient in accumulating anthocyanin metabolites than its *Solanum lycopersicum* counterpart in association with the *ANTHOCYANIN FRUIT* phenotype of tomato. Theor. Appl. Genet. **124**:295-307.

Shi, X., Gupta, S., and Rashotte, A.M. (2012). *Solanum lycopersicum* cytokinin response factor (SlCRF) genes: characterization of CRF domain-containing ERF genes in tomato. J. Exp. Bot. **63**: 973-982.

Sibley, C.R., Blazquez, L., Ule, J. (2016). Lessons from non-canonical splicing. Nat. Rev. Genet. **17**: 407-421.

Stam, R., Nosenko, T., Hörger, A.C., Stephan, W., Seidel, M., Kuhn, J.M.M., Haberer, G. Tellier, A. (2019). The de Novo Reference Genome and Transcriptome Assemblies of the Wild Tomato Species Solanum chilense Highlights Birth and Death of NLR Genes Between Tomato Species. (2019). G3 Oct **11**. pii: g3.400529.2019.

Stracke, R., Werber, M., and Weisshaar, B. (2001). The R2R3-MYB gene family in *Arabidopsis thaliana*. Curr. Opin. Plant Biol. **4**:447-456.

Syed, N.H., Kalyna, M., Marquez, Y., Barta, A., and Brown, J.W. (2012). Alternative splicing in plants - coming of age. Trends Plant Sci. **17**:616-623.

Journal Pre-proof

Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485**:635-641.

Wang, G., Chen, B., Du, H., Zhang, F., Zhang, H., Wang, Y., He, H., Geng, S., and Zhang, X. (2018). Genetic mapping of anthocyanin accumulation-related genes in pepper fruits using a combination of SLAF-seq and BSA. PLoS One **13**:e0204690.

Weits, D. A., Giuntoli, B., Kosmacz, M., Parlanti, S., Hubberten, H. M., Riegler, H., Hoefgen, R., Perata, P., van Dongen, J.T., and Licausi, F. (2014). Plant cysteine oxidases control the oxygen-dependent branch of the N-end-rule pathway. Nat. Commun. **5**:3425.

Xu, W., Dubos, C., and Lepiniec, L. (2015). Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. Trends Plant Sci. **20**:176-85.

