# Identification and characterization of regulatory proteins involved in anthocyanin biosynthesis in *Fragaria vesca* and *Rubus idaeus*

Dissertation

zur

Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

dem

Fachbereich Pharmazie der

Philipps-Universität Marburg

vorgelegt von

**Andrea Lorena Herrera Valderrama**

Aus **Villavicencio, Colombia**

Marburg/Lahn **Jahr 2019**

Erstgutachter: **Prof. Dr. Maike Petersen**

Zweitgutachter: **Dr. habil.  Stefan Martens**


Eingereicht am **21.11.2018**

Tag der mündlichen Prüfung: **18.01.2019**

Hochschulkennziffer: 1180

**Oral and poster presentations**

Oral Presentation: Effect of bHLH gene regulators of the anthocyanin pathway in wild strawberries on the polyphenol content of commercial strawberry fruit. XXVIIIth International Conference on Polyphenols. July 11-15, 2016. Vienna, Austria.

Poster: Regulation of anthocyanin-related biosynthetic genes by bHLH transcription factors in *Rubus idaeus* and *Fragaria vesca*. 8th International Workshop on Anthocyanins. September 16-19, 2015. Montpellier France.

Oral Presentation: bHLH transcription factors involved in the regulation of anthocyanin genes during fruit development in wild strawberry (*Fragaria vesca*) and red raspberry (*Rubus idaeus*). Postgraduate workshop of the section „Pflanzliche Naturstoffe" (Deutsche Botanische Gesellschaft). September 21-23, 2014. Nuernberg, Germany.

Poster and Oral Presentation: Activation of anthocyanin-related biosynthetic genes by bHLH transcription factors in *Rubus idaeus* and *Fragaria vesca*. 7th International Rosaceae Genomics Conference. June 22-25, 2014. Seattle, USA.

Poster: Analysis of anthocyanin-related transcription factors during fruit development in *Rubus idaeus* and *Fragaria vesca*. 7th International Workshop on Anthocyanins. September 09-11, 2013. Porto. Portugal.


Visits to external laboratories

-   Visit 1: November 2012 to September 2013
Laboratory of Breeding & Genomics research, Plant and Food Research, Auckland, New Zealand.
Objective: *In silico* identification of bHLH proteins and MYB repressor proteins in *Rubus idaeus* and *Fragaria vesca*?
Supervisor: Dr. Andrew Allan


-   Visit 2: June 2015 to August 2015
Laboratory of Biotechnology of Natural Products, TUM Technische Universität München. Freising, Germany. Funding provided by the COST Action FA1306.
Objective: Developing RNA interference constructs and training on fruit agroinfiltration technique to study the effect of bHLH proteins on color formation in strawberry fruit.
Supervisor: Dr. Wilfred Schwab

# ABSTRACT

Primary metabolites like carbohydrates, lipids, proteins, and nucleic acids, work as key components to sustain the plant. Besides those compounds, a wide array of so called 'secondary metabolites' are produced by the plant, which have been given much attention in the scientific literature due to their beneficial effect for the plant and the possible positive effect on human health. The number of identified plant secondary metabolites, exceeds 100.000 structures, which briefly can be classified as terpenoids, alkaloids, glucosides, sterols and, last but not least, phenylpropanoids.

*Fragaria vesca* and *Rubus idaeus*; woodland strawberry and commercial raspberry, respectively, are two of the most popular berries on the market. The characteristic red coloration of their fruits is caused by the presence of anthocyanin pigments, secondary metabolites from the class of polyphenols. Those compounds are mainly produced during the late stages of fruit maturation and are essential phenotypic features; making plant breeders around the world consider these plant metabolites as a trait to follow up.

The anthocyanin biosynthesis pathway is well studied in model plants. It is regulated at the transcriptional level by the well-known MBW complex. This complex is formed by the interaction of three different types of transcription factors (TFs): MYB, bHLH, and WD40, which have already been characterized in *Arabidopsis thaliana*, ornamental plants as *Antirrhinum majus* and *Petunia hybrida*, and even in some crops of major economic importance such as corn (*Zea mays*), soybean (*Glycine max*), and apple cultivars (*Malus domestica*). Nowadays, the level of complexity of the regulatory process of the anthocyanin biosynthesis pathway is becoming clear – one gene at a time. This regulation includes TFs, the promoter regions of the genes that are involved and the chromatin modifications necessary to carry out gene activation and consequent translation for the formation of each specific enzyme that will lead to the final anthocyanin formation inside the cells.

The recent sequencing and annotation of the genomes of strawberry and raspberry as well as the possibility of transformation and the high amount of health-promoting anthocyanins present in the berries potentially make these plants great model systems to study the regulation of anthocyanin biosynthesis. This study aims to identify the role of bHLH proteins from raspberry and strawberry involved in anthocyanin biosynthesis.

Based on the *A. thaliana* bHLH classification and phylogenetic studies reported, the genomes of *F. vesca* and *R. idaeus* were screened, and putative gene candidates were found for both species. Posterior sequence analyses based on protein primary structure and motif conservation were performed, and a total of 98 protein-coding sequences were found in *F. vesca* genome v1.0 and 90 sequences in the unreleased draft version of the *R. idaeus* genome. The *in silico* results obtained in chapter 3.1 provide three and two gene candidates for the woodland strawberry and raspberry, respectively: *Fv3-FV2G25270, Fv33-FV7G08120, Fv145-FV5G02910, Ri3 gene36602* and *Ri3-gene26116*.

After the identification of putative bHLH candidate genes, those genes were analyzed during fruit development and their function was studied *in vitro* and *in vivo* (chapter 4 and 5). The results of the study presented here forms the beginning of a possibility to breed new berry varieties with better traits, such as higher resistance to various stresses and a with a positive effect on the health of the consumer.

# ZUSAMMENFASSUNG

Primärmetabolite wie Kohlenhydrate, Lipide, Proteine und Nukleinsäuren sind Schlüsselbestandteile der Pflanzen und notwendig für deren Entwicklung. Neben diesen Primärmetaboliten produzieren Pflanzen aber auch sogenannte sekundäre Pflanzenstoffe, die auch für das Leben der Menschen eine große Bedeutung haben, z.B. als Geruchs-und Geschmacksstoffe oder als Arzneistoffe. Die Vielfalt dieser Pflanzenstoffe wird in vielen essbaren Pflanzen erforscht und belangt sich derzeit auf mehr als 100.000 Stoffe. Diese Untersuchungen haben gezeigt, dass das Vorkommen dieser Stoffe nicht nur nützlich für die Entwicklung der Pflanzen ist, sondern auch, dass diese Stoffe einen positiven Einfluss auf die Gesundheit des Menschen haben können. Zu den sekundären Pflanzenstoffen zählen Terpenoide, Alkaloide, Glucosinolate, Sterole und Polyphenole.

*Fragaria vesca* und *Rubus idaeus*, Wald-Erdbeere und Himbeere, sind zwei der beliebtesten Beerenfrüchte auf dem Weltmarkt. Die charakteristische rote Farbe der Früchte wird durch das Auftreten von Anthocyanpigmenten erzeugt, Sekundärmetabolite aus der Klasse der Polyphenole. Diese werden hauptsächlich während der Reifung der Beeren von der Pflanze produziert und Züchter weltweit benutzen das Vorkommen dieser Stoffe als typisches Merkmal in der Züchtung von Beerenfrüchten.

Der Biosyntheseweg der Anthocyane ist sehr ausgiebig untersucht worden in Modellpflanzen. Die Regulation des Biosyntheseweges erfolgt auf dem Level der Transkription durch einen Komplex von Transkriptionsfaktoren, dem sogenannten MBW- oder MYB-bHLH-WD40-Komplex. Die Proteine dieses Komplexes und ihre Funktionen wurden in *Arabidopsis thaliana* untersucht, in den Zierpflanzen *Antirrhinum majus* und *Petunia hybrida* sowie in Feldfrüchten, die für uns Menschen von hoher wirtschaftlicher Bedeutung sind, wie Mais (*Zea mays*), Sojabohne (*Glycine max*) und Apfelsorten (*Malus domestica*). Die Vielschichtigkeit der Regulierung dieses Biosyntheseweges wird immer offensichtlicher mit jedem Gen, das genauer untersucht wird. Es ist bekannt, dass Transkriptionsfaktoren, die Promoter-Regionen von involvierten Biosynthese-Genen sowie die Modifikation des Chromatins durch Chromatin-verändernde Proteine bei der Bildung von Anthocyanen in der Zelle eine Rolle spielen.

Die kürzlich erfolgte Sequenzierung und Genomannotation der Wald-Erdbeere und der Himbeere, die Möglichkeit der Transformation dieser Pflanzen sowie der extrem hohe Anteil an gesundheitsfördernden Anthocyanen in den Früchten, machen diese Pflanzen zu grossartigen Modell-Pflanzen, um die Regulierung des Anthocyan-Biosyntheseweges zu untersuchen. In der hier vorgelegten Studie wird das Ziel verfolgt, die Funktion der bHLH Proteine, die am Anthocyan-Biosyntheseweg beteiligt sind, zu identifizieren.

Basierend auf der Einteilung der *Arabidopsis* bHLH Proteine und basierend auf veröffentlichten phylogenetischen Studien wurden die Genome beider Pflanzen gescreent und putative Gen-Kandidaten wurden identifiziert. Mit der erfolgten Sequenz-Analyse basierend auf der Primärstruktur der Proteine und basierend auf Motiven/Domänen in der Protein-Tertiärstruktur, wurden 98 Sequenzen im *F. vesca* Genom v1.0 und 90 Sequenzen in der unveröffentlichten Entwurfsversion des *R. idaeus* Genoms gefunden, die für bHLH Proteine kodieren. Im Kapitel 3.1 werden die *in silico* Ergebnisse vorgestellt. Es wurden drei Kandidaten-Gene in *F. vesca* und zwei Kandidaten-Gene in *R. idaeus* gefunden, die für mögliche bHLH Proteine kodieren, die im Biosyntheseweg der Anthocyane eine Rolle spielen; *Fv3-FV2G25270*, *Fv33-FV7G08120*, *Fv145-FV5G02910*, *Ri3 gene36602* and *Ri3-gene26116*.

Nach der Identifizierung der bHLH Kandidaten-Gene wurden diese in verschiedenen Entwicklungsstadien in Früchten analysiert und ihre Funktion *in vitro* und *in vivo* untersucht (Kapitel 4 und 5). Die Ergebnisse unserer Untersuchungen bilden den Anfang für die Möglichkeit der Züchtung neuer Erdbeer- oder Himbeersorten mit verbesserten Eigenschaften wie Stressresistenz und einem verbesserten Effekt auf die Gesundheit der Verbraucher.

**TABLE OF CONTENTS**

# ABKÜRZUNGEN

| | |
|---|---|
| PAs | Proanthocyanidins |
| FAOSTAT | Food and Agriculture Organization Corporate Statistical Database |
| DFR | Dihydroflavonol 4-reductase |
| LDOX | Leucoanthocyanidin dioxygenase |
| ANS | Anthocyaninidin synthase |
| PAL | Phenylalanine ammonia-lyase |
| 4CL | 4-coumarate CoA ligase |
| C4H | Cinnamate 4-hydroxylase |
| CHS | Chalcone synthase |
| CHI | Chalcone isomerase |
| F3H | Flavanone 3-hydroxylase |
| TF | Transcription Factor |
| bHLH | basic Helix-Loop-Helix |
| ABA | Abscisic acid |
| PAP | Purple acid phosphatases |
| WBM | WD-40, bHLH, MYB |
| TT | Transparent testa |
| TTG1 | Transparent testa glabra 1 |
| DNA | Deoxyribonucleic acid |
| RNA | Ribonucleic acid |
| PIF | Phytochrome-interacting factors |
| HKG | Housekeeping gene |
| PCR | Polymerase chain reaction |
| qPCR | Quantitative qPCR |
| UHPLC | Ultra-High Performance Liquid Chromatography |
| TYR | Tyrosine |
| ACT | Actin |
| BLAST | Basic Local Alignment Search Tool |
| HMM | Hidden Markov Model |
| MUSCLE | Multiple Sequence Comparison by Log- Expectation |
| NAD | Nicotinamide adenine dinucleotide |
| ML | Maximum likelihood estimation |
| JTT | Jones–Taylor–Thornton model |
| PFR NZ | Plant Food and Research New Zealand |

# Chapter 1

General introduction

## GENERAL INTRODUCTION

**Strawberry & Raspberry: red soft fruits**

*Fragaria vesca* L. and *Rubus idaeus* L., or as commonly called wild strawberries and red raspberries, respectively, are both diploid members of the Rosaceae family. This plant family contains circa 3,000 of diploid and polyploid species, including important crops such as those belonging to the genera *Prunus* (almond and all stone fruits: peach, apricot, cherry and plum), *Malus* (apple), Pyrus (pear), *Rosa* (rose), *Rubus* (raspberry) and *Fragaria* (strawberry) (1, 2).

The Rosaceae family contains 73 genera, 1283 accepted taxa and the taxonomical classification for both species is (3, 4, 5):

Kingdom Plantae – Plants
Subkingdom Tracheobionta – Vascular plants
Superdivision Spermatophyta – Seed plants
Division Magnoliophyta – Flowering plants
Class Magnoliopsida – Dicotyledons
Subclass Rosidae
Order Rosales
Family Rosaceae

**Strawberries: genus *Fragaria***

Common strawberry fruits (*Fragaria x ananassa* L.) are an important commercial crop, widely grown in temperate regions all over the world, and strawberry is one of the most important berries of the market especially during the summer period of the northern hemisphere. According to the FAOSTAT report (year), around 6 million tons of berries were produced during the last three years (per year), and the worldwide strawberry production grew 49,8% in the previous 15 years, as can be seen in Figure 1 (2).

The strawberry fruit in reality is an aggregate accessory fruit; the botanical fruits are dry achenes that are embedded in the fleshy receptacle. The fleshy part is derived not from the plant ovaries but from the receptacle tissue of the flower (2, 5, 6).
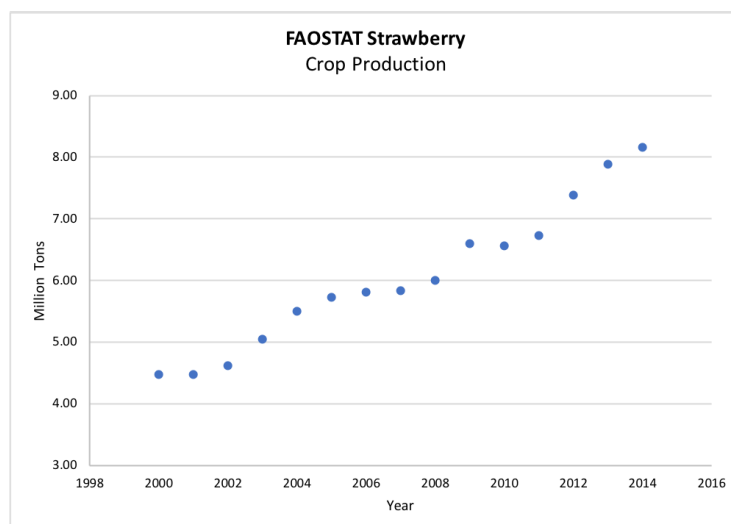
Figure 1: Commercial strawberry production (*Fragaria x ananassa*) of the past 10 years according to FAOSTAT (10).

The commercial strawberry *F.* x *ananassa* is an octoploid species (8 copies of each chromosome; 8n), and due to its complex genetic configuration, the scientific community decided to focus their genetic and genomic studies on the diploid woodland strawberry *F. vesca* as a model species for the entire genus (4, 5).

The so-called 'semperflorens' or 'alpine' forms of *F. vesca* ssp. *vesca* have been cultivated for centuries in European gardens, from the magnificent French gardens from Louis XV to the English gardens of the Royal Horticultural Society (4, 5).

Darrow (5) described these plants as follows: "*Fragaria* genus plants are erect, 10-30 centimeters high, with runners; leaves are thin, and light green with slender petioles, glabrous or becoming so above, lighter colored and lightly silky-hairy beneath, at least on the veins; leaflets are nearly sessile, rather small, relatively narrow, cuneate-ovate to rhombic-ovate with large sharp serrations; petioles and peduncles have few but generally spreading soft hairs; inflorescence is small, on usually tall inflorescences, equaling or exceeding the leaves; flowers are actinomorphic, white (sometimes tinged with pink), and usually 5-petalled about 1 centimeter in diameter, bisexual; fruit is hemispherical (in the type and seedling raised) flesh extremely soft, pulpy, generally aromatic to highly aromatic; seeds are small, raised, very prominent; calyx is reflexed, widely spreading" (2, 7).

