Chapter 5

Functional characterisation of a small family of MYB genes of tomato, shows that the Anthocyanin fruit locus encodes the \textit{SIAN2} gene

\textit{with Cornelis Spelt, Pierdomenico Perata, Silvia Gonzali, Simona Cardelicchio, Ronald Koes and Francesca Quattrrocchio}

Abstract
Anthocyanins are secondary metabolites that colour leaves, flowers and fruits and have antioxidant and nutraceutical properties. Unfortunately, most widely consumed vegetables and fruits, such as cultivated tomatoes, accumulate only very small amounts of these compounds. By interspecific crosses with wild \textit{Solanum} species, loci such as \textit{Anthocyanin fruit (Aft)}, have been introgressed into tomato. However, the nature of the dominant allele \textit{Aft}, which drives anthocyanin biosynthesis in the skin of the fruit, has remained unknown, despite its economical importance for the production of anthocyanin containing tomatoes. To identify the \textit{Aft} locus, we studied two tomato MYB genes, \textit{SIANT1} and \textit{SIANT2}, that are considered probable candidates for the \textit{Aft} locus. Comparison of the genomic fragment containing these two genes in control tomato and \textit{Aft} lines revealed the presence of numerous sequence polymorphisms in the promoter, introns and the coding sequence. \textit{SIANT2} is expressed in the \textit{Aft} line at a much higher level than \textit{SIANT1}, suggesting a major contribution of this gene to the “\textit{Anthocyanin fruit}” phenotype. Because ectopic expression of \textit{SIAN2} or \textit{SIANT1} allele from either a control (cv Alisa Craig) or \textit{Aft} lines equally activate strong production of anthocyanins in the whole plant, producing “deep purple” fruits, the \textit{Aft} phenotype is unlikely to be due to differences in the encoded proteins. However, experiments in which a genomic fragment of the \textit{SIAN2}$^{Aft}$ allele including its own promoter was introduced in cultivated tomato, highly induced expression of \textit{SIANT1} and structural anthocyanin genes resulting in an “\textit{Aft-like}” phenotypes. Taken together, these experimental evidences strongly indicate that \textit{SIAN2} is the \textit{Aft} locus.

Introduction
Anthocyanins are a group of water-soluble flavonoids derived from the phenylpropanoid biochemical pathway. They are wide spread pigments in the plant kingdom and confer red, purple, and blue pigmentation to most vegetative tissues, flowers, seeds, and fruits (Holton and Cornish, 1995). Besides their function in the attraction of pollinators and other agents for seed dispersal, anthocyanins also play, due to their chemical structure and antioxidant activity, a protective role in plant tissues. The biosynthesis of these compounds is induced -particularly in leaves and fruits- by high irradiance, low temperatures, drought or other environmental stresses (Gould, 2004). Anthocyanins have been shown to be also involved in the defence against pathogens, as ant herbivore and antimicrobial compounds (Karageorgou and Manetas, 2006).

Anthocyanins are ubiquitously present in plants and are therefore part of our diet (Gonzali et al., 2009). In the last decades, the antioxidant and health-promoting properties of these compounds have been documented in a still increasing number of studies showing that anthocyanins can improve sight and may have anti-inflammatory and anti-carcinogenic activities (Prior, 2003; Lila, 2004; Galvano et al., 2007; Kampa et al., 2007; Butelli et al., 2008; Wang and Stoner, 2008). Furthermore, their intake has been shown to stimulate protection against cholesterol-induced atherosclerosis and coronary heart disease (Prior, 2003). These findings stimulated the proliferation of studies aiming to increase the content of flavonoids and anthocyanins in crop plants (Boudet, 2007).
The biosynthesis of flavonoids starts from precursors originating from the shikimate and acetate-malonate pathways (Dooner et al., 1991; Awad et al., 2000). Many of the structural genes, encoding the enzymes of the pathway, and regulatory genes that control their transcription have been identified in several species, including Arabidopsis and petunia. The transcription of structural genes such as DIHYDROFLAVONOL REDUCTASE (DFR), which are specific for the anthocyanin pathway, is activated by a conserved triad of proteins with a MYB, basic-Helix-Loop-Helix (bHLH) and a WD40 domain respectively. In petunia and Arabidopsis the WD40 protein is expressed in virtually all tissues, whereas expression of the bHLH and MYB partners is confined to pigmented tissues. Consequently, forced expression of the bHLH and MYB gene is sufficient for ectopic expression of structural anthocyanins and the coloration of tissues that normally lack anthocyanins. Our recent study showed that in petunia the MYB proteins ANTHOCYANIN2 (AN2) and AN4 are sufficient to activate anthocyanin synthesis ectopically, as they can induce the transcription of the partner bHLH gene AN1 (chapter 6).

Tomato (*Solanum lycopersicum* L.), is a popular food plant whose fruits are among the most consumed vegetables worldwide. Although tomato plants can synthesize anthocyanins in vegetative tissues, showing that the biochemical pathway leading to anthocyanin accumulation is intact in this species, in the tomato fruits there are only very small amounts of flavonoids, mostly concentrated in the peel, where they accumulate during ripening. The typical red colour of cultivated tomatoes is due to massive accumulation of the carotenoid lycopene (Muir et al., 2001; Verhoeven et al., 2002; Le Gall et al., 2003; Torres et al., 2005; Bovy et al., 2007). The “black” or “purple” fruit colours that are common in some tomato heirloom varieties, originate from mutations affecting chlorophyll breakdown and carotenoid content, and are not related to anthocyanin production (Mes et al., 2008). However, anthocyanin pigmentation in fruits can be observed in many species closely related to tomato (e.g. *Solanum chilense*, *S. peruvianum*, *S. habrochaites*, *S. cheesmanii*, and *S. lycopersicoides*). The fact that anthocyanins still can be produced in some taxonomically close wild species suggests that the production of these compounds could have been lost during domestication.

To activate abundant anthocyanin synthesis in the tomato fruit it is sufficient to ectopically express DELILA and ROSEA (Butelli et al., 2008), which are bHLH and MYB regulators of the anthocyanin pathway in snapdragon flowers (Goodrich et al., 1992; Schwinn et al., 2006). This demonstrates that the absence of anthocyanins in normal tomato fruits is because the endogenous genes encoding the MYB and bHLH regulators are not expressed in this organ (Butelli et al., 2008; Gonzali et al., 2009). However, as consumers are often reluctant to accept genetically modified fruits or vegetables, at least within short time, plant breeding remains a more realistic approach for the enrichment of anthocyanins in tomato.