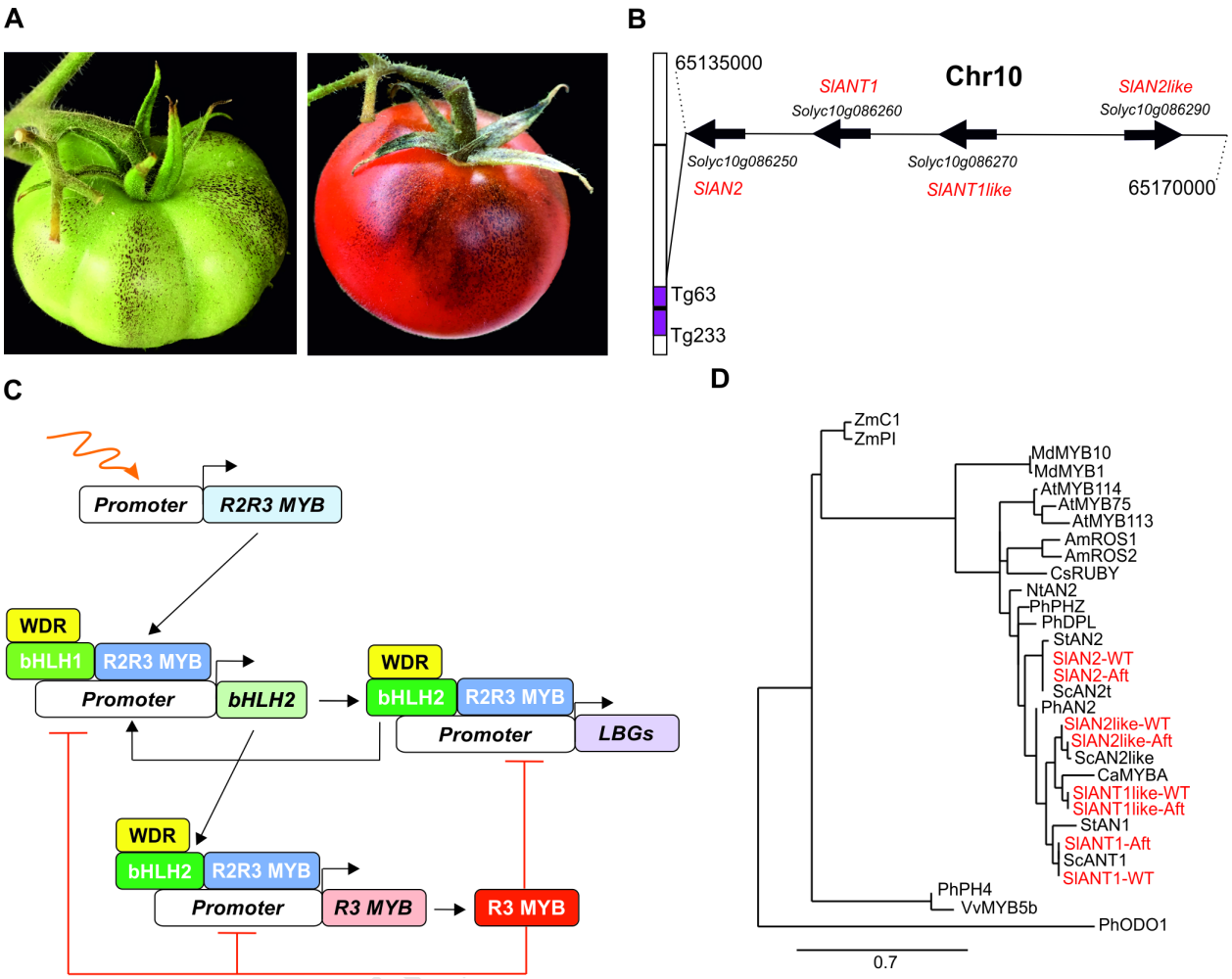
Yoo, S.D., Cho, Y.H., and Sheen, J. (2007). Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. Nat. Protoc. **2**:1565–1572.