**Crop origin**

The commercial strawberry *F.* x *ananassa* is a hybrid resulting from the cross of two American species: the North American *F. virginiana* and the South American *F. chiloensis*. The spontaneous cross among the American *Fragaria* species occurred when they were imported and maintained in European gardens during the XVII century. Both *Fragaria* ancestors were broadly cultivated by the native prehispanic indigenous communities, in South America by the Mapuches tribes and the Anishinaabe and Wampanoag

tribes on the Northern Continent. The geographical distance between the two species is the most probable reason for the lack of crossing events before the XVII century (5, 7).

Once the European breeders found the strawberries from the new world, many propagations and crossings were performed obtaining new varieties. An interesting trait for the breeders was the big fruit size observed in the Chilean strawberries (*F. chiloensis*), a quality feature that was lacking in the small European genotypes. Other crosses were focused on obtaining new varieties with intense color, aromas, and yield. Those trials were made between *F. vesca*, *F. moschata*, and *F. viridis* among other species. However, the only successful cross-species result was *F. x ananassa* (5, 7).

Antoine Nicolas Duchesne, a young enthusiastic French botanist, working on the Versailles gardens, listed in his work "L'Histoire Naturelle des Fraisiers" (1766) ten species and nine varieties of strawberry. Later, in 1781, in an article in the Encyclopédie Méthodique Botanique of Lamarck, he listed already twenty-five varieties of strawberries (5). This example represents a good illustration of the breeder's efforts in developing varieties for the growing European market as early as in the 18th century (5).

**Chemical composition and health benefits**

*F. vesca*, being a wild species with low production yield for the market requirements, has been studied in less proportion than the commercial hybrid *F. x ananassa.* Taking as a reference the commercial hybrid, these fruits are an excellent source of folates, potassium, and vitamin C (Table 1) (5, 7, 8).

Table 1. Nutritional facts for 100 grams of raw commercial strawberries *Fragaria x ananassa*. Source: Marlett1992 (8).

| Compound | Value 100 g | Units |
|---|---|---|
| Sugars | 4.9 | g |
| Proteins | 0.7 | g |
| Omega-3 fatty acids | 65 | mg |
| Omega-6 fatty acids | 90 | mg |
| Vitamin C | 58.8 | mg |
| Folate | 24 | µg |
| Calcium | 16 | mg |
| Phosphorus | 24 | mg |
| Potassium | 153 | mg |
| Water | 90.9 | g |

Strawberry fruits also contain elevated levels of phenolic compounds, including anthocyanins, condensed tannins also called proanthocyanidins (PAs), derivatives of various flavonoids, phenolic acids, and hydrolyzable tannins or ellagitannins (2, 7, 8). Studies on the nutritional impact of strawberry fruit consumption are increasing nowadays. The primary results of recent studies demonstrate that strawberry metabolites are involved in the prevention of particular kinds of cancers, anti-inflammatory events, and heart diseases (7). This activity is most probably due to the beneficial effects of secondary metabolites

and their antioxidant capacity, especially of the phenolic compounds. As shown in Table 2, strawberry fruits contain different types of polyphenols, pelargonidin and cyanidin being the most abundant anthocyanin types (Table 2) (5, 7, 9).

Table 2: Main polyphenols quantification on 100 g of commercial strawberries *Fragaria x ananassa*. Source: The Phenol-Explorer database (9).

| Compound | Value per 100 g | Unit |
|---|---|---|
| Cyanidin | 1.68 | mg |
| Petunidin | 0.1 | mg |
| Delphinidin | 0.3 | mg |
| Malvidin | 0 | mg |
| Pelargonidin | 24.9 | mg |
| Peonidin | 0.1 | mg |
| (+)-Catechin | 3.1 | mg |
| (-)-Epigallocatechin | 0.8 | mg |
| (-)-Epicatechin | 0.4 | mg |
| (-)-Epicatechin 3-gallate | 0.1 | mg |
| (-)-Epigallocatechin 3-gallate | 0.1 | mg |
| (+)-Gallocatechin | 0 | mg |
| Naringenin | 0.3 | mg |
| Apigenin | 0 | mg |
| Luteolin | 0 | mg |
| Isorhamnetin | 0 | mg |
| Kaempferol | 0.5 | mg |
| Myricetin | 0 | mg |
| Quercetin | 1.1 | mg |
| Proanthocyanidin dimers | 5.2 | mg |
| Proanthocyanidin trimers | 5.7 | mg |
| Proanthocyanidin 4-6mers | 23.3 | mg |
| Proanthocyanidin 7-10mers | 16.9 | mg |
| Proanthocyanidin polymers (>10mers) | 54.2 | mg |

## Raspberry: *Rubus idaeus*

The genus *Rubus* L. has between 700-800 species distributed over four of the five continents, growing at elevations from sea level to 4500 meters. Most *Rubus* species are perennial shrubs with biennial canes. Of all the species, three have a commercial importance: red raspberry (*R. idaeus* L.), black raspberry (*R. occidentalis* L.) and blackberry (*R.* sp., subgenus *rubus* L.), being raspberries and black raspberries part of the subgenus *Idaeo batus* (3, 10, 11). As seen in Figure 2, the raspberry production worldwide has an increasing tendency in the last years.

As described in Jennings (12): "Raspberry leaves are alternate, usually divided into 3-5 leaflets which are arranged pinnately, pedately, or less commonly palmately, but infrequently undivided. Flowers are in clusters, racemes or panicles, but are occasionally solitary, and are generally white but sometimes pink to rosy-purple. Each flower has five petals, five sepals, five bracts, numerous stamens, and several pistils clustered on a cone-shaped core known as a receptacle. Botanically, the fruits are not berries (although they are usually called berries), but are coherent aggregations of tiny drupelets. Fruits separate from the receptacle when picked with each raspberry resembling a hollow cone"; so basically, raspberries and blackberries are botanically separated by whether the receptacle of the fruit remains in it when the fruit is

picked in which case it is considered a blackberry, or if its remains on the plant leaving a hollow center to the fruit is considered a raspberry (12, 13).
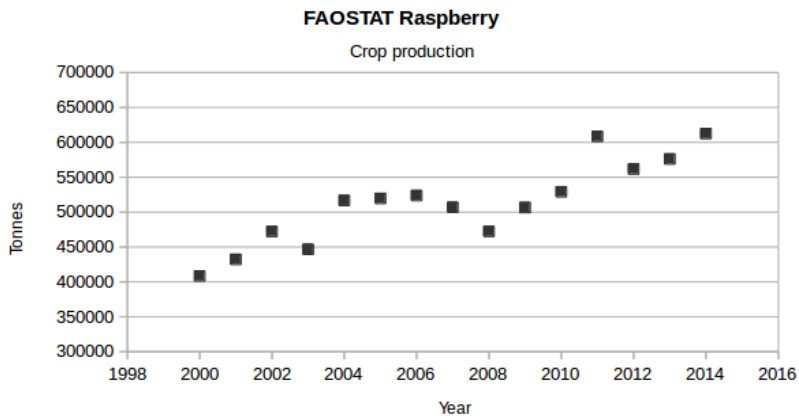


Figure 2: Commercial raspberries production (*Rubus idaeus*) of the past 10 years according to FAOSTAT (2).

**Origin**

The center of diversity of *Rubus* is believed to be in Northern China, where 250-300 species have been identified so far. While found on all continents except Antarctica, raspberries are most abundant in the Northern Hemisphere. The raspberry fruits have been cultivated before the fourth century. Since then, the plant has been genetically improved, first by the development of local varieties, and since about 1800 by deliberate hybridization among the American and European types (12, 13).

Raspberries were first introduced into cultivation in Europe nearly 450 years ago. By the early nineteenth century, more than 20 cultivars of red raspberry were grown in both England and the USA. English cultivars were then exported to the USA, where crosses between them and North American seedlings resulted in new improved varieties (13, 14).

Red raspberries are the most widely grown type of raspberry, while black raspberries are favorite only in some areas of the eastern USA and some parts of Europe. The progenies of black and red raspberries have purple fruits and canes; these types are widespread in eastern North America. Yellow-fruited *R. idaeus*, caused by a recessive mutation, is also grown on a limited scale for specialty markets due to their scarce yield (12, 14).

**Chemical composition and health benefits**

The main constituent of the raspberry fruit is water (ca. 87%). Of the remaining solids, 9% are soluble and the rest insoluble. Pectins compose 0.1 - 1.0% of the soluble fraction, but this amount decreases with

ripening due to hydrolysis. The main sugars are glucose, fructose and a smaller amount of sucrose. These are the major soluble component of the juice. A typical ripe raspberry fruit contains 5-6% sugar. Ascorbic acid (vitamin C) is the seventh largest component (Table 3) (13).

Table 3. Nutritional facts for 100 grams of raw commercial raspberries. Source: Marlett (8).

| Compound | Value 100 g | Units |
|---|---|---|
| Sugars | 4.4 | g |
| Proteins | 1.2 | g |
| Omega-3 fatty acids | 16.0 | mg |
| Omega-6 fatty acids | 249.0 | mg |
| Vitamin C | 26.4 | mg |
| Folate | 21.0 | µg |
| Calcium | 25.0 | mg |
| Phosphorus | 29.0 | mg |
| Potassium | 151 | mg |
| Water | 87.5 | g |

As in strawberries, in addition to being rich in the traditionally evaluated nutrients such as vitamin C, raspberries present elevated levels of polyphenols represented mostly by phenolic acids, flavonoids (flavonols, flavan-3-ols, anthocyanins), ellagitannins and PAs. A considerable amount of new research has been performed on variation patterns in the antioxidant capacity of *Rubus* species and hybrid crosses. As shown in Table 4, the main anthocyanin present in raspberries is cyanidin, followed by lower levels of petunidin, delphinidin and the cyanidin derivative peonidin (8, 9, 13).

Table 4: Main polyphenols quantification on 100 g of commercial raspberries. Source: The Phenol-Explorer database (9)

| Compound | Value per 100 g | Unit |
|---|---|---|
| Cyanidin | 45.77 | mg |
| Petunidin | 0.3 | mg |
| Delphinidin | 1.3 | mg |
| Malvidin | 0.1 | mg |
| Pelargonidin | 1 | mg |
| Peonidin | 0.1 | mg |
| (+)-Catechin | 1.3 | mg |
| (-)-Epigallocatechin | 0.5 | mg |
| (-)-Epicatechin | 3.5 | mg |
| (-)-Epicatechin 3-gallate | 0 | mg |
| (-)-Epigallocatechin 3-gallate | 0.5 | mg |
| (+)-Gallocatechin | 0 | mg |
| Naringenin | 0 | mg |
| Apigenin | 0 | mg |
| Luteolin | 0 | mg |
| Isorhamnetin | 0 | mg |
| Kaempferol | 0.1 | mg |
| Myricetin | 0 | mg |
| Quercetin | 1.1 | mg |
| Proanthocyanidin dimers | 11.8 | mg |
| Proanthocyanidin trimers | 5 | mg |
| Proanthocyanidin 4-6mers | 9 | mg |
| Proanthocyanidin 7-10mers | 1.1 | mg |
| Proanthocyanidin polymers (>10mers) | 0 | mg |

**Phenylpropanoids**

The number of predicted secondary plant metabolites exceeds 100.000 compounds (15), which briefly can be classified as terpenoids, alkaloids, sterols and last but not least phenylpropanoids. In general, many of the plant secondary metabolites are involved in plant responses to environmental influences, but also play a significant role in providing health-promoting compounds for humans (16). Studies of secondary plant metabolites are an essential part of any comprehensive research, and these studies help to reveal the richness and diversity of plant compounds in any major crop species.

In the case of berries of the Rosaceae family, the major types of secondary metabolites are the phenylpropanoids.

The phenylpropanoid-flavonoid biosynthesis pathway is one of the most intensively studied pathways in plant secondary metabolism. Phenylpropanoid compounds can have multiple functions for plant defense responses to light and UV stress, temperature (resistance to cold), water deficit and pathogen attack (16, 17).

In the last few years, phenylpropanoids, especially flavonoids, coming from unprocessed fruits and vegetables, were the subject of interest due to their possible beneficial effects on human health. The potential health benefits of dietary phenylpropanoids and in particular polyphenols have been proven in several studies on experimental models from human tissue culture to animal feeding assays. As an example, Mandave *et al.* (15) reported: "Treatment with strawberry extracts improved lipid profile, liver function, and serum creatinine and led to a significant increase in antioxidant status in diabetic rats" (15,16).

In the case of *Fragaria* and *Rubus* species, the red coloration of their fruits is due to the presence of anthocyanins. Anthocyanins are water-soluble pigments that are synthesized via the phenylpropanoid-flavonoid pathway. More specifically, they are a class of flavonoids initiated by the condensation of three malonyl-CoA molecules and one p-coumaroyl-CoA. Anthocyanins share the same upstream pathway with PAs, containing the typical C6-C3-C6 flavonoid skeleton. It is not until the latest steps that the anthocyanins are glycosylated and acylated through substitutions in their primary structure, generating their chemical diversity (16,18).

**Structural genes of the phenylpropanoid biosynthetic pathway**

The phenylpropanoid pathway starts from the shikimate pathway, responsible for the biosynthesis of folates and aromatic amino acids as phenylalanine, tyrosine, and tryptophan. The initial steps of the phenylpropanoid pathway are catalyzed by phenylalanine ammonia lyase (PAL), cinnamate 4-

hydroxylase (C4H), and p-coumaroyl CoA-Ligase (4CL). These enzymes are essential and provide the basis for all subsequent branches. The next three steps are catalyzed by chalcone synthase (CHS), chalcone isomerase (CHI), and flavanone 3ß-hydroxylase (F3H syn. FHT), respectively. As a result, dihydrokaempferol, characterized by a hydroxyl group at C4' in the B-ring, is produced. The subsequent hydroxylation of dihydrokaempferol at C3' is catalyzed by the flavonoid 3'-hydroxylase (F3'H) leading to the synthesis of dihydroquercetin (syn. Taxifolin) (16, 19, 20).

Most of all the structural genes of the phenylpropanoid biosynthetic pathway have been identified and characterized in *Arabidopsis thaliana*, and other important crop species such as *Zea mays*, *Vitis vinifera* and *Nicotiana tabacum* and even enzymes had been described in *Populus tremula* (poplar). The phenylpropanoid biosynthetic pathway can be divided into two steps: early and late (see Figure 3).
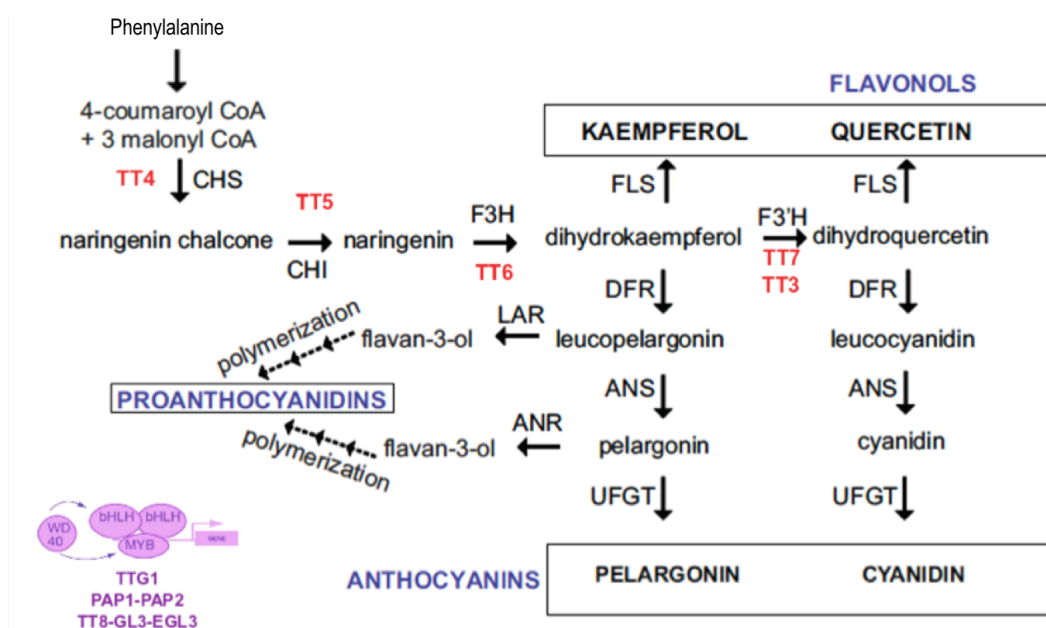


Figure 3: Overview of the *A. thaliana* phenylpropanoid pathway including the main enzymes in black color. tt mutants in capital red colors. End products are in capital letters: ANTHOCYANINS, PROANTHOCYANIDINS. Purple letters, represent transcriptional regulators (Modified from Falcone *et al.* 21).