A classical approach is to use wild *Solanum* species taxonomically close to *S. lycopersicum*, that have retained the ability to produce anthocyanins in fruits, for breeding with cultivated tomato (Jones et al., 2003). In this way it has been possible to transfer the ability to produce small amounts of anthocyanins in the peel of the fruit into cultivated tomatoes (Jones et al., 2003; Gonzali et al., 2009). The dominant locus *Aft* (*Anthocyanin fruit*) on chromosome 10, which triggers pigmentation upon stimulation by light, has been reported to originate from a cross of domesticated tomato with *S. chilense*, which is a wild tomato species with pigmented fruits (Georgiev, 1972). The dominant allele of *Aft* induces anthocyanin biosynthesis in immature green fruit, and subsequently anthocyanins are accumulate continuously throughout development (Mes et al., 2008). Further characterization of this tomato line showed that the *Aft* gene affects the expression of structural genes directly involved in anthocyanin production, such as FLAVONOID 3’5’-HYDROXILASE (F3’5’H), DFR, and LEUCANTHOXYANIDIN DIOXYGENASE (ANS) (Povero et al., 2011). Furthermore, the line (named *Aft*, accession number LA1966) containing this gene displays an elevated carotenoid content.
A recessive allele at another locus, Atroviolacea (Atv) located on chromosome 7, was introduced in tomato from a natural population of S. pimpinellifolium (Rick et al., 1968). Atv stimulates anthocyanin pigmentation in the whole tomato plant, particularly in vegetative tissues, in response to light stimulation (Mes et al., 2008) and the “SunBlack” line resulting from the cross between the atv and the Aft lines (double “mutant” atv Aft) shows intense coloration of vegetative organs and fruits. The analysis of transcriptomic changes in “SunBlack”, where Aft and atv alleles are both present, confirmed a synergistic effect of these two alleles in specifically activating anthocyanin biosynthetic genes (Povero et al., 2011).

Linkage studies have shown that the Aft locus co-segregates with two distinct genes located on chromosome 10, named SIAN2 (ANTHOCYANIN2; Boches and Myers, 2007) and SIAN1 (ANTHOCYANIN1; Sapir et al., 2008), both encoding R2-R3 MYB proteins and are highly similar to the regulatory anthocyanin genes AN2 of petunia (Quattrocchio et al., 1999) and ROSEA of snapdragon (Schwinn et al., 2006; Figure 1D).

In this study we analysed SIAN1 and SIAN2, to establish the contribution of each of these genes to the Anthocyanin fruit phenotype. We show here that SIAN2 and SIAN1 proteins can both activate anthocyanin biosynthesis in any tomato plant part when ectopically expressed. Moreover, transgenic Ailsa Craig (AC, our control line) tomato plants harbouring the SIAN2\textsuperscript{Atv} complete gene, phenocopy the Aft phenotype. Together with sequence comparison of the regulatory regions in Aft and control cultivars, these results strongly indicate that the SIAN2 gene is the Aft locus and that the difference between the AC and the Aft allele is in the regulation of the expression rather than in the protein functionality.

Results

SIAN1 and SIAN2 are adjacent genes

Several studies suggested that either SIAN1 or SIAN2, or both, might represent the Aft locus, because they co-segregate in crosses with the Aft phenotype (Boches and Myers, 2007; Povero et al., 2011). However, since the mapping populations were rather small, it cannot be ruled out that Aft is a separate locus in the same region. Analysis of the genomic regions containing SIAN1 and SIAN2 combined with the sequence information available on the website of the “Sol Genomic Network” (http://solgenomics.net/tools/blast/index.pl), enabled us to piece together the sequence of a single genomic fragment including both genes in the cultivated tomato control line AC (see material and methods) which harbours anthocyaninless fruits. The two genes turned out to be only ~2.7 kb apart (precisely 2.684 bp between the STOP codon of SIAN1 and the ATG of SIAN2; Figure 1A) and on the very same chromosome fragment we could identify yet another R2R3-MYB gene, located ~6 kb upstream of the SIAN1 coding region. We named this novel gene SIAN1-like because it is more
similar (84% identity at nucleotide level) to SIAN1 (Supplemental Data IA), than to SIAN2 (79% identity).

**Figure 1.** Structure and phylogenetic relationships of SIAN1 and SIAN2.

(A) Diagram showing the structure of genomic fragment from the “wild type” cultivar AC (LA2838A) containing SIAN1-like, SIAN1, and SIAN2. Black arrows indicate the orientation of the genes from 5’ to 3’, while the transcribed region of each gene is represented by grey-filled rectangles.

(B) PCR analysis of SIAN1 and SIAN2 alleles from AC and Aft tissues, starting from genomic DNA (gDNA) and/or cDNA, revealing a length polymorphism among SIAN2 alleles.

(C) Diagram showing the intron-exon structure (from ATG to STOP) of the SIAN1 and SIAN2 alleles of AC, Aft, and *S. chilense*. Exons are represented by solid black boxes. The insertion in the ScAN2 allele has many characteristics of a transposon, while the insertion in SIAN2AC does not have any of the characteristics of a mobile element.

(D) Phylogenetic tree showing the sequence similarity between the MYB domains of proteins involved in pigmentation in different plant species. The tree was constructed using the Neighbor-Joining (NJ) algorithm (MEGA 4.1 software). The GenBank accession number and the source of each sequence are described in the “Material and methods” section.
(Supplemental Data IIB). Thus the finding that both SlAN2 and SlANT1 co-segregate with the Aft locus is explained by their genomic contiguity.

Phylogenetic analysis of the three MYB genes in this small genomic region (Figure 1D) shows that they are, member of two distinct (although close) clades: SlANT1 and SlANT1-like are more related to AN4 of petunia, while SlAN2 is more related to AN2 of petunia. All three genes are very related to each other, indicating that they might originate from relatively recent duplications. This suggests that this gene cluster arose by two separate duplications. The first duplication occurred before the divergence of petunia and tomato lineage and gave rise to the SlAN2 and SlANT1 clade, followed later by a second duplication that gave rise to SlANT1-like.

Sequence polymorphisms in SlANT1 and SlAN2 of AC and Aft S. lycopersicum lines and S. chilense

To study sequence polymorphisms of the SlANT1 and SlAN2 genes in the AC, Aft and S. chilense genotypes, we amplified and sequenced the transcribed regions of these genes from the three different accessions. This showed that the SlAN2AC allele contained in the second intron an insertion of ~400 bp, when compared to SlAN2AT (Figure 1B and 1C; Supplemental Data IIA), while ScAN2 contains instead an insertion of 275 bp at a different position in the same intron (Figure 1C, Supplemental Data IIB). While the insertion in the SlAN2AC does not show any peculiarity, the sequence of the insertion in the ScAN2 allele presents the typical features of a transposon, such as a “Target Site Duplication (TSD)” of 8 bp and imperfect “Terminal Inverted Repeats (TIR)” (Figure 1C; Supplemental Data IIB). However, it appears that today this element is no longer mobile. These insertions in the second intron of SlAN2AC and SlAN2AT do not seem to have any effect on the splicing of the intron as PCR amplification of cDNAs resulted in amplicons of the same size (Figure 1B).

Besides the sequence length polymorphisms in the second intron of SlAN2AC, SlAN2AT and ScAN2, we could pinpoint several other differences in both the coding and non-coding regions (Figure 1C; Supplemental Data IIA-B). Some of the nucleotide substitutions result in amino acid substitutions in the encoded proteins (Supplemental Data IIC).