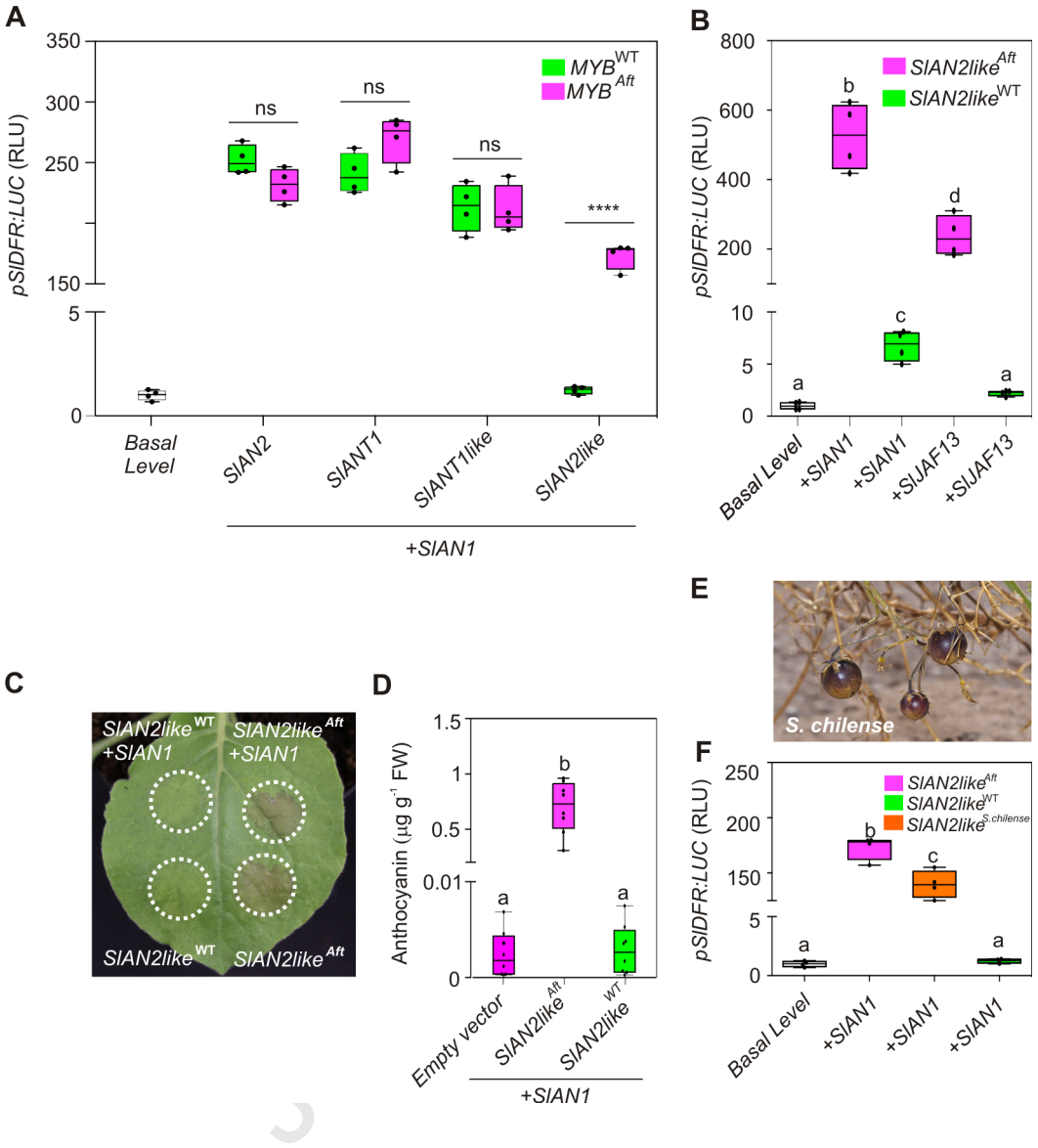
Zhang, F., Gonzalez, A., Zhao, M., Payne, C.T., and Lloyd, A. (2003). A network of redundant bHLH proteins functions in all TTG1-dependent pathways of Arabidopsis. Development **130**:4859–4869.

