University of Naples Federico II

Department of Agricultural Sciences
Division of Plant Genetics and Biotechnology
Ph.D. in Agrobiology and Agrochemistry
XXVIII Cycle

Investigating molecular mechanisms controlling phenylpropanoid production in potato and tomato

Vincenzo D'Amelia

Tutor:

Prof. Domenico Carputo

INDEX

ABSTRACT 4

CHAPTER 1. GENERAL INTRODUCTION 7

1.1 The importance of plant secondary metabolites 7

1.2 Flavonoids as antioxidants: role against environmental stresses 10

1.3 Flavonoids as antioxidants: role for human nutrition and food technology 14

1.4 Plant breeding, genomic and biotechnological approaches to study and use flavonoids 17

1.5 Aims of the research 20

1.6 References 23

CHAPTER 2. HIGH *AN1* VARIABILITY AND INTERACTION WITH BASIC HELIX-LOOP-HELIX CO-FACTORS RELATED TO ANTHOCYANIN BIOSYNTHESIS IN POTATO LEAVES 32

2.1 Introduction 32

2.2 Materials and methods 35
2.2.1 Plant materials and growth conditions 35
2.2.2 Total anthocyanin analysis 35
2.2.3 Nucleic acid extraction and molecular analysis of *AN1* 35
2.2.4 Cloning of *AN1* and basic helix-loop-helix (bHLH) genes from potato leaves 36
2.2.5 Gene expression analyses 37
2.2.6 Yeast two-hybrid assay 37
2.2.7 Bimolecular Fluorescent Complementation (BiFC) assay 38
2.2.8 Overexpression of potato transcription factors in tobacco plants 38
2.2.9 Bioinformatics and statistical analysis 38

2.3 Results 40
2.3.1 High *AN1* gene nucleotide variability and amino acid polymorphic sites 40
2.3.2 Isolation of *StJAF13* coding sequence 45
2.3.3 High *AN1* and *StJAF13* expression in red leaf genotypes 47

2.3.4 AN1 interacts with StbHLH1 and StJAF13 49
2.3.5 Co-expression of *AN1*, *StbHLH1* and *StJAF13* genes enhances anthocyanin accumulation in tobacco plants 52

2.4 Discussion **54**

2.5 Conclusions **58**

2.6 References **59**

2.7 Supplementary data Chapter 2 **65**

CHAPTER 3. AN ANCIENT MYB GENE DUPLICATION GENERATED A COLD-INDUCED REGULATOR OF ANTHOCYANIN BIOSYNTHESIS IN *SOLANUM COMMERSONII* **74**

3.1 Introduction **74**

3.2 Materials and methods **77**
3.2.1 Plant materials and growth conditions 77
3.2.2 Freezing-tolerance assay 77
3.2.3 Cold stress 77
3.2.4 RNA extraction and anthocyanin genes expression analyses 78
3.2.5 Metabolite analysis 78
3.2.6 Microscopy analysis 79
3.2.7 Yeast Two Hybrid analysis 80
3.2.8 Overexpression of potato transcription factors in tobacco plants 80
3.2.9 Bioinformatics and statistical analysis 81

3.3 Results **82**
3.3.1 Chilling temperatures increase the anthocyanin content in *S. commersonii* leaves 82
3.3.2 *ScAN2* is cold stress induced only in *S. commersonii* 84
3.3.3 *ScAN2* and *ScAN1* are paralog genes 86
3.3.4 Sc/StAN1 and Sc/StAN2 share interaction with the same bHLH co-partners 89
3.3.5 *ScAN2* is an anthocyanin inducer with a phenotypic pleiotropic effect 90
3.3.6 *ScAN2* induces accumulation of other phenolic compounds in the cell wall of transformed cells 94

3.4 Discussion **97**
3.4.1 Duplication and divergence expression: specific *ScAN2* up-regulation after cold stress 97
3.4.2 Different anthocyanin localization in *S. commersonii* and *S. tuberosum* leaves 99

3.4.3 Divergent binding ability reflects divergences in function 100
3.4.5 A possible divergence function of *ScAN2* connected to cold stress tolerance of *S. commersonii* 101

3.5 Conclusions 103

3.6 References 104

3.7 Supplementary data Chapter 3 110

CHAPTER 4. *SLMYB4* AND *SLELVIRA* NEGATIVELY INFLUENCE RUTIN AND CHLOROGENIC ACID CONTENT IN TOMATO FLESH 117

4.1 Introduction 117

4.2 Materials and methods 119
4.2.1 Plant Materials 119
4.2.2.Flavonol and caffeoyl quinic differentially expressed genes (DEGs) analysis 119
4.2.3 RNA and cDNA synthesis, cloning and sequence analysis of candidate genes 119
4.2.4 Virus induced gene silencing (VIGS) and transient over expression 120
4.2.5 HPLC analysis 120
4.2.6 Bioinformatics and statistical analysis 121

4.3 Results 122
4.3.1 Flavonol and caffeoyl quinic acid eQTLs hotspot characterization 122
4.3.2 *SlMYB4* is a tomato ortholog of Arabidopsis repressor *AtMYB4* 126
4.3.3 *SlELVIRA* mapping and characterization 128
4.3.4 Functional analysis of *SlMYB4* and *SlELVIRA* using VIGS 131

4.4. Discussion 133
4.4.1 *SlMYB4* is involved in flavonoid regulation in tomato flesh 133
4.4.2 *SlELVIRA,* a new characterized gene influencing flavonol and chlorogenic acid (CGA) accumulation in tomato flesh 134

4.5 Conclusions 136

4.6 References 137

4.7 Supplementary data Chapter 4 141

CHAPTER 5. GENERAL CONCLUSIONS 147

5.1 References 151

Abstract

Plant phenylpropanoids draw an outstanding interest due to their effect in both human nutrition and plant defence. In fact, compounds produced through the phenylpropanoid biochemical pathway represent indispensable elements for plants to face environmental stresses. Moreover, thanks to their strong antioxidant activity, these compounds are important components of functional food. In this work we focused on phenylpropanoid pathway, with a particular interest for the braches of flavonoids. In particular, the major aim of our research was to investigate the genetic mechanism controlling the production of flavonoids in order to clarify either their role in plant stress tolerance or to enhance the production of useful metabolites in plant derived foods. As regard the role of flavonoids in plant environmental tolerance, we focused on the ability of anthocyanin pigments to induce cold stress tolerance in potato. Since no molecular information was available on anthocyanin regulation in potato leaves, the first part of our research focused onto the determination of MYB/bHLH complex that influences anthocyanin production in vegetative tissues of the cultivated potato *Solanum tuberosum*. In particular, we found that *StAN1,* a gene codifying a MYB factor, displays intraspecific sequence variability in both coding/non-coding regions and in the promoter. In addition, expression analysis suggested that leaf pigmentation is associated with *StAN1* expression and that a bHLH, named *StJAF13*, acts as putative *StAN1* co-regulator for anthocyanin gene expression in leaves of red-leaf varieties. Functional analysis through protein/protein interaction and ectopic transgenic expression, further confirmed AN1/StJAF13 interaction complex to induce anthocyanin accumulation. Once characterized this anthocyanin complex, we compared the anthocyanin genes of the wild cold tolerant species *Solanum commersonii* with those of cultivated varieties under cold stress condition. As suggested by functional and metabolic analysis performed *Nicotiana benthamiana*, *ScAN2*, a paralog gene of *ScAN1*, evolved differently between cultivated and wild species. In *S. commersonii*, *ScAN2* seemed to keep a pleiotropic and ancestral function with respect to *ScAN1*,

inducing a multiple activation of several phenylpropanoid branches to respond to cold injury. Consistently, we found that *ScAN2* was up regulated after cold treatment only in the cold tolerant *S. commersonii*. In addition, metabolic and microscopy analyses suggested that *ScAN2* is connected to the production of phenolic compounds located on plasma membrane and cell wall of transformed tobacco cells. The third part of our work was mainly concentrated on the group of flavonols which are flavonoids with a strong anti-inflammatory activity as well as protective role against cardiovascular diseases. In this study, we tried to understand which were the genes that normally contribute to flavonol accumulation in tomato flesh. We choose an eQTLs approach to move in the intricate gene regulation architecture that may influence flavonol accumulation in tomato flesh. The material used was the introgression population developed from crosses between *S. pennellii* and the cultivated tomato *S. lycopersicum*, cultivar "M82". The most significant outcome from this research was the identification of two potential negative regulators we named *SlMYB4* and *SlELVIRA*. The function of these two genes was studied using VIGS (virus induced gene silencing) approach. The transient silencing of the *SlMYB4* and *SlELVIRA* resulted in an increase of flavonols as well as of chlorogenic acid in tomato flesh. This suggested a negative action of these two genes in flavonoid regulation. Ultimately, we believe that these studies may provide a new framework to explore how phenylpropanoid genes regulate the different branches of phenylpropanoid pathway to either increase plant tolerance to external stresses or to enhance the accumulation of human beneficial metabolites in important crops.

Chapter 1. General introduction

Chapter 1. General introduction

1.1 The importance of plant secondary metabolites

The population growth is expected to increase dramatically in the near future. According to the United Nation (UN) predictions, world population will raise from the actual 7.3 billion to 9.7 billion in 2050 (The Economist, 2015). What is more alarming of this prevision is that the growth rate is concentrated in developing countries where the current food availability is not enough to feed the new raised population. In these countries the less developed agriculture technologies, that are often associated to adverse environmental conditions, is a hard limit to improve (in terms of yield red quality) the crop productions. Moreover, the lack of food-chain infrastructure and storage technologies cause heavy pre-retail food waste. In developed countries, instead, the diet quality has become an important issue for an aging population that is getting more and more affected by life-style chronic diseases (Pillsbury et al., 2010). Functional foods and plant-based foods are drowning the public attention. In fact they are important sources either to invest against bad eating habits or to help an increasing vegetarian lifestyle with more nutrient vegetables and fruits. The challenges of feeding a changing world led the scientist to find some more efficient food production innovations (Godfray et al., 2010). Plant breeding and biotechnology have a fundamental role to implement the production for the new human needs. The plant breeding and biotechnology achievements can be summarized in: 1) more production and efficiency in the use of the agricultural resources; 2) plant protection against biotic and abiotic stress; 3) food bio-fortification to alleviate malnutrition or to face the modern need of nutrition and food technology processes. The main scope of plant breeding and biotechnology is consequently to gather in one genotype several advantageous characteristics (Borém & Fritsche-Neto, 2014). An interesting tool of the plant that can satisfy all the previous listed aims is the production of secondary metabolites. Plants are able to synthetize an enormous and variable quantity of molecules that are a key feature for plant

defence, nutrition and human health. Crop breeding has extensively used these wide metabolites variability to protect plants from pathogens (acting as antibiotic, antifungal and antiviral) or also to make them competitive for the space against other plants (acting as anti-germinative or toxic). Plant secondary metabolites are also important in defence of the plant against animals and insect thanks to their anti-feeding properties (Bourgaud et al., 2001). They also confer plant protection against abiotic stresses. They help plant to tolerate cold and heat stress or drought conditions. Secondary metabolites confer particular taste, odour and colours to the plants (Ramakrishna & Ravishankar, 2011). For this reason food technology is constantly looking for new natural food additives with beneficial effect on human health, with specific flavour or colour and also with preservative properties. The pharmaceutical and cosmetic industries are very much interested in plant secondary metabolites as bioactive compounds. In fact plant compounds have been used in traditional medicine for century and nowadays consumers are attracted by therapy with natural products (Calabrò, 2015). Table 1 shows examples of plant secondary metabolites used for different purposes.

Table 1. Examples of plant secondary metabolites used for different purposes: from plant defence to food technology.

Plant stress response			
Compound	**Plant**	**Type of response**	**Reference**
Sorbitol	Tomato	Salt stress	Tari et al., 2010
Polyamines	Rice	Salt stress	Ramakrishna & Ravishankar, 2011
Glucosinolates	Rapeseed	Drought stress	Jensen et al., 1996
Chlorogenic acid	Sunflower	Drought stress	Ramakrishna & Ravishankar, 2011
Flavonols	Soybean	UV-B stress	Yao et al., 2006
Anthocyanin	Red cabbage	Cold stress	Ahmad et al., 2015
Coumarin	Rapeseed	Several biotic stress	Dixon & Paiva, 1995
Glucosinolate	Crucifers	Several biotic stress	Dixon & Paiva, 1995
Nutritional properties for human health			
Compound	**Plant**	**Applications**	**Reference**
Anthocyanins	Berries	Hydrophilic antioxidant/chronic disease	Martin et al., 2013
Flavonols	Onion, Apple	Hydrophilic antioxidant/chronic disease	Soobrattee et al., 2006
Chlorogenic acid	Tomato, Potato, Coffee	Hydrophilic antioxidant/chronic disease	Islam, 2006
Vitamin C	Kiwi, Broccoli	Antioxidant/co-factor in several enzymatic reactions	Padayatty et al., 2003
Lutein	Spinach, Kale	Lipophilic antioxidant/decreases the risk of eye diseases	Rissanen et al., 2001
Lycopene	Tomato, Grapefruit, Papaya	Lipophilic antioxidant/stimulates the immune system	Martin et al., 2013
Food-technological properties			
Compound	**Plant**	**Applications**	**Reference**
Carotenoids	Tomato, Carrot	Colourant: yellow/orange	Blanc, 2002
Anthocyanins	Berries	Colourant: red, pink, purple, orange	Davies, 2009
Chlorophyll	Various	Colourant: green	Blanc, 2002
Resveratrol	Grape	Preservatives: antioxidant activity	Meyer et al., 2002
Capsaicin	Capsicum	Preservatives: antimicrobial activity	Naila et al., 2010

The high diversity and biological activity of plant secondary metabolites is extensively used to improve crops for different purposes (Kliebenstein, 2009). In fact, as discussed previously, the utility of these molecules is important for plant tolerance to environmental stresses and also for human nutrition and applications in food industries. In particular, three large classes of molecules are considered: terpenes, alkaloids and phenolics. These compounds derived from biochemical modification of primary metabolites (lipids, amino acid and carbohydrates) and are usually classified according to the biosynthetic pathway that produces them (Bourgaud et al., 2001). Terpenes are functionally important because they integrate to primary metabolism to produced hormones such as gibberellins and abscisic acid (Theis & Lerdau, 2003). Alkaloids are also involved in different plant physiological processes, but because of their pharmaceutical proprieties they are applied in different clinical uses (Shitan & Yazaki, 2007). Intensely studied are the phenolics and, in particular, the class of phenylpropanoids. In fact, these compounds showed a beneficial activity for a wide kind of purposes (Heim et al., 2002). In the following paragraph the antioxidant properties of phenylpropanoid compounds is analysed for their role in either plant protection or as food additives.

1.2 Flavonoids as antioxidants: role against environmental stresses

Crop yield increase is a very important goal to feed an over-increasing population. An important limit to this increment is given by adverse environmental conditions. In fact, it has been estimated that about half of the crop worldwide production is lost because of biotic and abiotic stress (Atkinson & Urwin, 2012). During their evolution, plants have developed sophisticated machineries to respond to both abiotic and biotic stress (Khraiwesh et al., 2012). The response usually starts with the perception of the specific stress signal, which leads to a chain of events and ends up with changes in plants' gene expression and metabolism (Dixon & Paiva, 1995). Phenylpropanoids represent a big class of secondary molecules that originated from deamination of phenylalanine and proceed with a series of biochemical modification to yield a wide kind of different compounds (Figure 1). Some of the products of the pathway can also be considered of primary

importance. For example, the monolignols are important blocks for lignin biosynthesis and consequently fundamental for plant survival. In the last decades the interest in this pathway has grown because it gives rise to molecules with strong activity in defence against either biotic or abiotic stress and also against a combined action of them (Atkinson & Urwin, 2012). Pathogen and herbivore attacks can be faced by different phenylpropanoids. For example, coumarin is induced by wounds of feeding herbivores with toxic effect on the animals and insect physiology. Wounding attacks induce the production of chlorogenic acid, ferulate esters (through the caffeic acid pathway) leading to increase the polyphenolic barrier of suberin and lignin of cell wall (Vishwanath et al., 2015). High concentration of isoflavans, isoflavonoids, stibenes, coumarins, flavonols and aurones may act as phytoalexins. In fact these can be produced around the site of infection at toxic level for the pathogens (Lattanzio et al., 2006). Phenylpropanoids have also a relevant protective role against abiotic stress and the recent literature is becoming more and more interested in this important function (Ahmad & Prasad, 2011).

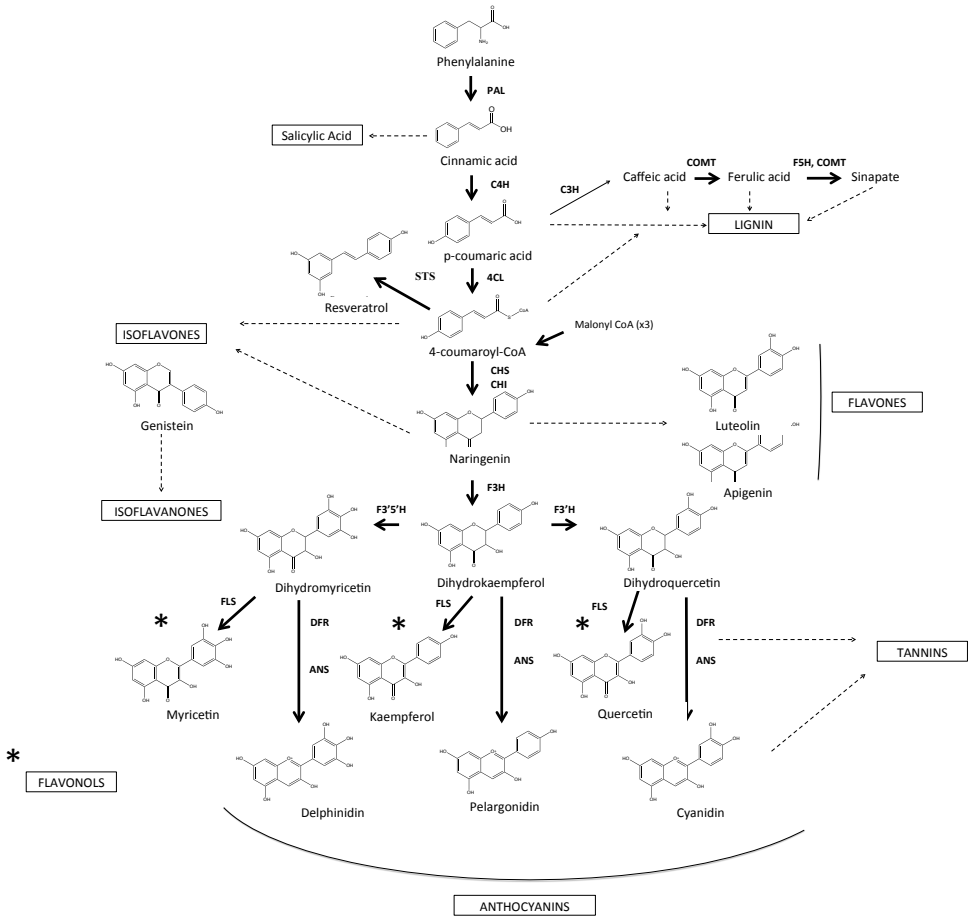

Figure 1. Phenylpropanoid pathway with particular relevance for anthocyanin and flavonoid branches. Enzymes involved in each step are indicated in uppercase letters. PAL, phenylalanine ammonia-lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate: CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; F3'H, flavonoid 3'-hydroxylase; F3',5'H, flavonoid 3',5'-hydroxylase; F3H, flavanone 3-hydroxylase; DFR, dihydroflavonol-4-reductase; ANS, anthocyanidin synthase; C3H, p-coumarate 3 hydroxylase; COMT, caffeic O-methyltransferase; F5H, ferulic acid 5-hydroxylase. Solid lines indicate biochemical steps defined by enzymes in the figure, while broken lines indicate biochemical multiple steps or steps undefined by enzymes in the figure.

Within the phenylpropanoids, the important group of flavonoids has long been reviewed as important compounds for plant-environment interaction (Pollastri & Tattini). The first step of flavonoids is directed by the activity of the enzyme chalcone synthase (CHS), which catalyse the production of flavones, flavonols, anthocyanin and tannins. In spite of different modes of action of these molecules, it has been hypothesized that their antioxidant property is the most important feature that allow flavonoids to face a wide and different kind of abiotic stresses. Flavonoids contribute to restore alternated cell homeostasis caused by an over-production of reactive oxygen species (ROS) in several stress situations (Di Ferdinando et al., 2012). ROS are reactive molecules that derive from partial reduction of oxygen. ROS can be listed according to the degree which oxygen is reduced: superoxide (O_2^-), hydrogen peroxide (H_2O_2) and hydroxyl radicals (OH^-). In not stress conditions plant metabolism normally produce ROS, which are physiologically maintained at level not toxic for the plant (Gill & Tuteja, 2010). Imbalance ROS production is caused by biotic (i.e. pathogen attack) and almost all the abiotic stresses (drought, salinity cold stress, UV-B, heavy metals, UV radiations, etc.), which alter the cell homeostasis with consequently oxidative damage and lethal effects for the plant (Sharma et al., 2012). High stress conditions usually inactivate antioxidant enzymes and, in these situations, the antioxidant function of flavonoids became particularly important (Hatier & Gould, 2008). The flavonoid intracellular and histological distribution, together with the antioxidant degree of a their specific chemical form, strongly suggests the primary use of these metabolites as ROS-quenchers. The vacuolar distribution of flavonoids indicates an activity in reducing H_2O_2 that usually escape by chloroplast under stress condition (Di Ferdinando et al., 2012). Other evidences show that flavonoids may be biosynthesized or accumulated in the chloroplast where they may act directly against the over-produced ROS in photoinibition condition (Agati et al., 2007).

It has been studied that different abiotic stress conditions induce specific type of flavonoids structures or that these structure are specifically produced by the same

stress on different plants. In particular, dihydroxy B-ring chemical forms are preferred to monohydroxy because the presence of an extra free hydroxyle (-OH) on the C-3' of the B-ring contributes to a stronger scavenger ability (Tattini et al., 2004). For example, in several plant species, quercetin 3-O- and luteolin 7-O-glycosides are much more produced than kaempferol 3-O- or apigenin 7-O-glycosides during either UV-B or drought stress (Di Ferdinando et al., 2012). The antioxidant activity of anthocyanins, highly produced under UV-B and cold stress, may also be considered predominant with respect to osmotic and screen roles. In fact, anthocyanin UV-screening features is mainly attributed to few acylated forms (Harborne & Williams, 2000), while the cold induction of anthocyanin, or flavonoids in general, seems to be due to the need of reducing the oxidative state caused by photoinibition condition that is enhanced during cold-stress (Steyn et al., 2002). Also the glycosylation of the molecules is considered a clue of their main antioxidant activity. In fact, though the antioxidant activity of these molecules is reduced with glycosylation, the glycosylated forms are more soluble in cellular milieu and are much more preserved from the autoxidation (Figure 2; Pearse et al., 2005; Di Ferdinando et al., 2012).

1.3 Flavonoids as antioxidants: role for human nutrition and food technology

It is well known that a not correct diet can cause several chronic diseases including obesity, type 2 diabetes, cardiovascular diseases and even cert type of cancers are heavily influenced by diet (Martin et al., 2013). Several experimental evidences showed that plant-based food prevent or reduce the risk of chronic diseases thanks to the chemopreventive properties of some metabolites produced by plants (Matkowski, 2008). The so-called "French Paradox" is a typical example that describes the beneficial effects of plant polyphenols, given by a moderate use of red wine, on a dairy fat diet typical of French people (Martin et al., 2011). Consequently researches on plants are beginning more and more addressed into the identification of health-promoting metabolites. In particular researches are focused into the genic control of these metabolites to improve

crops with more nutritional attributes. Many of these functional compounds from plant can be considered mainly for their antioxidant characteristics. Flavonoids are one of the most important classes of metabolites that naturally occur as phenolic antioxidants and that are widely present in the human diet. In fact, they contribute to the antioxidant properties of healthy vegetables and fruits as well as red wine and chocolate (Pandey & Rizvi, 2009). These beneficial effects are particularly due to the scavenger activity toward reactive oxygen and nitrogen species (RONS) that are usually underpin many degenerative and chronic pathologies. A diet rich of flavonoid may directly contribute to reduce either the effects or the production of RONS (Martin et al., 2011).

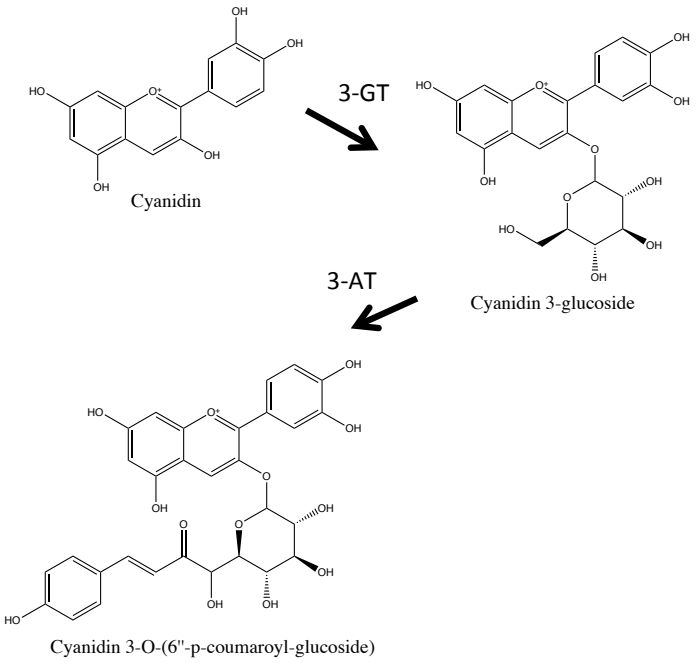

Figure 2. Chemical structures of two different cyanidin decorations mediated by 3-O-Glucosyltransferase (3-GT) and 3-O- Acetyltransferase (3-AT).

The outstanding characteristics of flavonoids are not just limited to nutritional properties indeed they can be helpful compounds for food storage and industry. The Royal Society of London coined the term of "sustainable intensification".

“Sustainable intensification” indicates modern agriculture that looks to all the aspect linked environment and food chain to produce more food for people (Davies et al., 2009). One of the most important aspects of the “sustainable intensification” is the reduction of food waste. In fact, roughly 30 to 40% of food in both the developed and developing worlds is lost to waste with different causes behind (Godfray et al., 2010). Consequently, there is an increasing need for continuing research in postharvest storage technologies and food preservatives. Modern consumers are anyway worried towards the use of chemical preservatives and consumers’ interest has started to be focused on natural products. Hence, industries are paying more emphasis on the use of plant compounds with antioxidant and antimicrobial properties. A success of flavonoid in this field is for example the transgenic tomato line “Del/Ros” (Butelli et al., 2008). This line accumulates high quantity anthocyanin in all the fruit, contributing not only to create an important “functional food”, but also to double the shelf life of tomato in a typical storage conditions. Further, the high antioxidant capacity induced by the accumulation of these metabolites gives tomato a reduced susceptibility to *Botrytis cinerea* (Zhang et al., 2013; 2015).

The use of food colorants as additives in the food industry is another factor that determinates the acceptability of consumers. Anthocyanins are for example widely used as food colourants (Markakis, 2012). However, although plants produce several types anthocyanins, the common commercial preparations do not contain many of them (e.g. 3- and 3,5-diglucosides of cyanidin, delphinidin and malvidin). They have limited colour stability with regard to pH, and are therefore limited in their food applications (Davies, 2009). This has led to research into sources of more complex, and thus more stable forms of anthocyanins, that usually have high levels of acylation (Bakowska-Barczak, 2005).

1.4 Plant breeding, genomic and biotechnological approaches to study and use flavonoids

Plant breeding and biotechnological studies frequently investigate and use interspecific variation in plant secondary metabolites to improve a crop. Given the diverse activities of secondary metabolites, this may be done to improve either agronomic or nutritional traits (Kliebenstein, 2009). Crops have been selected for years, producing consequent genetic bottlenecks and a reduced variability of metabolites that can be accumulated in their tissues. Plant breeding is based upon the production of genetic variability followed by efficient selection of newly produced genotypes. Variation can be produced in several ways, e.g. through crossing parental genotypes or exploring high variable germoplasm of plant kingdom. Plant secondary metabolites frequently have extensive intraspecific variation in both structure and content that is easily accessible (Kliebenstein, 2009). In particular, the wild ancestors of most crops are an important source of genetic variability, as they conserve a very large gene pool with respect to cultivated relatives. To access to this intraspecific and interspecific variation there is the need of a deep analysis of all available germoplasm to identify the desired trait (Glaszmann et al., 2010). Since metabolic traits are often a heritable character, the study of gene diversity at DNA level among plant population, wild ancestor or produced thought parental crossing is an important tool for breeding. Several molecular methods have been developed to detect such DNA variation and to assist traditional plant breeding (Keurentjes et al., 2006). Since 80s, molecular markers assisted selection (MAS) has highly helped plant breeding to enhance metabolic content in crop (Fernie & Schauer, 2009).

The variability of secondary metabolites between plants can be exhibited not just in presence or absence, but more often, in a quantitative variation. Further, the secondary metabolite accumulation is particularly influenced by interaction of genes with the environment and within a specific phonological stage. Different approaches can be used to undersell this variability but particularly interesting is the expression of quantitative trait loci (eQTL). The idea basing this approach is

that differences in gene transcript levels are also significantly important to influence plant phenotypes (Wentzell et al., 2007). The eQTL is a single chromosomic region that may influence the expression of a subset of genes controlling complex traits. This is particularly advantageous for metabolic traits because the regulation of a single metabolic pathway is characterized by a complex mechanism involving several structural genes (Fernie & Schauer, 2009) and also different regulatory genes that control the activation of the pathway during specific phonological stage or environmental context. For breeding purposes, the identification of a master gene within an eQTL hotspot with a major effect on a set of genes of the same biological process and pathway is particularly important (Ranjan et al., 2016). Genomic approaches are very helpful to study the architecture basing the regulation of an eQTL. Large- genome and RNA sequencing (RNA-seq) allows an easily identification of hotspot eQTL (Majewski & Pastinen, 2011). Further, the sequence information available in digital databases makes it easy to map and characterize gene within an eQTL. Other genomic approaches at proteome and metabolome level are tool for reverse genetic and biochemistry that can help breeding program to select and quickly produce crops with specific secondary metabolites (Ehlting et al., 2009). Another option to quantitatively increase plant metabolites is given by tissue culture and chromosome doubling. Plant cell and tissue cultures are one of the most used methods to enhance the production of different and useful secondary metabolites. In this system is very high the possibility to have somaclonal variability or can be induced using a direct mutagenesis approaches. Therefore, the variability obtained can be easily exploited to select high-producing strains (Chattopadhyay et al., 2002). These approaches are important to obtain on demand high yields of desired metabolites suitable for commercial purposes.

Another way to increase in both quantity and quality the production of plant secondary metabolites is the chromosome doubling. It has been studied that plants have used whole genome or single gene duplication to enhance the ability to produce different metabolites to tolerate different environmental conditions

(Moore et al., 2014). Through the use of colchicine it is possible to increase the level of cell ploidy and multiply the dosage of genes that control to secondary pathway (Dhawan & Lavania, 1996; Lavania et al., 2012).

When more metabolic variation is available from other species, transgenic manipulation is a suitable tool for crop manipulation and many biotechnological techniques have been developed to modify the production of plant secondary metabolites (Kliebenstein, 2009; Itkin & Aharoni, 2009). These techniques (commonly known under the name of bioengineering) are aimed to either study or improve the production of specific metabolites (Itkin & Aharoni, 2009). A common bioengineering method used to exploit the protective role of plant secondary metabolites is given by inducible promoters to express target genes in specific condition or in a specific plant tissue. In this way it is possible to avoid deleterious effect caused by a constitutive expression (Itkin & Aharoni, 2009). Moreover, it allows producing a specific metabolite only when it is necessary (i.e. stress conditions) or where they are required. Another important tool of the bioengineering are the transcription factors (TFs). Many secondary metabolites are highly controlled at transcriptional level and the use of TFs allows controlling more genes involved in a metabolic pathway (Itkin & Aharoni, 2009). In this way just altering the expression of a single gene a multiple step control can be obtained with a specific production of one or more type of metabolites. In the last decades many innovations were aimed into gene silencing. An innovative use has been done with Virus Induced Gene Silencing (VIGS). The techniques allows to silence transiently and quickly specific genes in various part of the plant without regenerating steps (Dinesh-Kumar et al., 2003). In silencing techniques one of the limits for successful results is the genetic redundancy that may, in some cases, interferes with the gene silencing. An option to avoid these problems is the use of chimeric repressor TF. In this case a regulatory protein is fused to a repressor motif (EAR-motif) that usually induces a changing in TF activity: from positive to negative regulator (Itkin & Aharoni, 2009). A very recent bioengineering strategy regards the genome editing using artificial nucleases. Clustered regularly

interspaced short palindromic repeat associate to nuclease Cas9 (CRISPR/Cas9) system is recently developed tool for the introduction of site-specific double-stranded DNA breaks in a target DNA sequence. Targeted genome editing has a very bi potential to accelerate basic research as well as plant breeding (Bortesi & Fischer, 2015).

1.5 Aims of the research

The genetic control of flavonoids is highly orchestrated at transcription level. Three regulatory classes of genes, belongings to MYB, bHLH and WD40 TFs take part alone or in a complex to regulate in spatial, temporal way and often under specific *stimuli*, the expression of the structural genes (Hichri et al., 2001). The possibility to identify a regulatory gene is definitely desirable. In fact, this means that with just one gene it is possible to regulate multiple steps in a single pathway (Grotewold, 2008). Although the phenylpropanoid pathway is one of the best studied, many gaps still remain as regards the genetic hierarchy. It controls either the tissue specific accumulation or the activation in response to a stress. This is particular true for crops. These gaps translate in limited possibilities to manipulate plant metabolites. For example, the manipulation of metabolites to defend plants against abiotic stresses is less developed with respect to biotic constraints (Kliebenstein, 2009). This is the case of anthocyanins, whose accumulation generally correlates with cold tolerance (Janská et al., 2010). Albeit several hypotheses have been made, it is still unclear which kind of mechanism is behind this tolerance (Tahkokorpi, 2010). In light of the lack of information available in this important research area, the first part of our thesis investigated anthocyanin production in leaves and in response to cold injury. This first part of our work was carried out on potato, with two main objectives being pursued:

- to identify the main factors regulating anthocyanin gene expression in cultivated potato (*S. tuberosum*) leaves, that represent the part of the plant that is more exposed to environmental stresses. In potato, little information is available on the genetic control of anthocyanins. Hence, we

characterized the MYB/bHLH complex that participates to anthocyanin regulation in leaves. Genomic, transcriptional and functional approaches, like Yeast Two Hybrid (Y2H), Bimolecular Fluorescence Complementation (BiFC) and transient overexpression, were used in this research (Chapter 2).

- to understand the role of anthocyanins in plant cold tolerance. In this case a comparative approach between the cultivated potato and the wild cold tolerant *S. commersonii* was carried out. Toward this goal, the genomic information available for both species (Potato Genome Sequencing Consortium, 2011; Aversano et al., 2015) has been of particular help. Functional analysis, combined with metabolic and microscopy investigations, better clarified how anthocyanins and the general phenylpropanoid pathway work in cold response (Chapter 3).

The lack of information we described above has a negative impact also to obtain plant-derived functional foods. Tomato is one of the crop extensively used to be enriched in beneficial metabolites (Raiola et al., 2014). Many studies focused on enhancing anthocyanins or flavonols in the fruit through either genetic engineering (Butelli et al., 2008; Luo et al., 2008) or breeding approaches (Povero et al., 2011). Unfortunately, the genetic regulation underlying the tissue specific accumulation of flavonoids is still unclear. This prevents enhancing useful metabolite levels in fruit flesh through conventional methods. This is the case of flavonols, which are naturally present only in traces in the flesh. Therefore, to help breeding to increase the level of these metabolites in tomato, in Chapter 4 the following objective was pursued:

- to identify, based on information from RNA sequencing data, a specific eQTL that influences flavonol structural genes and to assess potential negative regulators of flavonol accumulation in tomato flesh. For this purpose, *S. pennellii* introgression lines (ILs) were used as plant materials

and transient silencing (using VIGS- virus induced gene silencing method) and metabolic studies were carried out to verify gene functions.

1.6 References

Agati, G., Matteini, P., Goti, A., & Tattini, M. (2007). Chloroplast-located flavonoids can scavenge singlet oxygen. *New Phytologist*, *174*, 77-89.