It is striking that the polymorphisms in these three alleles do not support the idea that SlAN2AT, is introgressed from S. chilense. For example, as the S. chilense allele has a (possibly defective) transposon insertion that was lost by excision after the introgression into the Aft line, than SlAN2AT should contain a typical transposon footprint at the empty donor site, which is not the case. Moreover, there are many other polymorphisms seen between SlAN2AT and ScAN2 that are difficult to fit with the idea that SlAN2AT is an introgression from S. chilense. On the other hand, SlAN2AT can also not derive from the SlAN2AC allele, as this has a peculiar insertion in a different position. The conclusion of the sequence comparison of the different alleles is that the SlAN2AT allele has an unknown origin.

As recent studies suggested that the amino acid substitutions observed in SlANT1AT could potentially affect the function of the protein, and explain (at least in part) the Aft phenotype (Sapir et al., 2008) we also analysed in detail the sequence of this gene. Surprisingly, the sequence that we determined for SlANT1AT was identical to the sequence that Sapir et al. (2008) reported incorrectly as ScANT1. Furthermore, we found several differences between the sequence of ScANT1 and the one of SlANT1AT, which make the two alleles easily distinguishable from each other (Figure 1C-D; Supplemental Data IIE).

Finally, also for AN1, the observed sequence polymorphisms do not support the idea that SlANT2AT originates as an introgression from S. chilense.

Phylogenetic analysis of the SlAN2 and SlANT1 genes from tomato and S. chilense
The alignment based on the MYB domains of several R2R3-MYB factors (Supplemental Data III) involved in anthocyanin biosynthesis in different species (using the MEGA 4.1 software) revealed that SIA1 and SIA2 belong to two different, but very closely related, clades (Figure 1D) and SIA1 and SIA1-like are putative orthologs of AN2 from petunia, while SIA2 appears orthologous to petunia AN4. This suggests that SIA1 and SIA2 arose by gene duplication before the last common ancestor of petunia and tomato, followed by a second (later) duplication in the SIA1 lineage.

The MYB domains of SIA1AC and SIA1Aft group together, distant from ScANT1 (Figure 1D), while SIA2Aft and ScAN2 are more related to each other than to SIA2AC, although distances are in all cases very small.

**SIA2 is up-regulated in Aft tomatoes.**

To assess whether expression levels of the regulators SIA2 and SIA1 are significantly different in the anthocyanins accumulating Aft line compared to AC, we performed quantitative transcripts analysis. We analysed tomato fruits from AC and Aft plants at three different stages of development and ripening (mature green, turning red, and red). Anthocyanin quantification showed accumulation of pigments during fruit ripening in Aft plants, while, no significant anthocyanin accumulation was measured in AC tomatoes (Figure 2A), or in the flesh of the fruit from both genotypes analyzed (data not shown), consistent with previous results (Povero et al., 2011).

Figure 2. Anthocyanin production and expression of SIA1 and SIA2 genes in AC and Aft tomatoes. (A) Anthocyanin levels in tomato peel from AC and Aft fruits (data are means of three replicates ± SD). For each genotype, anthocyanin levels were measured in mature green, turning red and red stages of ripening. (B) SIA1 and SIA2 mRNA levels in tomato peel of AC and Aft fruits at mature green, turning red, and red stage. Relative expression levels of SIA1 and SIA2 were measured by quantitative RT-PCR. Data are means of three replicates ± SD.

The transcripts for SIA1 and SIA2 during fruit ripening showed a peak of expression for SIA2 and, to a lesser extent, for SIA1 in Aft fruits at the green stage (Figure 2B) while expression of both genes was negligible in AC peel (Figure 2B) or in the flesh of both genotypes (data not shown), where anthocyanins are not produced. We could not detect significant amounts of SIA1-like transcripts in any of the samples analyzed (data not shown). This suggests that SIA1-like could be a non-functional gene copy. As previously reported (Povero et al., 2011), SIA1 and SIA2 expression level declines during the turning to red and red ripening phases (Figure 2B).

Interestingly, the amount of SIA2 mRNA in Aft fruits was much higher than that of SIA1 (Figure 2B), suggesting that the SIA2 gene might play a more important role than SIA1 in producing the Aft phenotype.

**Ectopic expression of SIA2 results in “deep purple” tomatoes**
To assess whether differences in the protein sequence of SIANT1 and SIAN2 from AC and the Aft lines could be the molecular basis for the Aft phenotype, we prepared expression constructs in which the entire genomic fragments containing the coding sequence (from ATG to Stop codon including introns) of SIAN2AC, SIAN2Aft, SIANT1AC and SIANT1Aft were fused to the constitutive Cauliflower Mosaic Virus 35S promoter. All four these constructs were transformed in AC plants via Agrobacterium-mediated transformation.

We obtained 12 independent transgenics for the 35S:SIAN2AC construct, 14 transgenics for 35S:SIAN2Aft, 11 for the 35S:SIANT1AC and 13 for the 35S:SIANT1Aft constructs. In all the transgenic lines anthocyanin synthesis was enhanced, resulting in strong pigmentation in vegetative organs (stems, leaves, roots; Figures 3A-B and 5A) and flowers (Figures 3C and 5A).

In fruits, intense and homogeneous pigmentation was observed both in the peel and in the flesh. Purple anthocyanins were easily visible during the “green” stage of development (Figures 4A-B and 5A). At later stages, the increase in carotenoids gave rise to a “red/black” colour either in the skin or in the flesh of the fruit (Figure 4B). Furthermore, synthesis of anthocyanins was also enhanced in the embryos within the seeds of these transgenic tomatoes (Figure 4B).

In conclusion, SIAN2 ectopic expression in tomato results in “deep purple” anthocyanin pigmentation in vegetative and reproductive tissues as well as at different developmental stages. Both 35S:SIAN2AC and 35S:SIAN2Aft induced a significant overproduction of anthocyanins in transgenic tomato plants. This suggests that both proteins encoded by these two genes are equally functional despite their polymorphisms in amino acid sequences and in the sequence of the second intron (Figure 1B-C, Supplemental Data IIA). Therefore, it is unlikely that either the insertion in the second intron of SIAN2AC or the substitutions in the encoded protein suppresses the activity of this allele gene compared to SIAN2Aft.

The molecular and phenotypic analyses of 35S:SIANT1AC and 35S:SIANT1Aft are still underway, but preliminary observations (Figure 5B) indicate that there are no strong differences in the effect of the two alleles in driving anthocyanin accumulation when these are expressed from the CaMV35S promoter. Taken together, these experimental results indicate that the SIAN2 and SIANT1 proteins from the AC and the Aft lines do not exhibit clear differences in activity and that if the different pigmentation patterns of the two genotypes has to be ascribed to one or both these genes, it must originate from differences in their regulatory regions as the encoded proteins are functionally very similar.