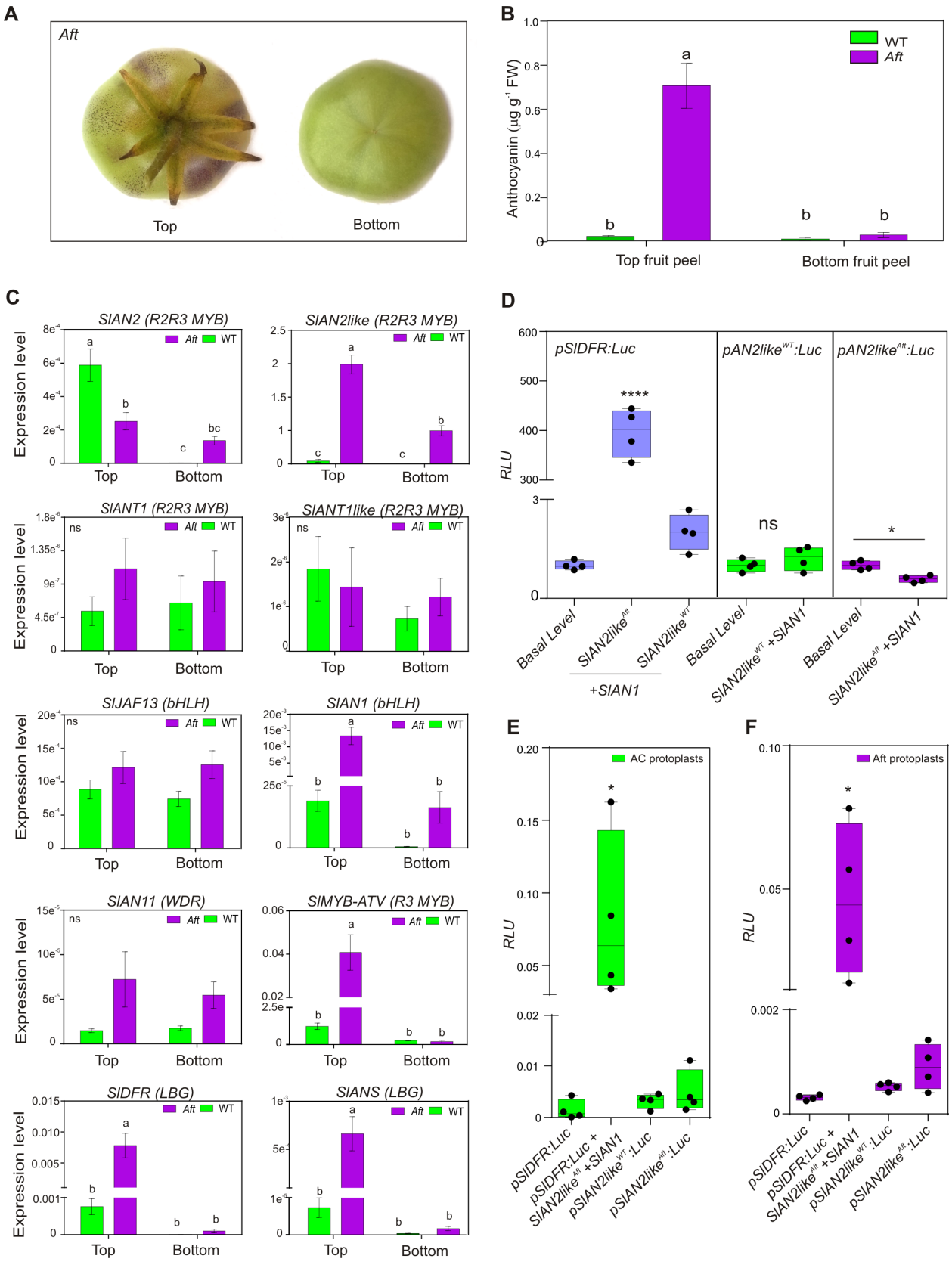
Zhong, S., Lin, Z., Fray, R.G., and Grierson, D. (2008). Improved plant transformation vectors for fluorescent protein tagging. Transgenic Res. **17**:985-989.