Ahmad, P., & Prasad, M. N. V. (Eds.). (2011). *Abiotic stress responses in plants: metabolism, productivity and sustainability*. Springer Science & Business Media.

Ahmed, N. U., Park, J. I., Jung, H. J., Hur, Y., & Nou, I. S. (2015). Anthocyanin biosynthesis for cold and freezing stress tolerance and desirable color in *Brassica rapa*. *Functional & Integrative Genomics*, *15*, 383-394.

Atkinson, N. J., & Urwin, P. E. (2012). The interaction of plant biotic and abiotic stresses: from genes to the field. *Journal of Experimental Botany*, *63*, 3523-3543.

Aversano, R., Contaldi, F., Ercolano, M. R., Grosso, V., Iorizzo, M., Tatino, F., ... & Carputo, D. (2015). The *Solanum commersonii* genome sequence provides insights into adaptation to stress conditions and genome evolution of wild potato relatives. *The Plant Cell*, *27*, 954-968.

Bakowska-Barczak, A. (2005). Acylated anthocyanins as stable, natural food colorants–a review. *Polish Journal of Food and Nutrition Sciences, 14*, 107-116

Blanc, P. J. (2002). Natural Food Colorants. *Encyclopedia of Life Support Systems, EOLSS Publishers, Oxford,(on line)*.

Borém, A., & Fritsche-Neto, R. (Eds.). (2014). *Omics in Plant Breeding*. John Wiley & Sons.

Bortesi, L., & Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnology Advances*, *33*, 41-52.

Bourgaud, F., Gravot, A., Milesi, S., & Gontier, E. (2001). Production of plant secondary metabolites: a historical perspective. *Plant Science*, *161*, 839-851.

Butelli, E., Titta, L., Giorgio, M., Mock, H. P., Matros, A., Peterek, S., ... & Martin, C. (2008). Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nature Biotechnology*, *26*, 1301-1308.

Calabrò, S. (2015). Plant Secondary Metabolites. In *Rumen Microbiology: From Evolution to Revolution* (pp. 153-159). Springer, India.

Chattopadhyay, S., Farkya, S., Srivastava, A. K., & Bisaria, V. S. (2002). Bioprocess considerations for production of secondary metabolites by plant cell suspension cultures. *Biotechnology and Bioprocess Engineering*, *7*, 138-149.

Davies, B., Baulcombe, D., Crute, I., Dunwell, J., Gale, M., Jones, J., ... Toulmin, C. (2009). Reaping the Benefits: Science and the sustainable intensification of global agriculture. *London: Royal Society*, 1- 86

Davies, K. M. (2009). An introduction to plant pigments in biology and commerce. In *Plant Pigments and their Manipulation* (pp. 1-22). Blackwell Publishing, Oxford.

Dhawan, O. P., & Lavania, U. C. (1996). Enhancing the productivity of secondary metabolites via induced polyploidy: a review. *Euphytica*, *87*, 81-89.

Di Ferdinando, M., Brunetti, C., Fini, A., & Tattini, M. (2012). Flavonoids as antioxidants in plants under abiotic stresses. In *Abiotic Stress Responses in Plants* (pp. 159-179). Springer, New York.

Dinesh-Kumar, S. P., Anandalakshmi, R., Marathe, R., Schiff, M., & Liu, Y. (2003). Virus-induced gene silencing. *Methods Molecular Biology*, *236*, 287-293.

Dixon, R. A., & Paiva, N. L. (1995). Stress-induced phenylpropanoid metabolism. *The Plant Cell*, *7*, 1085-1097.

Ehlting, J., Hamberger, B., Ginglinger, J. F., & Werck-Reichhart, D. (2009). Genome Wide Approaches in Natural Product Research. In *Plant-derived Natural Products* (pp. 475-503). Springer, US.

Fernie, A. R., & Schauer, N. (2009). Metabolomics-assisted breeding: a viable option for crop improvement?. *Trends in Genetics*, *25*, 39-48.

Gill, S. S., & Tuteja, N. (2010). Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant Physiology and Biochemistry*, *48*, 909-930.

Glaszmann, J. C., Kilian, B., Upadhyaya, H. D., & Varshney, R. K. (2010). Accessing genetic diversity for crop improvement. *Current Opinion in Plant Biology*, *13*, 167-173.

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., ... & Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people. *Science*, *327*, 812-818.

Grotewold, E. (2008). Transcription factors for predictive plant metabolic engineering: are we there yet?. *Current Opinion in Biotechnology*, *19*, 138-144.

Harborne, J. B., & Williams, C. A. (2000). Advances in flavonoid research since 1992. *Phytochemistry*, *55*, 481-504.

Hatier, J. H. B., & Gould, K. S. (2008). Foliar anthocyanins as modulators of stress signals. *Journal of Theoretical Biology*, *253*, 625-627.

Heim, K. E., Tagliaferro, A. R., & Bobilya, D. J. (2002). Flavonoid antioxidants: chemistry, metabolism and structure-activity relationships. *The Journal of Nutritional Biochemistry*, *13*, 572-584.

Hichri, I., Barrieu, F., Bogs, J., Kappel, C., Delrot, S., & Lauvergeat, V. (2011). Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *Journal of Experimental Botany*, *62*, 2465–2483.

Islam S. (2006). Sweetpotato (*Ipomoea batatas* L.) leaf: its potential effect on human health and nutrition. *Journal of Food Science*. *71*, 13–21.

Itkin, M., & Aharoni, A. (2009). Bioengineering. In *Plant-derived Natural Products* (pp. 435-473). Springer, US.

Janská, A., Maršík, P., Zelenková, S., & Ovesna, J. (2010). Cold stress and acclimation–what is important for metabolic adjustment?. *Plant Biology*, *12*, 395-405.

Jensen C.R., Mogensen V.O., Mortensen G., Fieldsend J.K., Milford G.F.J., Andersen M.N., & Thage J. H. (1996). Seed glucosinolate, oil and protein contents of field-grown rape (*Brassica napus* L.) affected by soil drying and evaporative demand. *Field Crops Research*, *47*, 93-105.

Keurentjes, J. J., Fu, J., De Vos, C. R., Lommen, A., Hall, R. D., Bino, R. J., ... & Koornneef, M. (2006). The genetics of plant metabolism. *Nature Genetics*, *38*, 842-849.

Khraiwesh, B., Zhu, J. K., & Zhu, J. (2012). Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, *1819*, 137-148.

Kliebenstein, D. J. (2009). Use of secondary metabolite variation in crop improvement. In *Plant-derived Natural Products* (pp. 83-95). Springer, US.

Lattanzio, V., Lattanzio, V. M., & Cardinali, A. (2006). Role of phenolics in the resistance mechanisms of plants against fungal pathogens and insects. *Phytochemistry: Advances in Research*, *661*, 23-67.

Lavania, U. C., Srivastava, S., Lavania, S., Basu, S., Misra, N. K., & Mukai, Y. (2012). Autopolyploidy differentially influences body size in plants, but facilitates enhanced accumulation of secondary metabolites, causing increased cytosine methylation. *The Plant Journal*, *71*, 539-549.

Luo, J., Butelli, E., Hill, L., Parr, A., Niggeweg, R., Bailey, P., ... & Martin, C. (2008). AtMYB12 regulates caffeoyl quinic acid and flavonol synthesis in tomato:

expression in fruit results in very high levels of both types of polyphenol. *The Plant Journal*, *56*, 316-326.

Majewski, J., & Pastinen, T. (2011). The study of eQTL variations by RNA-seq: from SNPs to phenotypes. *Trends in Genetics*, *27*, 72-79.

Markakis, P. (Ed.). (2012). *Anthocyanins as food colors*. Elsevier, UK.

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

Martin, C., Zhang, Y., Tonelli, C., & Petroni, K. (2013). Plants, diet, and health. *Annual Review of Plant Biology*, *64*, 19-46.

Matkowski, A. (2008). Plant in vitro culture for the production of antioxidants—a review. *Biotechnology Advances*, *26*, 548-560.

Meyer, A.S., Suhr, K.I., Nielsen, P., Holm, F., (2002). Natural food preservatives. In *Minimal Processing Technologies in the Food Industry* (pp. 124-174). CRC Press LLC, Florida.

Moore, B. D., Andrew, R. L., Külheim, C., & Foley, W. J. (2014). Explaining intraspecific diversity in plant secondary metabolites in an ecological context. *New Phytologist*, *201*, 733-750.

Naila, A., Flint, S., Fletcher, G., Bremer, P., & Meerdink, G. (2010). Control of biogenic amines in food—existing and emerging approaches. *Journal of Food Science*, *75*, 139-150.

Padayatty, S.J., Katz, A.,Wang, Y., Eck, P., Kwon, O., …& Levine, M. (2003). Vitamin C as an antioxidant: evaluation of its role in disease prevention. *Journal of the American College of Nutrition. 22*, 18–35.

Pandey, K. B., & Rizvi, S. I. (2009). Plant polyphenols as dietary antioxidants in human health and disease. *Oxidative Medicine and Cellular Longevity*, *2*, 270-278.

Pearse, I. S., Heath, K. D., & Cheeseman, J. M. (2005). Biochemical and ecological characterization of two peroxidase isoenzymes from the mangrove, *Rhizophora mangle*. *Plant, Cell & Environment*, *28*, 612-622.

Pillsbury, L., Miller, E. A., Boon, C., & Pray, L. (Eds.). (2010). *Providing Healthy and Safe Foods As We Age: Workshop Summary*. National Academies Press, Washington.

Pollastri, S., & Tattini, M. (2011). Flavonols: old compounds for old roles. *Annals of Botany*, *108*, 1225-1233.

Potato Genome Sequencing Consortium. (2011). Genome sequence and analysis of the tuber crop potato. *Nature*, *475*, 189-195.

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

Raiola, A., Rigano, M. M., Calafiore, R., Frusciante, L., & Barone, A. (2014). Enhancing the health-promoting effects of tomato fruit for biofortified food. *Mediators of Inflammation*, *2014*, 1-16.

Ramakrishna, A., & Ravishankar, G. A. (2011). Influence of abiotic stress signals on secondary metabolites in plants. *Plant Signaling & Behavior*, *6*, 1720-1731.

Ranjan, A., Budke, J. M., Rowland, S. D., Chitwood, D. H., Kumar, R., Carriedo, L., ... & Sinha, N. R. (2016). eQTL in a Precisely Defined Tomato Introgression Population Reveal Genetic Regulation of Gene Expression Patterns Related to Physiological and Developmental Pathways. *BioRxiv*, 040592.

Rissanen T.H., Voutilainen S., Nyyssonen K., Lakka T.A., Sivenius J.,..& Salonen J.T. (2001). Low serum lycopene concentration is associated with an excess incidence of acute coronary events and stroke: the Kuopio Ischaemic Heart Disease Risk Factor Study. *British Jorunal of Nutrition*, 85, 749–54

Sharma, P., Jha, A. B., Dubey, R. S., & Pessarakli, M. (2012). Reactive oxygen species, oxidative damage, and antioxidative defense mechanism in plants under stressful conditions. *Journal of Botany*, *2012*, 1-26.

Shitan, N., & Yazaki, K. (2007). Accumulation and membrane transport of plant alkaloids. *Current Pharmaceutical Biotechnology*, *8*, 244-252.

Soobrattee, M.A., Bahorun, T., Aruoma, O.I. (2006). Chemopreventive actions of polyphenolic compounds in cancer. *Biofactors, 27*, 19–35.

Steyn, W. J., Wand, S. J. E., Holcroft, D. M., & Jacobs, G. (2002). Anthocyanins in vegetative tissues: a proposed unified function in photoprotection. *New Phytologist*, *155*, 349-361.

Tahkokorpi, M. (2010). Anthocyanins under drought and drought-related stresses in bilberry (*Vaccinium myrtillus* L.). *Oulu: Acta Universities Ouluensis*.

Tari I, Kiss G, Deer AK, Csiszar J, Erdei L, Galle A, … & Simon , L.M. (2010). Salicylic acid increased aldose reductase activity and sorbitol accumulation in tomato plants under salt stress. *Biologia Plantarum*, *54*, 677-83.

Tattini, M., Galardi, C., Pinelli, P., Massai, R., Remorini, D., & Agati, G. (2004). Differential accumulation of flavonoids and hydroxycinnamates in leaves of *Ligustrum vulgare* under excess light and drought stress. *New Phytologist*, *163*, 547-561.

The Economist. *Global population forecasts*, January, 2010.

Theis, N., & Lerdau, M. (2003). The evolution of function in plant secondary metabolites. *International Journal of Plant Sciences*, *164*, 93-102.

Vishwanath, S. J., Delude, C., Domergue, F., & Rowland, O. (2015). Suberin: biosynthesis, regulation, and polymer assembly of a protective extracellular barrier. *Plant Cell Reports*, *34*, 573-586.

Wentzell, A. M., Rowe, H. C., Hansen, B. G., Ticconi, C., Halkier, B. A., & Kliebenstein, D. J. (2007). Linking metabolic QTLs with network and *cis*-eQTLs controlling biosynthetic pathways. *PLoS Genet*, *3*, 1687- 1701.

Yao, Y. A., Zu, Y. Q., & Li, Y. (2006). Effects of quercetin and enhanced UV-B radiation on the soybean (*Glycine max*) leaves. *Acta Physiologiae Plantarum*, *28*, 49-57.

Zhang, Y., Butelli, E., De Stefano, R., Schoonbeek, H. J., Magusin, A., Pagliarani, C., ... & Martin C. (2013). Anthocyanins double the shelf life of tomatoes by delaying overripening and reducing susceptibility to gray mold. *Current Biology*, *23*, 1094-1100.

Zhang, Y., De Stefano, R., Robine, M., Butelli, E., Bulling, K., Hill, L., ... & Martin C. (2015). Different Reactive Oxygen Species Scavenging Properties of Flavonoids Determine Their Abilities to Extend the Shelf Life of Tomato. *Plant Physiology*, *169*, 1568-1583.

Zhao, D., & Tao, J. (2015). Recent advances on the development and regulation of flower color in ornamental plants. *Frontiers in plant science*, *6,* 261-276.

Chapter 2. High *AN1* variability and interaction with basic helix-loop-helix co-factors related to anthocyanin biosynthesis in potato leaves

D'Amelia V., Aversano R., Batelli G., Caruso I., Castellano M. M., Castro-Sanz B., Chiaiese P., Fasano C., Palomba F., Carputo D.

Published in: *The Plant Journal, 80, 527-540*

Chapter 2. High *AN1* variability and interaction with basic helix-loop-helix co-factors related to anthocyanin biosynthesis in potato leaves

2.1 Introduction

Three major classes of molecules confer colour to plants: anthocyanins, carotenoids and chlorophylls (Tanaka et al., 2008). Among them anthocyanins are of particular interest for their well documented beneficial effects on plant physiological processes and human health (Stintzing & Carle, 2004; De Pascual-Teresa & Sanchez-Ballesta, 2008). Anthocyanin biosynthesis is primarily controlled through regulation of genes encoding the structural enzymes of the phenylpropanoid metabolic pathway (Hichri et al., 2011a). Expression of the structural genes is tightly controlled by the ternary complex 'MBW' (Patra et al., 2013). This complex is composed of MYB and basic helix–loop–helix (bHLH) transcription factors, together with WD40 repeat proteins that regulate flavonoid spatiotemporal production in conjunction with the promoters of structural genes (Lin-Wang et al., 2010; Feller et al., 2011). Plant MYBs are a large gene family whose members have many different functions and represent key factors activating specific downstream genes (Takos et al., 2006). The subfamily R2R3 MYB is the largest group present in higher plants and the most extensively studied. It possesses a structurally conserved DNA-binding domain consisting of up to two imperfect repeats, R2 and R3 (Jin & Martin, 1999). The R3 repeat provides a platform for protein–protein interaction, especially with the bHLH co-factor (Grotewold et al., 2000). In this complex the affinity between MYB and the *cis*-element of the target gene may be partly influenced by the bHLH partners (Hichri et al., 2011b). R2R3 MYBs play important roles in tissue-specific anthocyanin accumulation in many plants (Gao et al., 2013). They include *AN2* in petunia (*Petunia* × *hybrida*), *ROSEA1*, *ROSEA2*, and *VENOSA* in snapdragon (*Antirrhinum majus*), *C1* and *P1* in maize (*Zea mays*) and *PAP1* in Arabidopsis (Grotewold et al., 1991; Sainz et al., 1997; Quattrocchio et al., 1999; Borevitz

et al., 2000; Schwinn et al., 2006). The effects of bHLH co-factors in different tissues of the same species remain unclear. Recently, there has been increasing interest in understanding the molecular mechanisms regulating phenylpropanoid production in potato (*Solanum tuberosum*) tubers. It has been reported that the production of red and purple anthocyanins is controlled by *R* and *P* loci, while *D* (the developer locus) is responsible for tissue-specific anthocyanin accumulation in tuber skin (Jung et al., 2009). The *D*, *R* and *P* loci have been mapped to chromosomes 10, 2 and 11, respectively (Jung et al., 2009). Furthermore, their structural and regulatory function in the anthocyanin biosynthetic pathway has been elucidated. It has been reported that *R* encodes a dihydroflavonol 4-reductase (DFR), *P* a flavonoid 3′,5′-hydroxylase (F3′5′H). The *D* locus cosegregates with an ortholog of petunia *AN2*, an R2R3 MYB transcription factor (De Jong et al., 2004; Jung et al., 2009; Zhang et al., 2009a,b). Jung et al. (2009) designated this gene *AN1* and identified two different allelic forms: *AN1-777* and *AN1-816*. While our understanding of *AN1's* involvement in anthocyanin regulation in tubers has expanded, the role in leaves has received scant attention. Here production of anthocyanins could be important for plant defence mechanisms (Gould, 2004). In potato the additional advantage of a high leaf anthocyanin content may be related to a protective role against herbivorous insects (Schaefer & Rolshausen, 2005). In the case of aphids, this would have important implications also for virus spread. In leaves, it has been shown that the potato loci conditioning anthocyanin accumulation are tightly linked to each other and to locus *D* (De Jong, 1991; Jung et al., 2009). Further studies showed that the constitutive expression of *AN1* or *StMTF1* (another potato MYB gene) causes increased accumulation of anthocyanins in foliage (Rommens et al., 2008; Jung et al., 2009). Recently, Payyavula et al. (2013) showed that *AN1* expression was inducible by a sucrose treatment in plantlets of the cultivar Purple Majesty, suggesting that environmental conditions affect *AN1* transcript abundance in vegetative tissues.

The aim of our study was to identify the main factors regulating anthocyanin gene expression in potato leaves. Target genes believed to be involved in tuber pigmentation were selected for genetic and functional analysis. Sequence analysis of different potato genotypes showed an extensive intraspecific nucleotide sequence variation of *AN1*, both in the predicted promoter and coding sequence, where potential protein polymorphisms were identified. Expression analysis suggested that leaf pigmentation is associated with *AN1* expression and that *StJAF13* (previously named *StbHLH2;* Payyavula et al., 2013) acts as putative *AN1* co-regulator for anthocyanin gene expression in leaves of the red leaf variety 'Magenta Love,' while a concomitant expression of *StbHLH1* in 'Double Fun' may also contribute to anthocyanin accumulation in leaves of this cultivar. Protein interaction of AN1 with both StbHLH1 and StJAF13 was detected using yeast two-hybrid and, in the case of StJAF13, further confirmed using bimolecular fluorescence complementation (BiFC) assays. Stable co-transformation of *AN1* and either *StbHLH1* or *StJAF13* in tobacco (*Nicotiana tabacum*) produced a stronger pigmentation with respect to single *AN1* overexpression. These findings indicate that in potato leaves *StJAF13* enhances *AN1* activity in anthocyanin production.

2.2 Materials and methods

2.2.1 Plant materials and growth conditions

Used in this study were tetraploid (2n = 4x = 48) potato (*Solanum tuberosum*) commercial varieties employed as parents in the breeding programs carried out in Portici. They included 'Double Fun,' 'Magenta Love' (both with purple leaves), 'Blue Star,' 'Violet Queen,' 'Flamenco,' 'Briosa' 'Assergi,' 'Adora,' 'Blondy,' 'Carmine,' 'Désirée,' 'Pukara,' 'Ilaria,' 'Silvy' and 'Spunta.' Two *S. tuberosum* haploids (2n = 2x = 24) named 'COI25' and 'DEI23' were also analysed. Plants were micro propagated *in vitro* on Murashige and Skoog (MS) medium (Sigma-Aldrich, http://www.sigmaaldrich.com) with 1% (w/v) sucrose and 0.8% (w/v) agar, and incubated at 24°C, exposed to an irradiance of 200 µmol m^{-2} sec^{-1} and under a 16 h/8 h (light/dark) photoperiod. Three plants of each genotype were transplanted to styrofoam trays filled with sterile soil. Plants were maintained in a growth chamber at 25°C ± 2 under a 16 h/8 h (light/dark) photoperiod at 220 µmol m^{-2} sec^{-1} irradiance provided by a cool, white-fluorescent tube (Philips, http://www.philips.com). Young leaf samples were collected from each replicate after one month and stored at −80°C before analysis. Each sample used for nucleic acid extraction and anthocyanin analysis consisted of a pool of three replicates powdered in liquid nitrogen.

2.2.2 Total anthocyanin analysis

Total anthocyanin content was estimated with the pH-differential spectrum method as described by Zhang et al. (2012). One hundred mg of powdered samples of tobacco (*Nicotiana tabacum*) shoots and potato leaves were used for this analysis.

2.2.3 Nucleic acid extraction and molecular analysis of *AN1*

Genomic DNA was extracted using DNeasy Plant Mini Kit (Qiagen) according to the manufacturer's instructions. Four overlapping primer pairs were designed based on the consensus sequence of *AN1-777*, *AN1-816* alleles and their

respective mRNA sequences. Primer pairs were used to amplify a region from nucleotide 14 to nucleotide 1045 of the *AN1* genomic sequence (Supplementary data Table S1). PCR reactions were performed using GoTaq DNA Polymerase (Promega) as reported by manufacturer. PCR products were analysed on 2% (w/v) agarose gel and amplicons obtained were gel-purified with QIAquick Gel Extraction Kit (Qiagen). Purified products were sequenced and aligned against alleles *AN1-777* and *AN1-816*. A region between nucleotides –1545 and –2154 from *AN1* start codon was amplified using three pairs of primers designed on PGSC (Supplementary data Table S1). Total DNA-free RNA was purified using the Spectrum Plant Total RNA Kit and On-Column DNase I Digestion Set (Sigma-Aldrich), following manufacturer's instructions. One µg of total RNA was reverse transcribed to complementary DNA (cDNA) using oligo-dT(20) primer and SuperScript III reverse transcriptase (Invitrogen) in 20 µl of final reaction according to the manufacturer's instructions.

2.2.4 Cloning of *AN1* and basic helix-loop-helix (bHLH) genes from potato leaves

The coding sequences (CDS) of potato *AN1* and *JAF13* were amplified from cDNA of 'Magenta Love,' while *bHLH1* from 'Blue Star' using Phusion High-Fidelity DNA Polymerase (Thermo-Scientific). Gateway *att*B primers were used to obtain *att*B-flanked PCR. All primer pairs used are listed in Supplementary data Table S1. *Att*B-flanked PCR products were cloned in pDON207 (Invitrogen) to obtain entry clones. Two different CDSs were cloned for both *StJAF13* and *StbHLH1* using different start codons of the transcript sequence annotated on Potato Genome Sequencing Consortium (PGSC). They were named *StJAF13*, *StJAF13*Δ264, *StbHLH1* and *StbHLH1*Δ58. *StJAF13* cloned sequence started from nucleotide 298 to nucleotide 2559. *StbHLH1* corresponded to sequence annotated by Payyavula et al. (2013). *StJAF13*Δ264 and *StbHLH1*Δ58 corresponded to CDSs predicted on the PGSC database. All the obtained clones were sequenced.

2.2.5 Gene expression analyses

Conventional PCR was carried out on 'Double Fun' cDNA to amplify fragments of anthocyanin synthase (*ANS*), dihydroflavonol 4-reductase (*DFR*), *StbHLH1*, *StJAF13* and *AN1*. The fragments obtained were cloned in pGEM-T Easy (Promega). Plasmids obtained were normalized to a concentration of 25 ng/µl and 10-fold serial dilutions (ranging from 10−2 to 10−8) were used to construct standard curves. The concentration of each plasmid dilution was measured using the Qubit 2.0 Fluorometer (Invitrogen) and the corresponding copy number for each concentration (copies/µl) was calculated as reported by Whelan et al. (2003). The synthesised cDNA of each potato genotype was diluted five times in sterile water. One microliter of cDNA from each genotype and of plasmid dilution were used for quantitative RT-PCR (qRT-PCR) analysis using 0.3 µm of each primer pair. All reactions were run in triplicate using QuantiFast SYBR Green PCR Kit (Qiagen) in a final volume of 20 µl of reaction. qRT-PCR was carried out using the Rotor-Gene 6000 (Corbett) and cycle conditions indicated by QuantiFast SYBR Green PCR Kit handbook (Qiagen). Gene expression analysis was carried out using Rotor-Gene 6000 software. The standard curves were used to calculate the copy number of molecules per µl of the corresponding target genes in each potato genotype. Primer pairs used for qRT-PCR analysis are listed in Supplementary data Table S1.

2.2.6 Yeast two-hybrid assay

The CDS of *AN1* was cloned in frame in pGADT7 (Clontech) between *Eco*RI and *Xho*I restriction sites. *StJAF13*, *StbHLH1*, *StJAF13*Δ264 and *StbHLH1*Δ58 were inserted into pGBKT7 (Clontech) using *Xma*I and *Sal*I. All plasmids were sequenced to ensure that no mutations had been introduced. Bait and prey plasmids were co-transformed into yeast strain AH109 according to Bai & Elledge (1997). Co-transformants were grown overnight in liquid culture and an equal amount of cells for all co-transformations was spotted on media with and without histidine and adenine to check for bait and prey interaction.

2.2.7 Bimolecular Fluorescent Complementation (BiFC) assay

The p35S*::AN1*::*Nt-YFP*, *p35S::StJAF13::Ct-YFP*, *p35S::Nt-YFP* or *p35S*::Ct-YFP constructs were obtained from the corresponding entry clones using Gateway recombination technology (Invitrogen). BiFC experiments were performed by co-transfecting different combinations of the constructs into 2-week-old *Nicotiana benthamiana* leaves via *Agrobacterium tumefaciens* transformation as described in Payyavula et al. (2013) using 150 µm acetosyringone and setting the final OD (optical density) of each *Agrobacterium* suspension to 0.3. Imaging was conducted three days after infiltration with a Leica TCS SP8 confocal laser scanning and a × 40 water immersion objective. For nuclei staining, leaf samples were mounted into a 4′,6-diamidino-2-phenylindole (DAPI) staining solution (0.01% silwet/300 nm DAPI). The software package provided by the manufacturer was used for projections of serial optical sections and image processing.

2.2.8 Overexpression of potato transcription factors in tobacco plants

The *StJAF13*, *StbHLH1* and *AN1* genes from entry clones were cloned in the 35SCaMV expression cassette of pGWB411 (Nakagawa et al., 2007) using Gateway recombination technology (Invitrogen). *A. tumefaciens* strain LBA4404 transformed with each expression vector was used for agroinfiltration in fully expanded leaves of *N. benthamiana* as reported by Payyavula et al. (2013). Stable genetic transformation was carried out by co-cultivation of *N. tabacum* leaf explants with *A. tumefaciens* in accordance with Horsch et al. (1985).

2.2.9 Bioinformatics and statistical analysis

Amino acid and nucleotide sequences were collected from the National Center for Biotechnology Information (NCBI) and PGSC database. Alignments and phylogenetic trees were performed with geneious software v6.0.6 (Biomatter). blastp and blastx programs (http://www.ncbi.nlm.nih.gov/blast) were used to perform homology researches in GenBank. Analysis of variance (ANOVA) on

qPCR data was carried out using xlstat-pro 7.5.3 software (Addinsoft, http://www.xlstat.com). Duncan's test was performed to compare mean values.

2.3 Results

2.3.1 High *AN1* gene nucleotide variability and amino acid polymorphic sites

High variability was found in the *AN1* genomic sequence of 17 potato varieties and several variants were identified (Table 1). In the coding sequence (CDS), the frequency of all sequence variants per bp was 7%, with 58 sequence variants identified. Exon 3 presented the highest number of sequence variants (38), 30 of which were due to SNPs (single nucleotide polymorphisms) and eight to indels. Analysis of exons 1 and 2 revealed a total of 8 and 12 sequence variants, respectively, all attributable to SNPs. Exon 2 displayed a frequency of polymorphic sites per bp (about 9%) higher than that of exons 1 and 3 (about 6%). As regards the analysis of the nucleotide coding for domains (NCD), R2 and R3 exhibited a similar number of sequence variants (10 and 11, respectively). All the variants were due to SNPs, with only one indel detected in R3. The variable region (VR) showed 37 sequence variants (30 SNPs and seven indels), with an average indel length of 10 bp. In the intronic region 72 sequence variants and the highest frequency of polymorphic sites per bp (15%) were found. In particular, intron 1 displayed 25 sequence variants, comprising 20 SNPs and five indels (on average 7.6 bp long), while intron 2 showed 39 SNPs and eight indels (on average 3.1 bp long) and a frequency of polymorphic sites per bp lower than that of intron 1 (14% versus 19%).

Table 1. Summary of sequence variants obtained analysing all fragments amplified from potato genotypes and using the alleles *AN1-777* and *AN1-816*

	CDS				NCD[b]			introns		
	exon 1	exon 2	exon 3	total	R2	R3	VR[c]	intron 1	intron 2	total
Length of sequences (bp)	123	130	550	803	162	150	471	103	279	382
Frequency of all sequence variants (%)[a]	6.5	9.2	6.9	7.2	6.2	7.3	7.9	24.3	16.8	18.8
All sequence variants (no.)[a]	8	12	38	58	10	11	37	25	47	72
Frequency of polymorphic sites per bp (%)	6.5	9.2	5.5	6.2	6.2	6.7	6.4	19.4	14.0	15.4
Nucleotide substitutions per sites (no.)	8	12	30	50	10	10	30	20	39	59
Indels (no.)	0	0	8	8	0	1	7	5	8	13
Average indel length (bp)	0	0	15	15	0	48	10.3	7.6	3.1	4.8
Frequency of indels per bp (%)	0	0	1.5	1.0	0	0.7	1.5	4.9	2.9	3.4

[a]SNPs and indels.
[b]NCD: nucleotide coding for domains.
[c]VR: variable region.

To investigate the variants found in CDS, *in silico* analyses were carried out. Sequence alignments between *AN1* fragments of each potato genotype and the *AN1* GenBank reference sequences (*AN1-777* and *AN1-816*) provided evidence that indels within exon 3 were due to a deletion of 39 bp in *AN1-777* allele, confirming a previous report (Jung et al., 2009). In the same site we found deletions in 'Spunta' and 'Assergi' (Supplementary data Figure S1) consisting of 48 and 52 nucleotides, respectively. The latter overlapped the same region of a 39 bp deletion already described for allele *AN1-777*, but located 16 bp downstream. Table 2 summarises potential missense mutations found with respect to *AN1-777* and *AN1-816* sequences. Some of them were genotype-specific: in 'Adora', for example, isoleucin was substituted with threonine due to a single nucleotide mutation (T to C) at position 71 and isoleucin with valine as a consequence of a single mutation from A to G. 'Silvy,' 'Double Fun,' 'Magenta Love' and 'Flamenco' displayed a common mutation at position 949 (C to G), producing the substitution of isoleucine with methionine. Within exon 3 we found three additional mutations causing amino acid substitutions on predicted protein sequence. In particular, two of them (at nucleotide positions 1095 and 1166) resulted in a substitution from polar amino acids (threonine and serine, respectively) to the apolar alanine. Figure 1 displays the consequences of the amino acid substitutions previously presented and their location on the R2R3 MYB predicted protein. The double consecutive substitution (GG to CA) on exon 2 found in 'Magenta Love' and 'Silvy' altered the conservative regularly spaced tryptophan repetition in the helix turn helix structure. In fact, this apolar residue was replaced with the amino acid proline. Amino acid motif analysis showed that the deletion of 48 nucleotides found in the first part of exon 3 in 'Spunta' caused a loss of 16 amino acids in the last part of R3 domain. This means that 'Spunta' has an allele without the conservative protein motif ANDV described by Lin-Wang et al. (2010). This identifier motif was found in all the other fragments analysed as well as in the annotated alleles.

Table 2. Summary of missense mutations found with respect to *AN1-777* and *AN1-816* consensus sequence

Exon	Nucleotide positions	Missense mutations	Amino acid substitutions	Genotypes
1	71	T>C	I→T	Adora
	88	A>G	I→V	Adora
2	323-326	TTGG >ACCA	DW→EP	Silvy
	323-326	TTGG >ACCA	DW→EP	Magenta Love
3	858	G>A	C→Y	Adora
	949	C>G	I→M	Silvy
	949	C>G	I→M	Double Fun
	949	C>G	I→M	Magenta Love
	949	C>G	I→M	Flamenco
	1095	A>G	T→A	Spunta
	1095	A>G	T→A	Magenta Love
	1114	T>G	D→E	Magenta Love
	1114	T>G	D→E	Briosa
	1114	T>G	D→E	Double Fun
	1114	T>G	D→E	Spunta
	1166	T> G	S→A	Double Fun
	1166	T> G	S→A	Silvy
	1166	T> G	S→A	Magenta Love

AN1 CDS was cloned from red leaf 'Magenta Love.' CDS and predicted amino acid analyses confirmed the presence of all the previous substitutions found for 'Magenta Love' with respect to *AN1-777* and *AN1-816* alleles. The complete cloned CDS showed a substitution at the nucleotide position 26 (not identified with primer pairs used for *AN1* analysis). It was characterized by a transition mutation (C to T) causing a substitution from polar serine to apolar phenylalanine in the amino terminal part of the predicted protein (Figure 1). Translated BLAST

(tBLASTx) was run in the National Center for Biotechnology Information (NCBI) database to find homologies with 'Magenta Love' AN1 predicted protein. Two proteins (named AN1 and transcription factor R2R3 MYB) with a sequence identical to that of AN1 cloned from 'Magenta Love' were identified. Phylogenetic analysis picked out other *S. tuberosum* R2R3 MYB proteins (Supplementary data Figure S2). A similarity of 94% was found with CAI1 (corresponding to StMTF1), another potato MYB protein, clustering in a separate group along with tomato (*S. lycopersicum*) and wild tomatoes ANT1. We found two additional potato MYBs grouped with the tomato SlAN2. In particular, MYBA1 shared a similarity of 90% with AN1 cloned from 'Magenta Love'. MYBA1 corresponds to the translated sequence of *StAN3* indicated as a possible *AN1* pseudogene (Jung et al., 2009).

Regulation of *AN1* gene expression is fundamental towards anthocyanin accumulation (Payyavula et al., 2013). Sequencing of the 5′ upstream region was therefore carried out to identify potential polymorphisms between red and green leaf varieties. Sequence analysis revealed the presence in the candidate *AN1* promoter of a simple motif repeat (TA) 36 at −1969 bp from the ATG. BLASTn analysis of sequence immediately surrounding the TA motif showed an identity of 95–96% with the retrotransposons of the potato SINE (short interspersed elements) family named SolSV (Wenke et al., 2011). A region between nucleotides −2077 and −1895 of the *AN1* promoter was aligned with the SINE SolSV_St3. The TA motif showed high similarity with the tail of the putative SINE retrotransposons (Supplementary data Figure S1). Through sequence analysis we found that in 'Silvy' there were 8 TA repeats, as in SolS-V_St3. By contrast, in red leaf 'Magenta Love' and 'Double Fun' the number of TA repeats was 16.