The late steps of the anthocyanin pathway include steps from dihydroflavonols through leucoanthocyanidins to anthocyanidins as well as the further modifications of anthocyanidins. The steps from dihydroflavonols to anthocyanidins are consecutively catalyzed by dihydroflavonol 4-reductase (DFR) and anthocyanidin synthase (ANS, also called leucoanthocyanidin dioxygenase, LDOX).

Specific glycosylation, acylation, and methylation are modifications that change the biological properties of the conjugated molecules and are necessary for the proper stability, solubility, and localization of these molecules. These chemical modifications occur on anthocyanidins, proanthocyanidins, flavones, flavan 4-ols, etc. *A. thaliana* Transparent Testa (tt) mutant lines, have the enzymatic gene or transcriptional regulators altered, causing an effect on flavonoid synthesis (16, 19).

## Genetic regulation

## Transcription factors and promoter regions

A significant proportion of protein-encoding genes is dedicated to the control of gene expression. For example, the genome of *A. thaliana* includes 27,416 protein-coding genes (TAIR10, http://arabidopsis.org), of which 6% (more than 1700) encode transcription factors (TFs). The function of a few TFs has remained conserved between plants and animals (separated by over a billion years of evolution). Examples include members of the E2F family, which control core cell-cycle functions. However, most of the other TFs have significantly diverged in function since the separation of plants and animals, and approximately 45% of *Arabidopsis* TF belongs to families that are specific to plants (22, 23).

Work performed on model plants pinpointed the tight regulation of the flavonoid biosynthetic pathway during plant development. It is now established that the transcriptional regulation of the structural genes is controlled by MYB and basic helix-loop-helix (bHLH) TFs, together with WD40 proteins. Particular attention has hitherto been devoted to MYB, as demonstrated by the reported publications (17, 24, 25). Herein, the recent advances in the knowledge of the transcriptional regulation of the flavonoid pathway are discussed, with a particular focus on bHLH TFs (23).

## Genetic regulation of the phenylpropanoid pathway

The flavonoid biosynthesis pathway has been proven to be regulated at the level of transcription of genes encoding enzymes of the biosynthetic steps by the well-known MBW complex (Figure 4). The interaction of three different TFs forms this complex: MYB, bHLH, and the WD40, which has already been characterized in the model plant *A. thaliana*, flower plants as *Antirrhinum majus* and *Petunia hybrida*, and on some mayor economical crops as corn (*Z. mays*), soybean (*Glycine max*), grapevine (*V. vinifera*), apple (*Malus domestica*), and even commercial strawberry (*F. x ananassa*). Nowadays the level of complexity that the whole regulatory process requires is being examined - one gene at the time. The regulatory process includes the promoter regions and the chromatin modifications, which are necessary to carry out the gene activation and consequent translation for the formation of each specific enzyme that will lead to the final flavonoid and/or anthocyanin formation inside the cells (20, 26).
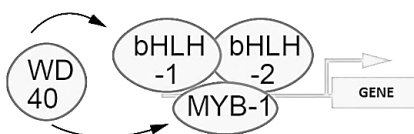


Figure 4: Schematic diagram representing the proposed MBW complex interaction during gene transcription.

**MYB proteins**

The first MYB TFs involved in regulation of the flavonoid pathway were identified in 1987 in maize and comprised C1 (Colorless 1) and Pl1 (Purple leaf 1), in addition to P1 (27). MYB TFs are characterized by the so-called N-terminal MYB domain, consisting of one to three imperfect repeats of almost 52 amino acids, R1, R2, and R3. Domains R2 and R3 can be easily recognized in Figure 5. While the MYB domain is involved in DNA binding and dimerization, the C-terminal region regulates target gene expression (activation or repression). MYB proteins carry out diverse regulatory functions inside the cells like ABA-response, cell cycle, biotic and abiotic stress and especially phenylpropanoid metabolism (23, 27, 28).
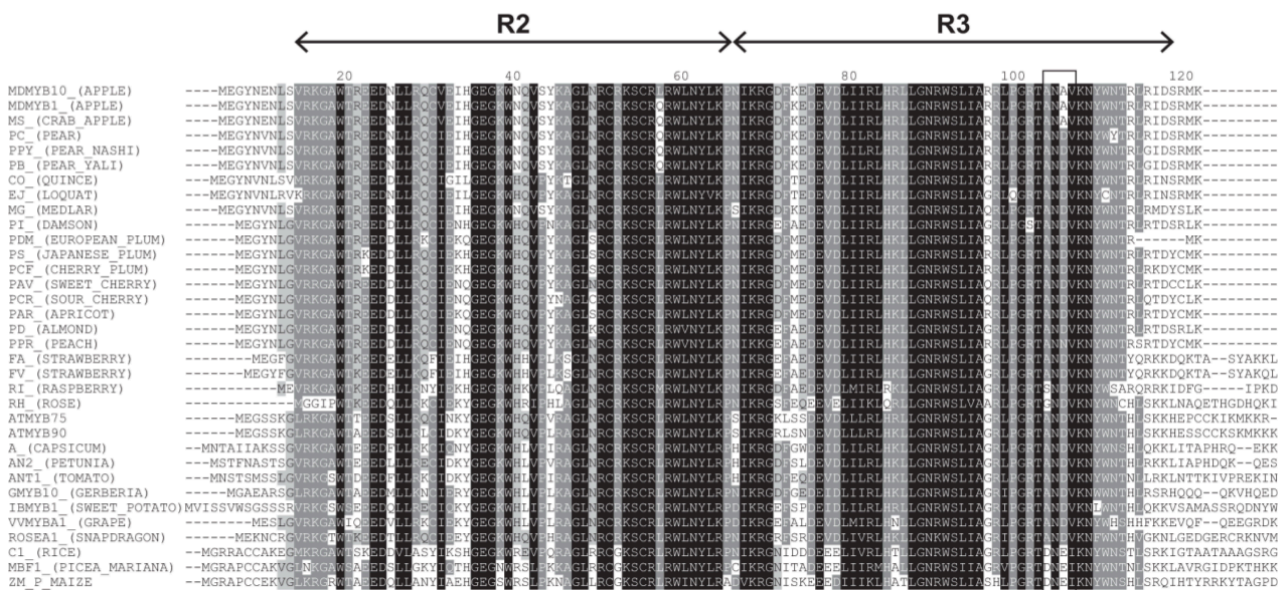


Figure 5: Amino acid sequence alignment of the R2 R3 region of reported MYB anthocyanin regulators. Source: Lin-Wang *et al.* 2010 (30).

MYB TFs regulating the phenylpropanoid pathway have been widely investigated and identified in diverse crops, ornamental and model plants, and according to the classification of Stracke *et al.* (28), these particular MYB proteins belong to subgroups one to seven of MYB family proteins.

Four MYB TFs, PAP1/MYB75, PAP2/MYB90, MYB113, and MYB114 with relatively high sequence similarities, have been identified to control anthocyanin biosynthesis in vegetative tissues in *A. thaliana*. All these four genes are R2R3-MYB proteins that contain two imperfect repeats in the MYB domain (see Figure 5; 8, 28).

One of the best-studied protein, the R2R3 MYB protein AtPAP1 (Production of Anthocyanin Pigmentation 1), is a master regulator of anthocyanin biosynthesis. *PAP1* in floral/fruit tissues is expressed at the highest level in comparison with its homolog expression in leaves. Another example is ZmPericarp1 (P1), an R2R3-MYB TF which can control the accumulation of various phenylpropanoids by activating a subset of phenylpropanoid biosynthetic genes in company with the C1 like R2R3-MYB (27).

**WD40-repeat proteins**

WD40 or WDR (WD repeat) proteins are involved in many eukaryotic cellular processes including cell division, vesicle formation and trafficking, signal transduction, RNA processing, and regulation of transcription in plants (23). They notably participate in chromatin remodeling, through modifications of the histone proteins, and can thus influence transcription. WD40 proteins have not any catalytic activity reported, but instead seem to be a docking platform for other proteins (23, 29, 30).

However, WD40 proteins are essential for the activity of the regulatory complex, as was demonstrated by the study of Carey *et al.* (30), on the maize *pac1* mutants and the TRANSPARENT TESTA GLABRA1 (TTG1) protein in *A. thaliana*. Both show high amino acid sequence identity with other WD40 proteins from diverse species, making gene complementation studies possible (23, 29, 30, 31).

Multiple experiments have demonstrated that *TTG1* is constitutively expressed in all tissues throughout the entire development of plants; besides, its expression does not respond to alteration of environmental conditions tested (19, 20, 32). All current data have shown that TTG1 has a central role in the WBM regulatory complexes to regulate epidermal cell fate and metabolic specificity leading to the production of anthocyanins and proanthocyanidins (33).

**bHLH family**

bHLH proteins are a group of TFs that regulate many essential physiological and developmental processes in eukaryotic cells. As described before, bHLH proteins are a big family of TFs present in the plant kingdom and are crucial regulators of the cell developmental processes in mammals and other animals (27).

The bHLH domain contains approximately 60 amino acids with two functionally distinctive regions, the basic region, and the helix-loop-helix (HLH) domain. The basic region consists of a stretch of ~13 mainly basic amino acids and can bind to the CANNTG sequence (E-box), where N corresponds to any nucleotide. Three amino acids in the basic region are highly conserved and directly involved in binding DNA (22, 23, 27).

The bHLH family has more than 140 genes annotated in *A. thaliana*, and 240 genes annotated in the *M. domestica* genome, including for example the proteins Phytochrome Interacting Factor3 (PIF3) and stomatal development genes (SPCH, MUTE and FAMA). Phylogenetic analyses have classified the diversity of bHLH proteins into many distinct groups. The main plant bHLH study was done by Heim et at. (22), where all *A. thaliana* bHLH protein sequences were analyzed based on their motifs and specific HxExR DNA binding domain, as previously done in human bHLHs by Atchley and Fitch (34). A total of 12

major groups or subfamilies were identified and classified for the model species *A. thaliana*, with proteins clustered on the tree having similar biological functions (27, 34, 35, 36).

Various phylogenetic analyses using *A. thaliana* and *Oryza sativa* bHLH proteins have been performed. According to these studies, bHLH proteins were classified into 15-25 subgroups with most groups containing additional conserved domains outside the bHLH domain. The first plant bHLH protein identified was the maize protein ZmRED1 (ZmR) and it has been shown to function as a co-regulator in phenylpropanoids biosynthesis (22).

In *A. thaliana*, four R-like bHLH proteins (GL3, EGL3, TT8, and AtMYC1) are present, and all seem to participate to a different extent in anthocyanin production. The *tt8* and *gl3* single mutants produce a significant amount of anthocyanin in hypocotyls and cotyledons of 5-day old seedlings, but neither *gl3/egl3* nor *gl3/egl3/tt8* triple mutants show any anthocyanin pigmentation (29). This suggests that EGL3 is the major contributor to anthocyanin biosynthesis but that TT8 and GL3 also play a role (22).

## MYB-WD-bHLH in Rosaceae

The genetic regulation of the phenylpropanoid pathway have also been studied in the Rosaceae family, especially in crop species. Apple (*M. domestica*) has been the most studied species regarding the relationship among the MYB TF and the red coloration in specific varieties.

In the last years, several publications have reported part of the regulatory and hierarchy network in which bHLH proteins are responsible for the correct maintenance of the transcription balance and consequent secondary metabolite formation (26); in the Rosaceae family the *M. domestica* MYB10 has been shown to be a potent activator of the whole pathway (31) and some studies for the commercial strawberry (*F. x ananassa*) confirm the same results on the role of FvMYB10. However, the part of the corresponding bHLH genes for the sequenced species *F. vesca* and their full transcription regulation network has not been established.

As an example, the gene *MdMYB10* was isolated from red-fleshed apples from the cultivar 'Red Field' in 2007 (31) and subsequently it was identified as the apple orthologue regulator, based on sequences from other species such as *Arabidopsis PAP1*. As described by Espley *et al.* (31): "differences in the activity of this gene (*MdMYB10*) that determine the color differences between the cortex tissues in these varieties" (31), a higher gene expression from *MdMYB10* is correlated with higher expression of the genes encoding the biosynthetic enzymes of the anthocyanin pathway, and an accumulation of anthocyanins in the fruit flesh. From this apple MYB discovery, the homologs from other Rosaceae species have been identified and characterized (17, 31).

*M. domestica MdTTG1* gene is the apple WD40 protein that promotes the accumulation of anthocyanins, and as the name suggests it is the homolog gene to *A. thaliana TTG1*. It was also proven that the apple WD40 protein MdTTG1 interacts with bHLH but not with MYB proteins to regulate anthocyanin accumulation (38).

Additional studies on *Prunus persica* have also been made, showing that a regulatory complex composed of MYB, bHLH3 and WD40 controls promoters of polyphenol biosynthetic steps in peach fruit, activating the transcription of several pathway genes (25). The flavonoid regulation in peach flowers was analyzed by Zhou *et al.* (39), and it was detected by transcriptomic means that not only one MYB but four MYB homologous genes were active in this tissue: *PpMYB9*, *PpMYB10*.2, *PpMYBPA1*, and the gene *Peace* (31).

Table 5: List of TFs involved in phenolic metabolism identified in the Rosaceae family members according to literature, table modified from (24).

| Species | MYBs | bHLH | WD40 |
|---|---|---|---|
| Strawberries | 4 | 2 | 1 |
| Pear | 2 | 2 | 1 |
| Nectarine | 2 | 3 | 0 |
| Apple | 6 | 2 | 1 |
| Cherry | 2 | 2 | 1 |
| Plum | 3 | 2 | 1 |
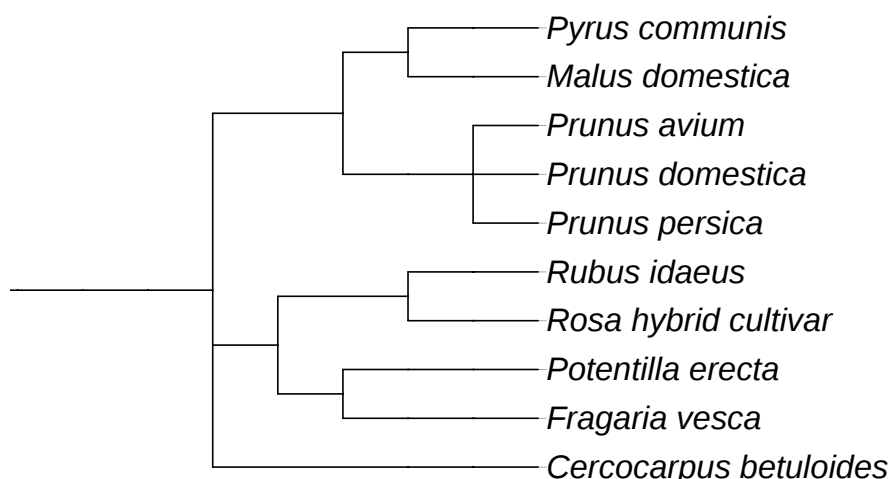


Figure 6: Phylogenetic trees based on the full NCBI taxonomy (scientific names) of the main crop fruit trees and bush species in the Rosaceae family. Phylogenetic tree generated by the website tree generator phyloT based on the NCBI taxonomy nucleotide and protein sequence databases (40).

In plum (*Prunus salicina*), three MYB homologs, and three bHLH expressed in a correlative way with the phenylpropanoid pathway genes were found (41). The reported results of Wei *et al.* (42) in sweet cherry

(*Prunus avium*) confirm the results of Starkevič *et al.* (43), where the bHLH and MYB correlation was also observed, Table 5 summarizes these number of proteins identified in the Rosaceae family (41, 42, 43).

In the Chinese sand pear (*Pyrus pyrifolia* cv. Aoguan), the *PyMYB10* gene is an ortholog of red-fleshed apple *MdMYB10* gene (18). Later, the same gene, *PyMYB10*, presented a homolog called *PyMYB10.1*. Interestingly, the *PyMYB10* gene was more activated by the *AtDFR* promoter in *N. benthamiana* leaves compared to the empty vector control, and, as the authors mention, it has been reported that distinct anthocyanin structural genes may not be regulated by a single MYB protein or bHLH protein in several plants (44).

In peach, three bHLH genes belonging to the clade of *Arabidopsis TT8*, *GL3* and *MYC1* were found in the peel of several varieties. The expression of these bHLHs was found to be tissue-specific, and *PpbHLH3* was the most expressed transcript.

In commercial strawberries, the MYB-bHLH-WD40 regulatory complex genes involved in the proanthocyanidin biosynthesis were identified and characterized. The role of the R2R3-MYB homologs FaMYB9 and FaMYB11 was confirmed by several experimental procedures; additionally FabHLH3 was confirmed to be a functional homolog of AtTT8 (45).