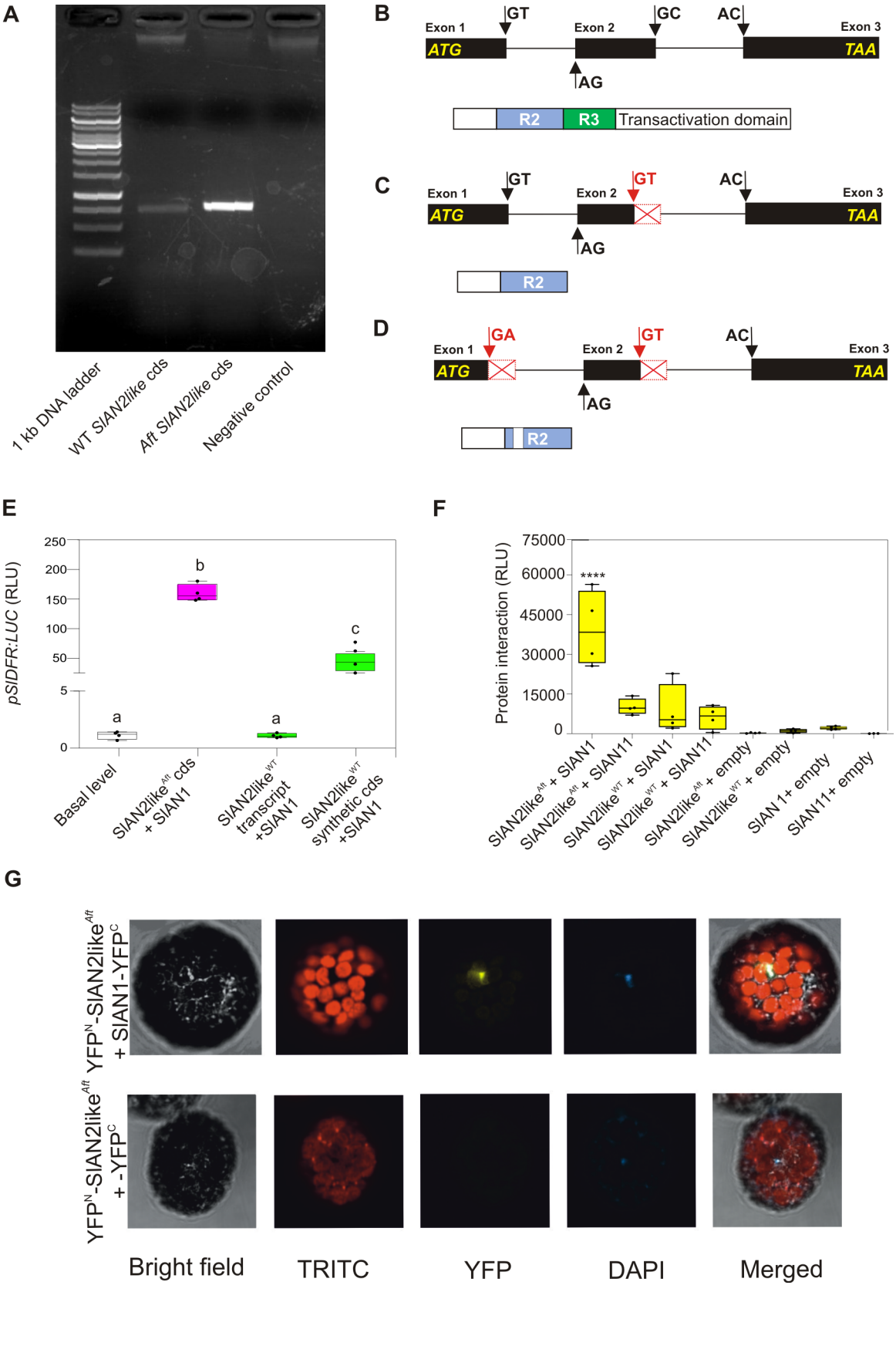
Zimmermann, I.M., Heim, M.A., Weisshaar, B., and Uhrig, J.F. (2004). Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. Plant J. **40**:22-34.

Journal Pre-proof


A
B
65135000
Chr10
SlANT1
Solyc10g086260
SlAN2like
Solyc10g086290
Solyc10g086250
SlAN2
Solyc10g086270
SlANT1like
65170000
Tg63
Tg233
C
Promoter
R2R3 MYB
WDR
bHLH1
R2R3 MYB
Promoter
bHLH2
WDR
bHLH2
R2R3 MYB
Promoter
LBGs
WDR
bHLH2
R2R3 MYB
Promoter
R3 MYB
R3 MYB
D
ZmC1
ZmPl
MdMYB10
MdMYB1
AtMYB114
AtMYB75
AtMYB113
AmROS1
AmROS2
CsRUBY
NtAN2
PhPHZ
PhDPL
StAN2
SlAN2-WT
SlAN2-Aft
ScAN2t
PhAN2
SlAN2like-WT
SlAN2like-Aft
ScAN2like
CaMYBA
SlANT1like-WT
SlANT1like-Aft
StAN1
SlANT1-Aft
ScANT1
SlANT1-WT
PhPH4
VvMYB5b
PhODO1
0.7