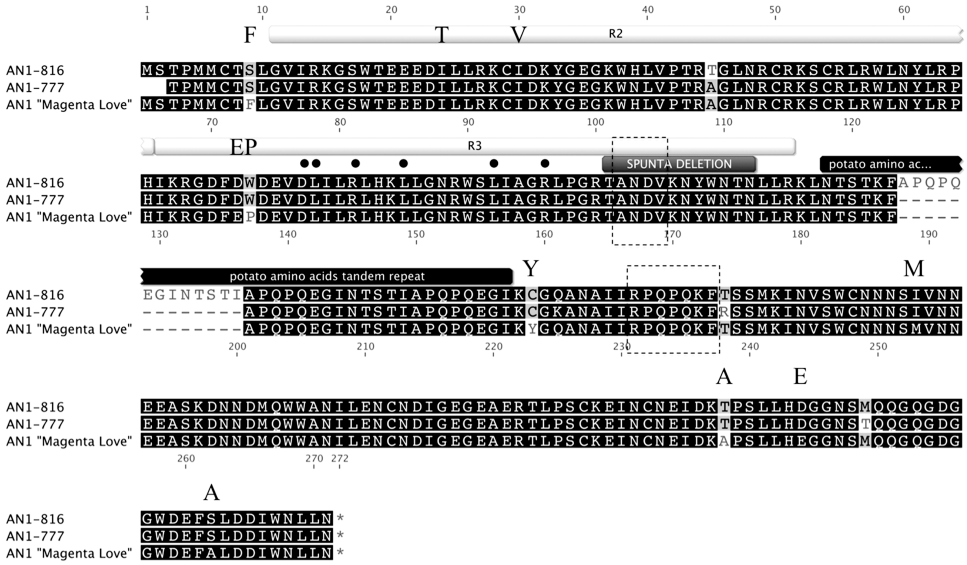

Figure 1. Alignment of the predicted amino acid sequence of *AN1-777*, *AN1-816* and *AN1* sequence from 'Magenta Love'. Shading of the alignment was carried out with the Geneious software and indicates the percentage of similarity between residues: 100% of similarity (black shading), 80–100% of similarity (dark-grey shading), 60–80% of similarity (light-grey shading), less than 60% of similarity (white shading). Capital letters above the alignment indicate the predicted amino acid substitutions of different potato genotypes. White boxes indicate MYB R2 and R3 domains. The site of predicted 'Spunta' deletion is marked. A typical potato amino acid tandem repeat is highlighted by a black box. Broken line boxes indicate conserved motif [A/S/G]NDV within R2R3 domain of dicot anthocyanin promoting MYBs and conserved motif [R/K]Px [P/A/R]xx [F/Y] typical of Arabidopsis anthocyanin promoting MYBs. Dark circles indicate amino acid residues involved in MYB interaction with bHLH in Arabidopsis.

2.3.2 Isolation of *StJAF13* coding sequence

Two potato bHLH genes (*StJAF13* and *StbHLH1*) were identified on Potato Genome Sequencing Consortium (PGSC) database based on homology with other species (Payyavula et al., 2013). These genes are putatively involved in regulation of anthocyanin biosynthesis in potato. The potato *StJAF13* sequence was blasted onto the NCBI database. We found an identity of 97% with *S. lycopersicum*

GLABRA3-like, 90% with *N. tabacum JAF13-like* a and b and 87% with that of *P. x hybrida PhJAF13*. Transcript alignments showed that annotated potato *StJAF13* CDS (1083 bp) was 800 bp shorter than its homolog (1883 bp). To investigate the real dissimilarity of the coding sequence length between potato and other species, *StJAF13* CDS sequence was cloned from complementary DNA (cDNA) of 'Magenta Love' beginning from the common start codon annotated for petunia *PhJAF13* and tomato *GLABRA3-like* CDS. The obtained *StJAF13* sequence was aligned against PGSC *JAF13* transcript and genomic sequences. We found that the CDS of *StJAF13* presented a deletion of 381 nucleotides with respect to the annotated sequence. This caused the elimination of a premature stop codon and the production of a CDS of 1.881 bp. Therefore, a revised genomic structure for potato *StJAF13* is proposed, as reported in Supplementary data Figure S3. No alternative splicing was found in the flanking regions of the deletion. In all genotypes a single amplification product of about 181 bp was amplified from their respective cDNAs using primers flanking the deletion. No products of 562 bp (indicating the presence of a 381 bp insertion) were found (Supplementary data Figure S4). Phylogenetic analysis showed that the StJAF13 translated sequence is grouped with tomato GLABRA3-like, tobacco JAF13-like a and b and petunia PhJAF13 (Supplementary data Figure S5). By contrast, in the other branch of the same subclade we found many putative anthocyanin regulators of the genus *Ipomoea*. The same relation with genus *Ipomoea* was also found for the bHLH1 translated sequence (Supplementary data Figure S6). StJAF13 sequence was aligned with phylogenetically related proteins (Figure 2): GLABRA3-like of tomato, PhJAF13 of petunia, JAF13-like b of tobacco, MYC-RP of *Perilla frutescens*, DELILA2 of *A. majus* and bHLH of *I. nil*. The amino terminal part, comprising approximately the first 200 amino acids is involved in the interaction with MYB partners (Hichri et al., 2011a). Our data showed that this region is highly conserved among the seven different species. JAF13 predicted protein annotated on PGSC database lacks this portion. The region located between MIR (MYB-interacting region) and bHLH domain was less conserved among the sequences aligned, except for tomato GLABRA3-like and StJAF13, which shared

similar sequences. In addition, GLABRA3-like and StJAF13 showed an identical bHLH domain.

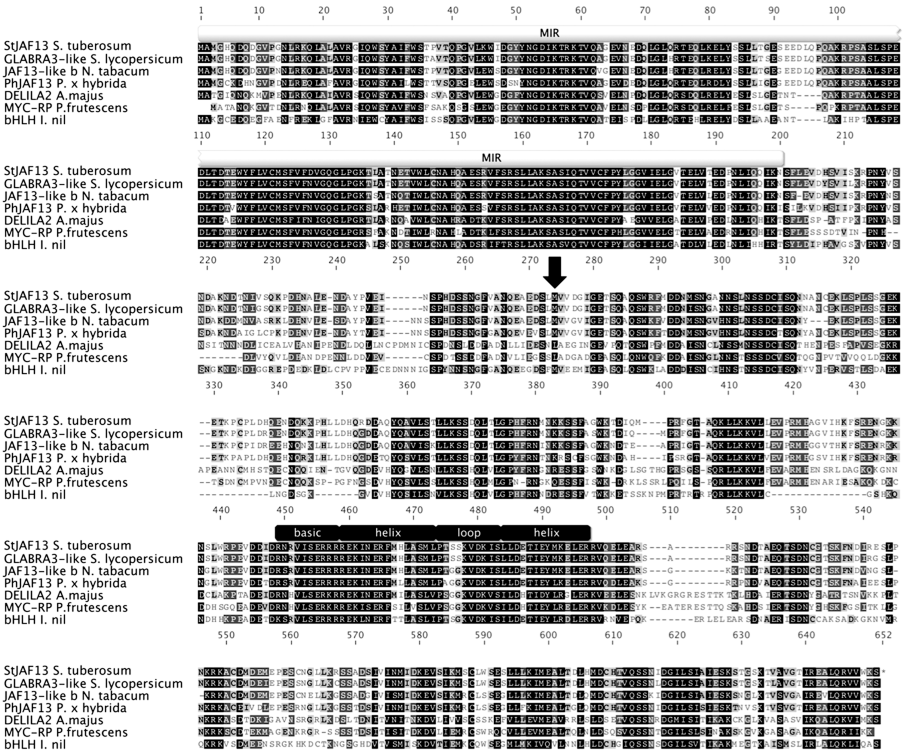

Figure 2. Alignment of StJAF13 with GLABRA3-like (*S. lycopersicum*), JAF13-like b (*N. tabacum*), PhJAF13 (*P. x hybrida*), DELILA2 (*A. majus*), MYC-RP (*P. frutescens*) and bHLH (*I. nil*). Shading of the alignment was carried out with the Geneious software and indicates the percentage of similarity between residues: 100% of similarity (black shading), 80–100% of similarity (dark-grey shading), 60–80% of similarity (light-grey shading), less than 60% of similarity (white shading). MIR: MYB-interacting region; bHLH: basic helix-loop-helix domain; The black arrow indicates the start of StJAF13Δ264 protein (M at site 274) annotated in the PGSC database.

2.3.3 High *AN1* and *StJAF13* expression in red leaf genotypes

Anthocyanin content was analysed to determine the amount of these pigments in leaves of potato genotypes. Red leaf 'Double Fun' and 'Magenta Love' showed the highest total anthocyanin content (17 mg g^{-1} and 13 mg g^{-1} FW, respectively) (Figure 3). On the same genotypes expression analysis of anthocyanin structural

and regulatory genes was carried out. Absolute quantification of dihydroflavonol 4-reductase (*DFR*) and anthocyanin synthase (*ANS*) genes expression is reported in Figure 4 (a). Compared with the other genotypes, 'Double Fun' and 'Magenta Love' showed a significantly higher expression of the two genes. The former genotype displayed the highest value for *DFR* and *ANS*: about 1750 copies/µl and 238 copies/µl, respectively. In 'Magenta Love' the number of copies/µl detected was about 612 for *DFR* and 15 for *ANS*. In all the other genotypes, expression of both genes was close to 0 copies/µl. Only in 'Violet Queen,' 'Flamenco' and 'Blue Star' the expression of *DFR* was between 20 and 100 copies/µl. A similar trend of gene expression was observed for *AN1* and *StJAF13* genes (Figure 4b). As for *AN1*, about 3000 copies/µl and 1200 copies/µl were detected in 'Double Fun' and 'Magenta Love,' respectively. As for *StJAF13*, 48 copies/µl were found in 'Double Fun' and 17 in 'Magenta Love.' Different results were found for *StbHLH1* (Figure 4b). Indeed, no expression was detected in red leaf 'Magenta Love,' whereas estimated copies/µl in 'Double Fun' was 57.

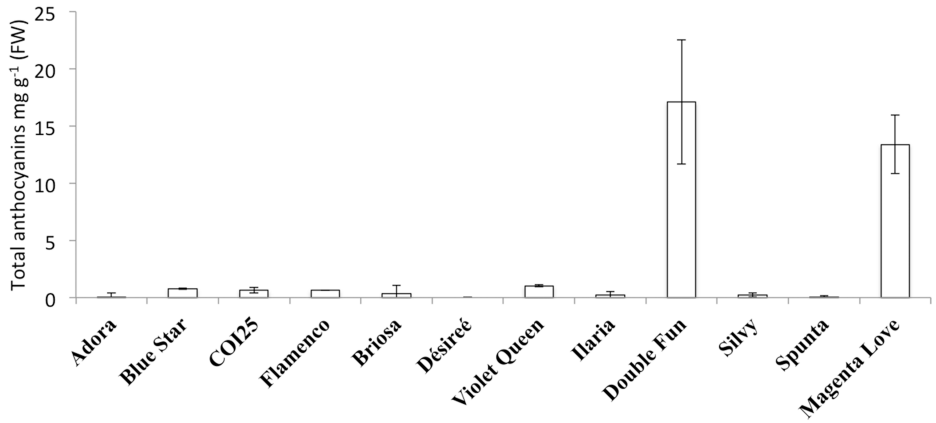

Figure 3. Total anthocyanin content (mg g^{-1} FW) in leaves of 12 varieties of potato.Each value represents the mean of three determinations (± standard deviation (SD)).

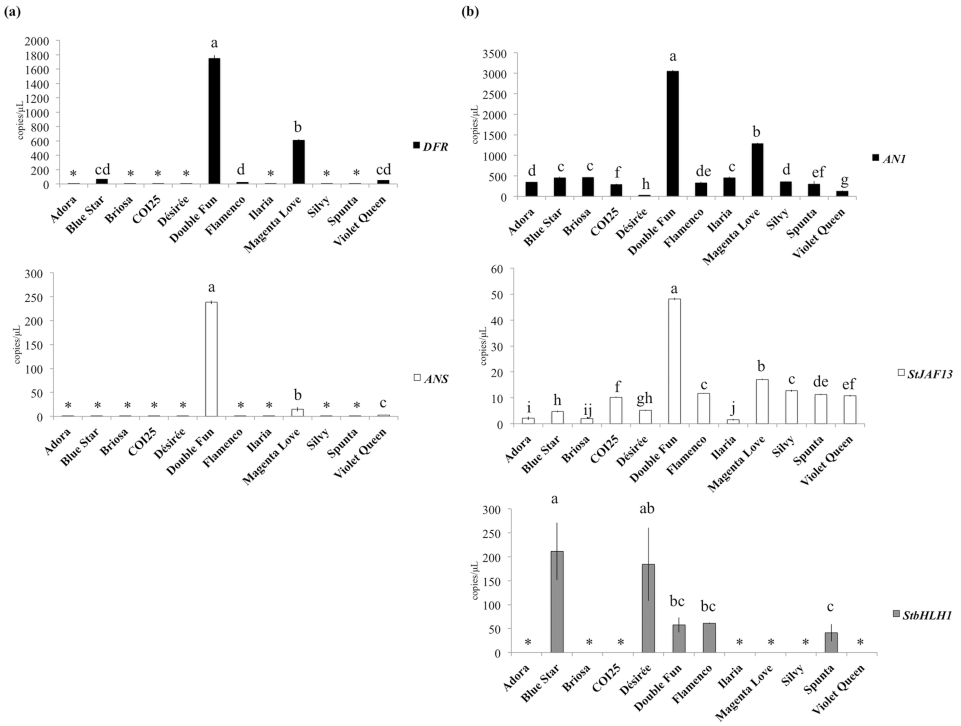

Figure 4. Gene expression analysis in leaves of potato genotypes as monitored by absolute qRT-PCR.(a) Expression of *DFR* and *ANS* anthocyanin structural genes.(b) Expression of *AN1*, *StJAF13* and *StbHLH1* transcription factors. Each value represents the mean of three determinations (± SD). Means denoted by the same letter did not differ significantly at $P \leq 0.05$ according to Duncan's multiple range test. Genotypes denoted by * showed no detectable levels of expression.

2.3.4 AN1 interacts with StbHLH1 and StJAF13

AN1 and *StJAF13* showed largely overlapping expression profiles, with higher levels detected in genotypes with red leaves. It is therefore conceivable that StJAF13 could be an AN1 interacting partner in leaves. The interaction between AN1 and StJAF13 was therefore tested in the yeast two-hybrid system (Figure 5 b, c). StbHLH1, previously shown to be strongly expressed in red- and purple-tuber varieties (Payyavula et al., 2013) was also tested. The coding sequences of *StJAF13* and *StbHLH1* were fused with the binding domain of GAL4 and

transformed into yeast strain AH109 together with the prey *AN1* fused to GAL4 activation domain. Yeast growth on selective media lacking histidine and adenine was observed, indicative of an interaction between AN1 and both StJAF13 and StbHLH1. No interaction was detected when the truncated versions of StJAF13 and StbHLH1, amplified using primers on the predicted translation initiation site as indicated in PGSC, were used (StJAF13Δ264, StbHLH1Δ58). This provided evidence that amino acids 1–264 of StJAF13 and 1–58 of StbHLH1 are required for the interaction with AN1 in yeast. To confirm the results obtained with the yeast two-hybrid, BiFC studies were performed. The protein AN1 fused upstream of the N-term fragment of YFP and StJAF13 fused to the C-terminal fragment of YFP were expressed by agroinfiltration in *N. benthamiana* leaves in combination or along with the respective controls. As shown in Figure 5 (d), fluorescence corresponding to reconstituted YFP was observed in the nucleus of co-transformed cells with the AN1 and StJAF13 constructs. No fluorescence was detected when these proteins were expressed along with only the N-term or the C-term of the YFP moieties nor when StJAF13Δ264 was used in combination with AN1. Together, our results confirmed that AN1 interacts with StbHLH1 and StJAF13 and that amino acids at the N-term of bHLHs might be required for the interaction to take place.

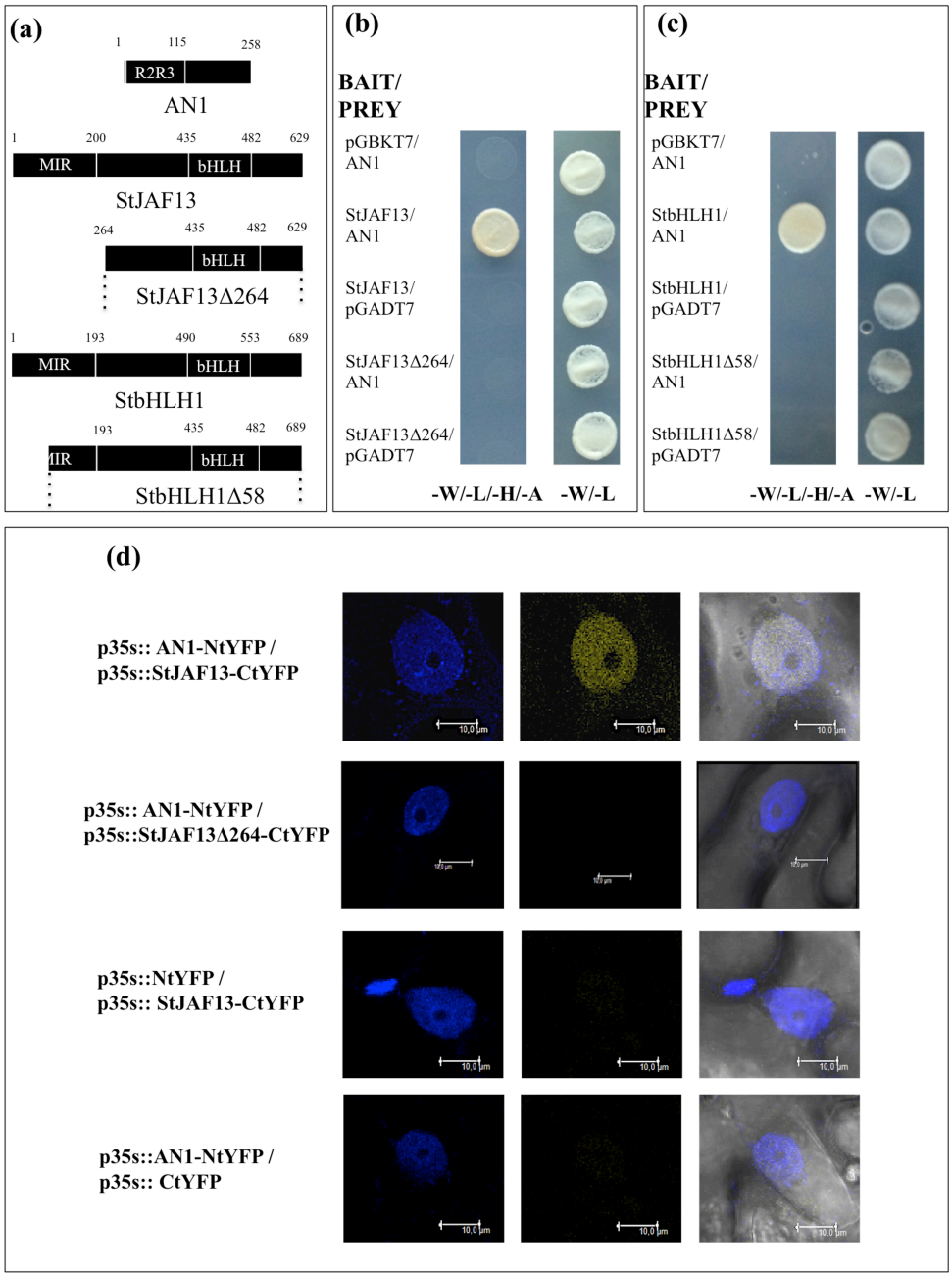

Figure 5. AN1 interacts with StJAF13 and StbHLH1.(a) Diagrams of the proteins used for the yeast two-hybrid and BiFC assay. R2R3, MYB repeat domains; MIR, MYB-interacting region; bHLH, basic helix-loop-helix domain.(b, c) AN1 interacts with StJAF13 (b) and StbHLH1 (c) in a yeast two-hybrid assay. StJAF13 and StbHLH1 were cloned in the bait plasmid pGBKT7 and co-transformed in

yeast with the prey AN1. The pGBKT7/AN1pGADT7 and pGBKT7bHLH/pGADT7 combinations were used as negative controls. Truncated fragments of StJAF13 and StbHLH1 lacking 264 and 58 amino acids at the N-term respectively (StJAF13Δ264 and StbHLH1Δ58), did not interact with AN1. Yeast cells grown for three days on synthetic complete media lacking tryptophan and leucine (–W/–L) and on selective media lacking tryptophan, leucine, histidine and adenine (–W/–L/–H/–A) are shown.(d) Analysis of the AN1 and StJAF13 interaction by BiFC in *N. benthamiana* leaves. Constructs expressing different combinations of AN1 fused to N-terminal and StJAF13 fused to C-terminal YFP fragments were agroinfiltrated together or along with control constructs. DAPI staining, reconstituted YFP fluorescence and merge images for a representative field for each combination are shown.

2.3.5 Co-expression of *AN1*, *StbHLH1* and *StJAF13* genes enhances anthocyanin accumulation in tobacco plants

To validate the interaction of AN1 with StJAF13 and StbHLH1 and to assess the impact of this interaction on anthocyanin biosynthesis, both transient and stable nuclear genetic transformations were performed. Transient expression affected tissue pigmentation in cells agroinfiltrated with the expression vectors p35s::*AN1*, and p35s::*AN1* plus p35S::*StJAF13* or p35s::*StbHLH1*. Anthocyanin accumulation was visibly enhanced in leaves co-infiltrated with *AN1* and *StJAF13* (Figure 6 a). Stable transgenic tobacco plants overexpressing *AN1*, *bHLHs* or a combination of *AN1* and *bHLHs* were generated. Transgenic plants showed differential phenotypes depending on the transgenes inserted (Figure 6 b). Overexpression of either *StJAF13* or *StbHLH1* alone produced shoots with no pigmentation variation, while the constitutive expression of *AN1* positively affected the pigmentation (Figure 6 b). Furthermore, co-expression of *AN1* with both bHLHs in tobacco cells produced shoots with an enhanced pigmentation. Measurements of anthocyanin levels of transgenic plants demonstrated that *AN1* overexpression increased the anthocyanin content in shoots and that this effect is potentiated by the co-expression with *StJAF13* and *StbHLH1* (Figure 6 d). These assays

confirmed the important role of *AN1* along with *StJAF13* and *StbHLH1* in regulating leaf anthocyanin pathway.

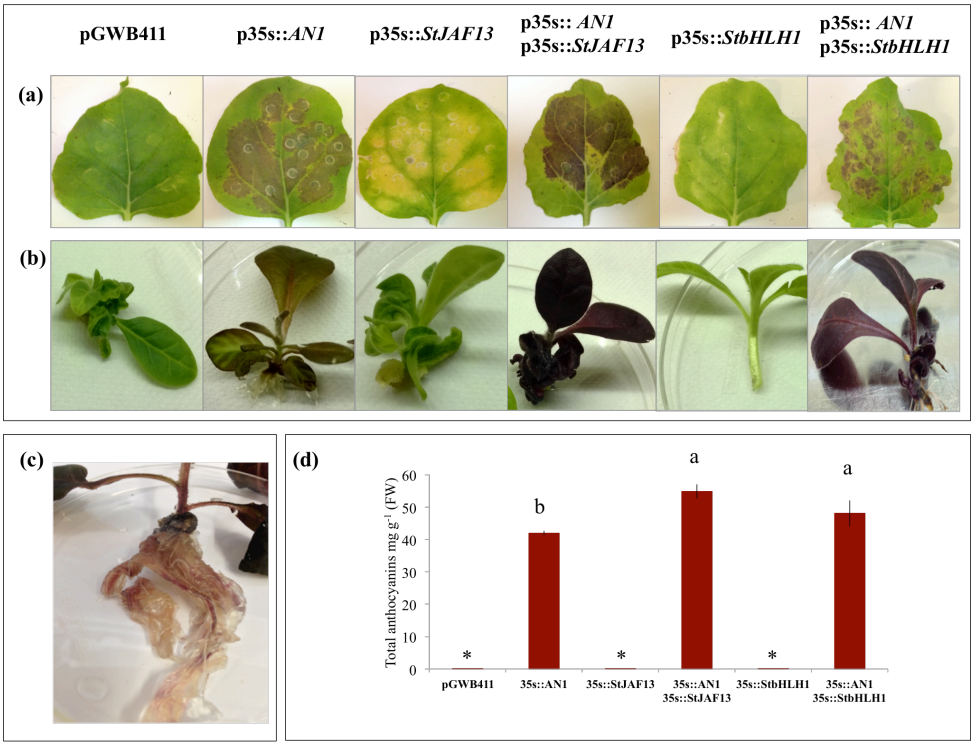

Figure 6. The effects of overexpression of *AN1*, *StbHLH1* and *StJAF13* in *Nicotiana* spp.

(a) Leaves of *N. benthamiana* after agroinfiltration with pGWB411 empty vector, *AN1*, *StJAF13*, *StbHLH1* or a combination of *AN1* with *StJAF13* or *StbHLH1*.

(b) Stable transgenic plants transformed with pGWB411 empty vector, *AN1*, *StJAF13*, *StbHLH1* or a combination of *AN1* with *StJAF13* or *StbHLH1*.

(c) Roots of transgenic plants overexpressing *AN1*.

(d) Total anthocyanin content (mg g^{-1} FW) in transformed tobacco plants. Each value represents the mean of three determinations (`SD). Means denoted by the same letter did not differ significantly at $P \leq 0.05$ according to Duncan's multiple range test. Genotypes denoted by * showed no detectable level of anthocy- anin content**.**

2.4 Discussion

The cultivated potato *S. tuberosum* is the fourth most important crop worldwide. It is also classified as sensitive to environmental stresses, which affect tuber yield and quality. The presence of high levels of anthocyanins in tissues exposed to stress conditions could be an important advantage for plant resistance: in leaves these pigments can act as UV-B filters, protect DNA from oxidative damage, increase resistance to pathogens thanks to their antimicrobial activity (Solfanelli et al., 2006; Van Oosten et al., 2013). In addition, since insects may show preferences for green leaves for food or oviposition and do not possess red colour receptors, leaf anthocyanins participate in defence mechanisms against herbivores (Schaefer & Rolshausen, 2005; Karageorgou & Manetas, 2006; Chittka & Döring, 2007). Lovdal et al. (2010) pointed out that a high level of anthocyanins and flavonoids in plants may reduce the need for pesticide treatments, and hence can be an interesting target for plant breeding. So far in potato only the molecular mechanisms and genes that control anthocyanin accumulation or biosynthesis in the tubers have received much attention (Jung et al., 2009; Zhang et al., 2009a,b; Stushnoff et al., 2010; Payyavula et al., 2013; Tai et al., 2013). All these works emphasised the role of the *AN1* gene, coding for a R2R3 MYB, in controlling the expression of structural genes involved in the anthocyanin pathway, especially in the tuber periderm (Jung et al., 2009). As there are few studies on other potato tissues, the research we carried out focused on the leaves. We analysed the *AN1* genomic sequence in a set of potato genotypes displaying different leaf pigmentations. Indeed, as reported by Schwinn et al. (2006), striking effects on the phenotype could be caused by small changes in MYB sequence. We found a high variability in both coding and non-coding sequences of *AN1*. Our results were consistent with the *AN1* sequence variability already reported by Jung et al. (2009) but also allowed us to explore its allelic diversity. In particular, a number of indels in the second intron and the third exon were found. These indels could explain results of Jung et al. (2009), who found bands of different sizes when they amplified a portion of these regions from different tetraploid potatoes. Most of the

variants we found were either silent mutations or were located in intronic regions. By contrast, others may have a functional meaning. The deletion found in 'Spunta', for example, caused the loss of important residues for strong (K–N) and weak (T–N) interaction with DNA bases (Hichri et al., 2011b). Similarly, the single amino acid substitutions we detected may have a potentially functional effect, producing structural protein modification. In grape, Hichri et al. (2011b) found that a single residue mutation in the R2 domain modified the protein interaction properties of MYB together with its DNA-binding affinities. Our transcription analysis confirmed that *AN1* expression is correlated with anthocyanin production in leaves. Genotypes with high leaf anthocyanin content had the highest number of copies of both *AN1* mRNA and *DFR*/*ANS* mRNA. The presence of a putative SINE retrotransposon element in *AN1* promoter may help to explain differences in *AN1* expression between our green and red leaf varieties. In fact, it is known that the presence of transposable elements affects *MYB* expression (Kobayashi et al., 2004; Walker et al., 2007; Telias et al., 2011; Butelli et al., 2012; Lisch, 2013). Therefore it is possible that the different length of TA motifs in 'Magenta Love' and 'Double Fun' with respect to 'Silvy' may influence *AN1* transcription levels. In apple it has been demonstrated that the presence of a 23-bp repeat motif causes an increase in MYB10 transcription levels, producing effects on leaf and fruit flesh colour (Espley et al., 2009). Research on *AN1* promoter is in progress to clarify the effects of TA motif on *AN1* expression level.

Previous reports have demonstrated that potato loci affecting tissue-specific accumulation of anthocyanins, including locus *Pw*, *Pd* and *Pv* (Kessel & Rowe, 1974; Garg et al., 1981; De Jong, 1991), are strongly associated to locus *D* (coding for *AN1*). Based on analysis of PCR fragments obtained through the amplification of *AN1* in several cultivars and breeding clones, Jung et al. (2009) suggested that two or more of the classical loci may correspond to *AN1* sequence variants. In light of this and considering the high variation both in *AN1* expression and nucleotide sequence here reported, we can speculate that *AN1* is a key gene responsible for differences in anthocyanin biosynthesis not only in the tuber but

also in the leaves. In sweet potato a mechanism of anthocyanin biosynthesis common to different tissues was also suggested (Mano et al., 2007). A similar hypothesis has been formulated for the apple (*Malus x domestica*) *Rni* locus that controls the red flesh phenotype but may be an allelic variant at the *MYB10* locus that cosegregates with the red foliage phenotype (Chagné et al., 2007; Espley et al., 2009). We found a strong association of *StJAF13* expression with anthocyanin production in leaves: red leaf genotypes 'Double Fun' and 'Magenta Love' showed a high expression of *AN1* and *StJAF13*. This was correlated with the expression levels of late anthocyanin structural genes (*ANS* and *DFR*). In tomato leaves a positive correlation between some secondary metabolites, including anthocyanins, and the expression of *ANT1* and *SlJAF1*3 (corresponding to *GLABRA3-like*) was also observed (Lovdal et al., 2010). The expression of *StbHLH1*, proposed as *AN1* co-regulator candidate in tubers (De Jong et al., 2004; Payyavula et al., 2013), was virtually undetectable in leaves of 'Magenta Love,' suggesting that in this genotype *StJAF13* is the main bHLH-type AN1 co-factor in leaves. By contrast, in the other red leaf genotype 'Double Fun' here analysed, the levels of expression found for *StbHLH1* may still be enough to trigger or contribute to anthocyanin production. It is therefore possible that the relative contribution to anthocyanin accumulation of StbHLH1 and StJAF13 could vary depending on the tissue (e.g. leaves or tubers), the genotype and/or environmental conditions. A similar mechanism was described in snapdragon, where R2R3 MYB *ROSEA* determines the pattern and the level of pigmentation in both lobes and tubes, while bHLH *DELILA* is required in both corolla tubes and lobes of the flowers, whereas bHLH *MUTABILIS* is required in lobes if *DELILA* is not functional (Schwinn et al., 2006; Petroni & Tonelli, 2011).

StbHLH1 and StJAF13 interact with AN1 and the N-terminal portion of bHLHs is required for the interaction to take place. Our results are in accordance to those reported in petunia by Quattrocchio et al. (2006). The authors suggested that the N-terminal of PhJAF13 was sufficient for interaction with AN2 (homolog of potato AN1). They also observed that the *PhJAF13* expression pattern perfectly

overlapped with that of *DFR* and *AN2* (Quattrocchio et al., 1998). The same authors found that the co-bombardment of *AN2* and *PhJAF13* in leaf cells induced activity of the *DFR* promoter, while *AN2* alone was less efficient and *PhJAF13* alone was insufficient for its activation. We can speculate that, depending on the potato transcription factor transformed, the expression of tobacco anthocyanin structural genes was differentially controlled. We observed an increase in anthocyanin accumulation in the co-presence of the over-expressed *bHLHs* and *AN1*, and no pigmentation variation with only *StJAF13 or StbHLH1*. Similarly, Butelli et al. (2012) observed that orange (*Citrus sinensis*) *RUBY* promoted a stronger pigmentation of transformed tobacco plants when co-expressed with snapdragon bHLHs. As postulated for tuber flesh (De Jong et al., 2004), *AN1* may be fundamental but not sufficient for the complete pigmentation of the leaves, and the interaction with *bHLHs* can have a crucial importance to improve the affinity with the promoter *cis*-element of the structural genes. Based on the results reported herein, it seems that the production of anthocyanins is associated to two combined mechanisms. One is linked to *AN1* expression, that is correlated with pigmentation intensity. The other to *AN1* allelic sequences, that could influence the mechanism of specific tissue production. It is also possible that, as hypothesised by Jung et al. (2009) for tuber flesh, the allelic configuration of different loci may influence the phenotype when *AN1* is constitutively expressed. In this scenario, a single amino acid substitution could cause an alteration of the interaction between MYB proteins with the co-partners, resulting in a variety of different pigment accumulation.

2.5 Conclusions

In conclusion, we found that high sequence variation characterizes *AN1* in potato, both in the gene body and in the promoter, and that high leaf anthocyanin content is associated to a high expression of *AN1* and *StJAF13* in 'Magenta Love.' We also demonstrated that AN1 protein physically interacts with StbHLH1 and StJAF13, and we located this latter interaction in the cell nucleus. Overexpression of *AN1* together with either *StJAF13* or *StbHLH1* in tobacco led to a stronger pigmentation as compared to plants where only bHLHs or *AN1* where expressed. All together, our results suggest that in leaves of 'Magenta Love' AN1 and StJAF13 can form a functional complex that drives anthocyanin biosynthesis. Future studies would be usefully spent further investigating the MBW complex, to characterize the potential pleiotropic functions of *AN1* and *StJAF13* and to better clarify the role of *StbHLH1* in potato leaves. In the tetraploid cultivated potato, the introgression of traits like high leaf pigmentation may require several generations of crosses and selection. For this reason, our results may provide the basis to identify genes responsible for anthocyanin biosynthesis, facilitating the selection of progeny with a high level of anthocyanins in leaves.

2.6 References

Bai, C., & Elledge, S. J. (1996). Gene identification using the yeast two-hybrid system. *Methods in Enzymology*, *273*, 331-347.

Borevitz, J. O., Xia, Y., Blount, J., Dixon, R. A., & Lamb, C. (2000). Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *The Plant Cell*, *12*, 2383-2393.

Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., ... & Martin, C. (2012). Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *The Plant Cell*, *24*, 1242-1255.

Chagné, D., Carlisle, C. M., Blond, C., Volz, R. K., Whitworth, C. J., Oraguzie, N. C., ... & Gardiner, S. E. (2007). Mapping a candidate gene (MdMYB10) for red flesh and foliage colour in apple. *BMC Genomics*, *8*, 212-223.

Chittka, L., & Döring, T. F. (2007). Are autumn foliage colors red signals to aphids?. *PLoS Biology*, *5*, 1640–1644.

De Jong, H. (1991). Inheritance of anthocyanin pigmentation in the cultivated potato: a critical review. *American Potato Journal*, *68*, 585-593.

De Jong, W. S., Eannetta, N. T., De Jong, D. M., & Bodis, M. (2004). Candidate gene analysis of anthocyanin pigmentation loci in the Solanaceae. *Theoretical and Applied Genetics*, *108*, 423-432.

de Pascual-Teresa, S., & Sanchez-Ballesta, M. T. (2008). Anthocyanins: from plant to health. *Phytochemistry Reviews*, *7*, 281-299.

Espley, R. V., Brendolise, C., Chagné, D., Kutty-Amma, S., Green, S., Volz, R., ... & Allan, A. C. (2009). Multiple repeats of a promoter segment causes transcription factor autoregulation in red apples. *The Plant Cell*, *21*, 168-183.

Feller, A., Machemer, K., Braun, E. L., & Grotewold, E. (2011). Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *The Plant Journal*, *66*, 94-116.

Gao, Z., Liu, C., Zhang, Y., Li, Y., Yi, K., Zhao, X., & Cui, M. L. (2013). The promoter structure differentiation of a MYB transcription factor RLC1 causes red leaf coloration in empire red leaf cotton under light. *PloS One*, *8*, e77891.

Garg, K. C., Tiwari, S. P., & Sharma, K. P. (1981). Inheritance of leaf pigmentation in dihaploid-Phureja hybrids of potato. *JIPA, Journal of the Indian Potato Association 8*, 31-34.

Gould, K. S. (2004). Nature's Swiss army knife: the diverse protective roles of anthocyanins in leaves. *BioMed Research International*, *2004*, 314-320.

Grotewold, E., Athma, P., & Peterson, T. (1991). Alternatively spliced products of the maize P gene encode proteins with homology to the DNA-binding domain of myb-like transcription factors. *Proceedings of the National Academy of Sciences*, *88*, 4587-4591.

Grotewold, E., Sainz, M. B., Tagliani, L., Hernandez, J. M., Bowen, B., & Chandler, V. L. (2000). Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R. *Proceedings of the National Academy of Sciences*, *97*, 13579-13584.