**Ectopic expression of SIAN2 is sufficient for ectopic expression of SIANT1 and the HLH gene SIAN1**

It has been shown for several species that activation of anthocyanin structural genes is dependent on the combination of a WDR, a MYB and a bHLH factors (Koes et al., 2005) and that the ectopic expression of both the MYB and bHLH factors is necessary and sufficient for ectopic anthocyanin synthesis (see chapter 2). Hence it is remarkable that the ectopic expression of only the MYB protein SIAN2 or SIANT1 is sufficient to induce
**Figure 3.** Strong anthocyanin pigmentation in vegetative tissues and flowers of plants overexpressing SlAN2AC and SlAN2AP.

(A) Comparison of vegetative tissues from AC and transgenic tomato plants. Transformed leaves and stems show a homogeneous dark colouration, and even roots are red under normal growth conditions.

(B) Pictures of the upper and abaxial sides of wild type and transgenic leaves. Purple anthocyanins are visible on both sides of the leaves expressing 35S:SlAN2AC and 35S:SlAN2AP.

(C) Pigmentation of transgenic flowers. Reddish-purple pigmentation in transgenic anthers is shown, whereas normal yellow pigmentation can be observed in transgenic petals (see “overview” and “stage 2” rows). However, at stage 1, some anthocyanin pigmentation can be observed even in petals.

Anthocyanin synthesis in the fruits (Figures 4A-B and 5A; Mathews et al., 2003), while expression of the bHLH anthocyanin activator DELILA is not (Mooney et al., 1995). To assess whether SlAN2 activates the expression of its bHLH and/or WD40 partner, we identified the tomato homologs of the bHLH genes AN1 and JAF13 and the WD40 gene AN11 of petunia, and analysed their expression in the transgenic 35S:SlAN2AC and 35S:SlAN2AP tomato lines.
Figure 4. Tomatoes transformed with the constructs 3SS:SlAN2AC or 3SS:SlAN2Aft, and analysis of the expression of different genes in these transgenic plants.

(A) “Deep purple” phenotype of 3SS:SlAN2AC and 3SS:SlAN2Aft tomatoes harvested at the green stage of ripening (plants were grown at normal greenhouse conditions).

(B) Phenotype of 3SS:SlAN2AC and 3SS:SlAN2Aft tomatoes harvested at the green and red stages of ripening. Purple/black anthocyanins are visible, even in the flesh of fruits and in transgenic embryos.

(C) Analysis of expression of genes encoding MYB (SlAN2 and SlANT1), bHLH (SlAN1 and SlJAF13), and WD40 (SlAN11) anthocyanin regulators, together with DFR, a late anthocyanin biosynthetic gene. Relative expression levels, measured by quantitative RT-PCR in leaves, petals, anthers, fruit peel, and fruit flesh are shown. Data are means of three replicates ± SD.
Figure 5. Preliminary phenotypic analyses of distinct tomato transgenic lines. 
(A) Pictures of AC plants transformed with 35S:SIA NT1^AC and 35S:SIA NT1^AN. Both the transgenes can activate anthocyanin biosynthesis in different organs, including roots, fruits (where purple “anthocyanin spots” can be observed), anthers and leaves. 
(B) Vegetative tissues of atv plants transformed with AN2^AN transgene showing typical intense pigmentation, as described for the atv genotype.

As expected SIA NT2 mRNA is expressed in all parts of these transgenic plants (Figure 4C) confirming that the transgene is (constitutively) expressed, consistent with the observed increased pigmentation (Figure 4A-B). However, the relative level of SIA NT2 expression in independent 35S:SIA NT2^AC and 35S:SIA NT2^AN transformants displayed some quantitative differences, presumably due to the different insertion sites of the transgenes. Related variation in expression is observed for DFR, which is however highly expressed in all transgenic samples, as compared to the same tissues in control AC plants lacking the transgene.

Interestingly, also the mRNA levels of the MYB gene SIA NT1 and the bHLH gene SIA N1 are increased by the ectopic expression of SIA NT2. This would suggest a role of “master gene” for SIA NT2, as activator of other anthocyanin regulators (Figure 4C), which might have interesting applications as it indicates that ectopic expression of a (certain) MYB alone can be sufficient to drive accumulation of anthocyanins in all parts of the plant.

Finally, SJA F13 and SIA N11 were expressed in all tissues of untransformed AC plants that we analysed and their expression was not clearly enhanced by ectopic expression of SIA NT2. The wide expression pattern of SIA N11 in tomato parallels that of AN11 in petunia (devVetten et al., 1997) indicating that also in tomato it is not the limiting factor that determines the pattern of pigmentation.
Figure 6. Phenotype and expression analyses of plants transformed with the construct AN2Aft transgene. (A) "Anthocyanin fruit-like" (Aft-like) phenotype of AN2Aft transgene tomato fruits harvested at the green and turning red stage of ripening, compared to untransformed AC control and Aft tomatoes. (B) Expression analysis of genes encoding selected MYB (SIAN2 and SIAN1), bHLH (SIAN1 and SIJAF13), WD40 (SIAN11), all putative regulators of anthocyanin biosynthesis in tomato, and DFR, a late anthocyanin biosynthetic gene. Relative expression levels were measured by quantitative RT-PCR in leaves and fruit peel. Data are means of three replicates ± SD.
Transfer of a genomic fragment containing SIANZ\textsuperscript{At} gene restores the “Antocyanin fruit” phenotype in AC tomatoes

As we did not find functional differences in the proteins encoded by SIANZ\textsuperscript{AC} and SIANZ\textsuperscript{At} genes that could account for the enhanced expression of SIANZ\textsuperscript{At}, we considered the possibility that this results from differences in promoter activity. Such a promoter activity might stem from cis-regulatory changes in the SIANZ promoter, or from increased activity of an upstream trans-acting factor. To gain more insight on this, we carried out a functional complementation experiment. A genomic fragment containing the entire SIANZ\textsuperscript{At} gene including ~2.4 kb sequence upstream the ATG was cloned in a T-DNA vector then introduced into Alisa Graig via Agrobacterium.

We obtained 12 transgenic AC plants containing the ANZ\textsuperscript{At} transgene and 5 of them displayed an “Aft-like phenotype”, with purple spots visible in the peel of the fruit, especially at green stage (Figure 6A), and on the stem (Figure 6B).

To confirm that this phenotype is due to the expression of the SIANZ\textsuperscript{At} allele in the AC background, to dissect the effects of the transgene, and to confirm activation of other transcription factors of the MYB/bHLH/WD40 complex, we analysed the expression of SIAN2, SIAN1, DFR, and the putative bHLH and WD40 regulators SIAN1, SIAF13, and SIAN11. We assessed the expression levels of these genes in leaves and fruit peel of the transgenic plants and control plants lacking the transgene.

These results show that SIANZ\textsuperscript{At} plays a major role in specifying the Aft phenotype, although do not exclude a contribution for other factors in the genome of the Aft line. The study of the effect of the SIANZ\textsuperscript{At} transgene shows that the “Aft-like” phenotype conferred is associated with differences in the strength and/or pattern of SIANZ\textsuperscript{At} expression, as compared to SIANZ\textsuperscript{AC}, which are at least in part due to differences in cis-regulatory elements of the SIANZ\textsuperscript{At} and SIANZ\textsuperscript{AC} alleles. Hence, the distinct pigmentation patterns of the Aft and AC lines are at least in part due differences in cis-regulatory elements of the SIANZ\textsuperscript{At} and SIANZ\textsuperscript{AC} alleles.