Lin-Wang *et al.* (46) were using the *F. vesca* genome sequence and identified the members of the MYB, bHLH, and WD40 protein families; with RNA-seq data and transgenic lines, they analyzed the specific roles of anthocyanin biosynthesis regulators. Among the primary results of their work, they established that the MYB10 has an auto-regulation capacity and they found that bHLH genes can have redundant functions in *F. vesca,* as they found that *FvbHLH33* expression does not affect anthocyanin concentrations in transformed knockout lines (46).

**AIMS**

Strawberries and raspberries are two of the most consumed berries on the market, being sold not only during the summer period of the northern hemisphere but all year around. Italy is one of the top producing countries for the European market, along with Germany and the Netherlands. Red coloration of berry fruits, due to the presence of anthocyanin pigments produced during the late stages of the fruit maturation process, is an important phenotypic feature of ripe berries. Recent studies focused on the beneficial health effects of consuming food with high content of polyphenols as they have influenced the decision of the breeders around the world to consider these plant metabolites as an important trait to follow up.

The overall aim of the study was to elucidate the effects of the bHLH proteins from closely related species of the cultivated strawberry, *F. vesca,* and *R. idaeus* involved in the general phenylpropanoid and anthocyanin pathways.

The following questions have been addressed in this thesis:

Aim 1: <u>Identification of bHLH proteins from raspberry and strawberry involved in anthocyanin biosynthesis</u>.

The recent publications of genome sequences from more and more plant species have allowed the discovery and identification of gene families involved in critical processes including the genetic regulation of the phenylpropanoid pathway. Taking advantage of the new genome release for the strawberry in 2011 and a draft version of the raspberry genome, one aim of this project was to identify and analyze in more detail the bHLH TFs involved in the phenylpropanoid pathway.

To achieve a correct identification of the bHLH proteins involved in anthocyanin biosynthesis in *F. vesca* and *R. idaeus,* the following in-silico analyses and experiments were performed:

- Literature search for characterized bHLH proteins in related species (*M. domestica, V. vitis, P. persica*) and the model species (*A. thaliana*).

- Whole genome mining for bHLH domains using available genomes.

- Manual data curation for wrongly annotated proteins.

- Sequence analysis and specific primer design for gene cloning.

- Cloning and sequencing of the bHLH candidate genes.

- Protein alignment based on Muscle algorithms.

- Phylogenetic tree construction.

- Identification of candidate genes involved in anthocyanin biosynthesis.

## Aim 2: Establishing the expression levels of candidate bHLH genes during fruit-development

The synthesis of phenylpropanoids and specifically red color formation due to anthocyanin accumulation has been previously described, making this pathway one of the most well-known pathways in literature. The genetic regulation has been widely explored in model plants and fruits (such as grapevine and apple among others). In the case of strawberry and raspberry, there is only little information available, making it crucial to study and unveil the genetic regulation involved in these two important crops.

Idea: To functionally characterize bHLH candidate genes involved in anthocyanin biosynthesis, to analyze the expression levels of these genes during fruit development and to compare these to the accumulation of anthocyanins during fruit development.

The following experiments where performed:

- RNA extraction and cDNA synthesis from a set of fruit developmental stages.

- Primer design for Housekeeping Genes (HKG) and pathway-specific genes.

- Primer validation and standardization.

- Real Time PCR (qPCR) on fruit and leaf tissues.

## Aim 3: Regulatory effect of strawberry and raspberry bHLH candidate genes on the anthocyanin and phenylpropanoid biosynthetic pathway.

In general, studies on the genetic regulation of the phenylpropanoid pathway by the MYB-bHLH-WD40 complex have been targeting the effect of the MYB proteins as the main character. However, the presence of WD40 and bHLH proteins is mandatory. Elucidating the role and specific function of the later proteins is a key step to understanding the whole pathway regulation correctly.

To further study the regulatory function of bHLH candidate genes from *F. vesca* and *R. idaeus*, vectors for gene over-expression and RNA silencing where constructed and integrated into different tissues of closely related species of cultivated strawberry and also model species for the study of anthocyanin biosynthesis (as *N. benthamiana, N. tabacum* and *A. thaliana*) through *Agrobacterium* infiltration (leading to transient transformation of the infiltrated tissue).

The following experiments where performed:

- Detection of CHS-promoter activation via luciferase assay in *N. benthamiana*.
- Anthocyanin synthesis induction and quantification in *N. tabacum* leaves.
- Complementation of *A. thaliana* bHLH mutant lines, and subsequent seed and seedlings analysis by UPLC-DAD.
- Transient expression of genes involved in the anthocyanin biosynthesis pathway in *F. x ananassa* fruits and quantification of polyphenols.

# REFERENCES

1. Daugelaite J., Driscoll A., Sleator R. (2013). An overview of multiple sequence alignments and cloud computing in bioinformatics. ISRN *Biomathematics. Vol* 2013, Article ID 615630.

2. Food and Agriculture Organization of the United Nations. (2012). FAOSTAT Database. Rome, Italy: FAO. Retrieved January 30, 2015 from http://faostat3.fao.org/home/.

3. Bushakra, J. M., Bryant, D. W., Dossett, M., Vining, K. J., VanBuren, R., Gilmore, B. S., … Bassil, N. V. (2015). A genetic linkage map of black raspberry (*Rubus occidentalis*) and the mapping of Ag4 conferring resistance to the aphid Amphorophora agathonica. TAG. Theoretical and Applied Genetics. 128 (8), 1631–1646.

4. Shulaev, V., Sargent, D. J., Crowhurst, R. N., Mockler, T. C., Folkerts, O., Delcher, A. L., … Folta, K. M. (2011). The genome of woodland strawberry (*Fragaria vesca*). *Nature Genetics*, 43 (2), 109–116.

5. Darrow M., (1966). The Strawberries: History, Breeding and Physiology. Holt, Rinehart and Winston San Francisco.

6. Giampieri F, Tulipani S, Alvarez-Suarez JM, Quiles JL, Mezzetti B, Battino M. 2012. The strawberry: composition, nutritional quality, and impact on human health. *Nutrition* 28, 9–19.

8. Marlett, A. (1992). Content and composition of dietary fiber in 117 frequently consumed foods. *Journal of the American Dietetic Association,* 92 2.

9. The Phenol-Explorer database, release 1.0, on polyphenol contents in foods, was developed at INRA, Unité de Nutrition Humaine (Centre de Recherche de Clermont-Ferrand, France) in collaboration with CIQUAL-AFSSA (Maisons-Alfort, France), the Wishart Research Group at the University of Alberta (Edmonton, Canada) and In Siliflo, Inc. (Edmonton, Canada).

10. Swanson *et al.* In Folta, K., & Kole, C. (2011). Chapter raspberries and blackberries, in book: G*enetics, genomics and breeding in fruit and vegetable crops*- berries. May 16, by CRC Press. 200 Pages.

11. Bushakra, J.M., Stephens, M.J., Atmadjaja, A.N., Lewers, K.S., Symonds, V.V., *et al.* (2012) Construction of black (*Rubus occidentalis*) and red (*R. idaeus*) raspberry linkage maps and their comparison to the genomes of strawberry, apple, and peach. *Theoretical and Applied Genetics*, 125: 311–327.

12. Jennings, D.L. (1988) Raspberries and Blackberries - their Breeding, Diseases, and Growth. New York: Academic Press.

13. Pritts, M. Raspberries and related fruit. Cornell University, New York. 2010.

14. Rafique, M. Z., Carvalho, E., Stracke, R., Palmieri, L., Herrera, L., Feller, A., … Martens, S. (2016). Nonsense mutation inside anthocyanidin synthase gene controls pigmentation in yellow raspberry (*Rubus idaeus* l.). *Frontiers in Plant Science*, 7, 1892.

15. Mandave, P., Khadke, S., Karandikar, M., Pandit, V., Ranjekar, P., Kuvalekar, A., & Mantri, N. (2017). Antidiabetic, Lipid Normalizing, and Nephroprotective Actions of the Strawberry: A Potent Supplementary Fruit. *International Journal of Molecular Sciences,* 18 (1), 124.

16. Winkel-Shirley, B. (2001): Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiology, 126 (2):485–493.

17. 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.

18. Feng, C., Chen,M., Xu, C.J., Bai, L., Yin, X.R. (2012). Transcriptomic analysis of Chinese bayberry (*Myrica rubra*) fruit development and ripening using RNA-Seq. *BMC Genomics, 1*3 (1), 19.

19. Vogt, T., (2010). Phenylpropanoid biosynthesis. Molecular Plant 3, 2-20.

20. Pourcel, L., Routaboul, J.-M., Kerhoas, L., Caboche, M., Lepiniec, L., & Debeaujon, I. (2005). Transparent testa10 encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in *arabidopsis* seed coat. The Plant Cell, 17 (11), 2966–2980.

21. Falcone Ferreyra, M. L., Rius, S. P., & Casati, P. (2012). Flavonoids: biosynthesis, biological functions, and biotechnological applications. Frontiers in Plant Science, 3, 222.

22. Heim, M., Jakoby, M., Werber, M., Martin, C., Weisshaar, B., Bailey, P. (2003). The basic helix–loop–helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. *Molecular Biology and Evololution. 20 (5)*: 735-747.

23. 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 (8), 2465–2483.

24. Liu, J., Osbourn, A., & Ma, P. (2017). Myb transcription factors as regulators of phenylpropanoid metabolism in plants. Molecular *Plant*, 8 (5), 689–708.

25. Ravaglia, D., Espley, R. V., Henry-Kirk, R. A., Andreotti, C., Ziosi, V., Hellens, R. P., … Allan, A. C. (2013). Transcriptional regulation of flavonoid biosynthesis in nectarine (*Prunus persica*) by a set of R2R3 MYB transcription factors. *BMC Plant Biology*, 13, 68.

26. Montefiori, M., Brendolise, C., Dare, A.P., Lin-Wang, K., Davies, K.M., …. Allan, A.C. (2015). In the Solanaceae, a hierarchy of bHLHs confer distinct target specificity to the anthocyanin regulatory complex. Journal of Experimental Botany 66, 1427–1436.

27. 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 (1), 94–116.

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

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

30. Carey, C.C., Strahle, J.T., Selinger, D.A., Chandler, V.L., (2004). Mutations in the pale aleurone color1 regulatory gene of the *Zea mays* anthocyanin pathway have distinct penotypes relative to the functionally similar TRANSPARENT TESTA GLABRA1 gene in *Arabidopsis thaliana*. *Plant Cell* 16, 450-464.

31. Van Nocker, S., Ludwig, P. (2003). The WD-repeat protein superfamily in *Arabidopsis*: conservation and divergence in structure and function. *BMC Genomics* 4, 50.

32. Rahim, M. A., Busatto, N., and Trainotti, L. (2014). Regulation of anthocyanin biosynthesis in peach fruits. *Planta* 204, 913–929.

33. Shi, M.-Z., & Xie, D.-Y. (2014). Biosynthesis and metabolic engineering of anthocyanins in *Arabidopsis thaliana*. *Recent patents on biotechnology*, 8 (1), 47–60.

34. Atchley, W. R., & Fitch, W. M. (1997). A natural classification of the basic helix–loop–helix class of transcription factors. *Proceedings of the National Academy of Sciences of the United States of America,* 94 (10), 5172–5176.

35. Pires, N., & Dolan, L. (2010). Origin and diversification of basic-helix-loop-helix proteins in plants. *Molecular Biology and Evolution*, 27 (4), 862–74.

36. Toledo-Ortiz, G., Huq, E., & Quail, P. H. (2003). The *arabidopsis* basic/helix-loop-helix transcription factor family. *The Plant Cell*, 15 (8), 1749–1770.

37. Espley, R. V., Hellens, R. P., Putterill, J., Stevenson, D. E., Kutty-Amma, S., & Allan, A. C. (2007). Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *The Plant Journal*, 49 (3), 414–427.

38. An X.-H., Tian, Y., Chen, K.-Q., Wang, X.-F., & Hao, Y.-J. (2012). The apple WD40 protein MdTTG1 interacts with bHLH but not MYB proteins to regulate anthocyanin accumulation. *Journal of Plant Physiology*, 169 (7), 710–717.

39. Zhou, H., Peng, Q., Zhao, J., Owiti, A., Ren, F., Liao, L., … Han, Y. (2016). Multiple r2r3-myb transcription factors involved in the regulation of anthocyanin accumulation in peach flower. *Frontiers in Plant Science*, 7, 1557.

40. Letunic, I., & Bork, P. (2011). Interactive Tree of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research*, 39 W475–W478.

41. Fang, Z.-Z., Zhou, D.-R., Ye, X.-F., Jiang, C.-C., & Pan, S.-L. (2016). Identification of candidate anthocyanin-related genes by transcriptomic analysis of "furongli" plum (*Prunus salicina lindl.*) during fruit ripening using RNA-seq. Frontiers in Plant Science, 7, 1338.

42. Wei, H., Chen, X., Zong, X., Shu, H., Gao, D., & Liu, Q. (2015). Comparative transcriptome analysis of genes involved in anthocyanin biosynthesis in the red and yellow fruits of sweet cherry (P*runus avium* l.). *PLoS ONE*, 10 (3).

43. Starkevič, P., Paukštytė, J., Kazanavičiūtė, V., Denkovskienė, E., Stanys, V., Bendokas, V., … Ražanskas, R. (2015). Expression and anthocyanin biosynthesis-modulating potential of sweet cherry (*Prunus avium* l.) myb10 and bhlh genes. *PLoS ONE*, 10 (5).

44. Feng, S., Sun, S., Chen, X., Wu, S., Wang, D., & Chen, X. (2015). PyMYB10 and PyMYB10.1 interact with bhlh to enhance anthocyanin accumulation in pears. *PLoS ONE*, 10 (11).

45. Schaart, J. G., Dubos, C., Romero De La Fuente, I., van Houwelingen, A. M. M. L., de Vos, … Bovy, A. G. (2013), Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (*Fragaria × ananassa*) fruits. *New Phytologist,* 197: 454–467.

46. Lin-Wang, K., McGhie, T. K., Wang, M., Liu, Y., Warren, B., Storey, R., … Allan, A. C. (2014). Engineering the anthocyanin regulatory complex of strawberry (*Fragaria vesca). Frontiers in Plant Science*, 5, 651.

# Chapter 2

General methods

# GENERAL METHODS

The following general methods where used

## Total plant RNA extraction

Fruit tissue was manually grounded by ceramic mortar and pestle under liquid nitrogen until a fine powder was obtained. RNA was extracted using the pine tree RNA extraction method reported by Lin-Wang (30). In brief, 100 mg of ground tissue was dissolved in 5 mL of extraction buffer in a Falcon tube and heated to 65°C for 5 min. An equal volume of chloroform: isoamyl alcohol (24:1) (5 ml) was added, and the suspension was mixed and centrifuged at room temperature at 13,000 g for 10 min to separate the phases. The top (aqueous) phase was transferred to a new tube, followed by a second centrifugation at 13,000 g for 10 min.

Finally, the aqueous phase was transferred to another centrifuge tube and the volume estimated. A ¼ volume 10 M LiCl was added to the supernatant, mixed, and incubated at 4°C overnight. The tube was centrifuged at 4°C at 18,000 g for 20 min. The RNA pellet was dissolved in 500 µl SSTE buffer (preheated at 65°C), and extracted with an equal volume of chloroform: isoamyl alcohol (24:1). After centrifugation for 10 min at 13,000 g, the top phase was transferred to a fresh tube, and two volumes ethanol (98%) were added to precipitate the RNA. Tubes were incubated at least 30 min at -70°C or 2 h at -20°C, followed by 20 min centrifugation at 4°C. The pellet was washed with 75% ethanol, dried and dissolved in 30 µl RNase free water.

Total RNA from leaf material was extracted using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, St. Louis, Missouri, USA) according to the manufacturer's manual, using 100 mg of tissue. Before RNA elution, an on-column DNase I treatment (Sigma-Aldrich) was performed.

Extraction Buffer
2% CTAB
2% PVP
100 mM Tris-HCl (pH 8.0)
25 mM EDTA
2.0 M NaCl

SSTE (keep at 65°C)
1.0 M NaCl

0.5 % SDS

10 mM Tris-HCl (pH 8.0)

## Plant genomic DNA extraction

The commercial kit NucleoSpin® Plant II from Macherey-Nagel (Duren, Germany) was used for 50 mg of frozen leaves according to the instruction provided by the manufacturer.

## Bacterial plasmid DNA extraction

Macherey-Nagel in column nucleic acid isolation was used for all the plasmid isolations employed in the following chapters of this thesis, in general, two different approaches were used:

a. Mini-prep plasmid purification: between 3 and 5 ml of bacterial culture was grown overnight at 37ºC and 220 rpm, and later collected and spun down in a 2 ml microcentrifuge tube. The pellet was processed according to the manufacturer's protocol.

b. Midi-prep plasmid purification: bacteria culture of 50-100 ml was grown overnight at 37ºC and 220 rpm, the cell pellet was collected by spinning down the culture in 50 ml plastic centrifuge tubes. The pellet obtained was handled according to the manufacturer's protocol.