Hichri, I., Barrieu, F., Bogs, J., Kappel, C., Delrot, S., & Lauvergeat, V. (2011). Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *Journal of Experimental Botany*, *62*, 2465–2483..

Hichri, I., Deluc, L., Barrieu, F., Bogs, J., Mahjoub, A., Regad, F., ... & Lauvergeat, V. (2011). A single amino acid change within the R2 domain of the VvMYB5b transcription factor modulates affinity for protein partners and target promoters selectivity. *BMC Plant Biology*, *11*, 117-130.

Horsch, R. B., Fry, J. E., Hoffmann, N. L., Eichholtz, D., Rogers, S. A., & Fraley, R. T. (1985). A simple and general method for transferring genes into plants. *Science*, *227*, 1229-1231.

Jin, H., & Martin, C. (1999). Multifunctionality and diversity within the plant MYB-gene family. *Plant Molecular Biology*, *41*, 577-585.

Jung, C. S., Griffiths, H. M., De Jong, D. M., Cheng, S., Bodis, M., Kim, T. S., & De Jong, W. S. (2009). The potato developer (D) locus encodes an R2R3 MYB transcription factor that regulates expression of multiple anthocyanin structural genes in tuber skin. *Theoretical and Applied Genetics*, *120*, 45-57.

Karageorgou, P., & Manetas, Y. (2006). The importance of being red when young: anthocyanins and the protection of young leaves of *Quercus coccifera* from insect herbivory and excess light. *Tree Physiology*, *26*, 613-621.

Kessel, R., & Rowe, P. R. (1974). Inheritance of two qualitative traits and a proposed genetic map for their linkage group in diploid potatoes. *Potato Research*, *17*, 283-295.

Kobayashi, S., Goto-Yamamoto, N., & Hirochika, H. (2004). Retrotransposon-induced mutations in grape skin color. *Science*, *304*, 982-982.

Lin-Wang, K., Bolitho, K., Grafton, K., Kortstee, A., Karunairetnam, S., McGhie, T. K., ... & Allan, A. C. (2010). An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae. *BMC Plant Biology*, *10*, 50-67.

Lisch, D. (2013). How important are transposons for plant evolution?. *Nature Reviews Genetics*, *14*, 49-61.

Løvdal, T., Olsen, K. M., Slimestad, R., Verheul, M., & Lillo, C. (2010). Synergetic effects of nitrogen depletion, temperature, and light on the content of phenolic compounds and gene expression in leaves of tomato. *Phytochemistry*, *71*, 605-613.

Mano, H., Ogasawara, F., Sato, K., Higo, H., & Minobe, Y. (2007). Isolation of a regulatory gene of anthocyanin biosynthesis in tuberous roots of purple-fleshed sweet potato. *Plant Pysiology*, *143*, 1252-1268.

Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., ... & Kimura, T. (2007). Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *Journal of Bioscience and Bioengineering*, *104*, 34-41.

Payyavula, R. S., Singh, R. K., & Navarre, D. A. (2013). Transcription factors, sucrose, and sucrose metabolic genes interact to regulate potato phenylpropanoid metabolism. *Journal of Experimental Botany*, *64*: 5115–5131.

Petroni, K., & Tonelli, C. (2011). Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Science*, *181*, 219-229.

Quattrocchio, F., Wing, J. F., Va, K., Mol, J. N., & Koes, R. (1998). Analysis of bHLH and MYB domain proteins: species-specific regulatory differences are caused by divergent evolution of target anthocyanin genes. *The Plant Journal*, *13*, 475-488.

Quattrocchio, F., Wing, J., van der Woude, K., Souer, E., de Vetten, N., Mol, J., & Koes, R. (1999). Molecular analysis of the anthocyanin2 gene of petunia and its role in the evolution of flower color. *The Plant Cell*, *11*, 1433-1444.

Quattrocchio, F., Verweij, W., Kroon, A., Spelt, C., Mol, J., & Koes, R. (2006). PH4 of petunia is an R2R3 MYB protein that activates vacuolar acidification through interactions with basic-helix-loop-helix transcription factors of the anthocyanin pathway. *The Plant Cell*, *18*, 1274-1291.

Rommens, C. M., Richael, C. M., Yan, H., Navarre, D. A., Ye, J., Krucker, M., & Swords, K. (2008). Engineered native pathways for high kaempferol and caffeoylquinate production in potato. *Plant Biotechnology Journal*, *6*, 870-886.

Sainz, M. B., Grotewold, E., & Chandler, V. L. (1997). Evidence for direct activation of an anthocyanin promoter by the maize C1 protein and comparison of DNA binding by related MYB domain proteins. *The Plant Cell*, *9*, 611-625.

Schaefer, H. M., & Rolshausen, G. (2006). Plants on red alert: do insects pay attention?. *BioEssays*, *28*, 65-71.

Schwinn, K., Venail, J., Shang, Y., Mackay, S., Alm, V., Butelli, E., ... & Martin, C. (2006). A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. *The Plant Cell*, *18*, 831-851.

Solfanelli, C., Poggi, A., Loreti, E., Alpi, A., & Perata, P. (2006). Sucrose-specific induction of the anthocyanin biosynthetic pathway in *Arabidopsis*. *Plant Physiology*, *140*, 637-646.

Stintzing, F. C., & Carle, R. (2004). Functional properties of anthocyanins and betalains in plants, food, and in human nutrition. *Trends in Food Science & Technology*, *15*, 19-38.

Stushnoff, C., Ducreux, L. J., Hancock, R. D., Hedley, P. E., Holm, D. G., McDougall, G. J., ... & Taylor, M. A. (2010). Flavonoid profiling and transcriptome analysis reveals new gene–metabolite correlations in tubers of *Solanum tuberosum* L. *Journal of Experimental Botany*, *61*, 1225–1238..

Tai, H. H., Goyer, C., & Murphy, A. M. (2013). Potato MYB and bHLH transcription factors associated with anthocyanin intensity and common scab resistance. *Botany*, *91*, 722-730.

Takos, A. M., Jaffé, F. W., Jacob, S. R., Bogs, J., Robinson, S. P., & Walker, A. R. (2006). Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. *Plant Physiology*, *142*, 1216-1232.

Tanaka, Y., Sasaki, N., & Ohmiya, A. (2008). Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *The Plant Journal*, *54*, 733-749.

Telias, A., Lin-Wang, K., Stevenson, D. E., Cooney, J. M., Hellens, R. P., Allan, A. C., ... & Bradeen, J. M. (2011). Apple skin patterning is associated with differential expression of MYB10. *BMC Plant Biology*, *11*, 93-106.

Van Oosten, M. J., Sharkhuu, A., Batelli, G., Bressan, R. A., & Maggio, A. (2013). The *Arabidopsis thaliana* mutant air1 implicates SOS3 in the regulation of anthocyanins under salt stress. *Plant Molecular Biology*, *83*, 405-415.

Walker, A. R., Lee, E., Bogs, J., McDavid, D. A., Thomas, M. R., & Robinson, S. P. (2007). White grapes arose through the mutation of two similar and adjacent regulatory genes. *The Plant Journal*, *49*, 772-785.

Wenke, T., Döbel, T., Sörensen, T. R., Junghans, H., Weisshaar, B., & Schmidt, T. (2011). Targeted identification of short interspersed nuclear element families shows their widespread existence and extreme heterogeneity in plant genomes. *The Plant Cell*, *23*, 3117-3128.

Whelan, J. A., Russell, N. B., & Whelan, M. A. (2003). A method for the absolute quantification of cDNA using real-time PCR. *Journal of Immunological Methods*, *278*, 261-269.

Zhang, Y., Cheng, S., De Jong, D., Griffiths, H., Halitschke, R., & De Jong, W. (2009). The potato R locus codes for dihydroflavonol 4-reductase. *Theoretical and Applied Genetics*, *119*, 931-937.

Zhang, Y., Jung, C. S., & De Jong, W. S. (2009). Genetic analysis of pigmented tuber flesh in potato. *Theoretical and Applied Genetics*, *119*, 143-150.

Zhang, B., Hu, Z., Zhang, Y., Li, Y., Zhou, S., & Chen, G. (2012). A putative functional MYB transcription factor induced by low temperature regulates anthocyanin biosynthesis in purple kale (*Brassica Oleracea* var. *acephala* f. tricolor). *Plant Cell Reports*, *31*, 281-289.

2.7 Supplementary data Chapter 2

Table S1. List of primers used in this study.

Figure S1. Single nucleotide substitutions and deletions in *AN1* exons with respect to reference alleles (*AN1-816* and *AN1-777*) present in GenBank database and *AN1* promoter analysis.

Figure S2. Phylogenetic analysis of MYB transcription factors which displayed similarity with the predicted protein of *AN1* cloned from "Magenta Love."

Figure S3. Genomic structure proposed for *StJAF13*; *JAF13* genomic sequence annotated on PGSC database; genomic sequence of *GLABRA3-like* of *S. lycopersicum*.

Figure S4. Agarose gel showing the products obtained from the amplification of exon 4 of JAF13.

Figure S5. Phylogenetic analysis of bHLH transcription factors that displayed similarity with the predicted protein of *StJAF13* cloned from "Magenta Love".

Figure S6. Phylogenetic analysis of bHLH transcription factors that displayed similarity with the predicted protein of *StbHLH1* cloned from "Blue Star".

Table S1. List of primer used in this study.

ACCESSION	Gene	Forward primer (5' to 3')	Reverse primer (5' to 3')	Used for
AY841128; AY841130	*AN1*	TGATGTGTACATCTTTGGGAGT	CCAAGCATGAATTTGATAACGA	
AY841128; AY841130		TGTAGACTGAGGTGGCTAAATTATCT	TTGAGGTCTTATTATGGCATTGG	*AN1* molecular analysis
AY841128; AY841130		GGAACACTAACCTTCTAAGGAAGC	TGGTGTTTTATCAATTTCATTGC	
AY841127; AY841129		ATATGCAATGGTGGGCAAAT	TCATATTACATACTTCACACTTGGAAT	
HG763862	*StJAF13*	CCAGATCAGCAAGACGATCA	GCTCTCGCTCCAAAGACAAC	qRT-PCR analysis
JX848660	*StbHLH1*	CCACCAAAGCCAGCTTTATC	ATCCGCTGGACAAATACCAG	
JX848659	*AN1*	AAGGAAAGGTTCATGGACTGA	TCCACTTCATCCCAATCAAAG	
HQ701728	*ANS*	GGGAAGTGGGTAACTGCAAA	TCTTCTCCTTTGGAGGCTCA	
AY289921	*DFR*	TGGACATGGGGTTTCAATTC	GCAACTGGTGCATTCTCCTT	
JX848659	*AN1*	ATGAGTACTCCTATGATGT (a)	CTAATTAAGTAGATTCCATATA (b,c)	Cloning in pDNOR207
HG763862	*StJAF13*	ATGCTATGGGACATCAAGATC (a)	TCAAGATTTCCAAACTACTCTC (b,c)	
HG763862	*StJAF13*Δ264	ATGGTGGTAGACGGTATTGGG (a)	TCAAGATTTCCAAACTACTCTC (b,c)	
JX848660	*StbHLH1*	ATGGAGATTATACAGCC (a)	TTAATTAGCTCTAGGGATTATC (b,c)	
JX848660	*StbHLH1*Δ58	ATGGAAGTTACTGCTGAAG (a)	TTAATTAGCTCTAGGG (b,c)	
JX848659	*an1*	AAAAGAATTCATGAGTACTCCTATGATG TG	ATATCTCGAGCTAATTAAGTAGATTCC	Yeast two hybrid
HG763862	*StJAF13*	ATTACCCGGGTATGGCTATGGGACATC	ATTTAGTCGACTCAAGATTTCCAAACT ACTC	
HG763861	*StJAF13*Δ264	ATAACCCGGGTATGGTGGTAGACGGTAT TGG	ATAAGTCGACTCAAGATTTCCAAACTA CTC	
JX848659	*StbHLH1*	ATTACCCGGGT ATGGAGATTATACAGCC	ATTTAGTCGACTTAATTAGCTCTAGGG	
JX848660	*StbHLH1*Δ58	CCCCGGGTATGGAAGTTACTGCTGAAG	ATTTAGTCGACTTAATTAGCTCTAGGG	
PGSC0003DMT400032139	*JAF13*	GATCTCACTGATACAGAGTGG	CAGTCTGGATAGATGCACTCT	*JAF13* exon 4 analysis
chr10:51745200..51749200	*AN1* promoter	CTGTTATTATCTGATAATG	GAAAGGGCCATATAATGTGC	*AN1* promoter analysis
		TGAGAAACATACCTGAAAACT	CAATCTGATGTGGTTGGATG	
		GTCACATCACTACACCACAT	GTAGTCCAAGGGAACCACCAC	

(a,b) AttB1 and attB2 sites were added respectively to forward and reverse primers ; (c) The stop codon was not included for C-terminal fusion as described in the Gateway Cloning Technology Instruction Manual.

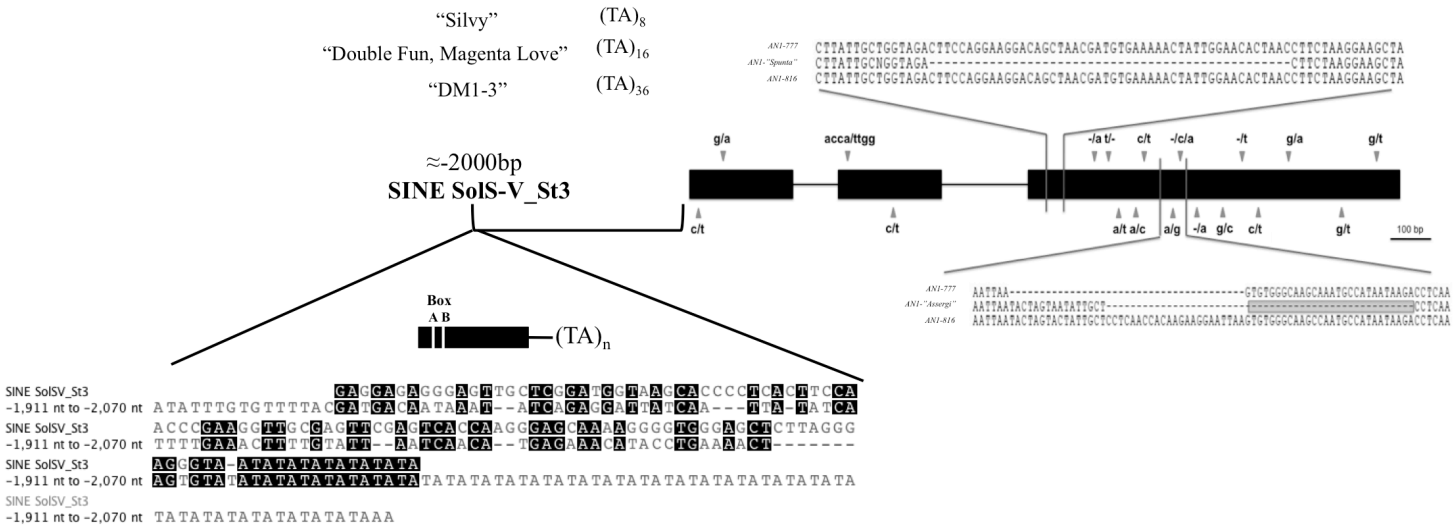

Figure S1. Single nucleotide substitutions and deletions in *AN1* exons with respect to reference alleles (*AN1-816* and *AN1-777*) present in GenBank database and *AN1* promoter analysis. On the right hand side: sequence variants found through the genomic analysis of 15 *S.tuberosum* varieties. The variants found are shown as compared to *AN1-777* and *AN1-816* consensus sequences and indicated first; the deletion in exon 3 (from nucleotide 670 to nucleotide 720 of *AN1* reference alleles) found in "Spunta" genotype; the rearrangements found from nucleotide 817 to nucleotide 884 of "Assergi". Grey box indicates 28 new nucleotides deleted with respect to reference *AN1* alleles. On the left hand side: simple motif repetition $(TA)_n$ in *AN1* promoter is the tail of a putative SINE (short interspersed element) transposon; different length in $(TA)_n$ motif between three genotypes, SINE-SolV_St3 and sequence annotated in PGSC database for *S.tuberosum* group Phureja DM1-3; general structure of a plant SINE; alignment of SINE-SolV_St3 with *AN1* promoter between nucletotides -2,077 and -1,895. Shading of the alignment was carried out with the Geneious software and indicates the percentage of similarity between residues: 100 % of similarity (black shading), 80 to 100 % of similarity (dark-grey shading), 60 to 80 % of similarity (light-grey shading), less than 60 % of similarity (white shading).

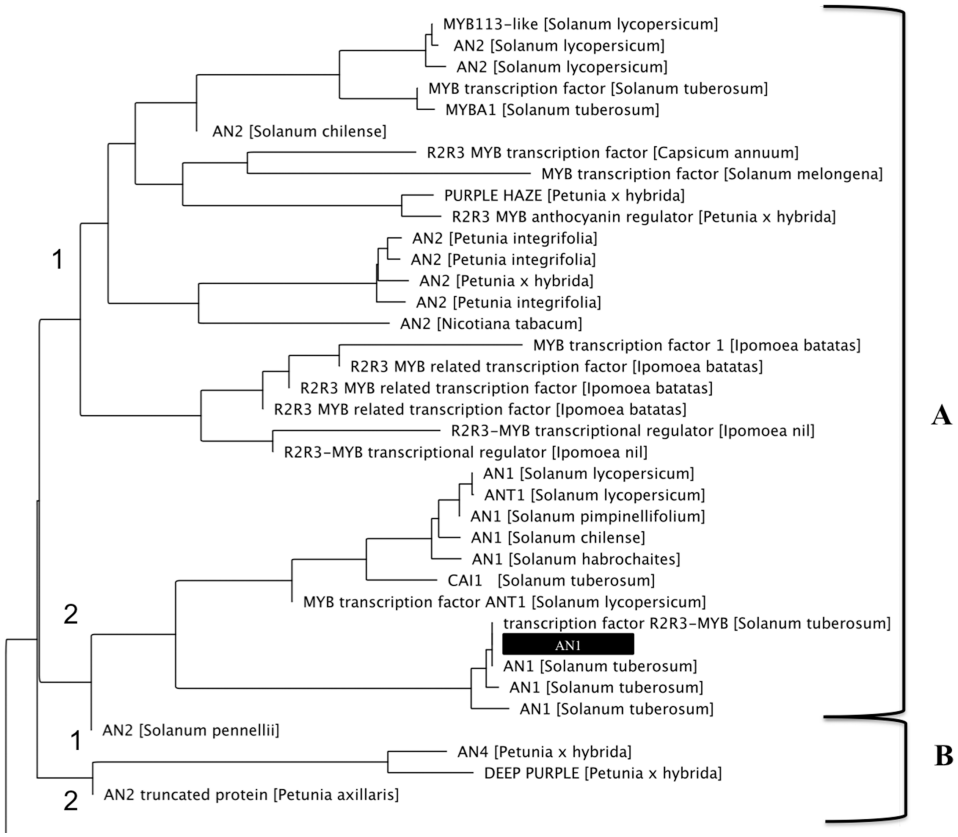

Figure S2. Phylogenetic analysis of MYB transcription factors that displayed similarity with the predicted protein of *AN1* cloned from "Magenta Love". The AN1 predicted protein of "Magenta Love" is indicated by a black box. Different clades and subclades are marked respectively by alphabetic letters and numbers.

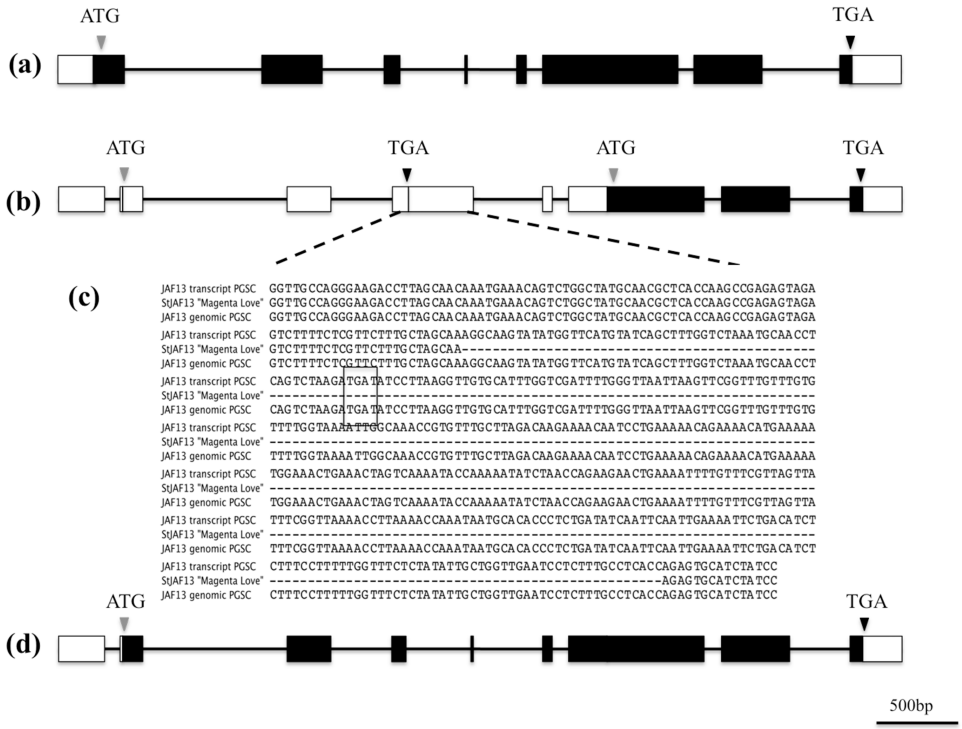

Figure S3. (a) The genomic structure proposed for *StJAF13*. (b) Genomic structure of the *JAF13* sequence annotated on PGSC database. (c) Particular of the alignment of the *StJAF13* coding sequence cloned from "Magenta Love" with the fourth exon of *JAF13* genomic and transcript sequence as annotated in the PGSC database; the premature stop codon on *JAF13* transcript sequence was marked by black box. (d) Genomic structure of the *GLABRA3-like* of *S. lycopersicum.* Exons and introns are drawn to scale. Exons are denoted by black boxes. Untranslated region (UTR) are denoted by white boxes. Upper case letters denote start and stop codons of different open reading frame (ORF).

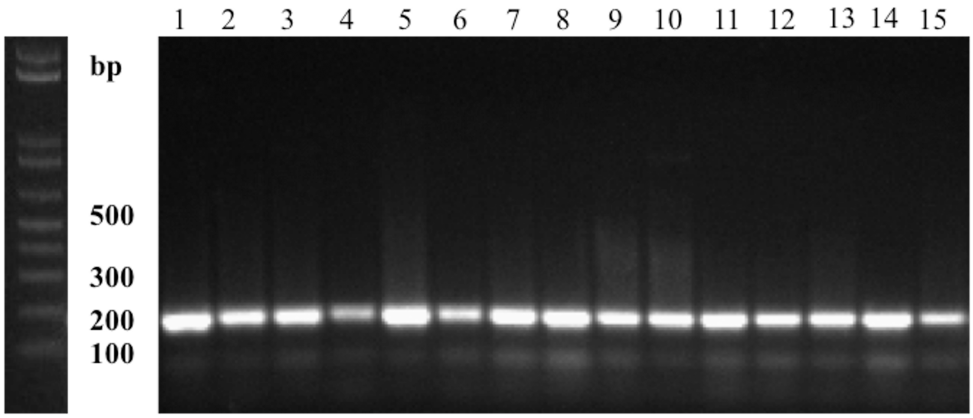

Figure S4. Agarose gel showing the products obtained from the amplification of exon 4 of JAF13. A single amplification product of 181 bp was obtained from cDNAs of the potato genotypes: (1) "Double Fun"; (2) "Blue Star"; (3) "COI25"; (4) "Désirée"; (5) "Magenta Love"; (6) "Pukara"; (7) "Spunta"; (8) "Violet Queen"; (9) "Ilaria"; (10) "Adora"; (11) "Briosa"; (12) "DEI23"; (13) "Silvy"; (14) "Blondy"; (15) "Flamenco".

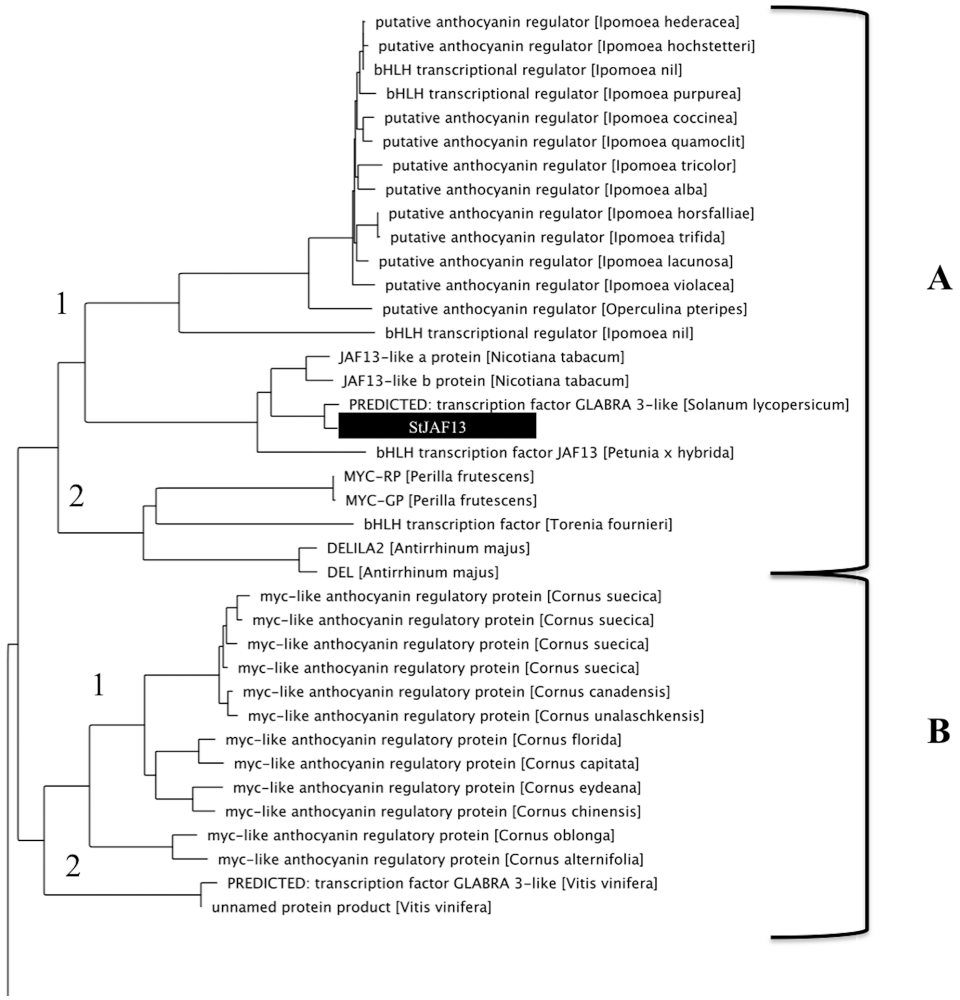

Figure S5. Phylogenetic analysis of bHLH transcription factors that displayed similarity with the predicted protein *StJAF13* cloned from "Magenta Love". The StJAF13 predicted protein of "Magenta Love" is indicated by a black box. Different clades and subclades are marked respectively by alphabetic letters and numbers.

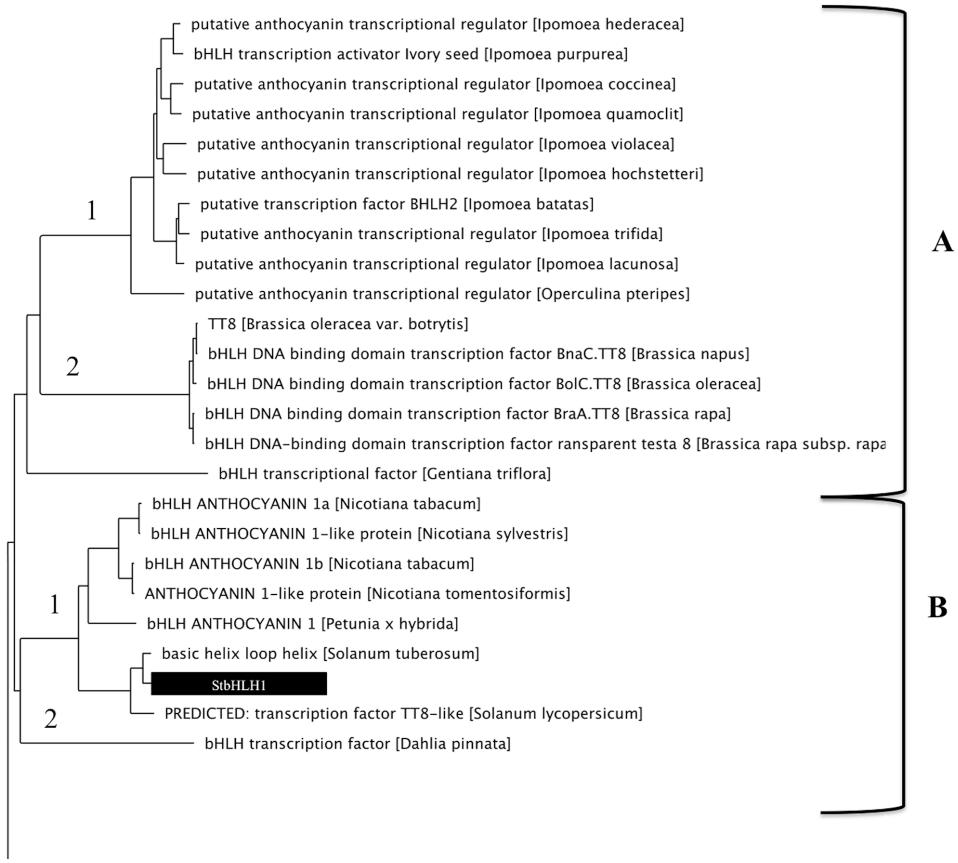

Figure S6. Phylogenetic analysis of bHLH transcription factors that displayed similarity with the predicted protein of *StbHLH1* cloned from "Blue Star". The StbHLH1 predicted protein of "Blue Star" is indicated by a black box. Different clades and subclades are marked respectively by alphabetic letters and numbers.

Chapter 3. An ancient MYB gene duplication generated a cold-induced regulator of anthocyanin biosynthesis in *Solanum commersonii*

Chapter 3. An ancient MYB gene duplication generated a cold-induced regulator of anthocyanin biosynthesis in *Solanum commersonii*

3.1 Introduction

Cold is a major abiotic stress and especially freezing events may severely reduce plant survival and crop yield worldwide. Therefore, improving tolerance to low temperatures is a major target for plant breeders. Several studies are being carried out to understand the mechanisms by which plants perceive cold stress signals and further how they translate the signals to the to cellular metabolism to activate adaptive responses (Mantri et al. 2012). Studies on the physiological and molecular mechanisms of tolerance to this environmental constrain have proven that it is based on the action of several genes and that it involves the reprogramming of both gene expression and protein translation (Bohnert et al., 2006). All these alterations usually flow in a severe metabolism change that allows plants to tackle the environmental stress. In particular, either the presence or the production at the right time and place of useful metabolites in response to an abiotic stress is often crucial for the plant to survive. Following cold stress, useful molecules as sugars and nitrogenous compounds (proline and glycine betaine) are usually synthetized because of their cryoprotectant proprieties (Ramakrishna & Ravishankar, 2011). Secondary metabolism is also highly affected by cold stress. Particularly important is the activation of the phenylpropanoid pathway. Several phenolic compounds are induced during the stress and the products are incorporated into the cell wall either as suberin or lignin (Ramakrishna & Ravishankar, 2011). Among them flavonoids, and in particular anthocyanins, are in general induced by low temperatures (Janská et al., 2010). Anthocyanins are fundamental pigments for plant physiological and reproductive processes. Gould, (2004) reported that anthocyanins can be regarded as "nature's Swiss army knife" due to their primary role in the tolerance to

stressors as diverse as drought, UV-B, and heavy metals, as well as in the defence against herbivores and pathogens. In this scenario anthocyanins can positively act both directly, as cell osmoregulators (Chalker-Scott, 2002), and indirectly enhancing the photosynthesis in cold environmental condition (Steyn et al., 2002). The cultivated potato (*Solanum tuberosum*) is sensitive to low temperatures and unexpected freezing events often cause production losses (Seppänen et al., 1998**)**. Breeding strategies are consequently aimed to introduce cold tolerance into the cultivated varieties, and a number of tuber-bearing *Solanum* species represents important genetic resource to improve this trait (Hijmans et al., 2003). *S. commersonii* has been reported to be the species with the highest resistance to low temperatures. It is also able to increase its tolerance upon exposure to low non-freezing temperatures (Aversano et al., 2015). In the last few years this species has attracted the attention of several research groups working in the field of proteomics (Folgado et al., 2013), transcriptomics (Carvallo et al., 2011) and breeding (Carputo, 2003). It has also been the first potato species whose genome sequence has been released (Aversano et al., 2015). However, the mechanism that allows *S. commersonii* to withstand cold exposure remains unclear. Estrada et al (1982) reported that an interesting phenotype of this and other cold tolerant potato species is the presence of anthocyanin in leaves. No study focused its attention on understanding anthocyanin involvement in *S. commersonii* tolerance also because no clues were available on the genetic control of these metabolites in the foliage of this species. In *S. tuberosum* the typical ternary regulation complex MYB/bHLH/WD40 that controls anthocyanin genes is being clarified. In particular, *StAN1* (a key gene of the complex that codify for R2R3 MYB) seems to be involved in tuber skin and flesh pigmentation (Jung et al., 2009; Payyavula et al., 2013). We have recently reported that in *S. tuberosum AN1* displays intraspecific sequence variability in both coding/non-coding regions and in the promoter, and that its expression is associated with high anthocyanin content in leaves (D'Amelia et al., 2014). Jung et al. (2009) found a *StAN1* homolog named *StAN2*. In spite of the high similarity between the two genes, there was no direct evidence of correlated *StAN2* and anthocyanin production in potato tissues (Jung

et al., 2009; André et al., 2009, Payyavula et al., 2013). Interestingly, Andrè et al (2009) reported that in potato tuber *StAN2* is highly expressed during drought stress and that its expression is connected with anthocyanin increase. They also hypothesized that *StAN2* may control the transcription of other genes in addition to those of the anthocyanin pathway.

Based on the high resistance and anthocyanin levels of *S. commersonii*, as well as on our recent results on anthocyanin regulation in potato reported in Chapter 2, the purpose of this research was to decipher anthocyanin regulation in *S. commersonii* and the connection between anthocyanin genes and stress response in this noteworthy potato species. Our functional and protein/protein interaction analysis provided evidence that an ancestral duplication generated the two MYB genes *ScAN1* and *ScAN2,* both able to induce the anthocyanin production. Further, we found that *ScAN2* was induced by cold stress only in the cold-tolerant *S. commersonii*. A key aspect that emerged from functional analysis was the divergent phenotype of *ScAN2,* suggesting that this MYB may activate additional metabolic pathway. Based on these results it is proposed that the duplication that originated *ScAN1* and *ScAN2* had as outcome the subfunctionalization of the two genes, with *ScAN1* selected mainly to fulfil anthocyanin production, while *ScAN2* conserved also the ability to regulate additional phenolic compounds in response to cold stress.

3.2 Materials and methods

3.2.1 Plant materials and growth conditions

Plants used in this study were the clone CMM1T of wild potato species *S. commersonii* (PI243503) and the commercial potatoes: "Blondy", "Désirée" and "Double Fun". The last already proved to have high anthocyanin content in leaf (D'Amelia et al., 2014). Plants were micro propagated *in vitro* on Murashige and Skoog (MS) medium (Sigma-Aldrich) with 1% (w/v) sucrose and 0.8% (w/v) agar, and incubated at 24 °C. Plants were hereafter transplanted to styrofoam trays filled with sterile soil and maintained in a growth chamber at 25°C ± 2 under a 16/8 h (light/dark) photoperiod at 220 μmol m^{-2} s^{-1} irradiance provided by a cool, white-fluorescent tube (Philips). Two different sets of each genotype were used for the freezing-tolerance assay and cold stress analysis respectively.

3.2.2 Freezing-tolerance assay

Full-expanded leaves were collected from three plants of each potato genotypes to test tolerance to cold as previously reported by Carputo et al. (2003). Total leachable solutes were determined after autoclaving for 15 min at 121 °C. The freezing temperatures causing 50% loss of total leachate from three independent experiments were averaged and the resulting mean was designated as the killing temperature (LT50).