Despite numerous attempts, we were not able to produce a similar construct containing the SIANZ\textsuperscript{AC} allele, to use as a control, because we could not amplify the complete 5’ flanking region of this gene. We could, however, amplify and sequence overlapping small fragments of the promoter from the SIANZ\textsuperscript{AC} allele (Figure 7). A comparison of the sequences obtained for the promoters ANZ\textsuperscript{AC} and ANZ\textsuperscript{At} revealed many sequence polymorphisms (Supplemental Data IVA), similar to the number of polymorphisms in the SIAN1 promoters from these tomato lines (Supplemental Data IVB). The ANZ\textsuperscript{AC} promoter sequence that we inferred in this way was identical to that in the (draft) genome sequence of the cultivar Heinze. However, these sequences do not suggest an explanation as to why the 2.5 kb SIANZ 5’ upstream sequence cannot be amplified in once piece from either Heinz or AC.

Hence, it is possible that these genomic regions contain some structure(s), like direct or inverted repeats, that block amplification of the entire promoter in one amplicon, and the same time cause miss-assembly of partial sequences.

Does the introduction of ANZ\textsuperscript{At} transgene in atv plants results in the “SunBlack” phenotype?
The recessive tomato *atv* phenotype consists in very intense anthocyanin pigmentation in the vegetative plant parts as consequence of exposure to light. Because *atv* *Aft* lines display a synergistic effect on the anthocyanin pathway resulting in the so called “SunBlack” phenotype (strongly pigmented fruits), we introduced the *AN2Atv* allele into *atv* tomato plants by *Agrobacterium*-mediated transformation, to test whether this would result in SunBlack tomatoes. The plants are at the moment still in the vegetative phase and they show the typical intense pigmentation of the stem described for the *atv* genotype (Figure 5B). It is necessary to wait for the production of fruit to see if these will be “SunBlack-like”.

Discussion

Anthocyanins and other flavonoids have been shown to have a health-promoting effect on animals and humans when present in the diet in relatively high amounts. Unfortunately, most of the fruit and vegetables in our daily diet are rather poor in these compounds, and this makes it interesting to unravel the genetic mechanisms at the basis of high anthocyanin accumulation in specific varieties of crops.

Anthocyanin pigments are normally absent from the fruits of cultivated tomatoes, where the red colour at ripening stage is due to the presence of carotenoids. Only some accessions, such as “Anthocyanin fruit” (named *Aft* in this study), can accumulate anthocyanins in the skin of the fruit. Here we show that the *Aft* phenotype is, at least in part, due to alteration in *SIAN2* and that this gene encodes a MYB protein that is sufficient to activate anthocyanin synthesis in a variety of tissues.

It was already known that both *SIANT1* and *SIAN2* map on chromosome 10 (De Jong et al., 2004; Boches and Myers, 2007; Sapir et al., 2008) and that polymorphisms in these two genes cosegregate with the *Aft* locus. We showed that both genes are part of a ~10 kb genomic region together with a third gene, *SIANT1-like*, that encodes a MYB protein with high similarity to *SIANT1*, but which seems not to be expressed (at least in the tissues we analyzed). Most-likely these highly similar and contiguous genes originated from duplication events. *SIANT1* and *SIAN2* belong to separate phylogenetic clades, *SIANT1* being more closely related to the petunia regulator of anthocyanin accumulation in anthers *PhAN4* (Povero et al., in preparation; chapter 6) whereas *SIAN2* is more similar to *PhAN2*, involved in anthocyanin production in petunia petals. This suggests that the duplication that gave rise to these genes occurred before the tomato and petunia lineages diverged.

Previous studies had shown that anthocyanin or flavonoid accumulation in different plant parts in tomato is induced by ectopic expression of genes encoding specific endogenous or exogenous MYB transcription factors, such as *SIANT1* (Mathews et al., 2003) and *AtMYB75/PAP1* from *Arabidopsis thaliana* (Zuluaga et al., 2008). However, the effect in the fruit was generally limited to weak/spotted colouration. Only the simultaneous overexpression of the MYB and bHLH anthocyanin regulators encoded by ROSEA and DELILA from Anthirrinum resulted in a sustained anthocyanin production in tomato fruits (Butelli et al., 2008), while expression of DELILA alone did only enhance pigmentation in some vegetative tissues but not in the fruit (Mooney et al., 1995).

These findings are in line with results in a variety of other species showing that the ectopic activation of structural anthocyanin genes requires the co-expression of both the MYB and bHLH factor. Ectopic expression of the WD40 protein from a transgene is not necessary as the encoding genes in petunia (*AN11*) and *Arabidopsis* (*TTG1*) are expressed in virtually all tissues.

In petunia (*Solanaceae* like tomato), the MYB factor (*PhAN2* in petals and *PhAN4* in anthers) interacts with two different bHLH transcription factors encoded by *PhAN1* and *PhJAF13* to induce transcription of anthocyanin genes. In petunia the constitutive expression either AN2 or AN4, is
sufficient for ectopic activation of the anthocyanin pathway, because AN2 and AN4 activate transcription of the bHLH gene AN1 in any plant tissue (Spelt et al., 2000; chapter 6).

In tomato, a similar mechanism of regulation can be envisaged as we can find all the orthologs of the mentioned petunia genes in the tomato genome. SIAN1 and SIAN2 are homologous to PhAN4 and PhAN2, and we also identified tomato orthologs of PhAN1, PhJAF13, and PhAN11. In a similar way the tomato bHLH gene SIAN1 can be ectopically activated by constitutive expression of the SIAN1 and SIAN2 alleles from either AC or the Aft line. This, together with the capability of SIAN2 to induce also the other tomato MYB gene SIAN1, suggests that SIAN2 acts as a master regulator of the pathway and can explain why constitutive expression of the MYB factor alone is sufficient for ectopic activation of the anthocyanin pathway making it a good candidate for the Aft locus. Moreover, this finding has important implications for the production of anthocyanin rich tomato fruits, as the up-regulation of this single gene is sufficient to reach the goal.

In AC plants expressing the ANZ^{transgene}, the fruits show a partial complementation, of the Aft phenotype. An accurate analysis of the phenotype reveals that in the Aft line, anthocyanin spots were more visible and widespread than in the transgenic phenocopy. Two factors, which are not mutually exclusive, can explain why the transgenic plants do not completely phenocopy the Aft line. First, the genomic fragment used for the complementation may lack some regulatory elements required for the full complementation of the Aft phenotype. Second, since the genetic background of the Aft line is unknown, and isogenic control lines for Aft mutation are not available among the S. lycopersicum varieties, we chose the cv. AC as host for our transformation studies. Hence unknown factors in the AC and Aft genetic background might be responsible for the different phenotypes. For example the Aft line is known to have a high-photoperiod response, which might have an enhancing effect on the level of anthocyanins. As AC has normal photoreponsiveness, this might explain why introduction of the SIAN2^{transgene} has a weaker phenotype.