Determination of the Concentration of Nucleic Acids

DNA and RNA yield and purity were checked by absorbance of UV light using the Thermo Scientific NanoDrop 8000 Spectrophotometer (Waltham, Massachusetts, USA). 2 μL of each sample was measured following the Nucleic Acid application tool from the machine software. For this quantification, the absorbance value at 260 nm was used. The nucleic acid purity was checked with the value of the 260/230 ratio, and only samples with values between 1.8 and 2.1 were kept.

## PCR reaction

For cloning, PCR reactions with Platinum® Taq DNA Polymerase High Fidelity (Life Technologies) was used according to the manufacturer's protocol.

| Component | 50-μL rxn | Final Concentration |
|---|---|---|
| 10X High Fidelity PCR Buffer | 5 μL | 1X |
| 10 mM dNTP mix | 1 μL | 0.2 mM each |

| | | |
|---|---|---|
| 50 mM MgSO$_4$ | 4 µL | 2 µM |
| 2 mM enzyme | 0.2 µL | 1 U |
| 10 µM forward primer | 1–2 µL | 0.2 -0.4 µM |
| 10 µM reverse primer | 1–2 µL | 0.2 -0.4 µM |
| Template (cDNA, DNA) | 1 µL | 1- 5 µM |
| H$_2$O (milliQ) | up to 50 µL | |

The touchdown PCR (TD-PCR) program was used as follows:

| Temperature ºC | Time | Cycles | |
|---|---|---|---|
| 94 | 3 minutes | | |
| | | | |
| 94 | 15 seconds | | |
| 62-52 | 30 seconds | 10 cycles | Tm  - 1°C every cycle |
| 68 | 1 minute/kb | | |
| | | | |
| 94 | 15 seconds | | |
| 56 | 30 seconds | 30 cycles | |
| 68 | 1 minute/kb | | |
| | | | |
| 68 | 5 minutes | | |

To test if the PCR reactions were performed successfully, 10 µL PCR products were loaded onto an agarose gel (from 1 to 1.8%, depending on the amplicon length).


**Constructs for *Agrobacterium***

Candidate genes were amplified from cDNA using specific primers that included attB sites (primer sequences in Table 11, attB sites are underlined). Next, the purified PCR product (purified using the NucleoSpin® Gel and PCR Clean-up from Macherey- Nagel), was ligated into the pDONR221 vector using BP Clonase and transformed into TOP10 cells following the manufacturer's instructions. Positive colonies were selected for plasmid isolation and sequencing. For each colony, a 3 ml culture was grown overnight at 37ºC in LB liquid media plus Kanamycin (50 µg/mL). Plasmid DNAs were purified using the NucleoSpin® Plasmid from Macherey- Nagel as described before.

Sequencing of the constructs was performed by Sanger sequencing using M13 Forward and Reverse primers. Subsequently, the genes that were cloned in pDONR221 were recombined

into the destination vector using LR Clonase (Gateway™ LR Clonase™ II Enzyme Mix, Thermo Fisher) following manufacturer's instructions, once again the positive colonies were selected on the corresponding media, plasmids were sequenced, and the presence of correct insert confirmed. The plasmids were later introduced into *Agrobacterium* strain GV3101.

**Solutions and media**

LB medium

| | |
|---|---|
| NaCl | 10 g/L |
| Yeast extract | 5 g/L |
| Tryptone | 10 g/L |
| Agar | 15 g/L |

pH 7 (adjusted with KOH)

SOC broth medium

| | |
|---|---|
| Tryptone | 20 g/L |
| Yeast extract | 5 g/L |
| NaCl (10 mM) | 0.5 g/L |
| KCl (10 mM) | 0.19 g/L |
| $MgCl_2$ (10 mM) | 0.95 g/L |
| $MgSO_4$ (10 mM) | 1.2 g/L |
| Glucose (10 mM) | 3.6 g/L |

YEB medium

| | |
|---|---|
| Sucrose | 5 g/L |
| Tryptone | 1 g/L |
| Yeast extract | 5 g/L |
| Beef extract | 5 g/L |
| $MgSO_4$ | 0.5 g/L |

MS 1⁄2 Half Strength

| | |
|---|---|
| MS basal salt mixture | 2,2 g/L |
| Sucrose | 15 g/L |
| Plant Agar | 7 g/L |

pH 5.6 – 5.8 (adjusted with KOH)

pH measurements were performed by using the pH meter WTW pH 7110 from Inolab® (Weilheim, Germany)

Chemical Reagents were purchased from Sigma-Aldrich, Fluka. Agar from Biokar diagnostic (Allonne, France)

**Bacterial strains**

*Escherichia coli*: DH5 alpha, TOP 10

*Agrobacterium tumefaciens*: GV3101, AGL0

**Plant materials**

*Nicotiana benthamiana*: cv Nb-1 (seeds origin Dr. Antje Feller)

*Nicotiana tobacco*: cv Virginia Gold (seeds origin Dr. Antje Feller)

*Arabidopsis thaliana*: cv Columbia, tt8 (seeds origin NASC -european *Arabidopsis* stock centre-)

*Fragaria vesca*: cv fragolina di bosco, Hawaii 4 (plants origin FEM, Vigalzano breeding facility)

*Rubus idaeus*: cv Heritage (plants origin FEM, Vigalzano breeding facility)

**Vectors:**

Cloning and sequencing purposes: pDONR221 (Invitrogen- Thermo Scientic Waltham, USA)
pGEMT-easy (Promega, Fitchburg, Wisconsin, USA)

Plant overexpression: pK7WG2D (Gateway technology; Gent University)
pHEX2 (Hellens *et al.* Ref 15 Chap 4)

Vector for plant RNA silencing: p9U10i (Hoffmann *et al.* Ref 19 Chap 5)

Plant Luciferase assay: pGreen (Hellens *et al.* Ref 7 Chap 4)
pSAK 277 (Espley, Ref 17 Chap 4)

**Antibiotics stock solutions and final working concentrations**

| Antibiotic | Stock Concentration | Solvent | Working Concentration |
|---|---|---|---|
| Carbenicillin | 100 mg/mL | $H_2O$ | 100 µg/mL |
| Kanamycin | 50 mg/mL | $H_2O$ | 50 µg/mL |
| Spectinomycin | 50 mg/mL | Water | 50 µg/mL |
| Tetracycline | 10 mg/mL | 70%EtOH | 5-10 µg/mL |
| Rifampicin | 50 mg/mL | DMSO | 2,5 µg/mL |

# Chapter 3

bHLH Family Phylogeny

# INTRODUCTION

## The bHLH family

The basic Helix-Loop-Helix domain (bHLH) is a highly conserved domain that forms a superfamily found in large numbers within plant, animal, and fungal genomes, and constitutes one of the largest families of TFs. This family is an essential part of many TFs involved in numerous regulatory processes, such as neurogenesis, cell differentiation and metabolism in mammals. In plants, their role includes environmental and stress response, organ development (like carpel, anther, and epidermal cells), and secondary metabolite biosynthesis of phenylpropanoids. In maize, the *R* gene was the first plant protein reported to possess a bHLH domain and to be involved in the phenylpropanoid/anthocyanin biosynthesis pathway (1, 2, 3, 4 ,5, 6).

Table 6: List of bHLH proteins reported in literature according to Plaza 2.5 (7)

| Species | bHLH reported |
|---|---|
| *Zea mays* | 201 |
| *Sorghum bicolor* | 174 |
| *Oryza sativa sp. japonica* | 164 |
| *Glycine max* | 392 |
| *Populus trichocarpa* | 202 |
| *Arabidopsis thaliana* | 155 |
| *Vitis vinifera* | 119 |
| *Malus domestica* | 241 |
| *Rubus occidentalis* | 90 |
| *Fragaria vesca* | 112 |
| *Rubus idaeus* | - |

Table 6a: bHLH subfamilies function and characterized genes (5, 8, 9)

| Subfamily | Function | Gene |
|---|---|---|
| Ia | Stomata differentiation | MUTE, FAMA, SPEECHLESS |
| Ib | Iron homeostasis and embryonic development | ORG2, ORG3 |
| III (a+c) | Regulation of iron uptake | FIT |
| IIIb | Stomata development and cold acclimatization | ICE1, SCRM2 |
| III (d+e) | Involved in abscisic acid, jasmonic acid and light signalling pathways | MYC2, JAI1 JIN1 |
| IIIf | Anthocyanin biosynthesis | GL3 EGL, TT8, ZmR, AmDEL |
| | trichome formation | PhJAF13, MdbHLH3, PhAN1 |
| IVa | Formation of ER body and derived structures | NAl1 |
| IVc | Metal homeostasis, auxin-conjugate metabolism | IRL3 |
| Va | Brassinosteroid signaling | BIMs |
| VII (a+b) | Light and gibberellin signalling | PIFs, PILs |
| VIIIb | Transmitting tract and stigma development | HECs |
| VIIIc (1) | Development of rhizoids and caulonemata | RHD6, RSL1 |
| VIIIc (2) | Root hair development | RLS1, RLS2, RLS3, RLS4 |
| XI | Fertilization and responses to phosphate deficiency stress | UNE12, PTF1 |
| XII | Brassinosteroid signalling, Blue-light receptor | BEEs, CIB1, CIBS |
| XIII | Root meristem | LHW |
| XIV | Gibberellin signaling | SAC51 |
| XV | Positive regulators of gibberellin signalling and light signal transduction | PRE1.6 KDR |
| Orphans | Anther development | AMS, DYT1 |
| | Development early embryo | MEEK |
| | Metabolic shade avoidance responses | PAR1, PAR2 |

**bHLH structure**

The bHLH domain-containing proteins are structurally heterogeneous. There are 21 subfamilies identified as shown in Table 6a. Each subfamily group has been proven to be involved in a particular function; as an example, the bHLH subfamily III (d + e) regulates the jasmonate signalling pathway, the subgroup Ia regulates stomata cell differentiation, and the protein members of subgroup IIIf are involved in the biosynthesis of phenylpropanoids (5, 10).

In addition to the basic DNA-binding and helix-loop-helix dimerization components, bHLH family proteins also contain additional domains: leucine zippers, structural protein motifs involved in protein-protein interactions or PAS (the PAS domain is named after three proteins containing it: *Drosophila melanogaster* Period (Per), the human aryl hydrocarbon receptor nuclear translocator (Arnt) and *D. melanogaster* Single-minded (Sim)). bHLH proteins can also contain specific sites for their interaction with other TFs such as MYB (5, 8, 10).

Besides the bHLH domain, bHLH proteins usually exhibit low conservation at the N-terminal region. However, groups of functionally related bHLH proteins share additional motifs among them. Some of these have been characterized in animals to determine specificity in DNA-

binding sequence recognition and dimerization activities, as responsible for the activation or repression of target genes or the binding to small molecules (1, 5, 8, 10, 11).

The bHLH domain is highly conserved and comprises approximately 60 amino acids with two functionally distinct regions. The basic region at the N-terminal is responsible for DNA interaction, generally containing five to six basic residues that facilitate DNA binding to the hexanucleotide E-box DNA motif CANNTG (5, 8, 12).

**bHLH IIIf group**

Heim *et al*. (8) described the protein structure of all the *Arabidopsis* bHLH proteins, based not only on the bHLH domain itself but also considering the additional conserved domains. The bHLH proteins that regulate anthocyanin biosynthesis in *A. thaliana* belong to group IIIf as can be seen in Table 6a (8). Proteins in that group contain certain conserved amino acids at the N-terminal, which are responsible for the protein–protein interaction with the R2R3-MYB proteins (5,  8,13, 9).

The proteins contained in the subgroup IIIf are involved in two different processes; phenylpropanoid/anthocyanin biosynthesis (AtbHLH0012/MYC1 and AtbHLH042/TT8) and trichome initiation (AtbHLH001/GL3). Group IIIf bHLH TFs involved in phenylpropanoid biosynthesis in other plant species include *Antirrhinum majus* DELILA, *Petunia hybrida* ANTHOCYANIN1 and JAF13, *V. vinifera* MYC1 and MYCA1 and *Z. mays* INTENSIFIER1 and Lc (5, 6, 8, 9, 13).

In addition to the described elements, the members of the IIIf group of bHLH proteins contain a conserved domain called ACT-like domain at the C-terminal. The ACT domain is named after three of the proteins that contain it: Aspartokinase, Chorismate mutase, TyrA (protein family of dehydrogenases dedicated to L-tyrosine (TYR) biosynthesis) (14).

The ACT domain is a structural motif of proteins of 70 – 80 amino acids long and is widely distributed in bacteria, animals, fungi, and plants; however, the ACT domain is not found in all organisms nor is it used consistently within species or in all proteins within a related function. The majority of ACT domain-containing proteins identified to date appear to interact with amino acids and are involved in some aspect of regulation of amino acid metabolism, including both metabolic enzymes and transcription regulators (13, 14, 15).

## MATERIAL AND METHODS

### Genome sequence

The woodland strawberry or *F. vesca* L. genome version 1.01 was obtained from the Genome Database for Rosaceae GDR website (www.rosaceae.org). The clone named *F. vesca* ssp. *vesca* 'Hawaii 4' was the selected genotype for the genomic studies of the Rosaceae community due to its previous reports on successful *Agrobacterium* transformation (16).

Raspberry genome draft version V 0.8 was obtained by collaboration with the raspberry genome consortium composed of Fondazione Edmund Mach, Plant & Food Research New Zealand and other institutions (personal communication with Dr. Dan Sargent). This genome draft is not yet released to the global scientific community.

### Local BLAST search

For the bHLH *in silico* identification, the "BLAST + stand-alone" application - Basic Local Alignment Search Tools, was implemented in the search for putative bHLH proteins. For both species, *F. vesca* and *R. idaeus*, databases were constructed using the "BLAST+ make blast DB" tool.

On each search, bHLH proteins previously reported were used as a query for the "tBLASTn" command: a protein query against a nucleotide database dynamically translated in all six frames. The accession numbers of the query proteins or translated products, in the GenBank database are: MdbHLH3 - ADL36597.1, MdbHLH33 - DQ266451, AtTT8 - CAC14865; ZmB - CAA40544, ZmLC- AAA33504, FxabHLH3 - AFL02463, FxabHLH33 - AFL02465 (17,18). The e-value cut off considered was 1e-99 (19, 20).

As an alternative search, the HMMER tool was employed for amino acid sequence analysis by comparing a profile-HMM with each genome database, the manually curated bHLH PF00010 pHMM (protein Hidden Markov Model) from PFAM (6). The cat *.fasta merged results from both searches *.fasta > combined.fasta command.

**Protein alignment**

After duplicate gene predictions were removed from the list obtained, the protein sequences (in the FASTA format) were aligned based on the Multiple Sequence Alignment MUSCLE: MUltiple Sequence Comparison by Log-Expectation. The option "maxiters 2" was used due to a large number of proteins to be analyzed (21).

**Phylogenetic analysis and tree construction**

Once the protein alignment was performed and properly edited by removing proteins not aligned in the bHLH motif area, a second alignment was made, and the result checked. Later, phylogenetic and molecular evolutionary analyses were conducted using MEGA version 6 (22), using the ML (Maximum Likelihood) method; additionally, to assess the reliability of the phylogenetic tree, a Bootstrap test was performed, with a number of 1000 replicates.

**Protein domain prediction**

The identification of protein domains in the candidate protein sequence was made through the Protein Structure Prediction Phyre 2.0 web tool. Phyre 2.0 (http://www.sbg.bio.ic.ac.uk/~phyre2/).

**Primer design**

Gene-specific primers were designed using Primer 3 (23) (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) to a stringent set of criteria, enabling application of universal reaction conditions: optimal length of PCR primers of 18-22 bp, GC content between 40-60%, and annealing temperature close to 58 ºC.

**bHLH protein homologs**

Literature search for characterized bHLH proteins in the closely related species apple (*M. domestica*) and model species (*A. thaliana*), reveals the highly conserved bHLH domain among them with identity values of up to 80%, while the identity of amino acids over the whole protein was only 20% (Data not shown).

Whole genome mining for bHLH domains was performed using available genomes of *F. vesca* v1.01 and *R. idaeus* draft version 0.8 as described in the material and method section of this chapter. As a reference, the genome sequence and predicted proteins of *M. domestica* version 2.0 were used.