3.2.3 Cold stress

Cold stress was performed as described by Evers et al. (2012) in three biological replicates. After two months of growth, one set of four plants was transferred to 7°C/2°C day/night with 12 h photoperiod at 200 μmol m^{-2} s^{-1} irradiance. A second set of plants was used as control and maintained in a growth chamber at conditions described before. Leaf samples were collected after 3 days of cold treatment and from the respective controls. Samples were immediately frozen in liquid nitrogen and stocked at -80°C.

3.2.4 RNA extraction and anthocyanin genes expression analyses

Total RNA was extracted from leaves of cold stressed plants and controls. RNA extraction was performed using the Spectrum Plant Total RNA Kit (Sigma-Aldrich) and following the manufacturer's guidelines. Genomic DNA contamination was eliminated using On-Column DNase I Digestion Set (Sigma-Aldrich) during RNA extraction. One µg of total RNA was reverse transcribed cDNA using oligo-dT(20) primer and SuperScript III reverse transcriptase (Invitrogen) in 20 µL of final reaction according to the manufacturer's instructions. The expression change of anthocyanin structural (*Sc/StDFR*, *Sc/StANS*) and regulatory genes (*Sc/StAN1*, *Sc/StAN2*, *Sc/StbHLH1* and *Sc/StJAF13*) was monitored using Real-Time RT-PCR. The experiment was carried out in biological triplicates using the 2x QuantiFast SYBR Green PCR Master Mix (Qiagen) and the ABI PRISM 7900HT Instrument (Life Science). Each 15-µl reaction comprised 300 nM of each primer and cDNA synthesized from 1 µg of total RNA (three replicates for each reaction). Cycle conditions indicated by QuantiFast SYBR Green PCR Kit handbook (Qiagen) were used. *Aprt* was used as endogenous control. Primer pairs used were listed in Supplementary data Table S1. A melting curve analysis of the PCR products was produced to verify their specificity and identity. In particular, different melting curves were observed for the two similar genes *Sc/StAN1* and *Sc/StAN2*. Results were then analysed using the ABI PRISM 7900HT Sequence Detection System Version 2.1 (SDS 2.1). The relative expression was estimated according to the ΔΔCt (Livak & Schmittgen, 2001).

3.2.5 Metabolite analysis

Total anthocyanin content was estimated using the pH-differential spectrum method as described by Zhang *et al*. (2012) and starting from fresh and freeze dried material for potato and transformed tobacco, respectively. On the same samples, UPLC-IT-ToF-MS was carried out to characterize phenolic compounds. Freeze-dried leaves (50 mg) were extracted with 2 of 80% MeOH on ice by

shaking overnight. Extracts were centrifuged (4000 rpm) at 4°C twice. Aliquots of the original sample were then diluted 6 times with 80% MeOH. The samples were analysed on a Nexera UPLC system attached to an ion-trap ToF mass spectrometer (Shimadzu). Separation was on a 100×2.1mm 2.6µ Kinetex XB-C18 column (Phenomenex) using the following gradient of acetonitrile versus 0.1% formic acid in water, run at 0.5mL.min^{-1} and 40ºC: 0.01 min, 2%; 11 min, 30%; 15 min., 90%; 16 min., 2%; 20 min., 2%. Phenolics were detected by collected UV/visible spectra from 200-600nm at 6.25Hz (0.16 sec time constant), and positive electrospray MS from *m/z* 200-2000 with a maximum ion-accumulation time of 20msec (automatic sensitivity control target of 70% optimal base-peak intensity). It also collected automatic (data-dependent) MS2 of the most abundant precursor ions, at an isolation width of *m/z* 3.0, 50% collision energy, and 50% collision gas. Spray chamber conditions were 250ºC curved desorbation line, 300ºC heat-block, 1.5L.min^{-1} nebulizer gas, and drying gas "on". The instrument was calibrated using sodium trifluoroacetate according to the manufacturer's instructions immediately prior to use. Identification of detected compounds was based on retention time, accurate masses of fragment ions, in combination with PDA absorbance spectra (recorded at 240–600 nm). Compound identification, peak comparisons and statistical analysis were carried out using Profiling Solution software (Shimadzu Corporation).

3.2.6 Microscopy analysis

Anthocyanin localization was studied using transverse hand-sections of the fresh material. Sections were mounted in 10% glycerol, and the histological location of red pigment analysed under bright field microscopy using Microscope Leica M205 FA. Transformed tobacco leaves were analysed under with a Leica SP5II confocal microscope equipped with a Leica HCX APO L U-V-I 63.0 x 0.90 UV water immersion objective. Fluorescence was detected by HyD3 in the range of 570 - 640 nm after excitation with the 488 nm line of an argon ion laser and at 561 nm with a diode-pumped solid-state laser. Fluorescent was recorded line by line with three- to six-fold averaging, depending on the background noise. The

LAS Advanced Fluorescence software was used for image acquisition and Fiji/ImageJ (version 1.50e) for image post-processing.

3.2.7 Yeast Two Hybrid analysis

The full-length coding sequences of ScAN1, ScAN2 and StAN2 were amplified using PCR and cloned in frame into pGADT7, prey plasmid, (Clontech, Mountain View, CA, USA) between EcoRI and XhoI restriction sites. ScJAF13 and ScbHLH1 were inserted into pGBKT7, bait plasmid, (Clontech) using SmaI and SalI. The bait plasmid StbHLH1pGBKT7 and StJaf13pGBKT7 were previously described (D'Amelia et al., 2014). All plasmids were sequenced to ensure that no mutations had been introduced. The bait and prey plasmids were co-transformed into yeast strain AH109 (Clontech, Mountain View, CA, USA) using the Lithium acetate/Polyethylene glycol method (Bai & Elledge, 1997). The self-activation test was performed prior to the test combinations of interest (Docimo et al., 2015). After verifying that the bait and prey plasmids when transformed alone conferred ability to grow on tryptophan or leucine, respectively, indicating presence of the plasmid, but not on media lacking three amino acids, which would have indicated self-activation, co-transformations to verify interactions were performed. Transformed colonies containing bait and prey plasmids were selected on synthetic drop-out medium lacking tryptophan and leucine (–W/–L). Co-transformants were grown overnight in liquid culture lacking tryptophan and leucine (–W/–L). For the interaction between bait and prey, an equal amount of cells was spotted on medium lacking tryptophan, leucine, histidine and adenine (–W/–L/–H/–A). Negative controls were also performed.

3.2.8 Overexpression of potato transcription factors in tobacco plants

The *Sc/StAN1* and *Sc/StAN2* coding sequences were cloned in the 35SCaMV expression cassette of pGWB411 (Nakagawa *et al*., 2007) using Gateway recombination technology (Invitrogen). *A. tumefaciens* strain GV3101 pmp 90, transformed with each expression vector, was used for agroinfiltration in fully expanded leaves of *N. benthamiana* as reported by Payyavula *et al*. (2013). Total

anthocyanin content was estimated using the pH-differential spectrum method (Zhang *et al*., 2012) using freeze dried materials.

3.2.9 Bioinformatics and statistical analysis

Amino acid and nucleotide sequences were collected from the National Center for Biotechnology Information (NCBI) and Potato Genome Sequencing Consortium (PGSC) database. Alignments and phylogenetic trees were performed with Geneious software v6.0.6 (Biomatters, http://www.geneious.com/). BLASTP and BLASTX programs (http://www.ncbi.nlm.nih.gov/blast) were used to perform homology researches in GenBank. Analysis of variance (ANOVA) was carried out using XLSTAT-Pro 7.5.3 software (Addinsoft, http://www.xlstat.com). Duncan's test was performed to compare mean values. Compound identification, peak comparisons and statistical analysis were carried out using Profiling Solution software (Shimadzu Corporation). Profiling Solution provided P-value. T-test Benjamini and Hochberg method (Benjamini & Hochberg, 1995) were used for peak comparisons. Promoter sequences of regulatory gene *ScAN1*/*ScAN2* and their respective potato orthologs were obtained using *S. commersonii* genome database (scaffold-4043) and PGSC database, respectively. The 1500 bps upstream the transcription start for each genes were analysed using the online software PlantCare (Lescot et al., 2002).

3.3 Results

3.3.1 Chilling temperatures increase the anthocyanin content in *S. commersonii* leaves

Figure 1 a reports the killing temperature (LT_{50}) of genotypes studied. CMM1T confirmed its ability to tolerate freezing condition, with a LT_{50} of -4.4 °C. Cultivated varieties "Blondy", "Double Fun" and "Désireé" were all less tolerant, with a LT_{50} > -3 °C. To investigate whether cold stress influences anthocyanin contents in potato leaves, we measured the difference (Δ_{AC}) in total anthocyanins between cold stressed genotypes vs. their respective controls (Figure 1b). CMM1T showed an increase in pigment amount of 1.8 mg g^{-1}. As far as *S. tuberosum* varieties are concerned, the increase in anthocyanin level was significantly lower than in *S. commersonii,* with red-leaf "Double Fun" showing the highest accumulation (about 0.5 mg g^{-1}). Since CMM1T and "Double Fun" were the only two genotypes with a significant presence of anthocyanins in leaves also in control conditions, pigment localization was further studied (Figure 2 a, b). CMM1T was pigmented for the entire lamina of the abaxial face and histological analysis showed that anthocyanin resided in epidermal cell layer (2 a). By contrast, "Double Fun" showed the presence of anthocyanin into adaxial face with a strong pigmentation in the palisade mesophyll area (Figure 2 b). The leaf anthocyanin profile of "Double Fun" revealed a wider composition of anthocyanins with respect to that of CMM1T. As shown by LC-MS-IT-ToF data, in the former genotype we identified di and tri-glycosylated forms of pelargonidin, malvidin, petunidin and peonidin with different cinnamic acid acylation. In CMM1T, the anthocyanin profile was reduced to the highest methylated anthocyanin malvidin. As in "Double Fun" malvidin was acylated with coumaryl and feruloyl groups (Supplementary data Table S2).

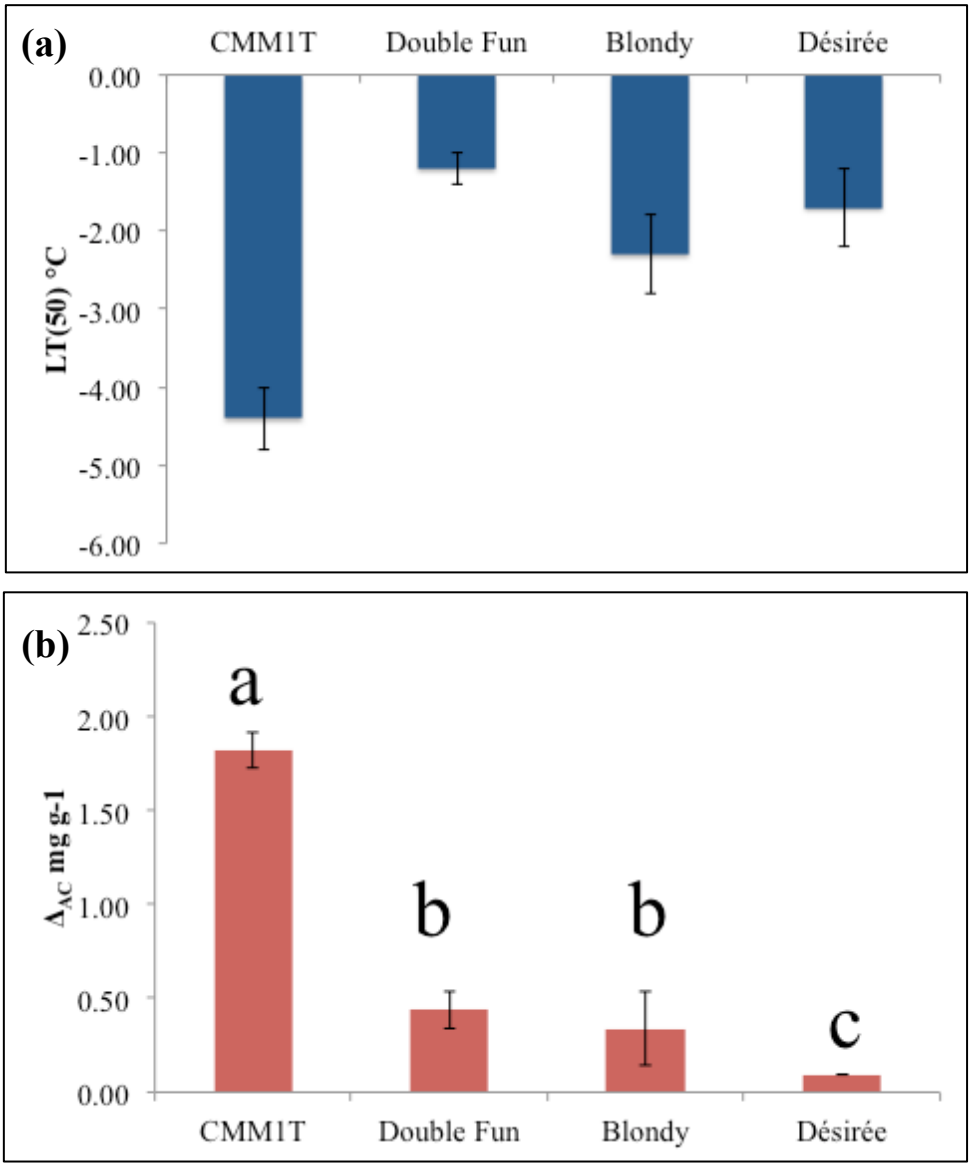

Figure 1. a) Killing temperature (LT50) in *S. commersonii* (CMM1T) and potato cultivated varieties following freezing stress. b) Difference in anthocyanin concentration (Δ_{AC}) between cold stressed genotypes vs. their respective controls. Each value represents the mean of three determinations (± SD). Means denoted by the same letter did not differ significantly at $P \leq 0.05$ according to Duncan's multiple range test.

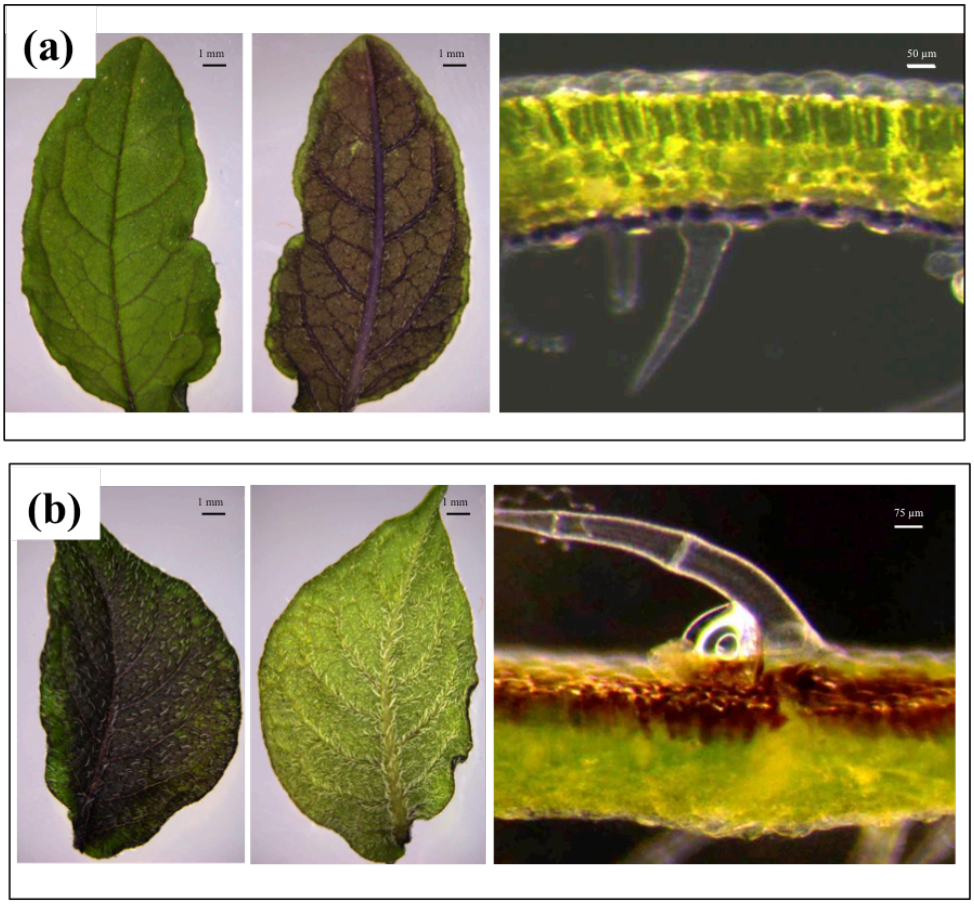

Figure 2. Leaf anthocyanin localization in *S. commersonii* CMM1T (a) and *S. tuberosum* "Double Fun" (b): from left to right: stereomicroscopic pictures of adaxial and abaxial face; freehand cross-sections of leaves, showing the different anthocyanin localization in the leaf of *S. commersonii* and *S. tuberosum*.

3.3.2 *ScAN2* is cold stress induced only in *S. commersonii*

To identify genes responsible for the cold-stress induced accumulation of anthocyanins we monitored the variation in the expression of genes related to the last steps of the biosynthetic pathway (Figure 3). In CMM1T, consistently with the anthocyanin increase, cold triggered the overexpression of the anthocyanin-related genes *DFR* and *ANS*, as well as those of the bHLH and MYB genes. Surprisingly, *ScAN1* expression was 17-fold lower than that of *ScAN2* (0.5 vs. 9

Fold Change, FC). In the cultivated genotypes a similar up-regulation was not observed. Indeed, in "Double Fun" a down-regulation of about – 3 FC was observed for *StAN2* and *StJAF13*, while a very weak expression variation was detected for the other genes. "Blondy" and "Désirée" shared a similar expression pattern, where *StAN1* was down-regulated of – 3 FC and the structural genes were moderately over-expressed with an increase of less than three fold.

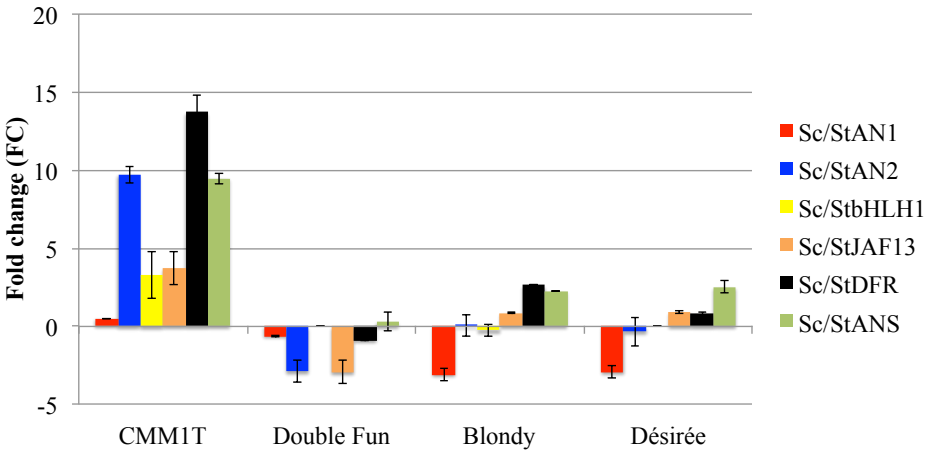

Figure 3. Quantitative real-time PCR analysis of the regulatory and structural anthocyanin genes. Bars represent the relative expression level monitored in leaves of each genotype after cold stress and in comparison to the respective controls.

3.3.3 *ScAN2* and *ScAN1* are paralog genes

To elucidate the genic structural relationship between *ScAN1* and *ScAN2*, their sequences were *in silico* mapped on the *S. commersonii* genome (Aversano et al., 2015; Figure 4 a). Both genes were present on the scaffold4043 and were 40.5 kb apart. Notably, 970 bp downstream of *ScAN1* sequence we found a duplicated region including a paralog of this gene (hereafter named *ScAN1-like*) and part of its 5' and 3' flanking regions (Figure 4 a). The duplicated portion of about 2.170 bp was present in tandem and shared a similarity of 99.3% with that harbouring *ScAN1*. About 9 kb upstream *ScAN2* we identified a sequence sharing the same structure of *ScAN1* and *ScAN2,* consisting of three exons and two introns (Figure 4 a). We named this sequence *ScMTF1*, as sequence-homology search in GenBank database revealed extensive similarity (95%) of its coding sequence (CDS) with the R2R3-MYB transcription factor *StMTF1*, annotated as inducer of chlorogenic acid and involved in the anthocyanin pathway (Rommens et al., 2008). A putative duplication event for *ScMTF1* was also observed. Indeed, a portion of 1.557 bp showed 85.5% of similarity with a region placed 6 kb upstream of *ScMTF1*. The paralog sequence (hereafter named *ScMTF1R2)* corresponds to the *ScMTF1* promoter and the its first exon. Overall, *ScMTF1R2* showed two exons and a coding sequence of only 126 bp. Consequently, *ScMTF1R2* was a truncated version of *ScMTF1*, sharing the 82.1% of identity for the domain R2. We also annotated different transposable elements (belonging to either LTR or NON-LTR classes) in between *ScMTF1*, *ScAN2* and *ScAN1*. All MYB genes analysed presented the same transcriptional orientation. Their CDSs showed an identity ranging from 60 to 71%, with the highest value found between *ScAN2* and *ScAN1,* especially when only the R2R3 nucleotide domains were aligned (83.3%). We did not find any homology when we aligned about 1500 bp of either the 5' or 3' flanking regions of *ScAN1* with that of *ScAN2*. The same happened when we tried with the *S. tuberosum* orthologs. We sought annotated motifs involved in stress response in 1500 bp promoter region of *Sc/StAN1* and *Sc/StAN2*. No remarkable differences were identified between the wild and

cultivated orthologs while we came across them within the same species. TC-rich repeats and an ABA-responsive element (ABRE), known *cis*-acting regulatory elements of osmotic, cold and general stress-responsive promoters (Yamaguchi-Shinozaki & Shinozaki, 2005; Sharma et al., 2014), were indentified for both genes. However, for *Sc/StAN2* they were closer to the ATG, in particularly at -179 and -110 bp for TC-rich repeat and ABRE, respectively (data not shown). Sintheny analysis with the ortholog region of the *S. tuberosum* Group Phureja (clone DM1-3 516 R44) reference genome revealed interesting structural differences. In particular, the duplication of the whole *StAN1* sequence was missing. Only the last 121 nucleotides of the third exon and a region of 1.9 kb at 3', were duplicated downstream *StAN1* (Figure 4 b). The presence of putative SINE elements and residues of TA-tail suggested that this region was a target of transposable elements insertions. *StMTF1* and *StMTF1R2* were not identified in the 30 kb region of the chromosome 10.

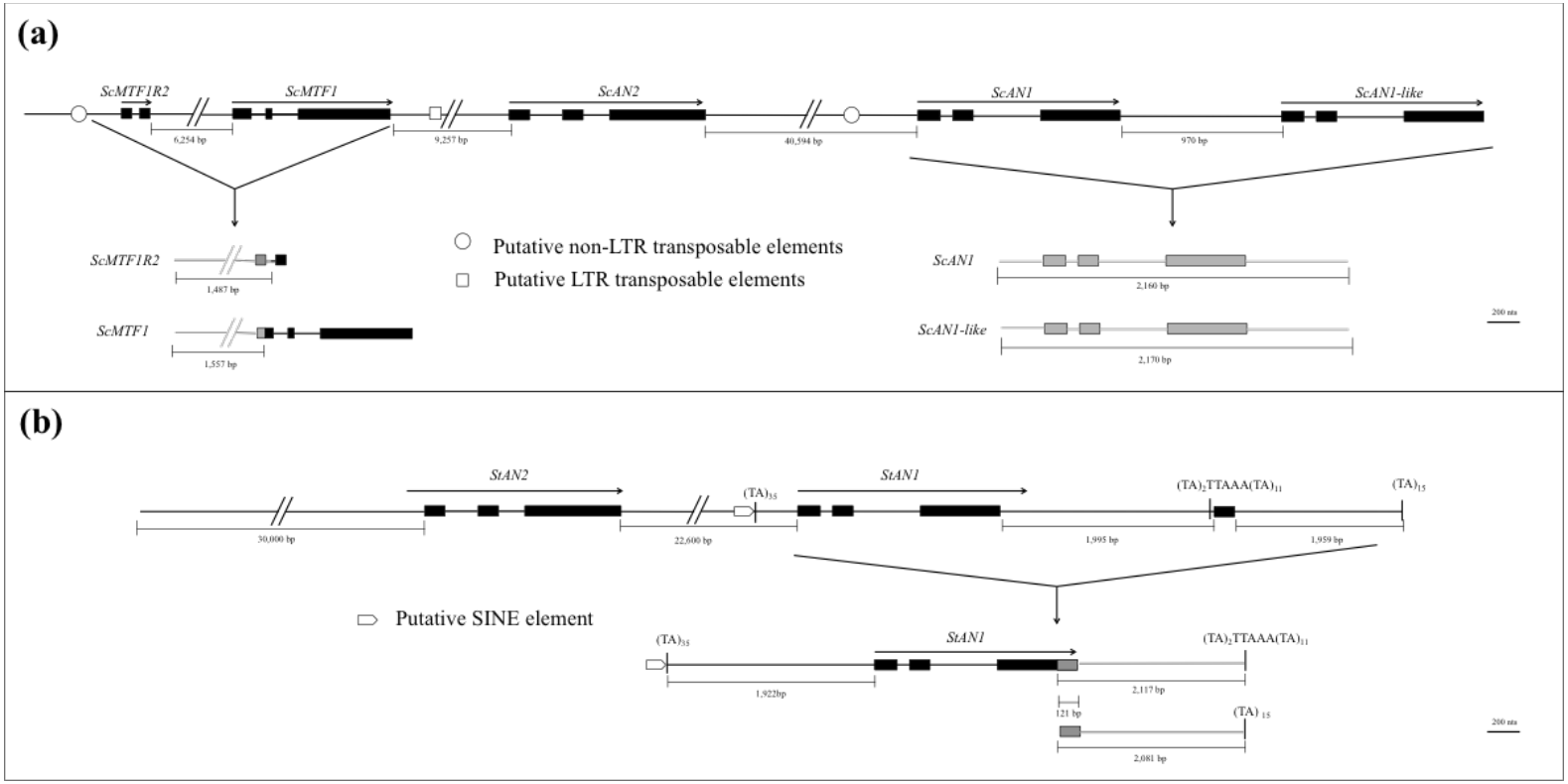

Figure 4. Diagrams of the genomic loci where *AN1* and *AN2* are located in *S. commersonii* CMM1T (a) and *S. tuberosum* DM1-3 516 R44 (b). Boxes represent exons, and the thin lines represent introns or intergenic regions. Arrows indicates secondary diagrams representing genomic regions where high nucleotide identity was analysed. In the secondary diagrams the grey boxes and lines represent region with high identity in exons, introns or intergenic region, respectively. Different putative transposable elements are also indicated.

3.3.4 Sc/StAN1 and Sc/StAN2 share interaction with the same bHLH co-partners

Yeast Two Hybrid system was performed to investigate if StAN1/StAN2 and the newly identified CMM1T orthologs share the same network of interaction. We fused in frame the entire coding sequence of the MYBs and bHLHs transcription factors with the AD or the BD of GAL4, respectively. Results showed that StAN2 interacts with StJAF13 and StbHLH1, previously shown to interact with StAN1 (Figure 5, D'Amelia et al., 2014). In fact, when yeast was co-transformed with either StAN2 and StJAF13 or StAN2 and StbHLH1, it was able to grow on selective media lacking tryptophan, leucine, histidine and adenine (–W/–L/–H/–A). The wild orthologs ScAN1 and ScAN2 also shared the same interaction ability. However, they were able to interact with ScbHLH1 only (Figure 5). In fact, yeast did not grow on the same selective media when transformed with either ScAN1/ScJAF13 or ScAN2/ScJAF13 combinations (data not shown). These results provided evidence that, similar to previous results obtained with StAN1, StAN2 interacts with StbHLH1 and StJAF13, while the MYBs cloned from *S. commersonii* were capable of interacting with ScbHLH1.

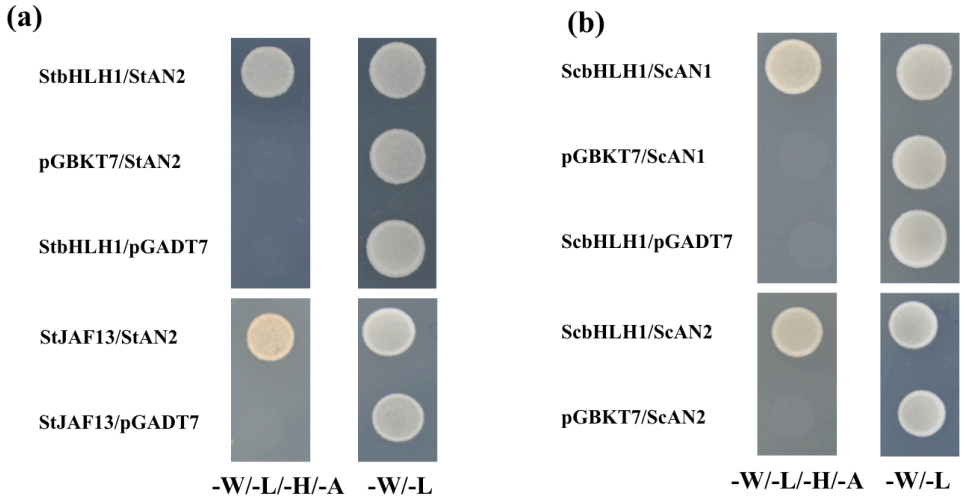

Figure 5. ScAN1 and Sc/StAN2 interact with bHLH co-partners in the yeast two-hybrid assay**.** Interaction of StbHLH1 and StJAF13 with StAN2 (A) and of ScBHLH1 with ScAN1 and ScAN2 (B). StAN2, ScAN1 and ScAN2 in pGADT7

were co-transformed with StbHLH1, StJAF13pGBKT7 or ScbHLH1 in yeast strain AH109. The *ScAN1*pGADT7/pGBKT7, *Sc/StAN2*pGADT7/pGBKT7, Sc/*StbHLH1*pGBKT7/pGADT7 and StJAF13pGBKT7/pGADT7 combinations served as negative controls. Yeast cells grown on synthetic media (–W/–L, Right) and on selective media (–W/–L/–H/–A, Left) are shown. Pictures were taken 3 days after incubation at 30◦C.

3.3.5 *ScAN2* is an anthocyanin inducer with a phenotypic pleiotropic effect

In silico and expression analysis suggested that *ScAN1* and *ScAN2* originated by the duplication of a common ancestor and that *Sc/StAN2* is potentially an anthocyanin inducer. Phylogenetic analysis showed that ScAN1 and ScAN2 protein sequences clusterized with anthocyanin promoting MYBs, presenting the closest homolgy with the Petunia AN2 branch (Supplementary data Figure S1). Nevertheless, the alignment between ScAN1 and ScAN2 a similarity of only 56.5% (Supplementary data Figure S2). It also indicated that AN2 protein sequence is highly conserved between *S. tuberosum* and *S. commersonii* In fact, only a single substitution between the two polar aminoacid asparagine (N) and serine (S) was observed in the last four amino acids (Supplementary data Figure S2). To test whether the difference in ScAN1 and ScAN2 protein sequences resulted in functional divergence, transient expression tests were performed in *N. benthamiana*. Overexpression of *ScAN1* induced anthocyanin accumulation in the whole agroinfiltrated area, confirming it is the *S. commersonii* ortholog of the potato *StAN1* (Figure 6 a). By contrast, measurement of anthocyanin level in leaves of *N. benthamiana* provided evidence that although *ScAN2* overexpression induced anthocyanin accumulation, its level was lower compared to *Sc/StAN1* (Figure 6 b). For all three constructs (*ScAN1*, *StAN1*, *ScAN2*) we found a big peak with UV/Vis-absorbance at 525 nm as compared to control (Supplementary data Figure S3). This peak displayed a monoisotopic mass of [M+H] = 611.158 m/z that correspond to Delphinidin-3-glucorhamnoside and, most likely to delphinidin 3-O-rutinoside (D3R): M = $C_{27}H_{31}O_{16}^{(+)}$ (Supplementary data Figure S3). In fact, D3R is the main anthocyanin produced by *N. benthamiana* (Outchkourov et al., 2014). Consistently with total anthocyanin quantification, chromatogram of

ScAN2 showed a very little anthocyanin increase with respect to *StAN1* and *ScAN1* (Supplementary data Figure S3).

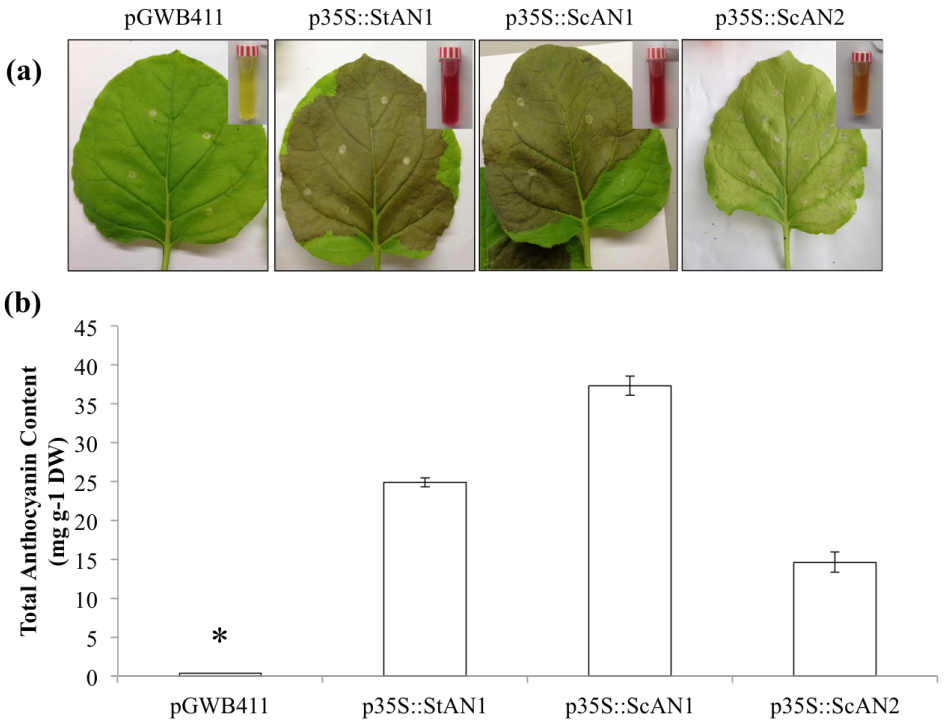

Figure 6. Transient overexpression assays in *N. benthamiana*. (a) Phenotypes obtained with of *StAN1*, Sc*AN1*, *ScAN1* and *ScAN2* overexpression. (b) Total anthocyanin content (mg g^{-1} FW) in transformed tobacco plants. Each value represents the mean of three determinations (`SD). * no detectable level of anthocyanins**.**

The expression of *ScAN2* under the constitutive promoter also led to a very divergent phenotype with respect to *ScAN1*, i.e. the suberification of the epidermal tissue that produced a paler colour of the leaf (Figure 6). The empty vector did not induce any divergent phenotype. The *ScAN2* phenotype was also obtained when we overexpressed the chimeric protein holding the Binding Domain (BD) of *ScAN2* (corresponding to the domain R2R3) fused with the

Activation Domain (AD) of *ScAN1* (corresponding to the C-terminal part of *ScAN1*, Figure 7 a, b). Conversely, when we used the BD of *ScAN1* we obtained a phenotype comparable to that of the entire *ScAN1* (Figure 7 a, b). This indicated that the divergent phenotype of *ScAN2* was caused by R2R3 domain and, consequently, by its binding ability. The alignment of the R2R3 domains (which hold the BDs) of ScAN2, ScAN1 with the snapdragon ROSEA1 suggested that the BD between ScAN2 (or StAN2) and ScAN1 (or StAN1) is almost conserved (Figure 8). In fact, ScAN1 and ScAN2 share 87.5% of identity for this region. In addition, in the R3 domain ScAN2 conserved the specific residued involved in the interaction with bHLHs. In the allignment we also added the snapdragon transcription factor ROSEA1. In fact, ROSEA1 was already tested to induce anthocyanin accumulation when overpexressed in *N. benthiamina* leaves (Outchkourov et al., 2014). We found few substitutions which may explain the divergent activity of the ScAN2 (or StAN2) BD. Within the domain R2R3, three interesting substitution were identified. They were chemically different and also specific for the BD of Sc/StAN2. In the R2 domain, the negative electrically charged D/E (position 31) and the non polar A (position 45) were substitued in ScAN2 with the polar amino acids Q and T, respectively. These substitutions were present in a region known to be involed in DNA binding. Within the R3 domain, the aromatic amino acid F in ScAN2 substituted in position 113 V and L of ROSEA1 and ScAN1, respectively (Figure 8).