The poor definition of the background of the line harbouring the Aft mutation, is a consequence of the unclear origin of the Aft locus itself. Although a number of papers suggest that the Aft locus was introgressed from S. chilense, a conclusive documentation of the origin of the Aft line is missing.

The analysis of the polymorphisms in the SIAN1 and SIAN2 genes, which co-segregate with the Aft locus, do not support the hypothesis of its origin from the S. chilense genome. The second intron of the SIAN2 gene, for instance, contains numerous polymorphisms, including the insertion of a putative ancient transposon in the S. chilense allele that is neither present in the AC allele (which contains instead a very different insertion at another position) nor in the Aft allele. The only document which reports the interspecific cross of S.lycopersicum and S. chilense is a “Report of the tomato genetic cooperative” (Georgiev, 1972), while there are no records about the breeding program that has followed. Therefore the origin of the chromosomal fragment containing these two genes (and most probably the Aft locus) remains unclear.

The analyses of the transgenic lines harbouring the ANT1^{transgene} as well as those transformed with the combination of ANT1^{transgene} and ANZ^{transgene} will give a final answer about the role of these two genes in the Aft phenotype. However, the fact that the expression of SIAN2 is sufficient to induce the expression of SIAN1 also in the AC background, already indicates that no specific differences in the promoter of SIAN1 between the AC and the Aft alleles can be responsible for the differences in expression level, and rather points towards the high expression of the SIAN2^{transgene} allele as responsible for a cascade of gene expression leading to anthocyanin accumulation in tissues where this is generally not occurring (e.g. in fruits).
The data we will get from the atv transgenics harbouring the AN2transgene construct, will not only help to further understand the effect of SlAN2 in the “SunBlack” tomato phenotype, will possibly also open the way to unravel the mechanism of action of the atv locus.

Material and methods

Plant material

Ailsa Craig (AC; accession number LA2838A), Aft (accession number LA1996), and S. chilenense (accession number LA1969) seeds were provided by the Tomato Genetic Resource Center (TGRC, University of California, Davis). Since near isogenic lines for Aft were not available among the S. lycopersicum varieties, we chose the cv. AC as a control tomato line for our analyses. We also considered other parameters, such as the morphology of the plant and the fruit, the size of mature tomatoes, and their ripening time (time from anthesis to different ripening stages). All of these characteristics are quite similar between AC and Aft, creating conditions to perform valid comparisons.

Plants were grown under normal greenhouse conditions (16 hours light, 8 hours dark, maximum temperature 30°C, minimum 20°C). Tomato plants from the different accessions were grown side by side. Fruits were collected at mature green (corresponding to 40 DPA for AC and 44 DPA for Aft), turning red (49 DPA for AC and 54 DPA for Aft), and red (60 DPA for AC and 64 DPA for Aft) stages of development and ripening. These stages correspond to mature green, breaker and red ripe, respectively, in the classification of (Giovannoni, 2004). At each of the three stages of development, a single fruit was harvested and equal peel samples were removed with a scalpel from each fruit. A total of three harvests taken from each genotype were analyzed, each harvest representing a biological replicate.

Sources of the sequences utilized for the phylogenetic tree

For the construction of the phylogenetic tree we used the MYB domain portion of the following sequences taken from these sources: SlANT1AC GenBank accession number AAQ55181.1 (Mathews et al., 2003). SlANT14ST GenBank accession number ABO26065.1 (Sapir et al., 2008). ScANT1 (this study). SlANT1-like (this study). StAN1 GenBank accession number AAX53089.1 (Jung et al., 2009). PhAN4 GenBank accession number ADQ00392.1 (chapter 6). StAN2 GenBank accession number AAX53091.1 (Jung et al., 2009). SlAN2AC (this study), SlAN24ST (this study), ScAN2 (this study). CaAN2 GenBank accession number CAE75745.1 (Borovsky et al., 2004). PhAN2 GenBank accession number ABO21074.1 (Quattrocchio et al., 1999). NtAN2 GenBank accession number ACO52470.1 (Pattanaik et al., 2010). VvMYBA1 GenBank accession number XP_002265406.1 (predicted by automated computational analysis). VvMYBA2 GenBank accession number BAD18978.1 (Kobayashi et al., 2004). VvMYBA3 GenBank accession number BAD18979.1 (Kobayashi et al., 2004). AmVNOSE1 GenBank accession number ABB83828.1 (Schwinn et al., 2006). AmROSEA1 GenBank accession number ABB83826.1 (Schwinn et al., 2006). AmROSEA2 GenBank accession number ABB83827.1 (Schwinn et al., 2006). AtPAP1 GenBank accession number AAG42001.1 (Borevitz et al., 2000). AtPAP2 GenBank accession number AAG42002.1 (Borevitz et al., 2000). PhPH4 GenBank accession number AAY51377.1 (Quattrocchio et al., 2006). AtMYB5 GenBank accession number AEE75369.1 (Salanoubat et al., 2000). PhOdorant1 GenBank accession number AAV98200.1 (Verdonk et al., 2005). PhMYB1 GenBank accession number CAA73836.1 (Avila et al., 1993). AmMIXTA GenBank accession number CAA55725.1 (Noda et al., 1994).
Genomic DNA (gDNA) amplification

Genomic DNA from tomato tissues was isolated according to the procedure followed by de Vetten et al. (1997). PCR reactions were performed using the “Phusion® High-Fidelity DNA Polymerase” (Finnzymes, Ratastie 2, 01620 Vantaa, Finland). Amplification of SlANT1 and SlAN2 genomic DNA and cDNA fragments for sequencing and for sequence length polymorphism detection was performed using the following pairs of primers:

**SlANT1:**
- SlANT1 _FW_: CACCATGAACGTACATCTATGTC
- SlANT1 _RV_: TTAAGTAGATCCATAAGTCATTTC

**SlAN2:**
- SlAN2 _FW_: CACCATGAATTCTCATGTGTC
- SlAN2 _RV_: TTAAGTAGATCCATAAGTCATTTC

Anthocyanin quantification

Anthocyanin extraction from the skin of tomato fruits was performed as described by (Solfanelli et al., 2006). The amount of anthocyanins was determined spectrophotometrically (A535) and expressed as mg of petunidin-3-(p-coumaroyl rutinoside)-5-glucoside per g, based on an extinction coefficient of 17,000 and a molecular weight of 934 (Butelli et al., 2008). Mean values were obtained from three independent replicates.