Using the tblastn command from the BLAST+ software, and considering the cutoff values for identity and e-value mentioned before, we obtained a number of 150 gene predictions for both species.
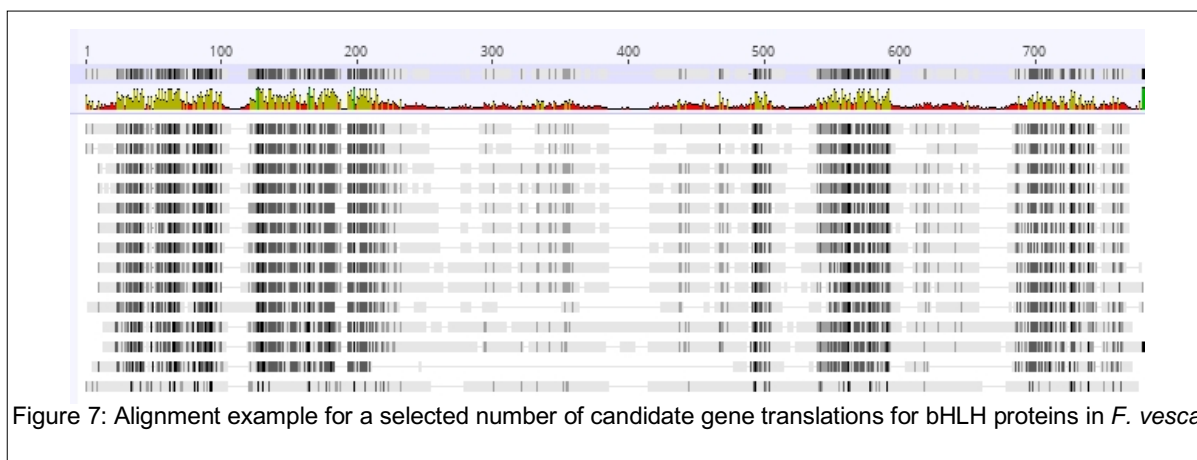


Figure 7: Alignment example for a selected number of candidate gene translations for bHLH proteins in *F. vesca*

Figure 7, shows an example of how candidate protein sequences align among them in specific regions, which can also be identified by their domains as the MYB interaction sites and the bHLH domain itself.

Next, the preliminary list was checked for truncated protein models, that contain less than 100 aa (amino acids) in their sequence, and the presence of the bHLH motif was confirmed by a mega query on PFAM (24). After this check, a final number of 98 bHLH candidates for *F. vesca* and 90 candidates for *R. idaeus* were found. All these protein sequences were used for the phylogenetic analyses.

Table 7: Genome description for *Fragaria, Rubus* and Apple.

|  | *Fragaria vesca* | *Rubus idaeus* | *Malus domestica* |
|---|---|---|---|
| Common name | Woodland strawberry | Raspberry | Apple |
| Chromosome number | 2n = 14 |  | 17 |
| Genome size | 240 Mb | 240 Mb approx | 742.3 Mb |
| Annotated gene models | 32831 | 36009 | 57.386 |
| bHLH candidates found | 98 | 90 | 219 |

Table 7, shows the number of bHLH proteins annotated in Rosaceae family members, and for all three species it is close to 0.5% of the total number of gene models predicted. This value of 0,5% as compared with the bHLH % values of other plant species already studied: 0.5% for *Arabidopsis* (25), 0.6% for *Z. mays* (10) and 0.4% for apple (26), considering this last species as the closest member for the strawberry and raspberry, we can expect around 100 bHLH genes in both species under study.

**Phylogenetic trees**

The 98 bHLH proteins found in *F. vesca* were aligned and compared versus the widely studied bHLHs from *A. thaliana*. As depicted in Figure 8, this robust tree was obtained after the total 284 sequences (98 for *F. vesca* and 186 for *A. thaliana*) were aligned and the distance pairwise was calculated between them. Some candidates where eliminated due their high pairwise distance values, in this case, two proteins were discarded, one from *F. vesca* that has been wrongly annotated (FvbHLH132) and a putative bHLH from *A. thaliana*.

For each phylogenetic tree, the Neighbor-Joining method was used to construct a matrix of pairwise distances estimated using a JTT model with distances corrected for multiple substitutions. All positions containing gaps and missing data were eliminated. The bootstrap consensus tree was calculated using 1000 replications, using the Bootstrap Test of Phylogeny of Felsenstein. The Bootstrap test is a re-sampling technique, in which the sampling distribution is estimated by repeatedly sampling from each sequence from the original data, with n replacements. There was a total of 150 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (22, 10).

**Fragaria vesca bHLH phylogenetic tree**



Figure 8: Phylogenetic tree constructed with the ML method using the bHLH transcription factor domain in *F. vesca* and *A. thaliana*. MEGA 6 software using bootstrap values based on 1000 iterations.

Is important to clarify that the primary purpose of this study is not the evolutionary aspects of the bHLH protein family but instead we are interested in the cluster formation between the protein members of the same bHLH clade as described in Heim *et al*. (8), in order to identify candidate proteins (and their gene sequences) involved in the regulation of phenylpropanoid compounds. The whole list of the bHLH protein members found in *F. vesca* can be found in annex 1.

Maximum likelihood - ML - computational tree analyses followed by 1000 bootstrap calculations were made to obtain the consensus tree shown in Figure 8. The ML method was used due to its mathematical reliability, and as observed in this unrooted tree, the 282 sequences gather on clusters that will be described next.

The phylogenetic tree observed in Figure 8, reveals the conserved relationship among the *Arabidopsis* bHLH protein sequences and the *F. vesca* ones. There is a visible cluster organization; as an example, the PIF proteins or group VIIa according to Heim *et al.* (8) can be recognized in the brown square (27); also the proteins belonging to the ICE family grouped as can be seen in the light blue area. Another clustered group was the brassinosteroid signaling components or BEE group. The members of the IIIf group, recognized as the bHLH involved in the phenylpropanoid regulation, congregating the genes MYC1, EGL3, EG1, TT8, and the equivalent homologs for *F. vesca*, is highlighted in red in Figure 8 (27, 28, 29).
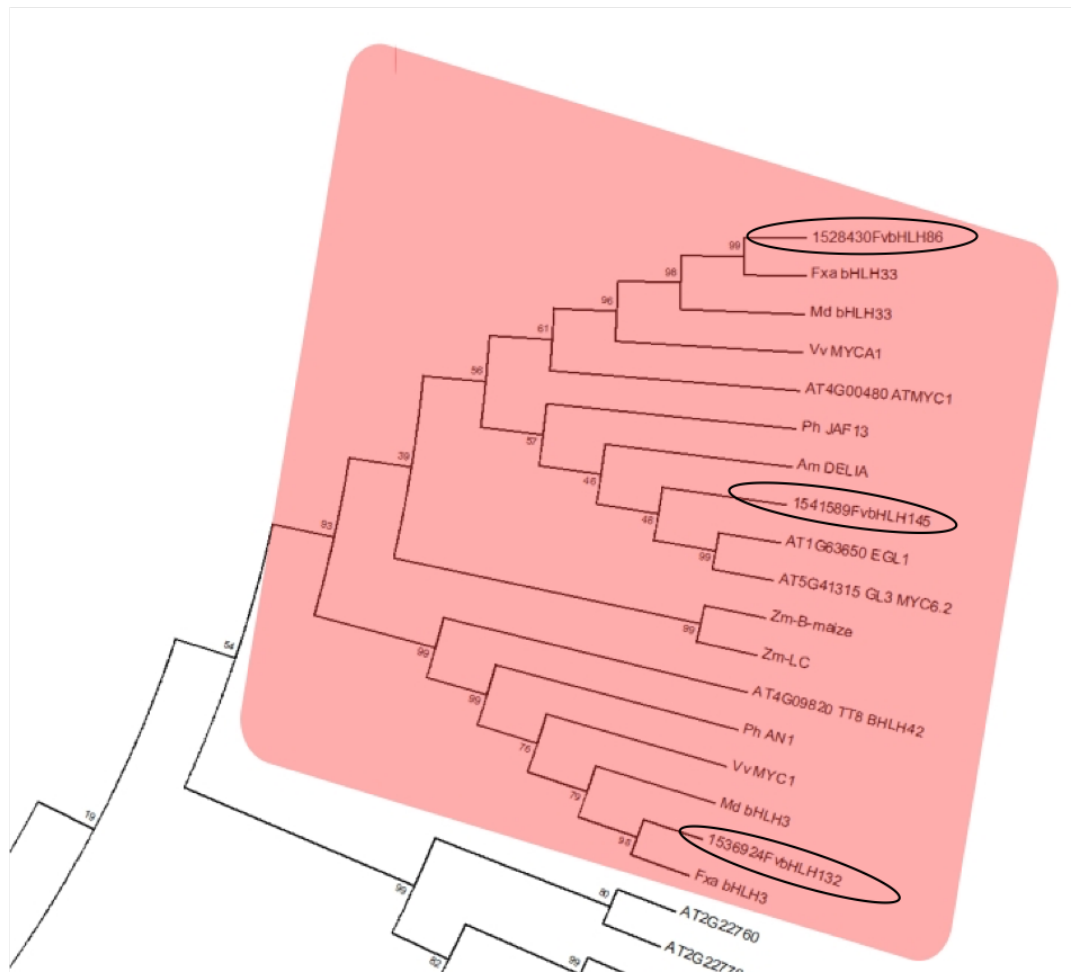


Figure 8.1: Zoom detail of the IIIf group from the phylogenetic tree constructed with the ML method using the bHLH transcription factor domain in *F. vesca* and *A. thaliana*. MEGA 6 software using bootstrap values based on 1000 iterations.

In the case of the group IIIf, where the bHLH phenylpropanoid regulators are found, for *F. vesca* we found 3 candidates identified as Fv_bHLH-132 or from now on called Fv bHLH 3, Fv_bHLH-86 or Fv bHLH33 and Fv_bHLH-145 (Figure 8.1).
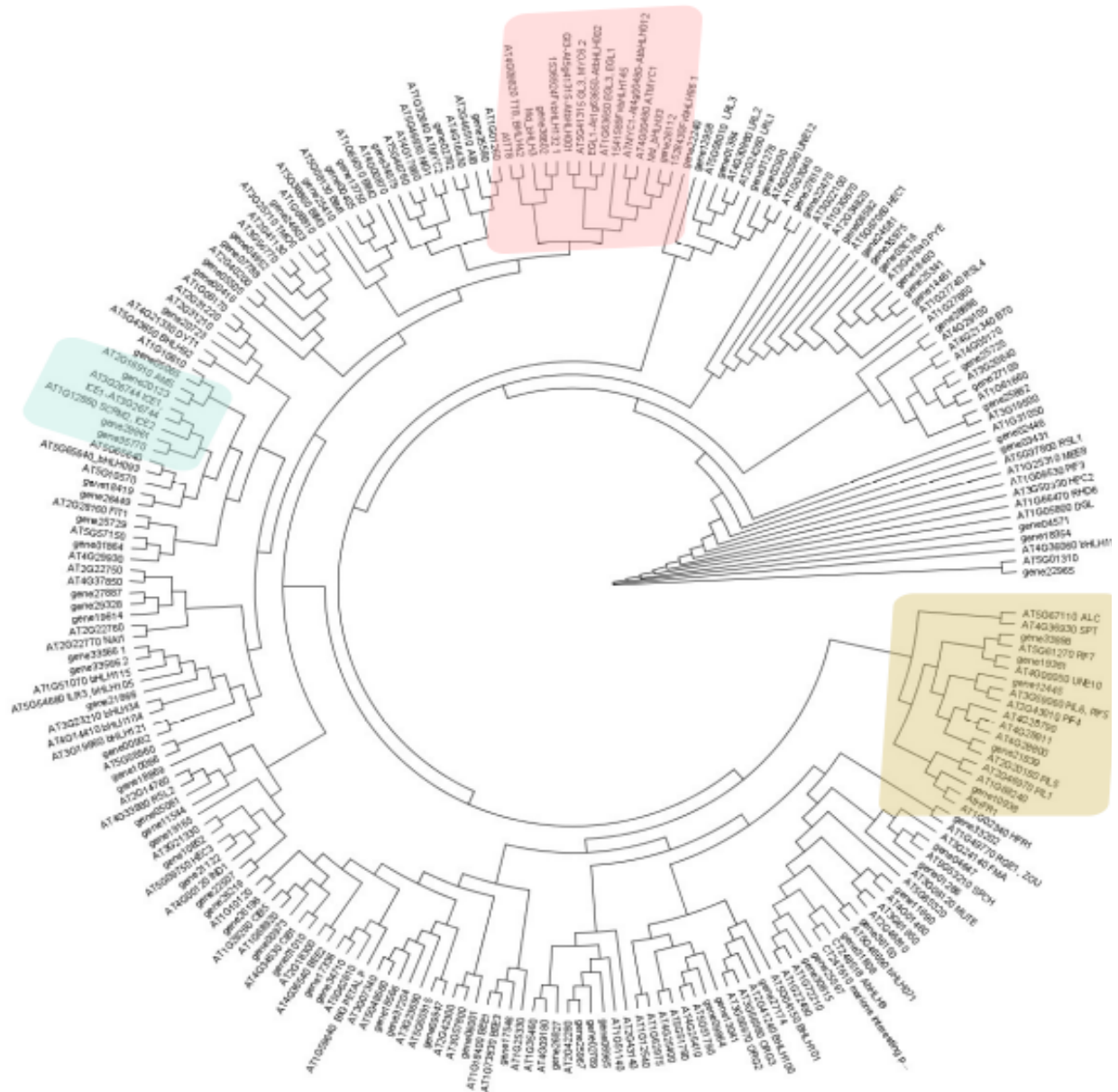
**Rubus idaeus bHLH vs Arabidopsis bHLH**



Figure 9: Phylogenetic tree constructed with the ML method using the bHLH transcription factor domain in *R. idaeus* and *A. thaliana*. MEGA6 software using bootstrap values based on 1000 iterations.

The *R. idaeus* bHLH candidate tree (Figure 9), was constructed using a reduced number of protein sequences, only 90 proteins were identified to contain the bHLH domain in the *R. idaeus* draft genome. This lower number of bHLHs represents a 40% reduction in the number of *R. idaeus* protein candidates compared to *F. vesca*, and it is 60% protein number reduction compared to the *Arabidopsis* sequences.

However, the family groups previously identified in Figure 8, like ICE, PIFs, and BEE proteins, were also found in the *R. idaeus* bHLH phylogenetic tree (Figure 9), with the corresponding

homologs. In the case of the group IIIf, where the bHLH phenylpropanoid regulators are found, there are only two *R. idaeus* candidate proteins that associate with the *Arabidopsis* and the *Fragaria* proteins.
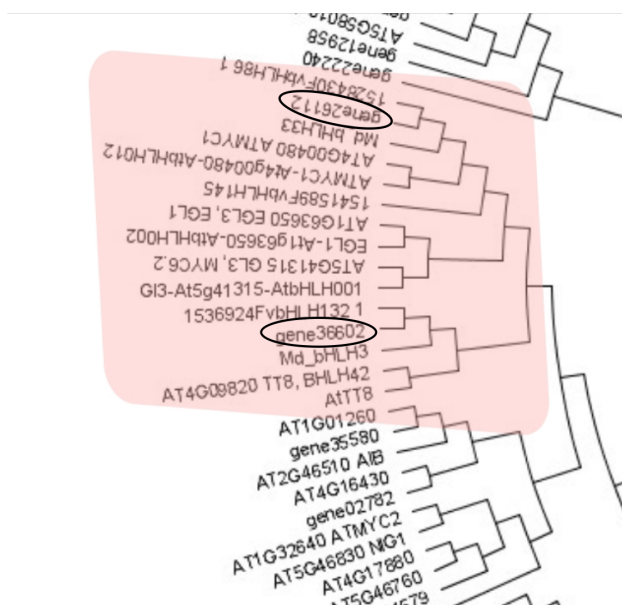


Figure 9.1: Zoom detail of the IIIf group from the phylogenetic tree constructed with the ML method using the bHLH transcription factor domain in *R. idaeus* and *A. thaliana*. MEGA6 software using bootstrap values based on 1000 iterations.

The bHLH candidate proteins presented an average size between 150 and 700 amino acids as shown in Table 8. Additionally, the percentage of bHLH candidates found in the genomes of strawberry and raspberry, was 0.3% and 0.2% respectively. These percentage values are low compared to other species as *A. thaliana*, and *M. domestica* where the percentage of bHLH proteins identified is around 0.4-0.5% (16, 26).

The relatively low number of proteins identified as bHLH putative in *R. idaeus* and *F. vesca*, can be explained mainly by two reasons: the first reason is due to the use of premature or 'draft' released versions of both genomes and secondly through the *ab initio* Hidden Markov Mode gene prediction method used on the gene annotation pipelines for those genomes (16).