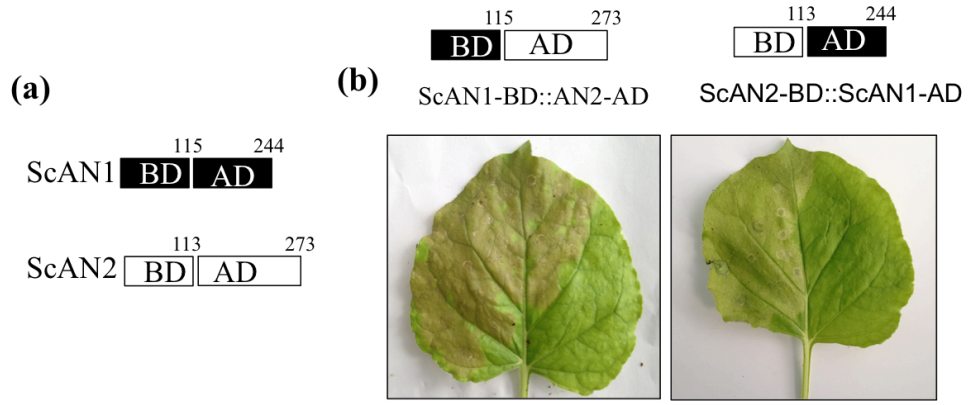

Figure 7. Transient overexpression assays in *N. benthamiana*. (a) Diagrams of the binding domain (BD) and activation domain (AD) of ScAN1 and ScAN2 in black and white, respectively. (b) Phenotypes obtained with overexpression of chimeric protein obtained switching the BD of ScAN2 with that of ScAN1 and vice versa (ScAN1-BD::AN2-AD and ScAN2-BD::ScAN1-AD) as indicated by diagrams above the figures.

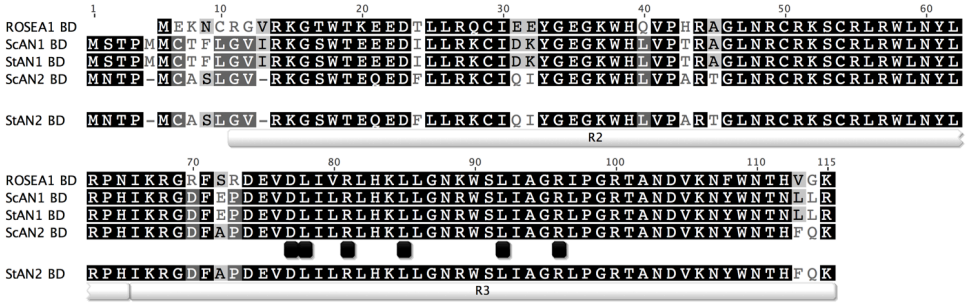

Figure 8. Alignment of the predicted binding domain (BD) of the MYB TF ROSEA (*Anthirrinum majus*), ScAN1, StAN1, ScAN2, StAN1. Shading of the alignment was carried out with the Geneious software and indicates the percentage of similarity between residues: 100% of similarity (black shading), 80–100% of similarity (dark-grey shading), 60–80% of similarity (light-grey shading), less than 60% of similarity (white shading). Black boxes indicate amino acid residues involved in MYB interaction with bHLH in Arabidopsis.

3.3.6 *ScAN2* induces accumulation of other phenolic compounds in the cell wall of transformed cells

The transient overexpression of *ScAN2* induced a lower D3R accumulation than *ScAN1*. A different visible phenotype was also produced. Further, we observed that the divergent phenotype of *ScAN2* with respect to *ScAN1* likely resided in the differences found in nucleotide sequence of R2R3 domains and in the protein binding ability. To investigate if *ScAN2* induced the production of other phenolic compounds, we used the aqueous-methanol extract of the transformed leaves to compare the global metabolic profile of *ScAN2* induction with respect to the empty vector pGWB411 (control). Several phenolic compounds were up regulated. Figure 9 shows a significantly (students *t*-test, $p < 0.05$) increased level of chlorogenic acid (CGA), methyl-chlorogenic acid (methyl-CGA) and also ferulic acid conjugated with polyamines (feruloyl-putrescine).

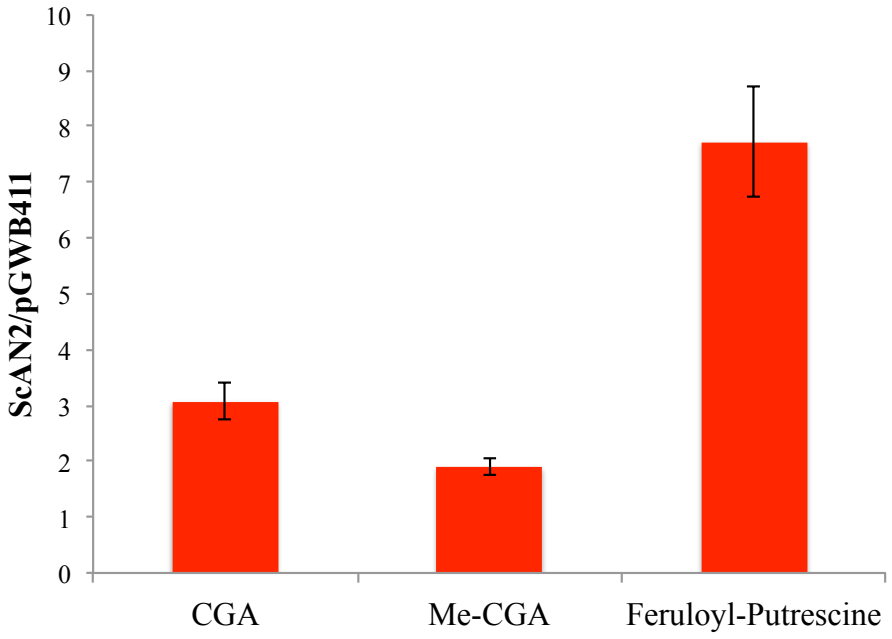

Figure 9. Ratios of the MS signals of significantly changed ($p < 0.05$) compounds in the *ScAN2* infiltrated leaves as compared to the empty vector pGWB411.

Since many of these phenolic compounds are autoflorescents when excited at appropriate wavelength (Hutzler et al., 1998), we localized these metabolites using confocal microscopy. We found a strong florescence in tissues infiltrated with *ScAN2* construct (Figure 10 a). This fluorescence localized in the cell wall and plasma membrane of epidermal tobacco cells. No fluorescence was detected in the empty vector pGWB411, while a very weak signal came from leaves infiltrated with *ScAN1* construct. Consistently, we observed the same results in agroinfiltrated leaves with the respective binding domain of either *ScAN2* or *ScAN1*. Another interesting outcome was the presence of fluorescent vesicles that we identified zooming in the epidermal cell wall (Figure 10 b). These were detected neither in the control nor in *ScAN1*.

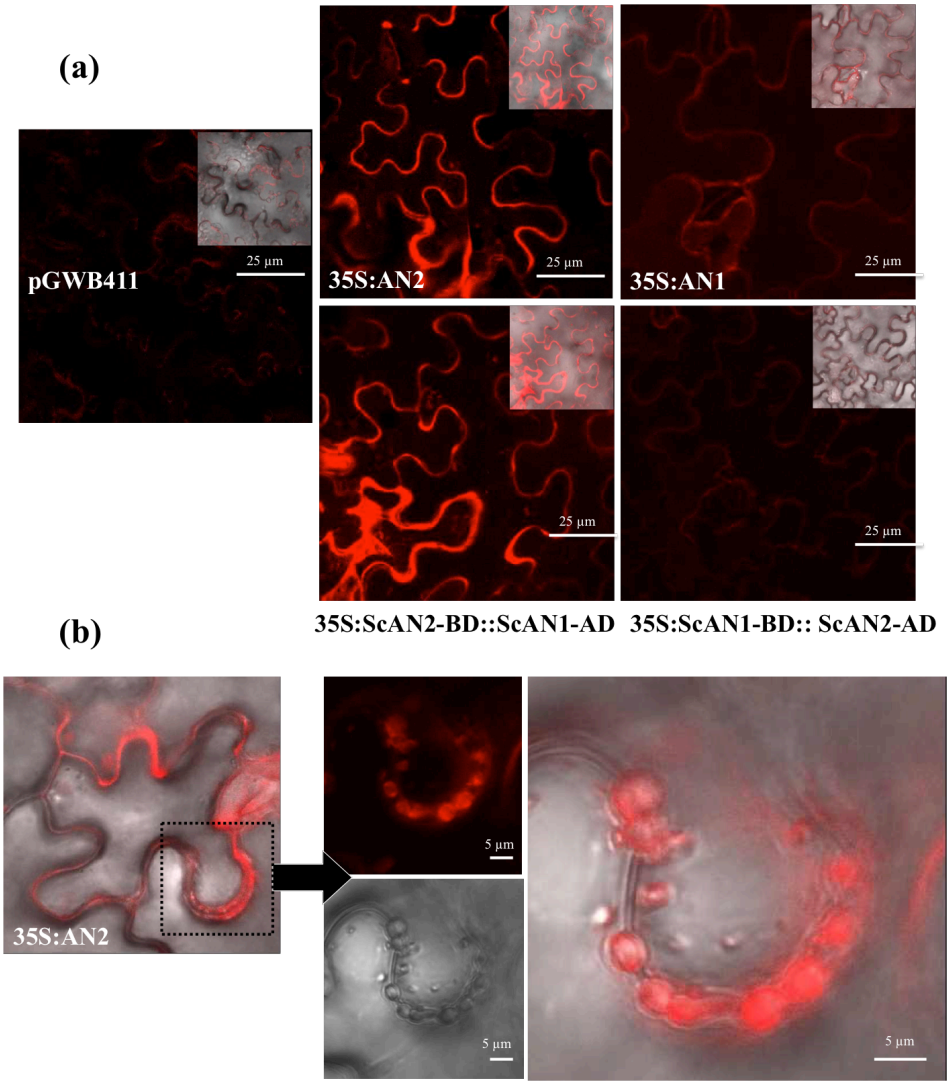

Figure 10. (a) Confocal microscopy images of *N. benthamiana* leaves after agroinfiltration with pGWB411 empty vector, (left); ScAN2 (upper image on the left); ScAN1 (upper image on the right); ScAN2-BD::ScAN1-AD (lower image on the left); ScAN1-BD::ScAN2-AD (lower image on the right). (b) Fluorescent vescicle identified in leaves infiltrated with ScAN2. All the samples were analysed with an excitation at 561 nm and emission bandwidth between 570 to 640 nm.

3.4 Discussion

Anthocyanin accumulation in vegetative tissues often coincides/correlates with cold tolerance. However, it is not clear if such tolerance is a direct or indirect consequence of cold-induced anthocyanin production. Anthocyanin accumulation and distribution in plant tissues are strictly controlled by the regulatory MBW (MYB, bHLH and WD40) ternary complex. *StAN1* belongs to R2R3 MYB TF family and is known to activate anthocyanin pathway in potato (Jung et al. 2009; Payyavula et al., 2013; D'Amelia et al., 2014). It has two homologs named *StMTF1* and *StAN2* (Jung et al., 2009). Though the genomic localization of *StMTF1* is not clear, the function of this gene has been described as chlorogenate inducer and its overexpression triggers anthocyanin accumulation in young transgenic potato plants (Rommens et al., 2008). By contrast, the function of *ScAN2* and its genetic relationship remains poorly characterized. Although Andrè and colleagues (2009) hypothesized that *StAN2* may control anthocyanin pathway and target genes of other pathways, no analyses were carried out to test its function so far.

3.4.1 Duplication and divergence expression: specific *ScAN2* up-regulation after cold stress

A very intriguing aspect related to *StAN2* expression is its association with a general production of phenolic compounds in tubers grown under drought conditions (André et al. 2009). Considering that cold and drought stresses have partially overlapping regulatory networks, we tested if *StAN2* had similar expression induction under cold treatment. We found that *ScAN2* was up regulated after cold stress in the cold tolerant *S. commersonii,* but not in *S. tuberosum*. In the wild potato species we also ascertained an anthocyanin increase triggered by cold treatment. These first results suggested that in *S. commersonii*, under cold stress, *ScAN2* acts as anthocyanin regulator in place of *ScAN1* or, alternatively, that *ScAN2* is the major regulator of anthocyanins in *S.*

commersonii, regardless the stress. The high identity (99%) between *ScAN1* and *ScAN2* and their physical localization, hinted that they are paralogs that arose from the duplication occurred in the last ~2.3 million years, after the divergence of potato and *S. commersonii* from their common ancestor (Aversano et al. 2015). The idea of duplication after the divergence of the two species was also suggested observing the same region in *S. tuberosum*. Here we found a different tandem duplication of the chromosomic portion that interested the last part of the *StAN1* third exon and its 3' flanking region. Gene duplication is very common in flowering plants, where it plays an important role in the evolution of diversity: paralog genes may in fact adopt part of the tasks of the parental gene (sub-functionalization) and thus potentially assume physiological differences. It is tempting to speculate that the different environmental conditions faced by the two species exerted positive selection for the maintenance of multiple copies of genes and over time led to differences in gene activity with adaptive values. In *S. commersonii*, in the same chromosomic region of *ScAN2*, we identified two *StMTF1* orthologs, namely *ScMTF1* and *ScMTF1R2*. The most likely scenario for these four MYB genes is that they derive from different chronological events of duplication, and in particular, *ScMTF1* is the result of specific S. *commersonii* retention. In fact, in *S. tuberosum* the syntenic regions *ScMTF1* and *ScMTF1R2* are not present. Based on the differences in the coding sequence of these genes, we can hypothesize that a first duplication from an ancestral progenitor occurred between *Sc/StAN1*, *Sc/StAN2* and *ScMTF1* before the differentiation of the two *Solanum* species. A second and more recent duplication may have concerned an entire region harbouring *ScAN1*, producing two genomic portions in tandem that showed an almost identical sequence. In summary, these results provided evidence that two intrachromosomal paralogs *ScAN2* and *ScAN1* have a different activation, with *ScAN2* being activated after cold treatment together with an anthocyanin over-production.

3.4.2 Different anthocyanin localization in *S. commersonii* and *S. tuberosum* leaves

To test the relationship between *ScAN2* and anthocyanin production in *S. commersonii,* we further investigated if *ScAN2* could be connected to different anthocyanin localization in the leaves of *S. commersonii*. In fact, it is known that structurally related MYB genes may differentially regulate tissue-specific production of anthocyanins in plant (Petroni & Tonelli, 2011). Clear examples are provided by the regulatory genes that control the snapdragon flower pigmentation. *ROSEA1, ROSEA2* and *VENOSA* are three MYB genes arisen from a duplication of a common ancestor. They differentially regulate the specific production of anthocyanin in different part of *Antirrhinum* flowers (Schwinn et al., 2006). In *Solanum*, it has been reported that *ScAN2* might correspond to the *Ul* (*Under-Leaf*) locus. This has been previously associated to the presence of anthocyanins in abaxial face of potato leaflets (Kessel & Rowe, 1974), a trait we observed in *S. commersonii* leaves. *Ul* was reported as linked to the other locus *Pw (Pigmented whorl)* (Kessel & Rowe, 1974). *Pw* has been previously associated to the phenotype "presence of anthocyanin in adaxial face", that we found in “Double Fun”. As discussed in our previous work (D’Amelia et al., 2014) the locus *Pw* may correspond to *ScAN1* (and consequently to the locus D). The physical association of *ScAN1* and *ScAN2* is consistent with the hypothesis that *ScAN2* corresponds to the locus *Ul*. The adaxial and mesophyllar localization of anthocyanins in “Double Fun”, together with the wide range of different and acylated anthocyanin forms we found in this genotype, suggest a potential role in light screen protection. In fact, da Silva et al. (2012) reported that acylation with aromatic organic acids increases anthocyanin absorbance in the UV-B spectra. Conversely, the presence of anthocyanins under leaf that we detected in *S. commersonii* is typical of plants growing in area where light is a limiting factor (Lee & Graham, 1986; Steyn et al., 2002). In this condition, the role of anthocyanins as UV-B screen is supposed to be less important. Nevertheless their protection against damage caused by photoinibition condition is still active

(Gould et al., 1995).

3.4.3 Divergent binding ability reflects divergences in function

Is the divergence between *ScAN2* and *ScAN1* just associated to a different expression pattern for anthocyanin production? Or has *ScAN2* a divergent function with respect to *ScAN1*? Our analysis supported the hypothesis that the function of *ScAN2* is not completely interchangeable with that of *ScAN1*. In fact, in spite of the fact that *ScAN2* was able to induce anthocyanin production in transgenic tobacco, the total anthocyanin amount was not comparable to that induced by *ScAN1* and, further, infiltrated tissue showed a different phenotypic aspect. Our functional study of the two R2R3 MYBs protein domains indicates that differences observed were mainly imputable to differences in the binding domain of the protein. This suggests that ScAN2, at least in *N. benthamiana*, has a different preference and/or affinity for the promoters of the anthocyanin biosynthesis or target genes. The difference between the BD of the two genes can be imputed to either an intrinsic capacity of the protein sequence to bind, or to the possibility of MYBs to interact with specific co-partners. We suppose that the difference in the *ScAN2* binding ability is likely caused by the BD sequence itself, rather than by the ability to interact with other tobacco endogenous bHLHs. In fact, yeast two-hybrids experiments showed that there were no differences in the interaction ability between ScAN2 and ScAN1 (or between the corresponding potato StAN1 and StAN2) with the potato bHLH co-partners. In *S. commersonii* the anthocyanin bHLHs (*ScJAF13* and *ScbHLH*) were also induced by cold treatment. So, this means that *ScAN2* is expressed together with the two bHLH co-partners in the same tissue and at the same time. Consequently, the variation identified in the BD of *ScAN2* with respect to *ScAN1* is likely the direct cause of difference in the binding ability of *ScAN2*. It is known that some R2R3 MYBs can recognise different sites in cis-elements called MBS1 and MBSII (Solano et al., 1997; Jin & Martin, 1999). In addition, it has been reported that substitutions in the amino acids present in the binding site of R2 or R3 domain can switch the site of binding from a promoter to another. Consequently, the few substitutions found

could not just alter the affinity with anthocyanin target but also switch to different target genes involved in other pathway (Solano et al., 1997; Hichri et al., 2011).

3.4.5 A possible divergence function of *ScAN2* connected to cold stress tolerance of *S. commersonii*

Up-regulation of *ScAN2* following cold treatment together with the findings of André et al. (2009) suggest a regulatory action of *ScAN2* in the pathway of phenolic compounds and a possible role in the tolerance to low temperatures. The paucity of sequence changes in the CDS of ScAN2 and StAN2 indicates that, after the divergence of the two species lineages, there were likely strong constraints on AN2 that conserved protein sequence. By contrast, a larger variation has been found between *ScAN1* and *StAN1*. This correlates with the *StAN1* sequence variability analysed in several potato varieties (D'Amelia et al., 2014, Liu et al., 2015). It can be therefore assumed that the conserved protein sequence of *Sc/StAN2* is connected to a constraint and important function. This is in accordance with the hypothesis that duplicated genes undergo to different fates depending on their function (Zhang et al., 2003). As suggested by our results, the presence of anthocyanins in *S. tuberosum* "Double Fun" is not a sufficient condition to increase tolerance to cold stress. The cold induction of *ScAN2* in *S. commersonii* could be connected not just to anthocyanin production but also to a metabolic secondary effect, as already discussed by André et al. (2006). From our investigation resulted that *ScAN2* indeed activates other branch of phenylpropanoid pathway and, in particular, the accumulation for chlorogenic acid, methyl-chlorogenic acid and ferulic acid conjugated with putrescine. Interestingly, it is known that cold stress induces these compounds (Ramakrishna & Ravishankar, 2011). Their presence could indicate the activation of the biochemical pathway that leads to monolignols. These latter compounds can be incorporated in to the cell wall either as suberin or lignin to increase cold resistance (Ramakrishna & Ravishankar, 2011; Le Gall et al., 2015). The role of polyamines (e.g. putrescine) in plant tolerance to several environmental stresses is widely demonstrated (Kusano et al., 2008). In particular, the increase of

putrescine conjugated with the ferulic acid can indicate a polyamine cell wall binding which is another direct adaptive response to osmotic stress (Hura et al., 2015). Consistently with these considerations, we observed a strong fluorescence localized in the cell wall and plasma membrane of only leaves infiltrated with *ScAN2* when we scan for phenolic compounds with confocal microscopy. Further, in the leaves transformed with *ScAN2* we also observed vesicles harbouring fluorescent cell wall like materials. This correlates with cold adaptative response as observed by Stefanowska et al., (2002) in *Brassica* cells. For these evidences we can assume that *ScAN2* is activated during cold (and osmotic stresses in general) to enhance the cell wall proprieties through the production of cell wall phenolics to enhance cell wall ability to tolerate the injury.

3.5 Conclusions

The comparison between *S. commersonii* and cold sensitive "Double Fun" revealed an intriguing aspect of the cold induced gene *ScAN2*. In fact, we found that *ScAN2* shares the same ancestor of *ScAN1* and consequently an involvement in anthocyanin regulation. This was not just confirmed by phylogenetic analysis but also by yeast two hybrids, that demonstrated that ScAN2 was able to interact with the same bHLH co-factor of ScAN1. Our overexpression experiments revealed that, at least in *N. benthamiana* system, the two genes had a different capacity to activate anthocyanin pathway. Further, we demonstrated that the phenotype of 35S:*ScAN2* is caused by different binding ability that, how suggested the metabolic analysis, is mainly connected to accumulation of cell wall protective compounds. Outchkourov et al. (2014) analysed a similar metabolic increase when overexpress the anthocyanin TF *ROSEA1* in *N. benthamiana* leaves. This could suggest a shared ability of the different anthocyanin gene orthologs to activate different branches of the same pathway. In particular, *ScAN2* showed a more specialized function to activate branches of phenylpropanoid pathway mainly linked to cold stress tolerance. Based on these results, we propose that the duplication between *ScAN1* and *ScAN2* produced a subfunctionalization or more likely an *escape from adaptive conflict* (EAC) (Des Marais & Rausher, 2008). In this scenario, after the duplication from an ancestral progenitor, one copy of the two genes maintained the ancestral function but also specialized in a new function. This would be the first case described for TFs and MYBs as an EAC. Future studies would clarify the pleiotropic effect of *ScAN2* and consequently its involvement for metabolic production and tolerance in potato.

3.6 References

André, C. M., Schafleitner, R., Legay, S., Lefèvre, I., Aliaga, C. A. A., Nomberto, G., ... & Evers, D. (2009). Gene expression changes related to the production of phenolic compounds in potato tubers grown under drought stress. *Phytochemistry*, *70*, 1107-1116.

Aversano, R., Contaldi, F., Ercolano, M. R., Grosso, V., Iorizzo, M., Tatino, F., ... & Carputo, D. (2015). The *Solanum commersonii* genome sequence provides insights into adaptation to stress conditions and genome evolution of wild potato relatives. *The Plant Cell*, *27*, 954-968.

Bai, C., & Elledge, S. J. (1996). Gene identification using the yeast two-hybrid system. *Methods in Enzymology*, *273*, 331-347.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289-300.

Bohnert, H. J., Gong, Q., Li, P., & Ma, S. (2006). Unraveling abiotic stress tolerance mechanisms–getting genomics going. *Current Opinion in Plant Biology*, *9*, 180-188.

Carputo, D. (2003). Cytological and breeding behavior of pentaploids derived from 3x× 4x crosses in potato. *Theoretical and Applied Genetics*, *106*, 883-888.

Carvallo, M. A., Pino, M. T., Jeknić, Z., Zou, C., Doherty, C. J., Shiu, S. H., ... & Thomashow, M. F. (2011). A comparison of the low temperature transcriptomes and CBF regulons of three plant species that differ in freezing tolerance: *Solanum commersonii*, *Solanum tuberosum*, and *Arabidopsis thaliana*. *Journal of Experimental Botany*, *62*, 3807-3819.

Chalker-Scott, L. (2002). Do anthocyanins function as osmoregulators in leaf tissues?. *Advances in Botanical Research*, *37*, 103-127.

D'Amelia, V., Aversano, R., Batelli, G., Caruso, I., Castellano Moreno, M., Castro-Sanz, A. B., ... & Carputo, D. (2014). High *AN1* variability and interaction with basic helix-loop-helix co-factors related to anthocyanin biosynthesis in potato leaves. *The Plant Journal*, *80*, 527-540.

da Silva Ferreira, P., Paulo, L., Barbafina, A., Elisei, F., Quina, F. H., & Maçanita, A. L. (2012). Photoprotection and the photophysics of acylated anthocyanins. *Chemistry–A European Journal*, *18*, 3736-3744.

Des Marais, D. L., & Rausher, M. D. (2008). Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature*, *454*, 762-765.

Docimo, T., Francese, G., Ruggiero, A., Batelli, G., De Palma, M., Bassolino, L., ... & Tucci, M. (2015). Phenylpropanoids accumulation in eggplant fruit: characterization of biosynthetic genes and regulation by a MYB transcription factor. *Frontiers in Plant Science*, *6*, 1233.

Estrada, N. (1982). Breeding wild and primitive potato species to obtain frost-resistant cultivated varieties. In *Plant cold hardiness and freezing stress. Mechanisms and crop implications* (pp. 615-634). Academic Press, USA.

Evers, D., Legay, S., Lamoureux, D., Hausman, J. F., Hoffmann, L., & Renaut, J. (2012). Towards a synthetic view of potato cold and salt stress response by transcriptomic and proteomic analyses. *Plant Molecular Biology*, *78*, 503-514.

Folgado, R., Panis, B., Sergeant, K., Renaut, J., Swennen, R., & Hausman, J. F. (2013). Differential Protein Expression in Response to Abiotic Stress in Two Potato Species: *Solanum commersonii* Dun and *Solanum tuberosum* L. *International Journal of Molecular Sciences*, *14*, 4912-4933.

Gould, K. S. (2004). Nature's Swiss army knife: the diverse protective roles of anthocyanins in leaves. *BioMed Research International*, *2004*, 314-320.

Gould, K. S., Kuhn, D. N., Lee, D. W., & Oberbauer, S. F. (1995). Why leaves are sometimes red. *Nature*, *378*, 241-242.

Hichri, I., Deluc, L., Barrieu, F., Bogs, J., Mahjoub, A., Regad, F., ... & Lauvergeat, V. (2011). A single amino acid change within the R2 domain of the VvMYB5b transcription factor modulates affinity for protein partners and target promoters selectivity. *BMC Plant Biology*, *11*, 117-130.

Hijmans, R. J., Jacobs, M., Bamberg, J. B., & Spooner, D. M. (2003). Frost tolerance in wild potato species: Assessing the predictivity of taxonomic, geographic, and ecological factors. *Euphytica*, *130*, 47-59.

Hura, T., Dziurka, M., Hura, K., Ostrowska, A., & Dziurka, K. (2015). Free and cell wall-bound polyamines under long-term water stress applied at different growth stages of× *Triticosecale Wittm*. *PloS One*, *10*, e0135002.

Hutzler, P., Fischbach, R., Heller, W., Jungblut, T. P., Reuber, S., Schmitz, R., ... & Schnitzler, J. P. (1998). Tissue localization of phenolic compounds in plants by confocal laser scanning microscopy. *Journal of Experimental Botany*, *49*, 953-965.

Janská, A., Maršík, P., Zelenková, S., & Ovesna, J. (2010). Cold stress and acclimation–what is important for metabolic adjustment?. *Plant Biology*, *12*, 395-405.

Jin, H., & Martin, C. (1999). Multifunctionality and diversity within the plant MYB-gene family. *Plant Molecular Biology*, *41*, 577-585.

Jung, C. S., Griffiths, H. M., De Jong, D. M., Cheng, S., Bodis, M., Kim, T. S., & De Jong, W. S. (2009). The potato developer (D) locus encodes an R2R3 MYB transcription factor that regulates expression of multiple anthocyanin structural genes in tuber skin. *Theoretical and Applied Genetics*, *120*, 45-57.

Kessel, R., & Rowe, P. R. (1974). Inheritance of two qualitative traits and a proposed genetic map for their linkage group in diploid potatoes. *Potato Research*, *17*, 283-295.

Kusano, T., Berberich, T., Tateda, C., & Takahashi, Y. (2008). Polyamines: essential factors for growth and survival. *Planta*, *228*, 367-381.

Le Gall, H., Philippe, F., Domon, J. M., Gillet, F., Pelloux, J., & Rayon, C. (2015). Cell wall metabolism in response to abiotic stress. *Plants*, *4*, 112-166.

Lee, D. W., & Graham, R. (1986). Leaf optical properties of rainforest sun and extreme shade plants. *American Journal of Botany*, *73*, 1100-1108.

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., ... & Rombauts, S. (2002). PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. *Nucleic Acids Research*, *30*, 325-327.

Liu, Y., Lin-Wang, K., Deng, C., Warran, B., Wang, L., Yu, B., ... & Allan A. C. (2015). Comparative transcriptome analysis of white and purple potato to identify genes involved in anthocyanin biosynthesis. *PloS one*, *10*, e0129148.

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. *Methods*, *25*, 402-408.

Mantri, N., Patade, V., Penna, S., Ford, R., & Pang, E. (2012). Abiotic stress responses in plants: present and future. In *Abiotic Stress Responses in Plants* (pp. 1-19). Springer, New York.

Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., ... & Kimura, T. (2007). Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *Journal of Bioscience and Bioengineering*, *104*, 34-41.

Outchkourov, N. S., Carollo, C. A., Gomez-Roldan, V., de Vos, R. C., Bosch, D., Hall, R. D., & Beekwilder, J. (2014). Control of anthocyanin and non-flavonoid compounds by anthocyanin-regulating MYB and bHLH transcription factors in *Nicotiana benthamiana* leaves. *Frontiers in Plant Science*, *5*, 1-9.

Payyavula, R. S., Singh, R. K., & Navarre, D. A. (2013). Transcription factors, sucrose, and sucrose metabolic genes interact to regulate potato phenylpropanoid metabolism. *Journal of Experimental Botany*, *64*, 5115-5131.

Petroni, K., & Tonelli, C. (2011). Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Science*, *181*, 219-229.

Ramakrishna, A., & Ravishankar, G. A. (2011). Influence of abiotic stress signals on secondary metabolites in plants. *Plant Signaling & Behavior*, *6*, 1720-1731.

Rommens, C. M., Richael, C. M., Yan, H., Navarre, D. A., Ye, J., Krucker, M., & Swords, K. (2008). Engineered native pathways for high kaempferol and caffeoylquinate production in potato. *Plant Biotechnology Journal*, *6*, 870-886.

Schwinn, K., Venail, J., Shang, Y., Mackay, S., Alm, V., Butelli, E., ... & Martin, C. (2006). A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus Antirrhinum. *The Plant Cell*, *18*, 831-851.

Seppänen, M. M., Majaharju, M., Somersalo, S., & Pehu, E. (1998). Freezing tolerance, cold acclimation and oxidative stress in potato. Paraquat tolerance is related to acclimation but is a poor indicator of freezing tolerance. *Physiologia Plantarum*, *102*, 454-460.

Sharma, M., Singh, A., Shankar, A., Pandey, A., Baranwal, V., Kapoor, S., ... & Pandey, G. K. (2014). Comprehensive expression analysis of rice Armadillo gene family during abiotic stress and development. *DNA Research*, *21*, 267-283.

Solano, R., Fuertes, A., Sánchez-Pulido, L., Valencia, A., & Paz-Ares, J. (1997). A single residue substitution causes a switch from the dual DNA binding specificity of plant transcription factor MYB. Ph3 to the animal c-MYB specificity. *Journal of Biological Chemistry*, *272*, 2889-2895.

Stefanowska, M., Kuraś, M., & Kacperska, A. (2002). Low Temperature-induced Modifications in Cell Ultrastructure and Localization of Phenolics in Winter

Oilseed Rape (*Brassica napus* L. var. *oleifera* L.) Leaves. *Annals of Botany*, *90*, 637-645.

Steyn, W. J., Wand, S. J. E., Holcroft, D. M., & Jacobs, G. (2002). Anthocyanins in vegetative tissues: a proposed unified function in photoprotection. *New Phytologist*, *155*, 349-361.

Yamaguchi-Shinozaki, K., & Shinozaki, K. (2005). Organization of *cis*-acting regulatory elements in osmotic-and cold-stress-responsive promoters. *Trends in Plant Science*, *10*, 88-94.

Zhang, J. (2003). Evolution by gene duplication: an update. *Trends in Ecology & Evolution*, *18*, 292-298.

Zhang, B., Hu, Z., Zhang, Y., Li, Y., Zhou, S., & Chen, G. (2012). A putative functional MYB transcription factor induced by low temperature regulates anthocyanin biosynthesis in purple kale (*Brassica Oleracea* var. *acephala* f. tricolor). *Plant Cell Reports*, *31*, 281-289.

3.7 Supplementary data Chapter 3

Table S1. List of primers used in this study.

Table S2. Anthocyanin profile from leaf extracts of *S. tuberosum* "Double Fun" and *S. commersonii* CMM1T.

Figure S1. Phylogenetic relationship analysis of ScAN2 and ScAN1 and known flavonoid MYB regulators from other species.

Figure S2. Alignment of the predicted amino acid sequence of *ScAN1, StAN1, ScAN2* and *StAN2.*

Figure S3. UPLC-IT-ToF-MS detection at 525 nm of extracts from pGWB411 empty vector (control), *ScAN1* and *ScAN2* infiltrated *N. benthamiana* leaves.

Table S1. List of primer used in this study.

GENE ID	Gene code	Forward primer (5' to 3')	Reveres primer (5' to 3')	Used for
Sotub04g030980.1.1	*Aprt*	GAACCGGAGCAGGTGAAGAA	GAAGCAATCCCAGCGATACG	qRT-PCR analysis
HG763862	*Sc/StJAF13*	CCAGATCAGCAAGACGATCA	GCTCTCGCTCCAAAGACAAC	
JX848660	*Sc/StbHLH1*	CCACCAAAGCCAGCTTTATC	ATCCGCTGGACAAATACCAG	
AY841131	*Sc/StAN2*	AGACCTCAACCTCGGAACTTCTCA	GTCCACCATTGAACTCCATCGTCT	
JX848659	*Sc/StAN1*	AAGGAAAGGTTCATGGACTGA	TCCACTTCATCCCAATCAAAG	
HQ701728	*Sc/StANS*	GGGAAGTGGGTAACTGCAAA	TCTTCTCCTTTGGAGGCTCA	
AY289921	*Sc/StDFR*	TGGACATGGGGTTTCAATTC	GCAACTGGTGCATTCTCCTT	
JX848659	*Sc/StAN1*	ATGAGTACTCCTATGATGT[a]	CTAATTAAGTAGATTCCATATA[b,c]	Cloning in pDNOR207
HG763862	*Sc/StJAF13*	ATGCTATGGGACATCAAGATC[a]	TCAAGATTTCCAAACTACTCTC[b,c]	
AY841131	*Sc/StAN2*	ATGAATACTCCTATGTGTGC[a]	TTAATTAAGTAGATTCCATAAG[b,c]	
JX848660	*StbHLH1*	ATGGAGATTATACAGCC[a]	TTAATTAGCTCTAGGGATTATC[b,c]	
JX848659	*Sc/StAN1*	AAAAGAATTCATGAGTACTCCTATGATGTG	ATATCTCGAGCTAATTAAGTAGATTCC	yeast two hybrid
HG763862	*Sc/StJAF13*	ATTACCCGGGTATGGCTATGGGACATC	ATTTAGTCGACTCAAGATTTCCAAACTACTC	
AY841131.1	*Sc/StAN2*	AAAAGAATTCATGAATACTCCTATGTGTGC	ATATCTCGAGTTAATTAAGTAGATTCCATAA	
JX848659	*StbHLH1*	ATTACCCGGGT ATGGAGATTATACAGCC	ATTTAGTCGACTTAATTAGCTCTAGGG	

(a,b) AttB1 and attB2 sites were added respectively to forward and reverse primers as described in the Gateway Cloning Technology Instruction Manual. (c) The stop codon was not included for C-terminal fusion as described in the Gateway Cloning Technology Instruction Manual.

Table S2. Anthocyanin profile from leaf extract of *S. tuberosum* "Double Fun" and *S. commersonii* CMM1T.