Expression analysis by quantitative RT-PCR

Total RNA was extracted from tomato samples using a “Spectrum™ Plant Total RNA Kit” (Sigma-Aldrich, St Louis, MO, USA) according to the manufacturer’s instructions. We treated the RNA with DNase using a “TURBO DNA free Kit” (Ambion, Austin, TX, USA). An “iScript cDNA Synthesis Kit” (Bio-Rad Laboratories, Hercules, CA, USA) was then used to reverse-transcribe one microgram of each sample into cDNA. Quantitative RT-PCR amplification (qPCR) was performed using an ABI Prism® 7300 Sequence Detection System (Applied Biosystems, Foster City, CA, USA). The primers used for the amplification of the reference, the regulatory and the structural genes analyzed, with relative GenBank accession numbers, are listed in Table S1. LeEF1A (Lycopersicon esculentum elongation factor 1-alpha, GenBank accession number X14449) and 18SrRNA (Sapir et al., 2008) were both used as reference genes. qPCR reactions were carried out using a Power SYBR Green PCR Master Mix (Applied Biosystems), 10 ng of cDNA template, and 4.5 pmol gene-specific primers in a final reaction volume of 15 µl. The relative quantitation of each individual gene expression was performed using the geometric averaging method (geNorm) (Vandesompele et al., 2002).

Cloning procedure, plasmid construction, and tomato transformation

SlAN2 and SlANT1 were amplified using the “Phusion® High-Fidelity DNA Polymerase” (Finnzymes, Ratastie 2, 01620 Vantaa, Finland), from total genomic DNA extracted from AC (to prepare the construct containing SlAN2<sup>AC</sup> and SlANT1<sup>AC</sup>) or Aft (to prepare the construct containing SlAN2<sup>Aft</sup> and SlANT1<sup>Aft</sup>). The primers used are the one described in the paragraph “Genomic DNA (gDNA) amplification”. The fragments were cloned into the Gateway vector pK7WG2 and the resulting plasmids were transformed into the Agrobacterium tumefaciens GV3101 strain.

To obtain the AN2<sup>Aft</sup> transgene and ANT1<sup>Aft</sup> transgene constructs we amplified the genomic fragments by “Phusion® High-Fidelity DNA Polymerase”, using DNA extracted from Aft tissues as
The primers used for the amplification contained attB-1 and -2 recombination sites (underlined):

- AN2\textsuperscript{ attB\_FW\_attB1}: GGGGACAAGTTTGTACAAAAAAGCAGGCTTGTACTCCTTTAGTCT
- AN2\textsuperscript{ attB\_RV\_attB2}: GGGGACCCTTTGTACAAGAAAAGCAGGCTTGTACCTTTAGTCT
- AN\textsuperscript{ attB\_FW\_attB1}: GGGGACAAGTTTGTACAAAAAAGCAGGCTTGTACTCCTTTAGTCT
- AN\textsuperscript{ attB\_RV\_attB2}: GGGGACCCTTTGTACAAGAAAAGCAGGCTTGTACCTTTAGTCT

Entry clones were produced in the Gateway vector pDONR221 P1-P2 and then recombined with vector pDEST pKGWD,0 (Invitrogen, Poststraße 22 D, 32584 Löhne, Germany). For tomato plant transformations, we followed the protocol published by Zuluaga et al. (2008).

**Production of small fragments for the sequencing of S\textit{ANT1} and S\textit{IAN2}**

In order to sequence the promoter and the genomic DNA of \textit{S\textit{ANT1}} and \textit{S\textit{IAN2}} from different accessions (AC, Aft, and \textit{S.chilense}), we carried out Phusion PCR analyses using distinct pairs of primers. We designed several pairs of primers in such a way that we obtained distinct overlapping fragments that were sequenced and assembled into a contig encompassing both genes (Figure 7).

**Figure 7.** Schematic representation of overlapping PCR fragments (black rows) used in order to sequence the promoter and genomic region of \textit{S\textit{ANT1}} and \textit{S\textit{IAN2}} (of different accessions). Each primer is identified with a number and the orientation: Fw=Forward Primer and Rv=Reverse Primer. The sequence of each primer is reported here below:

1Fw: CTCCCTAGCAGCAGCTAGCTAC and 1Rv: AGCCCTTTACGATCCAATTTGT
2Fw: TGGCGCAAGAAGAAGAATAA and 2Rv: CAGCAATGAAATAATCATCCTCAGCA
3Fw: GACAACTATATATCAAGGATGAGTCA and 3Rv: GTTCAATAATGGATGATATATA
4Fw: GATATAGAGTGAATACATAAGGAACA and 4Rv: GTTCAATAATGGATGATATATA
5Fw: TATGGAACGAGAGAAATATGCT and 5Rv: TTGGAAGAGAGGGAATGAAATAAAGAG
6Fw: TGGCGCAAGAAGAAGAATAA and 6Rv: TGGTGCAAAATAGTGTGCTC
7Fw: TATGGAACGAGAGAAATATGCT and 7Rv: GATGGAAGAGAGGGAATGAAATAAAGAG
8Fw: GCTTCTTTTGGTTCAGCAATA and 8Rv: GATGGAAGAGAGGGAATGAAATAAAGAG
9Fw: TGGCGCAAGAAGAAGAATAA and 9Rv: TGGTGCAAAATAGTGTGCTC
10Fw: AAGGGTTGGAAAGAAATAGG and 10Rv: AGTCTCAAGATGAGATCAAGATCCTCA
11Fw: AAGGGTTGGAAAGAAATAGG and 11Rv: TGGTGCAAAATAGTGTGCTC
11Rv: TGGTGCAAAATAGTGTGCTC