**ACT domain and phenylpropanoid biosynthesis regulation**

When looking into detail at the amino acid sequences of the bHLH proteins reported to be involved in the phenylpropanoid biosynthesis regulation, it can be noted that all the bHLH members of the clade IIIf from diverse plant species present a secondary motif in the C-terminal region, called ACT domain (Aspartokinase, Chorismate mutase, TyrA). Due to this particularity of the IIIf members, manual data curation for identification of wrongly annotated proteins, and secondary domain detection was confirmed by Phyre 2.0 (30).

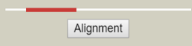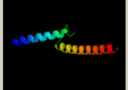| Template | Alignment Coverage | 3D Model | Confidence | % i.d. | Template Information |
|---|---|---|---|---|---|
| c4ywcA | Alignment | | 100.0 | 31 | **PDB header:**transcription regulator<br>**Chain: A: PDB Molecule:**transcription factor myc3;<br>**PDBTitle:** crystal structure of myc3(5-242) fragment in complex with jaz9(218-2 239) peptide |
| d1nlwe | Alignment | | 99.3 | 17 | **Fold:**HLH-like<br>**Superfamily:**HLH, helix-loop-helix DNA-binding domain<br>**Family:**HLH, helix-loop-helix DNA-binding domain |

Whole aa lenght bHLH 3

Figure 10: PHYRE 2.0 results indicating secondary motifs in the whole *F. vesca* bHLH 3 amino acids sequence. Protein bHLH 3 presents the ACT-like domain of the proteins belonging the clade IIIf.

The role of the ACT domain is at the moment not completely understood on the regulatory level, but just recently the importance of the ACT domain homodimer for *in vivo* function was revealed in the maize bHLH protein R (5, 14, 15, 31).

bHLH domain analyses of the candidate proteins obtained on the phylogenetic analyses were performed as described in the Material and Methods of this chapter. Briefly: Protein Structure Prediction Phyre 2.0 web tool was employed and made in two steps in Phyre 2.0 (30).

In the first step of the domain analyses, the whole protein sequence of each candidate was uploaded to the Phyre 2.0 online platform. In the case of *F. vesca*, the three candidates that presented high homology within the reported bHLH involved in the phenylpropanoid pathway regulation, and for *R. idaeus*, two candidate genes were analyzed. As a result, for every candidate protein the two main domains identified were the bHLH-MYC binding domain and the basic Helix Loop Helix DNA binding domain, these results can be seen in Figure 10 for *F. vesca* bHLH 3.

The second steps consisted in the analyses of exclusively the last 100 aa of the C-terminal region of each gene prediction, and this was made to detect the presence of the small ACT domain without the interference of other domains in the neighborhood. Figure 11 shows details of PHYRE 2.0 domain analysis of the last 100 aa from two *F. vesca* bHLH protein candidates. The upper part is a typical ACT domain structure, and lower portion shows a non-related anthocyanin regulator domain found in the protein model Fv_bHLH 42. This candidate protein was used as a negative control, due to the proximity of this candidate with the IIIf group in the phylogenetic protein analyses. Additionally, Fv_bHLH 42 is not expressed in the fruit tissue and does not have the ACT domain at the C-terminal region typical of the proteins of the bHLHs from IIIf group.
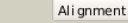
Figure 11: PHYRE 2.0 results indicating secondary motifs in the last 100 amino acids region of the C-terminal tail of two bHLH models from *F. vesca*. Protein bHLH 3 presents the ACT-like domain of the proteins belonging the clade IIIf.

In summary, the following can be concluded from the *in silico* search on the *F. vesca* and *R. idaeus* genomes, and the subsequent phylogenetic analyses:

 a) Three and two genes for each species, respectively, belong to the IIIf group,

 b) Protein sequences of these genes comply with the domain requirements for being considered in further analyses of TFs of the phenylpropanoid pathway.

Table 8 shows more detailed information about the candidate genes for phenylpropanoid biosynthesis regulation:

Table 8: List of bHLH candidates putatively involved in phenylpropanoid biosynthesis regulation.

| Species | Gen name | Gene Length (cds) | Protein Length (aa) |
|---|---|---|---|
| *Fragaria vesca* | Fv3 - FV2G25270 | 2172 bp | 723 |
| | Fv33- FV7G08120 | 1962 bp | 653 |
| | Fv145 -FV5G02910 | 1896 bp | 631 |
| *Rubus idaeus* | Ri3- gene36602 | 1912 bp | 637 |
| | Ri3- gene26116 | 1916 bp | 639 |

**Sequence analysis and specific primer design for gene cloning**

Considering the relationship between the *F. vesca* bHLH genes and the *Arabidopsis* bHLH genes, it was evident that *Fv3* and *At TT8* are homologs, *Fv33* at the same time is homolog

to *At MYC1*, Petunia *JAF 13* and *Vitis MYC 1*, and finally *Fv 145* is a homolog of *At GL3* as seen in Figure 8.1. Additionally, it can be concluded that all the *F. vesca* bHLH genes contain seven exons (Figure 12), and other members of the IIIf group share this particular structure as *A. thaliana TT8*, which contains the same number of introns as the homologs *Fv33*, and *Fv145*.

For the *Rubus* bHLH sequences, a complete analysis couldn't be performed due to the lack of the final genome data. Several attempts to sequence genomic regions of both candidate genes were unsuccessful due to technical problems, including the large regions to amplify and rich CG regions of the sequence. For the primer design, the analysis for *Fragaria* intron-exon structure was used as an example. As reported in Carretero-Paulet *et al*. (11) the *F. vesca* and *R. idaeus* intron pattern distribution is conserved within the previously reported proteins of the IIIf family, providing another proof of the reliability of our phylogenetic analysis (11, 32).
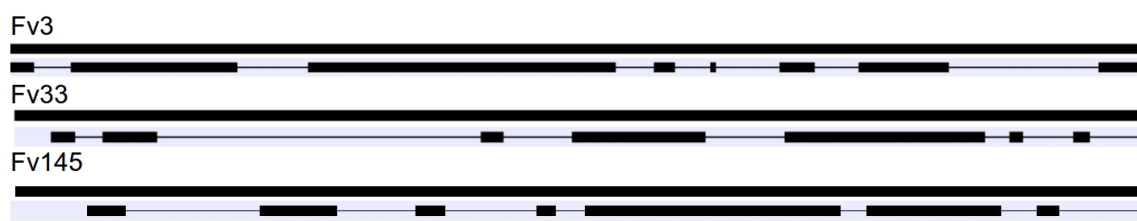


Figure 12: Intron exon structure of *Fragaria vesca* bHLH protein candidates of the group IIIf. Black boxes indicate the coding regions or exons, lines indicate non-coding regions Geneious 9.3 software.

The differences in the exon length seen in Fig 12, might be due either the wrongly annotated gene models used in these analyses, further confirmation of this genes intron-exons structure and length needs to be confirmed by re sequencing not only mRNA from different varieties but also genomic DNA (16).

**Protein alignment based on MUSCLE algorithms**

To get a better knowledge of the proteins specifically involved in the regulation of the polyphenol biosynthesis pathway, the earlier described protein sequences were analyzed again. This time including other IIIf protein representatives, not only from the Rosaceae family but others reported in the literature. A second MUSCLE alignment was made as shown in Figure 13.
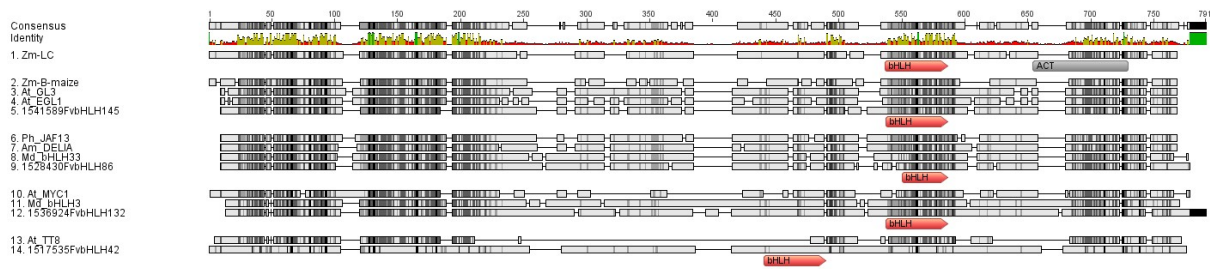
Figure 13: bHLH amino acid sequence of the three *F. vesca* candidates, and ten reported bHLH proteins from the group IIIf, the bHLH domain is highlighted in the red rectangle, and the ACT domain as a gray box at the first protein sequence. The upper line represents the consensus identity. (Geneious software 9.3).

The identity among the sequences is conserved; especially on the N-terminal region corresponding to the R2R3-MYB binding domain that is located in the first 100 amino acids, and the following bHLH domain situated between the amino acids 550-600 approx. The third domain ACT is also evident in the C-terminal end of the alignments.

## CONCLUSIONS AND REMARKS

Comparing the number of bHLH proteins from *A. thaliana* and the bHLH homolog proteins of some Rosaceae family members identified until now (*P. avium*, *F x ananassa*, and *P. persica*), there are three bHLH genes in each species fitting into the sub-family IIIf. Gene prediction analysis and identification of the bHLH encoding genes related to the phenylpropanoid pathway in *M. domestica* retrieved only two candidate genes to date (17, 33, 35). So the third protein in *M. domestica* is still missing, the reason for this possible incomplete identification of the genetic regulator involved is perhaps due to the big genome size of this species, plus the recent whole genome duplication event that this crop had undergone, making the available genome to cover only ~89 % of the non-repetitive portion of the genome, in addition to the high heterozygosity present in the genome itself. Also, it has to be considered that the possibility of a complete absence of the bHLH homolog in *M. domestica* can be due to the loss of this said regulator (17, 26).

Interestingly, in peach, Rahim *et al.* (33), described three bHLH proteins that had been reported to be involved in the anthocyanin regulation of the fruit, which led to consider the preliminary results found in this study on the *R. idaeus* species as incomplete, as only two candidate genes instead of the expected three bHLH as in *F. vesca* were identified. This result can be explained because the genome used is a preliminary draft version (33).

Table 9: List of Rosaceae bHLH candidates involved in phenylpropanoid biosynthesis regulation found in this thesis, reported in literature, and their equivalent *A. thaliana* homolog.

| *Arabidopsis thaliana* | *Malus domestica* | *Prunus persica* | *F x ananassa* | *Fragaria vesca* | *Rubus idaeus* |
|---|---|---|---|---|---|
| At TT8 | Md bHLH 3 CT151981 | Pp bHLH 3 ppa002884m | Fv bHLH 3 AFL02463 | Fv bHLH 3 like | Ri bHLH 3 like |
| AtMYC1 AEG74473.1 | Md bHLH 33 CT130287 | Pp bHLH 33 ppa002645m | Fv bHLH 33 AFL02465 | Fv bHLH 33 like | Ri bHLH 33 like |
| At GL3 | - | Pp bHLH GL3 ppa002762m | - | Fv bHLH 145 | - |

Finally, a specific group IIIf phylogenetic tree was calculated to clarify the intricate relationship between the candidate proteins. For this analysis, the sequences of reported bHLH proteins found in the literature were used: 16 proteins from widely studied plants like the *A. thaliana* MYC, TT8 and GL3, *Z. mays* and *Antirrhinum* (snapdragon flower), sequence from species with elevated production of phenylpropanoid compounds like *V. vinifera* and *P. hybrida,* the *R. idaeus* and *F. vesca*. Also, the sequence from Fv_bHLH_42 was included as this model protein exhibits a high percentage of similarity with Fv3 but does not have the ACT domain on the C-terminal region as mentioned before. Figure 14 represents the consensus ML obtained tree.

In the lower branches of the tree in Figure 14, the so-called bHLH 3 groups with *At* TT8 and homologs *Ph* AN1 and *Vv* MYC1 can be recognized. The second big cluster can be identified in the upper part of the tree, by the presence of the bHLH 33 group proteins and the MYC homologs. The third cluster that derives from the 33 cluster is formed by At_GL3 and At_EGL3 (trichome regulatory protein), and the well-known anthocyanin regulatory genes from *P. hybrida* JAF 13, and is in this small cluster were the closest homolog to *Fv_bHLH_145,* Pear *Pp*_0027062 is located (32, 33, 36).
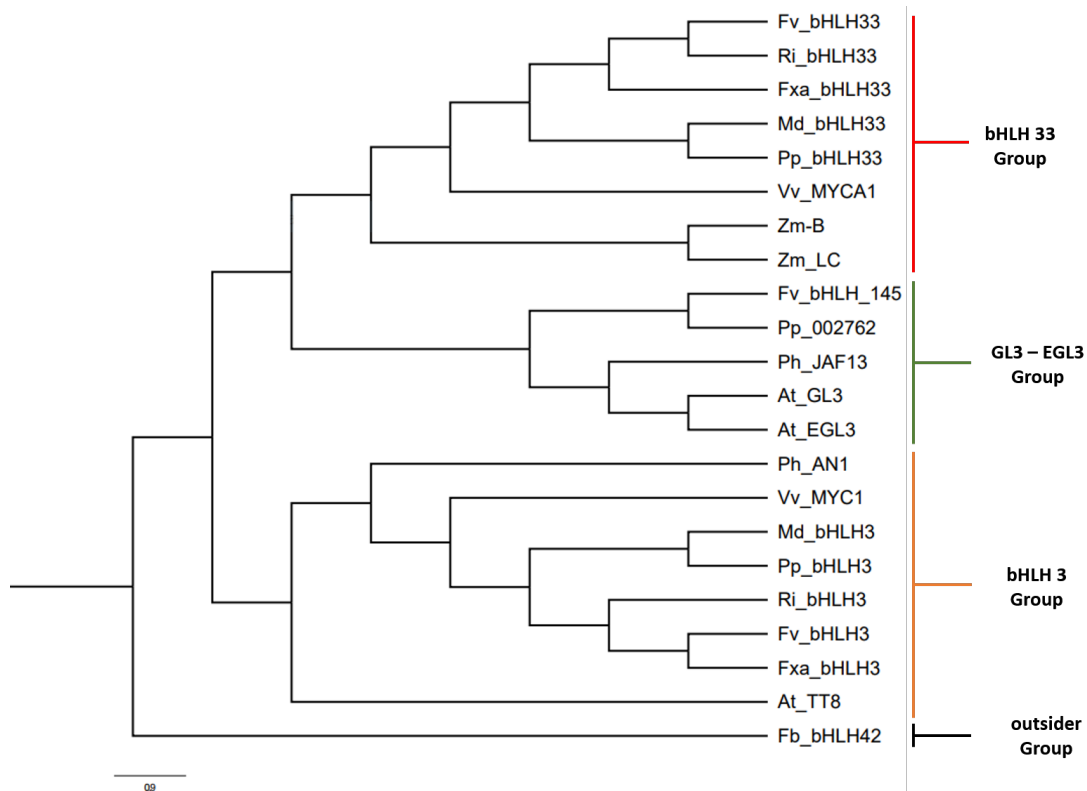
Figure 14: Molecular phylogenetic analysis by Maximum Likelihood method. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model. The analysis involved 22 amino acid sequences. There were a total of 150 positions in the final dataset. Evolutionary analyses were conducted in MEGA6. Proteins included in this analysis with accession number in parentheses are: MdbHLH3 (CN934367); MdbHLH33 (DQ266451), AtTT8, CAC14865; ZmB, CAA40544; ZmLC, AAA33504, FxabHLH3, AFL02463; FxabHLH33, AFL02465 (39, 40)

*A. thaliana* TT8 is located in the same cluster as *V. vinifera* Vv MYC1 (Myc anthocyanin regulatory protein). This result confirms the description of the phylogenetic relationship of *Vitis* bHLH proteins made by Hichri *et al.* (37), where it was proven that VvMYC1 is involved in the regulation of tannin and anthocyanin production. Other proteins located in this clade are also reported to have a phenylpropanoid regulatory effect in their species (37, 38).

# REFERENCES

1. Sailsbery, J. K., & Dean, R. a. (2012). Accurate discrimination of bHLH domains in plants, animals, and fungi using biologically meaningful sites. *BMC Evolutionary Biology*, 12 (1), 154.

2. Liu, W., & Chen, D. (2013). Phylogeny, Functional Annotation, and Protein Interaction Network Analyses of the *Xenopus tropicalis* basic Helix-Loop-Helix Transcription Factors. *BioMed Research International*, *2013*, 145037.

3. Sailsbery, J. K., Atchley, W. R., & Dean, R. A. (2012). Phylogenetic Analysis and Classification of the Fungal bHLH Domain. *Molecular Biology and Evolution*, *29* (5), 1301–1318.

4. Xie, X. B., Li, S., Zhang, R.-F., Zhao, J., Chen, Y.-C., Zhao, Q., … Hao, Y.-J. (2012). The bHLH transcription factor MdbHLH3 promotes anthocyanin accumulation and fruit coloration in response to low temperature in apples. *Plant, Cell & Environment,* 35 (11), 1884–1897.