Identified compound	Formula M+	[M+H]	Double Fun			CMM1T		
			RT (min)	Mean peak Area (mAU)	SE(+/-)	RT(min)	Mean peak Area (mAU)	SE(+/-)
Pelargonidin-Rhamnosyl-2Glucoside	$C_{33}H_{41}O_{19}(+)$	741.2237	2.819	8469865	291685	ND	0	
Peonidin-Rhamnosyl-2Glucoside	$C_{18}H_{30}O_{14}(+)$	771.2342	2.961	37705124	513008	ND	0	
Malvidin-Rhamnosyl-2Glucoside	$C_{35}H_{45}O_{21}(+)$	801.2448	3.096	23066712	560151	3.063	4877453	478343
Petunidin-Rhamnosyl-2Glucoside	$C_{34}H_{43}O_{21}(+)$	787.2291	2.686	2171937	75876	ND	0	
Peonidin-Coumaryl-3Glucoside_et.al	$C_{43}H_{49}O_{23}(+)$	933.2659	5.467	14058329	437371	ND	0	
Malvidin-Caffeoyl-Rhamnosyl-2Glucoside	$C_{44}H_{51}O_{24}(+)$	963.2765	5.566	13096240	570859	ND	0	
Pelargonidin-Coumaroyl-Rhamnosyl-2Glucoside	$C_{42}H_{47}O_{21}(+)$	887.2589	5.946	9100437	454232	ND	0	
Peonidin-Coumaroyl-Rhamnosyl-2Glucoside	$C_{43}H_{49}O_{22}(+)$	917.271	6.137	28559121	1001251	ND	0	
Malvidin-Coumaryl-Rhamnosyl-2Glucoside (I)	$C_{44}H_{51}O_{23}(+)$	947.2816	6.252	38194679	1142832	6.224	3620129	160163
Malvidin-Coumaryl-Rhamnosyl-2Glucoside (II)	$C_{44}H_{51}O_{23}(+)$	947.2816	6.5	7420057	186795	ND	0	
Peonidin-Coumaryl-Rhamnosyl-2Glucoside	$C_{43}H_{49}O_{22}(+)$	917.271	6.586	7481047	301554	ND	0	
Malvidin-Feruloyl-Rhamnosyl-2Glucoside	$C_{45}H_{53}O_{24}(+)$	977.2921	6.519	3755048	112884	6.502	3703487	360798

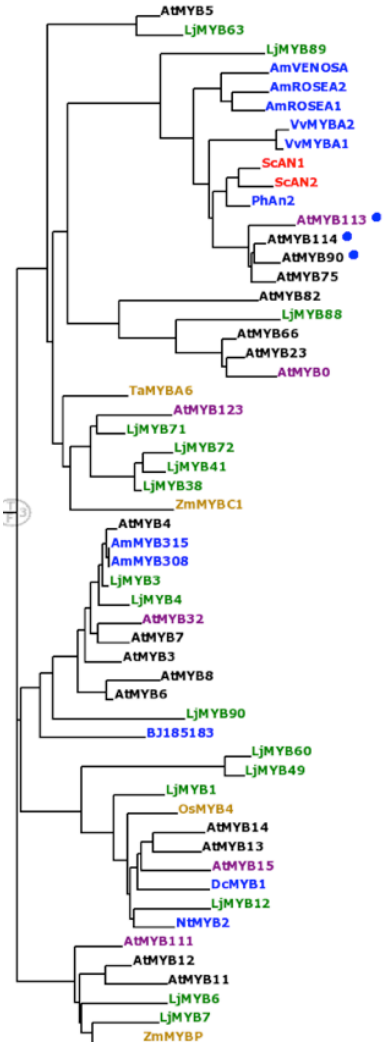

Figure S1. Phylogenetic relationship analysis of ScAN2 and ScAN1 and known flavonoid MYB regulators from other species

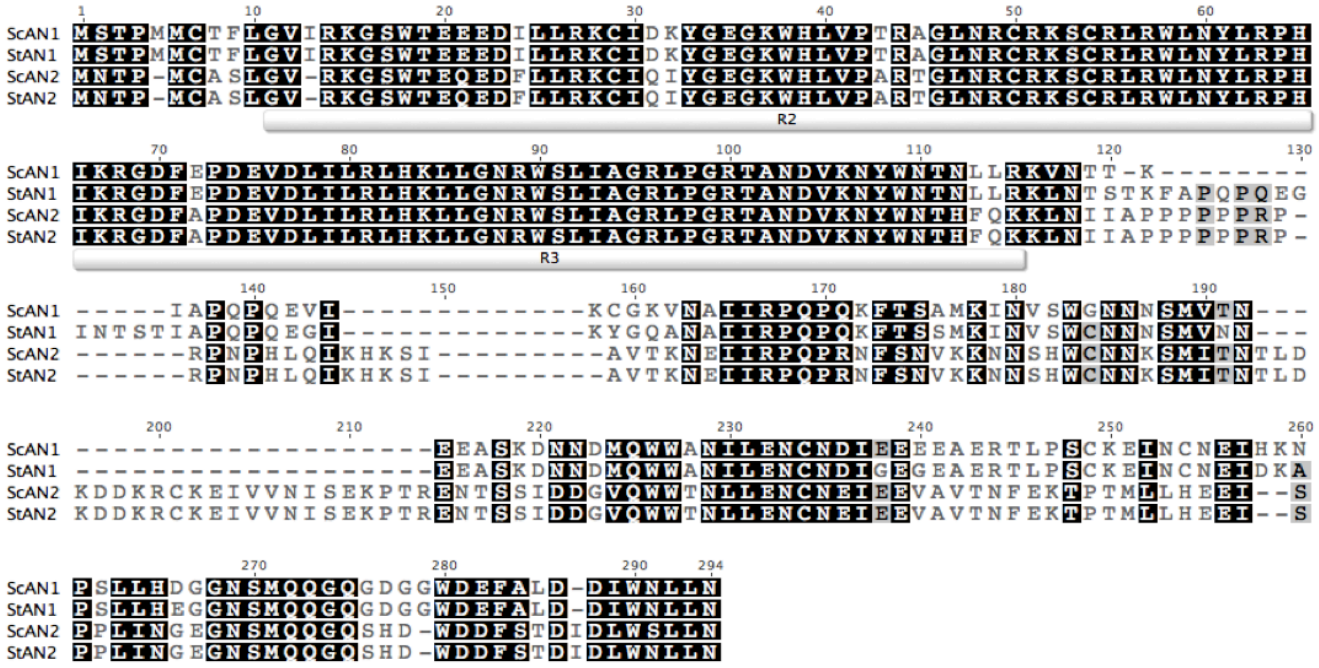

Figure S2. Alignment of the predicted amino acid sequence of *ScAN1, StAN1, ScAN2* and *StAN2*. Shading of the alignment was carried out with the Geneious software and indicates the percentage of similarity between residues: 100% of similarity (black shading), 80–100% of similarity (dark-grey shading), 60–80% of similarity (light-grey shading), less than 60% of similarity (white shading). White boxes indicate MYB R2 and R3 domains.

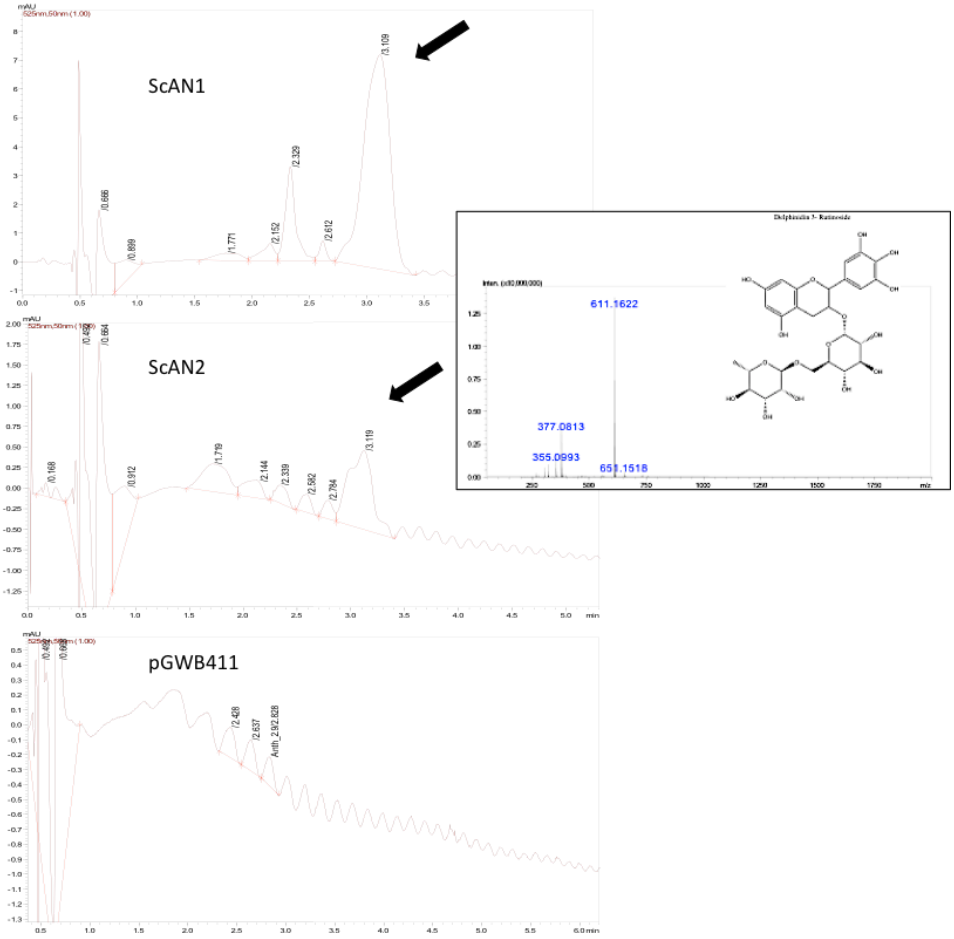

Figure S3. UPLC-IT-ToF-MS detection at 525 nm of extracts from *ScAN1*, *ScAN2* and pGWB411 empty vector (control) infiltrated *Nicotiana benthamiana* leaves. Inserted panel shows the detected accurate mass of the delphinidin-3-rutinoside (D3R) peak.

Chapter 4. *SlMYB4* and *SlELVIRA* negatively influence rutin and chlorogenic acid content in tomato flesh

Chapter 4. *SlMYB4* and *SlELVIRA* negatively influence rutin and chlorogenic acid content in tomato flesh

4.1 Introduction

Plants accumulate metabolites with nutritional and antioxidant properties that have an accepted beneficial effect on human health. Among these bioactive compounds, polyphenols has been highly investigated because associated with a reduced risk to develop chronic diseases (Pandey & Rizvi, 2009). The scientific interest for these molecules also resides in their use as natural additives to give colour, flavour, odour and storage stability to food (Cheynier, 2005). Tomato, as one of the most worldwide consumed crop, is the best candidate to be enriched in these beneficial compounds. In the last decades several strategies, based on metabolic engineering, have been used to create tomato lines rich in several bioactive polyphenols (Raiola et al., 2014). An excellent example of these transgenic lines is the high-flavonols and chlorogenic acid (CGA) producer "AtMYB12" (Luo et al., 2008). Flavonols, such as kaempferol and quercetin, are indeed well characterized to offer protection against major diseases, including coronary heart diseases and cancer (Martin et al., 2013). Further, the synergistic effects of flavonols with the over- accumulation of CGA, makes "AtMYB12" a very attractive functional food. In spite of these brilliant characteristics, consumers often do not accept genetically modified food and breeders are investing research to obtain not transgenic line with high nutritional properties (Povero et al., 2011). Unfortunately, in conventional ripe tomato fruit, flavonols preferentially accumulates in the skin, while only traces of rutin (quercetin-3-O-rutinoside) have been detected in the flesh (Bovy et al, 2002). In fact, it has been studied that the tomato endogenous *SlMYB12* is able to directly regulates the flavonol *CHS, CHI* and *FLS* genes, but it is nevertheless low expressed in the flesh (Ballester et al., 2010; Zhang et al., 2015). One explanation of this *SlMYB12* down regulation can reside in the fact that flavonols, for their "sunscreen" properties, are preferentially accumulated by plant only in tissues exposed to UV-

B. Consequently additional factors may influence *SlMYB12* in flavonol biosynthesis regulation (Tilbrook et al., 2013). In this study, we choose an eQTLs approach to move in the intricate gene regulation architecture that may influence flavonol accumulation in tomato flesh. The material used was the introgression population developed from cross between *Solanum pennellii* with the cultivated tomato *S. lycopersicum*, "M82" (Eshed and Zamir, 1995). In this material, the single discrete wild segment introgressed into the cultivated tomato background can operate as an eQTL. Using RNAseq data already available at John Innes Centre, we identified an eQTL influencing the expression of flavonol and caffeoyl quinic genes. Among differentially expressed genes (DEGs) found in the eQTL we characterized and identified, through silencing approaches, two regulatory genes with negative effect on flavonol and CGA content in tomato flesh. These genes can be target of genome editing approaches to create tomato line with high flavonol accumulation into the flesh.

4.2 Materials and methods

4.2.1 Plant Materials

The tomato introgression lines (*S. lycopersicum* M2 x *S. pennellii*) IL10-1, the sub-line IL10-1-1 and the parental genotype "M82" (Eshed and Zamir, 1995) were used for RNA extraction and gene cloning. For functional and metabolic analysis the high anthocyanin producing tomato line *del*/*ros1* was used. This transgenic line was previously obtained by fruit specific expression of the snapdragon (*Antirrhinum majus*) *DELILA* and *ROSEA1* in "Micro-Tom" (*S. lycopersicum*) background (Butelli et al., 2008). *Nicotiana benthamiana* plants were used for transient expression analysis. Seeds of the transgenic tomato lines were grown in green house with controlled temperature and light conditions: at 25–30°C (day)/18–20°C (night). *N. benthamiana* and tomato ILs were grown in controlled temperature condition (25°C) at natural light April/August at John Innes Centre, Norwich (UK).

4.2.2.Flavonol and caffeoyl quinic differentially expressed genes (DEGs) analysis

Using RNAseq data already produced and statistically analysed at the John Innes Centre (Norwich, UK) differentially expressed genes (DEGs) were analysed as ratio between absolute values of ILs and *S.lycopersicum* "M82". For each comparison, genes that did not show sufficient statistic significance were excluded from the DEG analysis. The DEGs were mapped on ILs using the bin map present in Chitwood et al., (2013) paper.

4.2.3 RNA and cDNA synthesis, cloning and sequence analysis of candidate genes

Total RNA was extracted using the Trizol reagent (Life Technologies, Carlsbad, CA, USA) following the manufacturer's protocol. Complementary DNA was synthesized from DNase-treated total RNA using the SuperScript™ III Kit (Life Technologies). Full length of the CDS of *SlMYB4* and *SlELVIRA* was amplified

from cDNA of tomato IL10-1, IL10-1-1, "M82" and "Micro-Tom" fruit skin and pericarp using attB primer (Supplementary data Table S1). *Att*B-flanked PCR products of each primer pair were cloned in pDON207 vectors (Invitrogen) to obtain entry clones through BP reaction. *SlMYB4* was sub-cloned in pGWB411 destination vector thorough LR reaction according to manual instructions (Invitrogen). For *SlELVIRA* two entry clones were produced and they were named pDONR207-*SlELIVRA* and pDONR207- *SlELVIRAb*.

4.2.4 Virus induced gene silencing (VIGS) and transient over expression

TRV-based silencing vectors pTRV1, pTRV2 and pTRV2- *DEL/ROS1* were prepared as reported by Orzaez et al., (2009). Roughly, 200-300 bp fragments of target genes were amplified from cDNA of Micro-Tom sub-cloned in pTRV2-*DEL/ROS1* through BP and LR reaction according Gateway manual (Invitrogen). Sequenced pTRV2-*DEL/ROS1*-*SlMYB4* and pTRV2-*DEL/ROS1*-*SlELVIRA* were electroporated in *Agrobacterium tumefaciens* strain GV3101:pMP90. Agroinfiltration was performed as described previously (Orzaez et al., 2009; Zhang et al., 2013). Agroinfiltrated fruit showing silenced anthocyanins were collected two weeks after breaker stage.

4.2.5 HPLC analysis

Freeze-dried tomato pericarp (100 mg) was extracted with 10mL 80% MeOH overnight on ice by shaking overnight. Extracts were centrifuged (4000 rpm) at 4 °C and supernatant was filtered through a 0.22 μm membrane filter (Millipore). Aliquots of the original sample were then diluted ten times with 80% MeOH. All samples were analysed on a Surveyor HPLC system (Thermo), using 10μL injections. Phenolics were separated on a 100Å ~2mm 3μ Luna C18(2) column (Phenomenex) using the following gradient of acetonitrile versus 0.1% formic acid in water, run at 300μL.min-1 and 30°C: 0 min, 1% ACN; 4 min, 1%ACN; 23 min, 30% ACN; 30 min, 70% ACN; 30.5 min, 1% ACN; 37 min, 1% ACN. Compounds were identified and quantified by direct comparisons with commercial standards (Sigma) for rutin (quercetin-3-O-rutinoside) and

chlorogenic acid.

4.2.6 Bioinformatics and statistical analysis

Amino acid and nucleotide sequences were collected from the National Center for Biotechnology Information (NCBI) and Sol Genomics and Arabidopsis (TAIR) database. Alignments and phylogenetic trees were performed with GENEIOUS software v6.0.6 (Biomatters, http://www.geneious.com/) or with Interspecies Transcription Factor Function Finder for Plants IT3F (http://jicbio.nbi.ac.uk/IT3F/). BLASTP and BLASTX programs (http://www.ncbi.nlm.nih.gov/blast) were used to perform homology researches in GenBank. Analysis of variance (ANOVA) on metabolic data was carried out using XLSTAT-PRO 7.5.3 software (Addinsoft, http://www.xlstat.com). Duncan's test was performed to compare mean values.

4.3 Results

4.3.1 Flavonol and caffeoyl quinic acid eQTLs hotspot characterization

Differentially expressed genes (DEGs) involved in both flavonoid and caffeoyl quinic acid biosynthesis were identified using (ILs) RNAseq data of fruit pericarp (flesh). The expression variation of these genes, which characterized the different ILs, allowed us to identify in IL10-1 an eQTL candidate to influence flavonoid biosynthesis (Table 1). The identified eQTL corresponded to the *S. pennellii* introgressed region harboured by the IL10-1. The sub-line IL10-1-1, harbour a shorter portion of the introgressed part of the line 10-1 was also considered in our analysis. Among the DEGs we looked for the tomato ortholog of Arabidopsis *AtMYB12*, a transcription factor known to be an inducer of both flavonol and caffeoyl quinic acid synthesis in transgenic tomato fruit (Luo et al., 2008). *SlMYB12* was more than 2 fold change up regulated in both IL10-1 and IL10-1-1 as compared to "M82" (Table 1). In our analysis we also included structural genes involved in the earlier step of phenylpropanoid pathway like different versions of phenylalanine ammonia-lyase family (we annotated as *SlPAL1, 2, 3, 4, 5*), 4-coumarate-CoA ligase (*Sl4CL* and *Sl4CL-like*), chalcone synthase (*SlCHS-1* and *SlCHS-2*) and chalcone isomerase (*SlCHI* and *SlCHI-like*; Table 1). These genes were generally up regulated in both ILs, though with some differences. For example, in IL10-1 only *SlPAL3* showed a 2 fold change, while *SlC4H* was highly expressed with an induction of 10 fold change with respect to "M82". Differently, two *SlPAL* versions, *SlPAL1* and *SlPAL3* were up regulated with 2 fold in the sub-line 10-1-1. The introgression of *S. pennellii* genomic portion in IL10-1 induced a high increase of 10-fold *SlC4H* expression. This expression was not induced when only a shorter portion of the same region was introgressed as shown in IL10-1-1. In fact, in this latter line, *SlC4H* did not show similar high increase. Conversely, we found that *SlCHS* was highly up regulated (with 13 and 9 fold change for the version *SlCHS-1* and *SlCHS-2*, respectively) for IL10-1-1. The genes codifying for the enzymes that specifically lead to flavonols showed a consistent expression patter for both IL10-1 and IL10-1-1. In fact, the expression of *F3'5'H*, codifying

the flavonoid 3'-5' hydroxylases, was not detected either in IL10-1 or IL10-1-1, while the flavonoid 3'- hydroxylases (*F3'H*) and flavonol synthase (*SlFLS*) genes were active in both ILs. Since flavonol biosynthesis was co-regulated with caffeoyl quinic acid synthesis by *AtMYB12* (Luo et al., 2008), in our analysis we inserted also the genes codifying for the Hydroxycinnamoyl-CoA:quinate hydroxycinnamoyltransferase (*SlHQT*), the cytochrome P450 p-coumarate-3-hydroxylase (*SlC3H*) and the hydroxycinnamoyl-CoA:shikimate/quinate hydroxycinnamoyltransferase (*SlHCT*). These gene were particularly up regulated in the sub-line 10-1-1, while in IL10-1 they showed a slightly down-regulation with respect "M82".

In Figure 1 a, b we report how the flux of the pathways changes depending on the portion introgressed. Based on the up regulation of the genes reported in Table 1, it can be easily observed that in the sub-line IL10-1-1 the flux is directed to chlorogenic acid (CGA) and not only to flavonol synthesis (Figure 1b). These regions were consequently considered as putative loci in which single regulators (such as TFs) may influence the expression of both flavonol and caffeoyl quinic acid genes. In the introgressed region of IL10-1 as well as in IL10-1-1 we screened for genes with a potential repressor activity against flavonol structural genes.

Table 1. Differentially expressed genes analysis of phenylpropanoid pathway (flavonoid and caffeoyl quinic acid) as monitored by RNAseq analysis of fruit pericarp of *S. pennellii* introgression lines IL10-1 and IL10-1-1. Each value represents the fold change measured by the ratio between absolute values (ILs/M82).

Gene	Gene ID	IL10-1/M82	IL10-1-1/M82	Chromosome	Reference
SlMYB12	Solyc01g079620	2.918660287	2.717703349	1	Solgenomic database
	Earlier steps of phenylpropanoid				
SlPAL1	Solyc09g007890	1.986434109	2.107881137	9	Solgenomic database
SlPAL2	Solyc09g007900	1.597749437	1.395948987	9	Solgenomic database
SlPAL3	Solyc09g007910	2.087947883	2.540716612	9	Solgenomic database
SlPAL4	Solyc09g007920	1.568357862	1.455654531	9	Solgenomic database
SlPAL5	Solyc10g086180	1.461706783	1.682713348	10	Solgenomic database
SlC4H	Solyc05g047530	10.41297935	1.507374631	5	Solgenomic database
Sl4CL	Solyc03g117870	1.315601955	0.951860138	3	Solgenomic database
	Flavonoid				
Sl4CL-like	Solyc06g035960	1.484234234	0.797297297	6	Solgenomic database
SlCHS-1	Solyc09g091510	3.469208211	13.06158358	9	Solgenomic database
SlCHS-2	Solyc05g053550	2.779605263	9.68656015	5	Solgenomic database
SlCHI	Solyc05g010320	1.047619048	0	5	Solgenomic database
SlCHI-Like	Solyc05g052240	2.001615509	5.598546042	5	Solgenomic database
SlF3H	Solyc02g083860	2.144211239	4.354096141	2	Solgenomic database/Groenenboom et al. (2013)
SlF3'5'H	Solyc11g066580	0	0	11	Solgenomic database/Groenenboom et al. (2013)
SlF3'H	Solyc03g115220	2.045918367	3.382653061	3	Solgenomic database/Groenenboom et al. (2013)
SlFLS	Solyc11g013110	1.642925089	4.073959572	11	Solgenomic database/Groenenboom et al. (2013)
Sl3-GT	Solyc10g083440	3.037542662	4.914675768	11	Solgenomic database/Groenenboom et al. (2013)
	Caffeoyl quinic acid				
SlHCT	Solyc03g117600	1.083906465	2.068775791	3	Solgenomic database
SlCH3	Solyc10g078240	0.592203898	2.808095952	10	Solgenomic database
SlHQT	Solyc07g005760	0.754828567	1.287168738	7	Solgenomic database

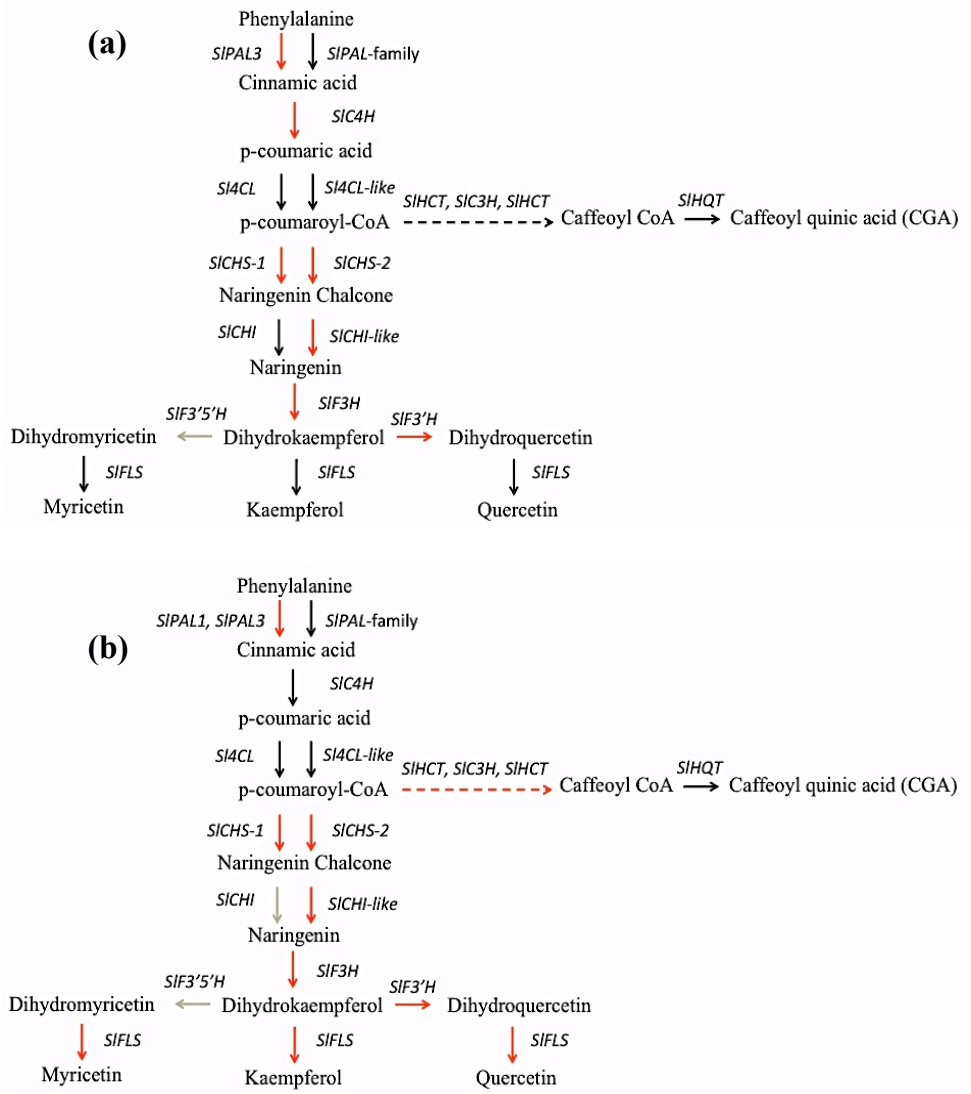

Figure 1. Differentially expressed genes analysis of phenylpropanoid pathway analyzed in IL10-1 *Vs* "M82" (a) and IL10-1-1 *Vs* "M82" (b). Red arrows indicate pathway steps in which the relative structural gene was up regulated with a fold change >2; white arrows indicate pathway steps in which structural gene expression was not detected.

4.3.2 *SlMYB4* is a tomato ortholog of Arabidopsis repressor *AtMYB4*

On the basis of expression data and gene annotation, we identified two main candidates. The first candidate was a MYB transcription factor we named *SlMYB4*. We mapped *SlMYB4* on IL10-1, in the region that does not overlap the introgressed part of IL10-1-1 (Supplementary data Figure S1 a). In IL10-1-1 this gene was down regulated (0.6 fold change), while in IL10-1 it did not show any variation with respect "M82" (Supplementary data Figure S1 b). *SlMYB4* was previously annotated as *SlTHM27* (Solyc10g055410; Lin et al., 1996). Based on the phylogenetic tree, SlMYB4 clustered with two R2R3-MYBs, named AtMYB4 and AmMYB308, that act as repressor in different branch of the phenylpropanoid pathway (Figure 2 a). Alignment of SlMYB4 with the Arabidopsis AtMYB4 and the allele cloned from IL10-1 (corresponding to *S. pennellii* ortholog SpMYB4) indicated that SlMYB4 and SpMYB4 have a conserved motif (Figure 2 b), named as *ethylene-responsive element binding factor-associated Amphiphilic Repression* (EAR-motif; Legay et al., 2007). The EAR-motif is typical for transcription factors with repressor activity. The DNA-binding R2R3 domain was highly conserved between the three proteins. The only difference we found in this region was the Threonine (T) substitution in position 31, which characterized SpMYB4, in place of an Alanine (A) found in SlMYB4 and AtMYB4 (Figure 2b). Ectopic expression *SlMYB4* in *N. benthamiana* leaves (Figure 2 c) produced a phenotype comparable to that of AtMYB4 in tobacco leaves (Jin et al. 2000; Supplementary data Figure S2). This was characterized in both experiments by the presence of white lesions similar to those observed for another repressor, AmMYB308 (Tamagnone et al., 1998; Jin et al. 2000).

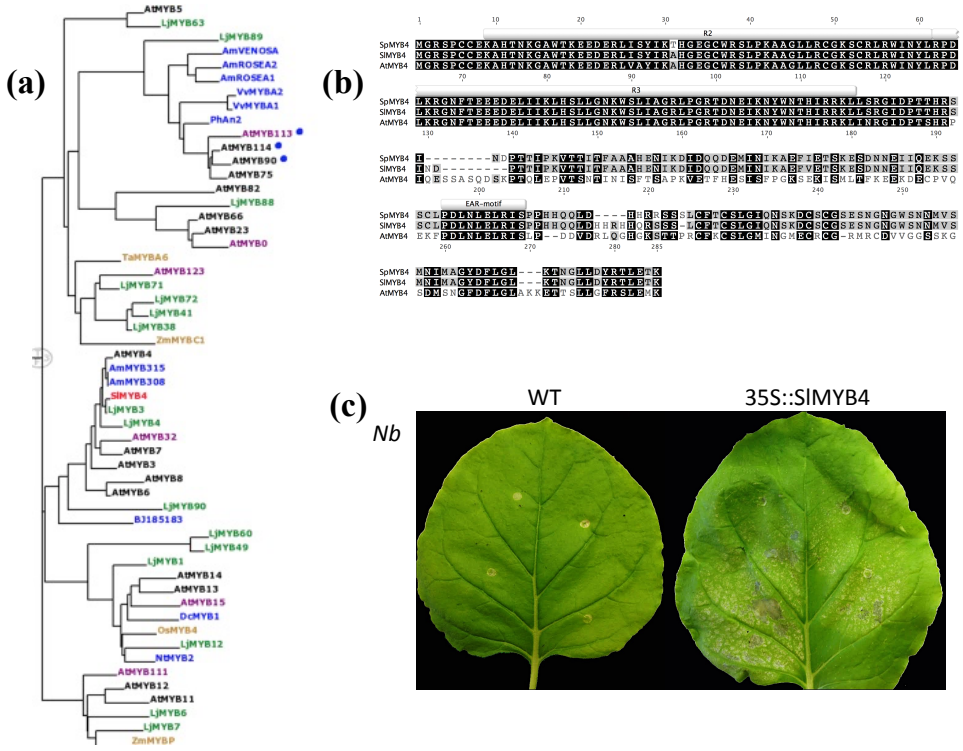

Figure 2. Analysis of SlMYB4. (a) Phylogenetic relationship analysis of tomato SlMYB4 and known flavonoid MYB regulators from other species. (b) Alignment of SpMYB4, SlMYB4 and AtMYB4. Shading of the alignment was carried out with the Geneious software and indicates the percentage of similarity between residues: 100% of similarity (black shading), 80–100% of similarity (dark-grey shading), 60–80% of similarity (light-grey shading), less than 60% of similarity (white shading). White boxes indicate either MYB R2 and R3 domains or the ethylene *responsive element binding factor-associated Amphiphilic Repression* (EAR-motif). (c) Phenotypic effects of *SlMYB4* overexpression in *N. benthamiana* (Nb). The overexpression caused the production of white lesions similar to those observed in *Jin et al.,* (2000) for AtMYB4 in tobacco plant (Supplementary data Figure S2).

4.3.3 *SlELVIRA* mapping and characterization

The second candidate identified was annotated as WD40 protein and we called it *SlELVIRA*. *SlELVIRA* was chromosomally mapped where the wild *S. pennellii* homolog portion was introgressed in IL10-1-1. *SlELVIRA* was consistently down regulated in both IL10-1 and IL10-1-1 (0.3 and 0.4, respectively) as compared to "M82". Since there were no functional annotations for SlELVIRA in tomato, we use BLASTp algorithm to search homolog proteins in *Arabidopsis thaliana* database. Results are summarized in Figure 3 by a phylogenetic tree. SlELVIRA clustered within a class WD40 transducin proteins. These proteins were functionally annotated as signal transducer in TAIR database. *SlELVIRA* CDS was amplified using the cDNA from "M82" fruit. On agarose gel, the amplified products showed, two discrete bands of about 2,6 Kb and 2,5 Kb, that were cloned separately and sequenced. The two sequences of *SlELVIRA* corresponded to of the database gene. Interestingly, the two cloned sequences exactly matched to each other expect for 91 bp, that were missing in shorter one that we called *SlELVIRAb* (Figure 4 a,b) The alignment of *SlELVIRA* and *SlELVIRAb* with their genomic sequence revealed that the 91 bp missing nucleotides of *SlELVIRAb* corresponded entirely to exon 2 (Figure 4 a,b). Hence, we considered *SlELVIRAb* a likely alternative spliced form of *SlELVIRA*. Besides, the amino acid prediction showed that the alternative splicing caused a frameshift mutation in the third exon and the formation of a premature stop-codon (Figure 3 b). SlELVIRAb was trunked of 360 amino acids in the C-terminal part and lacked of the most part of its WD40 active site (Supplementary data Figure S3 a, b). The amplification with a set of primer surrounding exon 2 suggested that the alternative splicing had tissue specificity in "M82" mature fruit (10 days after breaker stage). As shown in Figure 5, we only had one product of about 1109 bp when we amplified from cDNA of "M82" fruit skin, corresponding to the predicted product for *SlELVIRAb*. When we used the cDNA from peeled pericarp, two bands were obtained: one corresponding to *SlELVIRAb*, whereas the higher band (1200 bp) corresponded to *SlELVIRA* (Figure 5).