A
pSlDFR:LUC (RLU)
MYB^WT
MYB^Aft
ns
ns
ns
****
Basal Level
SlAN2
SlANT1
SlANT1like
SlAN2like
+SlAN1
B
pSlDFR:LUC (RLU)
SlAN2like^Aft
SlAN2like^WT
Basal Level
+SlAN1
+SlAN1
+SlJAF13
+SlJAF13
C
SlAN2like^WT +SlAN1
SlAN2like^Aft +SlAN1
SlAN2like^WT
SlAN2like^Aft
D
Anthocyanin (µg g-1 FW)
Empty vector
SlAN2like^Aft
SlAN2like^WT
+SlAN1
E
S. chilense
F
pSlDFR:LUC (RLU)
SlAN2like^Aft
SlAN2like^WT
SlAN2like^S.chilense
Basal Level
+SlAN1
+SlAN1
+SlAN1


A
Aft
Top
Bottom
B
Anthocyanin (μg g-1 FW)
WT
Aft
Top fruit peel
Bottom fruit peel
C
SlAN2 (R2R3 MYB)
SlAN2like (R2R3 MYB)
SlANT1 (R2R3 MYB)
SlANT1like (R2R3 MYB)
SlJAF13 (bHLH)
SlAN1 (bHLH)
SlAN11 (WDR)
SlMYB-ATV (R3 MYB)
SlDFR (LBG)
SlANS (LBG)
Expression level
ns
Top
Bottom
D
pSlDFR:Luc
pAN2likeWT:Luc
pAN2likeAft:Luc
RLU
****
ns
*
Basal Level
SlAN2likeAft
SlAN2likeWT
+SlAN1
SlAN2likeWT +SlAN1
SlAN2likeAft +SlAN1
E
AC protoplasts
RLU
pSlDFR:Luc
pSlDFR:Luc + SlAN2likeAft +SlAN1
pSlAN2likeWT:Luc
pSlAN2likeAft:Luc
F
Aft protoplasts
RLU
pSlDFR:Luc
pSlDFR:Luc + SlAN2likeAft +SlAN1
pSlAN2likeWT:Luc
pSlAN2likeAft:Luc


A
1 kb DNA ladder
WT SlAN2like cds
Aft SlAN2like cds
Negative control
B
Exon 1
GT
Exon 2
GC
AC
Exon 3
ATG
TAA
AG
R2
R3
Transactivation domain
C
Exon 1
GT
Exon 2
GT
AC
Exon 3
ATG
TAA
AG
R2
D
Exon 1
GA
Exon 2
GT
AC
Exon 3
ATG
TAA
AG
R2
E
pSlDFR:LUC (RLU)
250
200
150
100
50
5
0
a
b
a
c
Basal level
SlAN2like^Aft cds + SlAN1
SlAN2like^WT transcript +SlAN1
SlAN2like^WT synthetic cds +SlAN1
F
Protein interaction (RLU)
75000
60000
45000
30000
15000
0
****
SlAN2like^Aft + SlAN1
SlAN2like^Aft + SlAN11
SlAN2like^WT + SlAN1
SlAN2like^WT + SlAN11
SlAN2like^Aft + empty
SlAN2like^WT + empty
SlAN 1+ empty
SlAN11+ empty
G
YFP^N-SlAN2like^Aft + SlAN1-YFP^C
YFP^N-SlAN2like^Aft + -YFP^C
Bright field
TRITC
YFP
DAPI
Merged