where: 1Fw, 2Fw, 3Fw, 4Fw, 5Fw, 6Fw, 7Fw, 8Fw, 9Fw, 10Fw, 11Fw, 11Rv_b: TGGTGCAAAATAGTGTGCTC.
### Supplemental Table 1. Primers used for quantitative RT-PCR analyses.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Gene Description</th>
<th>Sequence ID</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>TaqMan probe</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EF1a</strong></td>
<td>Elongation factor-1-alpha</td>
<td>X14449 (GenBank)</td>
<td>TGCTTGCTTTTGC CCTTGGT</td>
<td>CGATTTCATCAT CCTAGGCTGGA</td>
<td>CTGCTGTA ACAAGAT GGATGC</td>
</tr>
<tr>
<td><strong>18S rRNA</strong></td>
<td>18S ribosomal RNA</td>
<td>XS1576 (GenBank)</td>
<td>GCGACGCATCAT TCAAATTTC</td>
<td>TCCGGAATCGAA CCCTAATTTC</td>
<td>CAACAGAT GGTCACTT ATTG</td>
</tr>
<tr>
<td><strong>SIANT1</strong></td>
<td>Anthocyanin 1</td>
<td>EF433416 (GenBank)</td>
<td>AAGTGGATCTCA TTTTGAGGCTTCA</td>
<td>TCCITCCC GGGA AATCCCA</td>
<td></td>
</tr>
<tr>
<td><strong>SIAN2</strong></td>
<td>Anthocyanin 2</td>
<td>FJ744761 (GenBank)</td>
<td>TTCCAGGAAGGA CAGCAAAC</td>
<td>AAGGCAGAGGA GAATGAGGA</td>
<td></td>
</tr>
<tr>
<td><strong>SIDFR</strong></td>
<td>Dihydroflavonol reductase</td>
<td>Z18277 (GenBank)</td>
<td>CAGGCAAGGG GACCTTCATTT G</td>
<td>GCACACATCTAG CCACATCGTA</td>
<td></td>
</tr>
<tr>
<td><strong>SIAN1</strong></td>
<td>Putative S. lycopersicum Anthocyanin 1</td>
<td>SL2.40ch09 (SolGenes)</td>
<td>CCTCTCTTGAC CCGGTTG</td>
<td>GCCTGTTGTCG ATCTTGG</td>
<td></td>
</tr>
<tr>
<td><strong>SIAN1</strong></td>
<td>Putative S. lycopersicum Anthocyanin 11</td>
<td>SGN-E235386 (SolGenes)</td>
<td>TCGGGGATCAC ATCCCAAGA</td>
<td>TCCATCAAGGT TGGAAGAC</td>
<td></td>
</tr>
<tr>
<td><strong>SUAF13</strong></td>
<td>Putative S. lycopersicum Johnandfrancesca 1 3</td>
<td>SGN-U215976 (SolGenes)</td>
<td>ATGAAGTGGAAC CGAAGAGA</td>
<td>TCCATCAAGCA AACAGAC</td>
<td></td>
</tr>
</tbody>
</table>
A. Clustal alignment of SIAN1-like and SIAN1 genomic sequences (spanning from the ATG to the Stop codon). Exonic regions are underlined, polymorphisms are highlighted in grey and white.
B. Clustal alignment of *SIANT1-like* and *SIAN2* genomic sequences (spanning from the ATG to the Stop codon). Exonic regions are underlined, polymorphisms are highlighted in grey and white.
A. Clustal alignment of *SIA2* genomic sequences (spanning from the ATG to the Stop codon) from AC and Aft (exonic regions are underlined, polymorphisms are highlighted in grey and white).
B. Clustal alignment of SIAN2 genomic sequences (spanning from the ATG to the Stop codon) from Aft and S. chilense (exonic regions are underlined, polymorphisms are highlighted in grey and white). "Target Site Duplication (TSD)" and "Terminal Inverted Repeats (TIR)", typical of a transposon, and present in the insertion in ScAN2, are marked/labelled.
C. Clustal alignment of SIAN2 aminoacid sequences from AC, Aft and S. chilense (polymorphisms are highlighted in grey and white).
D. Clustal alignment of SIANTI genomic sequences (spanning from the ATG to the Stop codon) from AC and Aft (exonic regions are underlined, polymorphisms are highlighted in grey and white).
E. Clustal alignment of SIAN1 genomic sequences (spanning from the ATG to the Stop codon) from Aft and S. chilense (exonic regions are underlined, polymorphisms are highlighted in grey and white).
F. Clustal alignment of SIAN1 aminoacid sequences from AC, Aft and S. chilense (polymorphisms are highlighted in grey and white).
SUPPLEMENTAL DATA III.

Clustal alignment of MYB domains of distinct R2R3-MYB transcription factors involved in anthocyanin biosynthesis in different species. Polymorphisms are highlighted in grey and white.
<table>
<thead>
<tr>
<th>Alignment of Sian2 promoter sequences from AC and Aft (polymorphisms in grey and white)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ProSian2 1192 AGG TAC ACA GAC CCA GCC AT TGA CTT AT TGA GAT CTC CAG TGT GAT A</td>
</tr>
<tr>
<td>ProSian2 1197 TAC AAG TAC GT CAG TGA GGT GGT GT AAG TAA TAA CGA GAG CAC AG</td>
</tr>
<tr>
<td>ProSian2 1252 TAC AAA TAC GT CAG TGA GGT GGT GT AAG TAA TAA CGA GAG CAC AG</td>
</tr>
<tr>
<td>ProSian2 1289 TAC AAA TAC GT CAG TGA GGT GGT GT AAG TAA TAA CGA GAG CAC AG</td>
</tr>
<tr>
<td>ProSian2 1326 TAC AAA TAC GT CAG TGA GGT GGT GT AAG TAA TAA CGA GAG CAC AG</td>
</tr>
<tr>
<td>ProSian2 1373 TAC AAA TAC GT CAG TGA GGT GGT GT AAG TAA TAA CGA GAG CAC AG</td>
</tr>
<tr>
<td>ProSian2 1430 TAC AAA TAC GT CAG TGA GGT GGT GT AAG TAA TAA CGA GAG CAC AG</td>
</tr>
<tr>
<td>ProSian2 1485 TAC AAA TAC GT CAG TGA GGT GGT GT AAG TAA TAA CGA GAG CAC AG</td>
</tr>
<tr>
<td>ProSian2 1542 TAC AAA TAC GT CAG TGA GGT GGT GT AAG TAA TAA CGA GAG CAC AG</td>
</tr>
<tr>
<td>ProSian2 1599 TAC AAA TAC GT CAG TGA GGT GGT GT AAG TAA TAA CGA GAG CAC AG</td>
</tr>
<tr>
<td>ProSian2 1656 TAC AAA TAC GT CAG TGA GGT GGT GT AAG TAA TAA CGA GAG CAC AG</td>
</tr>
<tr>
<td>ProSian2 1713 TAC AAA TAC GT CAG TGA GGT GGT GT AAG TAA TAA CGA GAG CAC AG</td>
</tr>
<tr>
<td>ProSian2 1770 TAC AAA TAC GT CAG TGA GGT GGT GT AAG TAA TAA CGA GAG CAC AG</td>
</tr>
<tr>
<td>ProSian2 1827 TAC AAA TAC GT CAG TGA GGT GGT GT AAG TAA TAA CGA GAG CAC AG</td>
</tr>
<tr>
<td>ProSian2 1884 TAC AAA TAC GT CAG TGA GGT GGT GT AAG TAA TAA CGA GAG CAC AG</td>
</tr>
<tr>
<td>ProSian2 1941 TAC AAA TAC GT CAG TGA GGT GGT GT AAG TAA TAA CGA GAG CAC AG</td>
</tr>
<tr>
<td>ProSian2 2008 TAC AAA TAC GT CAG TGA GGT GGT GT AAG TAA TAA CGA GAG CAC AG</td>
</tr>
<tr>
<td>ProSian2 2065 TAC AAA TAC GT CAG TGA GGT GGT GT AAG TAA TAA CGA GAG CAC AG</td>
</tr>
<tr>
<td>ProSian2 2122 TAC AAA TAC GT CAG TGA GGT GGT GT AAG TAA TAA CGA GAG CAC AG</td>
</tr>
<tr>
<td>ProSian2 2180 TAC AAA TAC GT CAG TGA GGT GGT GT AAG TAA TAA CGA GAG CAC AG</td>
</tr>
<tr>
<td>ProSian2 2237 TAC AAA TAC GT CAG TGA GGT GGT GT AAG TAA TAA CGA GAG CAC AG</td>
</tr>
<tr>
<td>ProSian2 2294 TAC AAA TAC GT CAG TGA GGT GGT GT AAG TAA TAA CGA GAG CAC AG</td>
</tr>
</tbody>
</table>

A.
B. Alignment of \textit{SANT1} promoter sequences from AC and Aft (polymorphisms in grey and white).
References