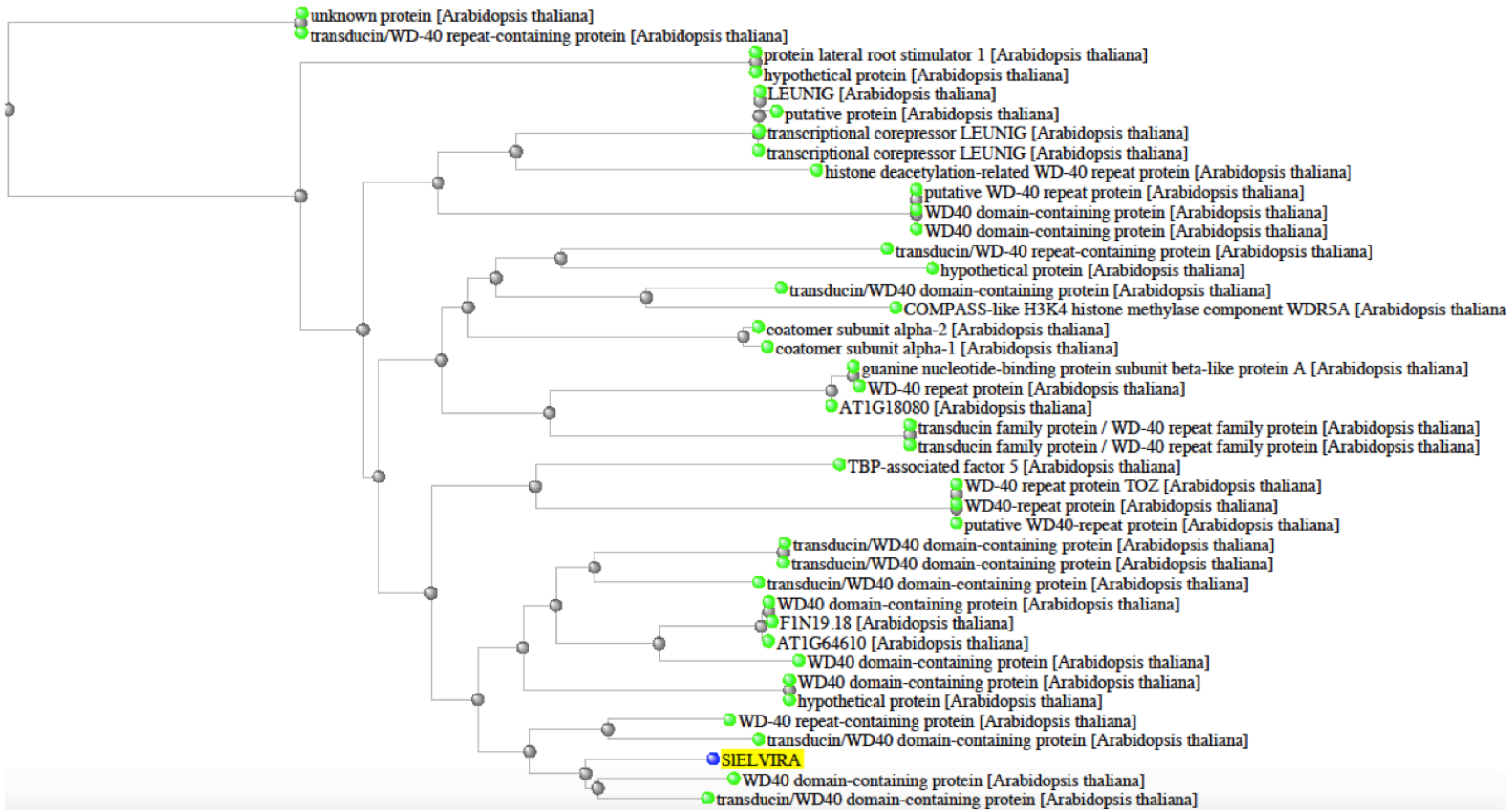

Figure 3. Phylogenetic relationship analysis of tomato SlELVIRA and homolog proteins from Arabidopsis database.

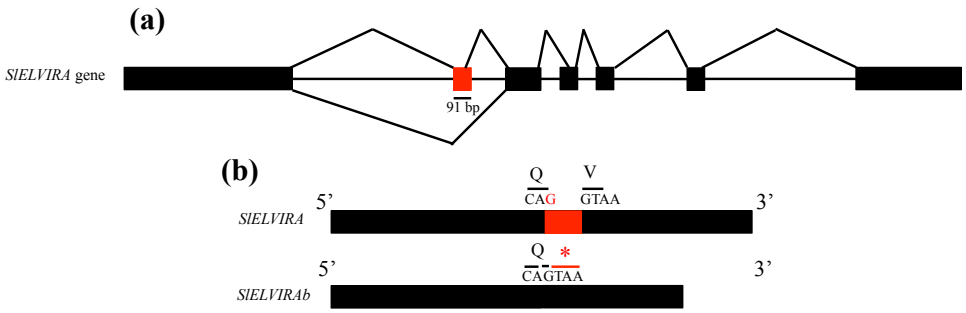

Figure 4. (a) Genomic structure of *SlELVIRA*. Black boxes indicate exons, while black straight line indicate introns. Red box indicates the missing exon 2 (91 bp) in *SlELVIRAb*. Curved lines indicate the possible exon junction in mature mRNA. (b) Two possible mature mRNA that can be produced by *SlELVIRA* mRNA splicing: *SlELVIRA* functional mRNA harbouring exon 2 and *SlELVIRAb* spliced form. The splicing of exon 2 produces a frameshift mutation and the generation of a premature stop codon (TAA).

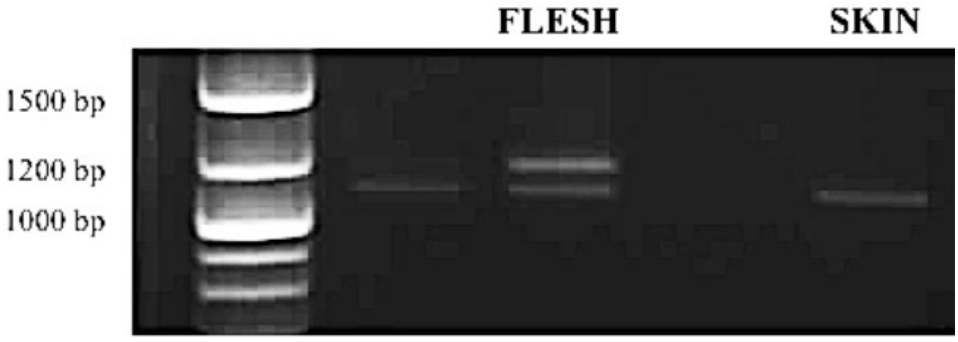

Figure 5. Tissue specificity of *SlELVIRAb* mRNA expression as analysed by RT-PCR. Amplification products were resolved on a 2% agarose gel.

4.3.4 Functional analysis of *SlMYB4* and *SlELVIRA* using VIGS

To study the function of the *SlMYB4* and *SlELVIRA* in flavonoid biosynthesis, VIGS (Virus induced gene silencing) method was carried out on *del/ros1* tomato fruits. Anthocyanin-monitored VIGS confirmed that the silencing occurred either in the control (VIGS-*DEL/ROS*) or in target gene constructs (VIGS-*SlMYB4,* VIGS-Sl*ELVIRA*) as demonstrated by the block of anthocyanin production in silenced tissue of tomato *DEL/ROS* Micro-Tom (Figure 6 a). Silencing was also confirmed by qPCR (data not shown). To determine whether the two identified genes had some repressor effects on flavonol and caffeoyl quinic synthesis, both rutin (quercetin-3-O-rutinoside) and CGA were quantified from the extract of silenced parts of fruit pericarp (Figure 6 b, c). *SlMYB4* and *SlELVIRA* silencing had a clear influence on rutin content, as evidenced by the increase (Δ) of rutin of about 2 mg g^{-1} (DW) *vs.* VIGS-*DEL/ROS* (Figure 6 c). VIGS-*SlMYB4* had no evident impact on CGA (chlorogenic acid) content. Conversely, when VIGS-*SlEVLIRA* construct was used, the CGA content increased of about 3 mg g^{-1} (DW) in silenced pericarp (Figure 6 c).

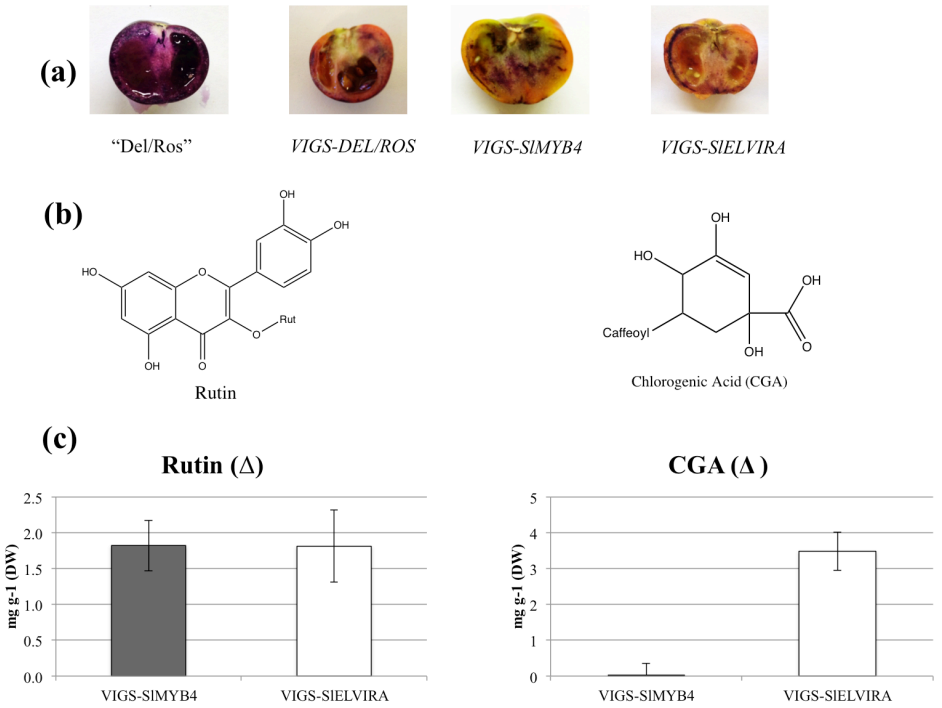

Figure 6. a) From left to right, phenotype of "Del/Ros", VIGS-*Del/Ros1*, VIGS-*SlMYB4*, and VIGS-Sl*ELVIRA*, Micro-Tom" fruits. Pictures were taken at two weeks after breaker stage. b) Molecular structure of rutin (quercetin 3-O rutinoside) and chlorogenic acid (CGA). c) Increase (Δ) of rutin and chlorogenic acid (CGA) content in silenced flesh *VIGS-SlMYB4* and *VIGS-SlELVIRA*, respectively.

4.4. Discussion

This study aimed at identifying regulatory genes associated with an eQTL with a specific impact on flavonols and chlorogenic acid. Hence, the transcript variation of genes controlling the flavonoid and caffeoyl quinic acid pathway were analysed in *S. pennellii* ILs (*S. pennellii* x *S. lycopersicum*). We identified an eQTL hotspot in IL10-1 on chr10, in a region physically replaced by *S. pennellii* homolog. We observed that the wild introgression caused an expression activation of flavonoid structural genes physically located in other chromosomes (expect for *SlPAL5*). Hence, this region may be better considered as a *trans*-eQTL and putatively harbour master regulatory genes (e.g TFs). The IL10-1 was compared with its sub-line IL10-1-1, because it still consistently conserved a general activation of flavonoid genes as well as genes involved in the caffeoyl quinic acid pathway. Consistently, in these same ILs Alseekh et al., (2015) found an increase of flavonoids and caffeoyl quinic acids compared to "M82". In these two lines we also observed up regulation of *SlMYB12* that is not generally expressed in tomato flesh (Ballester et al., 2010; Zhang et al., 2015). Based on these findings, we hypothesized the introgression caused that sort of misleading into network of regulation of flavonol and CGA pathway. Since *SlMYB12*, unlike flavonoid MYB TF, does not need bHLH co-factors (Zimmermann et al., 2004; Mehrtens, 2005), we decided to look for potential negative regulator of flavonols in the flesh. In particular, we characterized two genes with different characteristics, *SlMYB4* and *SlELVIRA*.

4.4.1 *SlMYB4* is involved in flavonoid regulation in tomato flesh

SlMYB4 was the first candidate we focused on. This gene was previously described as *SlTHM27* and codifies for an R2R3-MYB transcription factor with a regulation activity of phenylpropanoids in tomato skin (Lin et al., 1996; Adato et al., 2009). Performing VIGS test on *SlMYB4*, we demonstrated that its silencing induced an increase of rutin content in silenced tomato flesh, but not a parallel increase of chlorogenic acid. Similarly, in IL10-1, in which tomato *SlMYB4* was

replace by wild allele *SpMYB4*, the expression of *SlHCT*, *SlCH3* and *SlHQ* were not activated and even decreased compared to "M82". The T substitution of SpMYB4 in the binding domain, that is a region completely conserved between SpMYB4, SlMYB4 and AtMYB4, may explain a possible dysfunction of its activity. *AtMYB4* is known to be a repressor of hydroxycinnamic acid and phenylpropanoid metabolism in general, acting mainly on *C4H* gene expression in response to UV-B exposure (Jin et al. 2000). In IL10-1 we found that the transcripts of *C4H* was highly present; this fits with the idea of a SpMYB4 dysfunction. In this scenario, *SpMYB4* cannot stop the flux to flavonol genes like *CHS*, *CHI* and *FLS* that are induced by an active *SlMYB12*.

4.4.2 *SlELVIRA,* a new characterized gene influencing flavonol and chlorogenic acid (CGA) accumulation in tomato flesh

The WD40 protein, we named SlELVIRA was the second candidate with a repressor activity. Though it phylogenetically clustered with protein involved in signal transduction, to our knowledge neither this gene nor an Arabidopsis orthologous has been already characterized. Another characteristic that let us to think in a negative action of *SlELVIRA* with respect to flavonoid production was the spliced form *SlELVIRAb*. In fact, *SlELVIRAb* codifies for a trunked and, potentially, not functional protein. This form is likely the only form present in the skin of mature fruit that is the fruit tissue were flavonols are mainly produced. The transcript levels of *SpELIVRA* in the flesh were less abundant in IL10-1 and IL10-1-1 compared to that of "M82". Conversely, in the same genotype not only the flavonoid genes but also *SlHCT*, *SlCH3* and *SlHQ* were highly expressed. The silencing of *SlELVIRA* using VIGS showed an increase of rutin as well as a parallel increase of chlorogenic acid. Thus confirms the expression data observed in IL10-1-1. Further researches should be undertaken to investigate the role of *SlELVIRA*. A hypothesis is that *SlELVIRA* can have an activity similar of another WD40 transducin protein named *COP1* (Tilbrook et al., 2013). In Arabidopsis it is known that *AtMYB12* is actively UV-B induced through a UV-B transducers UVR8/COP1/HY5 complex (Tilbrook et al., 2013). Further, *AtCOP1* showed

similar splicing regulation and can influence AtMYB12 expression, as mediator of UV-B induction (Zhou et al., 1998;Torii et al., 1998). Hence, this led us to speculate that *SlELVIRA* may similarly act as *AtCOP1*.

4.5 Conclusions

In conclusion, this study has identified two genes with a putative major effect in an eQTLs influencing flavonol and chlorogenic acid accumulation in tomato flesh. Breeding programs aimed at enhancing the content of these healthy compounds in tomato flesh could be helped in the use of these eQTLs thanks to the information arisen from *SlMYB4* and *SlELVIRA* characterization. The molecular mechanism explaining how these genes work needs to be clarified in future studies mainly based on transcriptional and functional analysis. What kind and whether a relation exists between *SlMYB4*, *SlELIVRA* and *SlMYB12* regulation is particularly interesting to study.

4.6 References

Adato, A., Mandel, T., Mintz-Oron, S., Venger, I., Levy, D., Yativ, M., ... & Aharoni A. (2009). Fruit-surface flavonoid accumulation in tomato is controlled by a SlMYB12-regulated transcriptional network. *PLoS Genet*, *5*, e1000777.

Alseekh, S., Tohge, T., Wendenberg, R., Scossa, F., Omranian, N., Li, J., ... & Fernie A.R. (2015). Identification and mode of inheritance of quantitative trait loci for secondary metabolite abundance in tomato. *The Plant Cell*, *27*, 485-512.

Ballester, A. R., Molthoff, J., de Vos, R., te Lintel Hekkert, B., Orzaez, D., Fernández-Moreno, J. P., ... & Bovy, A. (2010). Biochemical and molecular analysis of pink tomatoes: deregulated expression of the gene encoding transcription factor SlMYB12 leads to pink tomato fruit color. *Plant Physiology*, *152*, 71-84.

Bovy, A., de Vos, R., Kemper, M., Schijlen, E., Pertejo, M. A., Muir, S., ... & van Tunen, A. (2002). High-flavonol tomatoes resulting from the heterologous expression of the maize transcription factor genes *LC* and *C1*. *The Plant Cell*, *14*, 2509-2526.

Butelli, E., Titta, L., Giorgio, M., Mock, H. P., Matros, A., Peterek, S., ... & Martin, C. (2008). Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nature Biotechnology*, *26*, 1301-1308.

Cheynier, V. (2005). Polyphenols in foods are more complex than often thought. *The American Journal of Clinical Nutrition*, *81*, 223-229.

Chitwood, D. H., Kumar, R., Headland, L. R., Ranjan, A., Covington, M. F., Ichihashi, Y., ... & Sinha, N. R. (2013). A quantitative genetic basis for leaf morphology in a set of precisely defined tomato introgression lines. *The Plant Cell*, *25*, 2465-2481.

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

Jin, H., Cominelli, E., Bailey, P., Parr, A., Mehrtens, F., Jones, J., ... & Martin, C. (2000). Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis. *The EMBO Journal*, *19*, 6150-6161.

Legay, S., Lacombe, E., Goicoechea, M., Brière, C., Séguin, A., Mackay, J., & Grima-Pettenati, J. (2007). Molecular characterization of EgMYB1, a putative transcriptional repressor of the lignin biosynthetic pathway. *Plant Science*, *173*, 542-549.

Lin, Q., Hamilton, W. D., & Merryweather, A. (1996). Cloning and initial characterization of 14 MYB-related cDNAs from tomato (*Lycopersicon esculentum* cv. Ailsa Craig). *Plant molecular Biology*, *30*, 1009-1020.

Luo, J., Butelli, E., Hill, L., Parr, A., Niggeweg, R., Bailey, P., ... & Martin, C. (2008). AtMYB12 regulates caffeoyl quinic acid and flavonol synthesis in tomato: expression in fruit results in very high levels of both types of polyphenol. *The Plant Journal*, *56*, 316-326.

Martin, C., Zhang, Y., Tonelli, C., & Petroni, K. (2013). Plants, diet, and health. *Annual Review of Plant Biology*, *64*, 19-46.

Mehrtens, F., Kranz, H., Bednarek, P., & Weisshaar, B. (2005). The Arabidopsis transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. *Plant Physiology*, *138*, 1083-1096.

Orzaez, D., Medina, A., Torre, S., Fernández-Moreno, J. P., Rambla, J. L., Fernández-del-Carmen, A., ... & Granell, A. (2009). A visual reporter system for virus-induced gene silencing in tomato fruit based on anthocyanin accumulation. *Plant Physiology*, *150*, 1122-1134.

Pandey, K. B., & Rizvi, S. I. (2009). Plant polyphenols as dietary antioxidants in human health and disease. *Oxidative Medicine and Cellular Longevity*, *2*, 270-278.

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

Raiola, A., Rigano, M. M., Calafiore, R., Frusciante, L., & Barone, A. (2014). Enhancing the health-promoting effects of tomato fruit for biofortified food. *Mediators of inflammation*, *2014,* 1-16.

Tamagnone, L., Merida, A., Parr, A., Mackay, S., Culianez-Macia, F. A., Roberts, K., & Martin, C. (1998). The AmMYB308 and AmMYB330 transcription factors from *Antirrhinum* regulate phenylpropanoid and lignin biosynthesis in transgenic tobacco. *The Plant Cell*, *10*, 135-154.

Tilbrook, K., Arongaus, A. B., Binkert, M., Heijde, M., Yin, R., & Ulm, R. (2013). The UVR8 UV-B photoreceptor: perception, signaling and response. *The Arabidopsis Book*, *11,* e0164.

Torii, K. U., McNellis, T. W., & Deng, X. W. (1998). Functional dissection of Arabidopsis COP1 reveals specific roles of its three structural modules in light control of seedling development. *The EMBO Journal*, *17*, 5577-5587.

Zhang, Y., Butelli, E., De Stefano, R., Schoonbeek, H. J., Magusin, A., Pagliarani, C., ... & Martin C. (2013). Anthocyanins double the shelf life of tomatoes by delaying overripening and reducing susceptibility to gray mold. *Current Biology*, *23*, 1094-1100.

Zhang, Y., Butelli, E., Alseekh, S., Tohge, T., Rallapalli, G., Luo, J., ... & Martin, C. (2015). Multi-level engineering facilitates the production of phenylpropanoid compounds in tomato. *Nature Communications*, *6,* 1-11.

Zhou, D. X., Kim, Y. J., Li, Y. F., Carol, P., & Mache, R. (1998). COP1b, an isoform of COP1 generated by alternative splicing, has a negative effect on COP1

function in regulating light-dependent seedling development in Arabidopsis. *Molecular and General Genetics MGG*, *257*, 387-391.

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

4.7 Supplementary data Chapter 4

Table S1. List of primers used in this study.

Figure S1. (a) Schematic representation of chromosome 10 in *S. pennellii* introgression lines IL10-1 and IL10-1-1. The *SlMYB4* and *SlELVIRA* position is also indicated. (b) Diagram showing the ratios of the expression in the pericarp of IL10-1 and IL10-1-1 as compared to "M82".

Figure S2. Phenotypic effects of *AtMYB4* overexpression in tobacco (Nt) from Jin et al., (2000)

Figure S3. a) Diagrams of the predicted proteins of SlELVIRA and SlELVIRAb. b) Alignment of the WD40 motifs of SlELVIRA and SlELVIRAb

Table S1. List of primers used in this study

Gene	Forward primer (5' to 3')	Revers primer (5' to 3')	Used for
SlMYB4	GCTGAATTCGTTGAAACAAGCA	ACTTTTCACTTCCACAACTGCA	VIGS:Cloning in pDONR207
	ATGGGAAGGTCACCTTGTTGTGAG	TCACTTAGTTTCCAAAGTCTATG	CDS cloning in pDONR207
SlELVIRA	TCATGATGGTCTCGACGAGG	CAATGACCAAACCCCAGCAG	VIGS:Cloning in pDONR207
	ATGAGCAAAGCAAGAGAAGG	TTAAATCCGAAATGGTAAGCC	CDS cloning in pDONR207
SlELVIRA	AGATGTTGAAATTGAGGAGAGT	ACAAGGGGAGCTGTCATTTG	*SlELVIRA* splicing analysis

(a,b) AttB1 and attB2 sites were added respectively to forward and reverse primers as described in the Gateway Cloning Technology Instruction Manual

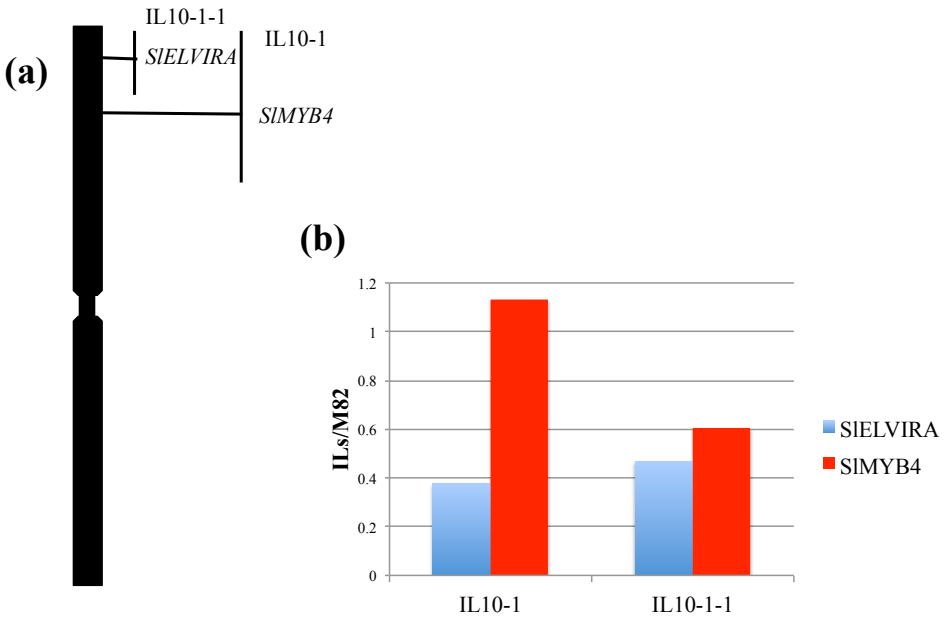

Figure S1. (a) Schematic representation of chromosome 10 in *S. pennellii* introgression lines IL10-1 and IL10-1-1. The *S. pennellii* introgressed regions are represented as thin lines in S. *lycopersicum* "M82" background on chromosome 10 (thick line). The *SlMYB4* and *SlELVIRA* position is also indicated. (b) Expression of *SlMYB4* and *SlELVIRA* in the pericarp of IL10-1 and IL10-1-1. Value is given as ratio ILs/M82.

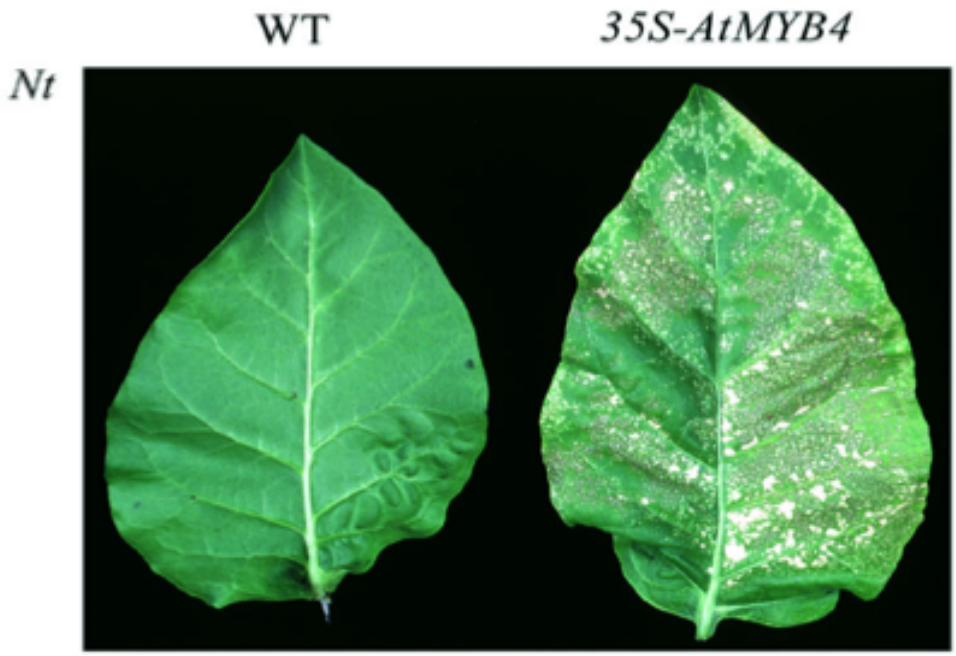

Figure S2. Phenotypic effect of *AtMYB4* overexpression in tobacco from Jin et al (2000), *The EMBO Journal*, *19*, 6150-6161.

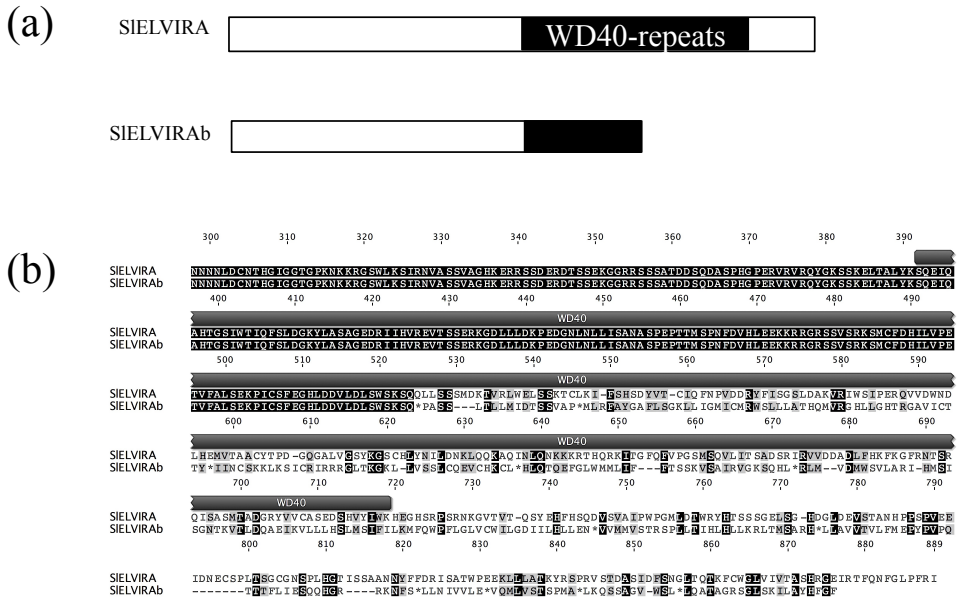

Figure S3. a) Diagrams of the predicted proteins of SlELVIRA and SlELVIRAb. b) Alignment of the WD40 motifs of SlELVIRA and SlELVIRAb

Chapter 5. General conclusions

Chapter 5. General conclusions

As extensively discussed in the general introduction of this thesis, plant secondary metabolites draw an outstanding interest due to their positive proprieties. In fact, the presence of these metabolites in plants is beneficial for both human consumption and plant defence (Korkina, 2007; Ramakrishna & Ravishankar, 2011). In this work we focused on phenylpropanoid pathway, with a particular interest for the braches of flavonoids. The compounds produced through this biochemical pathway represent indispensable elements for the plants. Indeed, molecules like hydroxycinnamyl alcohols are necessary for plant structure because they are the building block of lignin (Fraser & Chapple, 2011). Similarly important is their role in human nutrition. In fact, thanks to their strong antioxidant activity, they are one of the main components of the "functional food". In our work we obtained new and interestingly results that extend the knowledge on the genetic regulation of anthocyanin and flavonols in relation to plant defence and human nutrition, respectively. The study aimed to characterize the genetic control of these two classes of flavonoid in two important crops. The genetic control underlying the anthocyanin accumulation in response to plant stress was studied in potato while, in tomato was concentrated our attention for flavonols.

Since no molecular information was available on anthocyanin regulation in potato leaves, the first part of our research focused onto determination of MYB/bHLH complex that influence anthocyanin production in vegetative tissues of the cultivated potato *Solanum tuberosum* (Chapter 2). We found that *StAN1,* the gene codifying a MYB factor, displays intraspecific sequence variability in both coding/non-coding regions and in the promoter. In addition, its expression was associated with high anthocyanin content in leaves. Expression analysis provided evidence that leaf pigmentation is associated to *StAN1* expression and that *StJAF13* acts as putative *StAN1* co-regulator for anthocyanin gene expression in leaves of a red leaf variety named "Magenta Love", while a concomitant expression of *StbHLH1* may contribute to anthocyanin accumulation in leaves of

"Double Fun". Yeast two-hybrid experiments confirmed that AN1 interacts with StbHLH1 and StJAF13 and the latter interaction was verified and localized in the cell nucleus by bimolecular fluorescence complementation assays. In addition, transgenic tobacco (*Nicotiana tabacum*) overexpressing a combination of either *AN1* with *StJAF13* or *AN1* with *StbHLH1* showed deeper purple pigmentation with respect to *AN1* alone. This further confirmed AN1/StJAF13 and AN1/StbHLH1 interactions. Our findings demonstrate that the classical loci identified for potato leaf anthocyanin accumulation correspond to *AN1* and may represent an important step to expand our knowledge on the molecular mechanisms underlying anthocyanin biosynthesis in different plant tissues. These results have been published on *the Plant Journal*. They have been confirmed by following researches and became the bases for studies mainly associated to anthocyanin production in the tuber. Liu et al. (2015, 2016), for example, found that the *StAN1* alleles we characterized were also associated to a different colour tuber phenotype, either in their expression or in their ability to induce anthocyanins. They also verified that *StJAF13*, the bHLH gene we found to be associated to leaf pigmentation, has also an important role to induce anthocyanin in the tuber (Liu et al., 2016).

In the third chapter of the thesis we compared the anthocyanin genes of the wild cold tolerant species *Solanum commersonii* with those of cultivated varieties under cold stress condition. As suggested by functional and metabolic analysis in *Nicotiana benthamiana*, two paralog MYB genes evolved differently between cultivated and wild species. In *S. commersonii*, *ScAN2* seems to keep a pleiotropic and ancestral function with respect to *ScAN1*, inducing a multiple activation of several phenylpropanoid branches to response to cold injury. By contrast, *ScAN1* may be considered a more specialized copy connected to the anthocyanin biochemical pathway. Therefore, we can suppose that after duplication *ScAN1* underwent to a sub-functionalization more specialized to anthocyanin production. An important clue of *ScAN2* function was suggested by the metabolic and microscopy analysis. In fact, it resulted that *ScAN2* is able to induce the

production of phenolic compounds on plasma membrane and cell wall of transformed cells. This characteristic is connected to the ability of plant to increase tolerance to low temperatures through lignification and suberin deposition on cell wall. Consistently, we found that ScAN2 was up regulated after cold treatment only in the cold tolerant *S. commersonii*. Based on the fact that also André et al. (2009) analysed *ScAN2* up regulation after drought stress in potato, it is likely that *ScAN2* has a role connected with osmotic stress response. Ultimately, we believe that this study provides a framework to explore how the phenylpropanoid genes evolved to control different branches of phenylpropanoid pathway to tackle external stresses.

As discussed previously, another important characteristic of flavonoid is the beneficial effects they have on human health acting as free radical scavengers. Recently, it has been discovered that the overexpression of the transcription factor *AtMYB12* redirects the primary metabolism flux to enhance the production of bioactive compounds in tomato fruit (Zhang et al., 2015). A particularly evident effect of *AtMYB12* is the induction of flavonols that accumulate in tomato flesh (Luo et al., 2008; Zhang et al., 2015). In the fourth chapter of this thesis we tried to understand which were the genes that normally contribute to flavonol accumulation in tomato. The most significant outcome from this research was the identification of two potential negative regulators we named *SlMYB4* and *SlELVIRA*. From our results it is possible to hypothesize that a single amino acid substitution found in the *S. pennellii* allele (*SpMYB4*) may affect its function. This can explain why in the introgression lines harbouring *SpMYB4* an increase of flavonol gene expression in tomato flesh was found. More intriguing is *SlELIVRA*. In fact, no ortholog genes have been characterized before. We believe that the deleterious alternative splicing of *SlELVIRAb* in tomato skin may explain why flavonols are produced specifically in that tissue. A common aspect between *SlMYB4* and *SlELVIRA* is the putative association to UV-B response. In fact, it is know that *AtMYB4* controls the phenylpropanoid pathway and production of UV-protecting sunscreens metabolites such as flavonols and sinapate esters (Jin et al.,

2000). Similarly, the phylogenetic characteristic of SlELVIRA let us to suppose that it belongs to the group of COP1 proteins which are connected with the UV-B signal transduction. However, their role as repressor is particularly attractive to obtain enhanced flavonol lines. In fact, we believe that simply through their stable knock out it is potentially possible to increase rutin accumulation in the flesh, as tested by transient silencing (VIGS). *SlMYB4*, albeit it did not show a double influence in both flavonol and CGA content (as *SlELVIRA* had), is the best candidate to target. In fact, MYB TFs are particularly specific in target binding and may avoid pleiotropic effects when manipulated. Currently, within a collaboration with Prof. Cathie Martin from the John Innes Centre, a genome editing approach is being undertaken to knock out *SlMYB4* and, hopefully, to create a high producing flavonol tomato line.

5.1 References

André, C. M., Schafleitner, R., Legay, S., Lefèvre, I., Aliaga, C. A. A., Nomberto, G.,... & Evers, D. (2009). Gene expression changes related to the production of phenolic compounds in potato tubers grown under drought stress. *Phytochemistry*, *70*, 1107-1116.

Fraser, C. M., & Chapple, C. (2011). The phenylpropanoid pathway in Arabidopsis. *The Arabidopsis Book*, *9*, e0152.

Jin, H., Cominelli, E., Bailey, P., Parr, A., Mehrtens, F., Jones, J., ... & Martin, C. (2000). Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis. *The EMBO Journal*, *19*, 6150-6161.

Korkina, L. G. (2007). Phenylpropanoids as naturally occurring antioxidants: from plant defense to human health. *Cell Mol Biol*, *53*, 15-25.

Liu, Y., Lin-Wang, K., Deng, C., Warran, B., Wang, L., Yu, B., ... & Allan, A. C. (2015). Comparative transcriptome analysis of white and purple potato to identify genes involved in anthocyanin biosynthesis. *PloS one*, *10*, e0129148.

Liu, Y., Lin-Wang, K., Espley, R. V., Wang, L., Yang, H., Yu, B., ... & Allan A. C. (2016). Functional diversification of the potato R2R3 MYB anthocyanin activators AN1, MYBA1, and MYB113 and their interaction with basic helix-loop-helix cofactors. *Journal of Experimental Botany*, 67:2159-2176.

Luo, J., Butelli, E., Hill, L., Parr, A., Niggeweg, R., Bailey, P., ... & Martin, C. (2008). AtMYB12 regulates caffeoyl quinic acid and flavonol synthesis in tomato: expression in fruit results in very high levels of both types of polyphenol. *The Plant Journal*, *56*, 316-326.

Ramakrishna, A., & Ravishankar, G. A. (2011). Influence of abiotic stress signals on secondary metabolites in plants. *Plant Signaling & Behavior*, *6*, 1720-1731.

Zhang, Y., Butelli, E., Alseekh, S., Tohge, T., Rallapalli, G., Luo, J., ... & Martin, C. (2015). Multi-level engineering facilitates the production of phenylpropanoid compounds in tomato. *Nature Communications*, *6*, 1-11.