5. 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 (1), 94–116.

6. Ludwig, S. R., Habera, L. F., Dellaporta, S. L., & Wessler, S. R. (1989). Lc, a member of the maize R gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the myc-homology region. *Proceedings of the National Academy of Sciences of the United States of America*, *86* (18), 7092–7096.

7. Letunic, I., & Bork, P. (2011). Interactive Tree of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research*, 39 W475–W478.

8. Heim, M., Jakoby, M., Werber, M., Martin, C., Weisshaar, B., Bailey, P. (2003). The basic helix–loop–helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. *Molecular Biology and Evolution*. 20 (5), 735-747.

9. Pires, N., & Dolan, L. (2010). Origin and diversification of basic-helix-loop-helix proteins in plants. *Molecular Biology and Evolution*, 27 (4), 862–74.

10. Morgenstern B, Atchley WR. (1999). Evolution of bHLH transcription factors: modular evolution by domain shuffling. *Molecular Biology and Evolution*. 16:1654–1663.

11. Carretero-Paulet, L., Galstyan, A., Roig-Villanova, I., Martinez-Garcia, J. F., Bilbao-Castro, J. R., Robertson, D. (2010). Genome-wide classification and evolutionary analysis of the bHLH Family of transcription factors in arabidopsis, poplar, rice, moss, and algae. *Plant Physiology*, 2010, 153:1398.

12. Atchley, W.R. and Fitch, W.M. (1997). A natural classification of the basic helix-loop-helix class of transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 5172–5176.

13. Bai, Y., Pattanaik, S., Patra, B., Werkman, J. R., Xie, C. H., & Yuan, L. (2011). Flavonoid-related basic helix-loop-helix regulators, NtAn1a and NtAn1b, of tobacco have originated from two ancestors and are functionally active. *Planta*, 234 (2), 363–375.

14. Kong, Q., Pattanaik, S., Feller, A., Werkman, J.R., Chai, C., … Yuan, L. (2012). Regulatory switch enforced by basic helix-loop-helix and ACT-domain mediated dimerizations of the maize transcription factor R. *Proceedings of the National Academy of Sciences of the United States of,* 109: 2091–2097.

15. Grant G. (2006). The ACT domain: a small molecule binding domain and its role as a common regulatory element. *Journal Biological Chemistry*. 2006 Nov 10;281.

16. Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher A, *et al*. (2011). The genome of woodland strawberry (*Fragaria vesca*). *Nature Genetics* 43:109–11.

17. Espley, R. V., Hellens, R. P., Putterill, J., Stevenson, D. E., Kutty-Amma, S., & Allan, A. C. (2007). Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *The Plant Journal*, 49 (3), 414–427.

18. Montefiori, M., Brendolise, C., Dare, A.P., Lin-Wang, K., Davies, K.M., …. Allan, A.C. (2015). In the Solanaceae, a hierarchy of bHLHs confer distinct target specificity to the anthocyanin regulatory complex. *Journal of Experimental Botany* 66, 1427–1436.

19. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 3.

20. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*, 421.

21. Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32 (5), 1792–1797.

22. Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, 30 (12), 2725–2729.

23. Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., & Leunissen, J. A. M. (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research,* 35, 71–74.

24. Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., … Bateman, A. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, *44-48.*

25. The *Arabidopsis* Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Research*. 2008 Jan; 36

26. 11. Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., … Viola, R. (2010). The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nature Genetics*, 42 (10), 833–839.

27. Kunihiro, A., Yamashino, T., Nakamichi, N., Niwa, Y., Nakanishi, H., & Mizuno, T. (2011). Phytochrome-interacting factor 4 and 5 (pif4 and pif5) activate the homeobox athb2 and auxin-inducible iaa29 genes in the coincidence mechanism underlying photoperiodic control of plant growth of *Arabidopsis thaliana*. *Plant and Cell Physiology*, 52 (8), 1315–1329

28. Kurbidaeva A., Ezhova T., Novokreshchenova M., (2014). *Arabidopsis thaliana* ice2 gene: phylogeny, structural evolution and functional diversification from ICE 1. *Plant Science* 229, 10-22.

29. Cifuentes-Esquivel, N., Bou-Torrent, J., Galstyan, A., Gallemí, M., Sessa, G., Salla Martret, M., … Martínez-García, J. F. (2013). The bHLH proteins BEE and BIM positively modulate the shade avoidance syndrome in *Arabidopsis* seedlings. *Plant* J, 75: 989–1002.

30. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2015). The Phyre2 web portal for protein modelling, prediction and analysis. *Nature Protocols*, 10 (6), 845–858.

31. Chipman, D. M., & Shaanan, B. (2001). The ACT domain family. *Current Opinion in Structural Biology,* 11 (6), 694–700.

32. Spelt C., Quattrocchio F., Mol J., Koes R. (2000). anthocyanin1 of petunia encodes a basic helix-loop-helix protein that directly activates transcription of structural anthocyanin genes. *Plant Cell* 12: 1619–1632.

33. Rahim, M. A., Busatto, N., and Trainotti, L. (2014). Regulation of anthocyanin biosynthesis in peach fruits. *Planta* 204, 913–929.

34. Starkevič, P., Paukštytė, J., Kazanavičiūtė, V., Denkovskienė, E., Stanys, V., Bendokas, V., … Ražanskas, R. (2015). Expression and anthocyanin biosynthesis-modulating potential of sweet cherry (*Prunus avium l.)* myb10 and bHLH genes. *PLoS ONE*, 10 (5).

35. Bonaventure, J. & Ghouzzi, V. (2003). Structure and function of H-TWIST. *Expert Reviews in Molecular Medicine*. Vol. 5; 29.

36. Payne CT, Zhang F, Lloyd AM. (2000). GL3 encodes a bHLH protein that regulates trichome development in *arabidopsis* through interaction with GL1 and TTG1. *Genetics* 156:1349–1362.

37. 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 (8), 2465– 2483.

38. Nesi, N., Jond, C., Debeaujon, I., Caboche, M., and Lepiniec, L. (2001). The *Arabidopsis* TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. *Plant Cell* 13, 2099–2114.

39. Lin-Wang, K., McGhie, T. K., Wang, M., Liu, Y., Warren, B., Storey, R., … Allan, A. C. (2014). Engineering the anthocyanin regulatory complex of strawberry (*Fragaria vesca*). *Frontiers in Plant Science*, *5*, 651.

40. Schaart, J. G., Dubos, C., Romero De La Fuente, I., van Houwelingen, A. M. M. L., de Vos, R. C. H., … Bovy, A. G. (2013). Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (*Fragaria × ananassa*) fruits. *New Phytolist*, 197: 454–467.

# Chapter 4

bHLH candidates in fruit development and relationship with flavonoid pathway genes

## INTRODUCTION

Having identified the candidates of the bHLH IIIf family by the phylogenetic analyses shown in the previous chapter 3, I continued the studies on the following genes; three bHLH genes for *F. vesca* (Fv bHLH3, Fv bHLH33 and Fv bHLH145) and two bHLH genes for *R. idaeus* (Ri bHLH3 and Ri bHLH33).

In this chapter, we are going to follow up the fruit development process on both species especially their gene expression patterns to determinate, if our candidate genes are expressed in the fruit tissue and if their expression patterns change as the fruit develops and the colored anthocyanins and other phenylpropanoid compounds are synthesized.

Previous work on closely related species has been focused exclusively on the difference in gene expression between contrasting color cultivars and fruit tissue vs. vegetative tissue as leaves. During my Ph.D. research, I wanted to move a step forward on this subject and study how the bHLH genes change their expression during the fruit development compared to green leaves (1, 2, 3).

Later on, we continued with the candidate bHLH genes, by evaluating the effect of the overexpression of each gene in a heterologous plant system broadly used in the TF research, as the Renilla/Luciferase activity by the activation of a specific promoter.

### Promoter regions & transcription factors

Gene expression is the process where a gene is switched on, and it is the result of the complex interaction of environmental stimuli, *cis*-acting elements, and TFs among others. The *cis*-elements are regions of non-coding DNA that regulate the transcription of neighboring genes, as represented in Figure 15. For example, they can be located in the promoter regions of the upstream DNA sequences of a gene, and it is this particular region that contains specific sequences recognized and bound by proteins involved in the initiation of transcription (4, 5, 6).

Characterization of plant promoters and their interaction with TFs is often carried out by generation and analysis of transgenic plants expressing promoter/reporter gene fusions. The most commonly used reporter genes are the GUS reporter and the bioluminescence protein luciferase (5, 7).

Transient expression is a commonly used technique alternative to a stable transformation. As reported by several literature examples, a simple infiltration of *Agrobacteria* carrying plasmid constructs of DNA of interest into tobacco leaves, *in vivo* expression analysis of promoters and TFs can be conducted in as little as two to three days (5, 8, 9).
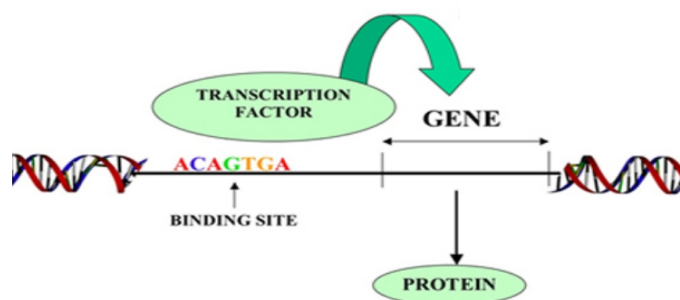


Figure 15: Simplified scheme representing the promoter region of the gene and the binding site for a TF resulting in the activation of the gene translation.

**Renilla luciferase assay**

The luciferase reporter assay (LUC/REN) is widely used as a tool to study gene expression and its regulation (activation or repression). It is a convenient, relatively inexpensive assay that gives quantitative measurements in a short time. There are several commercially LUC/REN kits available that include all the reagents necessary for the reaction to occur. One of the most common proteins used is luciferase from firefly (*Photinus pyralis*). Using this assay, a second luciferase reporter is used for normalization purposes. In the same expression vector are inserted the *LUCIFERASE* gene (controlled by a strong constitutive promoter like 35S), and the *RENILLA* gene controlled directly by the specific promoter to study (9, 10, 11, 12).

To perform the reporter assay, it is necessary to clone the regulatory region of the gene-of-interest upstream of the luciferase gene in one expression vector, transform *Agrobacterium* cells and then infiltrate the plant tissue with the bacteria cells suspended in an adequate infiltration media. Since the promoter region of a gene of interest is fused to the luciferase reporter gene, the luciferase activity can be directly correlated with the activation of the promoter this gene of interest (9, 10, 11).

## MATERIALS AND METHODS

**Plant material**:

Fruit and leaf tissues were collected from the varieties mentioned in the general material and methods section (2.2). The plants were grown and maintained at the facilities of the Fondazione Edmund Mach in Vigalzano at 533 meters altitude (Province of Trento, Italy). Between 10-12 berries were collected from five different plants of the same cultivar for RNA isolation.

Fruit tissue from different developmental stages was collected according to earlier studies carried out by the metabolomics group of the Fondazione Edmund Mach, and the New Zealand Plant and Food Research Institute. The time points selected for the woodland strawberry fruit development are depicted in Figure 16.



Figure 16: Overview of the different developmental stages of fruits of *F. vesca* that were harvested in this study.

TP1 and TP2 are the earliest stages of fruit development in *F. vesca* (TP meaning Time Point), they correspond to the flower bud before opening and fully opened flowers. TP3 is the time point where ovary tissue is beginning to expand, approximately 3-4 days after the flowers are fully opened. At TP4, fruits start to show a light green coloration with white achenes, 10 days after flowers have fully opened. At TP5, the fruit begins the color transformation to white/pink, this process occurs 2-2.5 weeks after the flower opening, from this point the fruit stops increasing in size. TP6 corresponds to fully mature fruits, where the berries have a completely red color.

In the case of raspberry, the three flower stages (TP1-TP3) were not collected. Instead, only the different developmental stages of fruit development equivalent to the strawberry material (TP4-TP6) were collected. These stages represent the white (growing phase), pink (turning color and maturation starting stage) and red fruit (mature berry), as visualized in Figure 17.

Figure 17: Overview of the different developmental stages of fruits of *R. idaeus* that were harvested in this study.

## RNA isolation and cDNA synthesis

Fruit tissue was manually ground by ceramic mortar and pestle under liquid nitrogen until a fine powder was obtained. RNA was extracted using the Pine-tree RNA extraction method reported by Lin-Wang (13). In brief, 100 mg of tissue was dissolved in 5 mL extraction buffer in a Falcon tube and heated to 65°C for 5 min. An equal volume of chloroform: isoamyl alcohol (24:1) (5 ml) was added and the suspension was mixed, and centrifuged at room temperature at 13,000 g for 10 min to separate the phases. The top (aqueous) phase was transferred to a new tube, followed by a second centrifugation at 13,000 g for 10 min.

Finally, the aqueous phase was transferred to another centrifuge tube, and the volume estimated, a ¼ volume of 10 M LiCl was added to the supernatant and mixed, and incubated at 4°C overnight. The tube was centrifuged at 4°C at 13,000 g for 20 min, and the RNA pellet was dissolved in 500 µl of SSTE buffer (preheated at 65°C) (see general Material & Methods), and extracted with an equal volume of chloroform: isoamyl alcohol. After centrifugation for 10 min at 13,000 g, the top phase was transferred, and two volumes of ethanol (98%) were added to precipitate the RNA. Tubes were incubated at least 30 min at -70°C or 2 h at -20°C, followed by 20 min of centrifugation at 4°C. The pellet was washed with 75% ethanol, dried and dissolved in 30 µl of RNase free water.

Total RNA from leaf material was extracted using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich) according to the manufacturer's manual, using 100 mg of tissue. Before RNA elution, an on-column DNAse I treatment (Sigma-Aldrich) was performed.

## cDNA synthesis

1 µg of total RNA was used for the DNA synthesis using the SuperScript® VILO cDNA Synthesis Kit and Master Mix in a final volume of 22 µL. The integrity of the obtained cDNA was tested with actin primers.

**Primer design for Real-Time quantitative PCR**

The qPCR primers were designed to amplify fragments smaller than 250 bp from the coding region of each transcript.

Each primer was tested twice: first under standard PCR conditions, and later in a cDNA pool mix of all the samples (1:20) to confirm the specificity of each primer set and evaluate the primer efficiency.

Table 10. Sequences of the qPCR primers used in this study and their amplification product sizes.

| PRIMER | SEQUENCE | AMPLIFIED FRAGMENT |
|--------|----------|--------------------|
| Fv3_qFor | CACCTCACTGCCATCCGCAACAA | 168 bp |
| Fv3_qRev | GGGAGAGGACGTCCGTTACCGTT | |
| Fv33_qFor | ATGGCCAATGGGACTCAAATC | 132 bp |
| Fv33_qRev | TTACCAGCAATTTTCCAAAG | |
| Fv145_qFor | ATGGGTTCTAGGCCCCAGAACCAGGAG | 168 bp |
| Fv145_qRev | CGGCTCAGTCACACCCAGCTCTA | |
| Ri3_qFor | TCGGTTGAGAAAGGGGACCCCAC | 117 bp |
| Ri3_qRev | AAGGGCACCAACGACCGGAGTAT | |
| Ri33_qFor | CCTTGGAACATTCATCAGCA | 118 bp |
| Ri33_qRev | CCTCTGCACCTCACAAATCA | |

**qPCR set up**

For the qPCR analysis, Master Mix from BioRad was employed: using 2 µl of 20x diluted cDNA in a 50 µl reaction volume qPCR conditions were 3 min at 95°C, followed by 40 cycles of 5 s at 95°C, 5 s at 60°C, and 10 s at 72°C, followed by 65°C to 95°C melting curve detection. The qPCR efficiency of each primer set was measured analyzing the standard curve of a cDNA serial dilution of each gene.

***Agrobacterium* Infection**

*Agrobacterium tumefaciens* is the causative agent of crown gall, a disease of dicotyledonous plants characterized by a tumorous phenotype. *Agrobacterium* is a gram-negative bacterium

with a predominantly saprophytic lifestyle that has evolved to genetically engineer plant cells and use them to produce compounds that they can utilize as a carbon/nitrogen source (8).

During the infection by *A. tumefaciens*, a piece of DNA is transferred from the bacterium to the plant cell. This DNA is a copy of a segment called the T-DNA I (transferred DNA I), and it is carried on a specific plasmid, the tumor-inducing or Ti-plasmid. Due to this particular property of *A. tumefaciens*, this microorganism is widely used as a tool for genetic engineering, not only in plants but also in other organisms. Specific strains have been developed for the transformation of animal cells (8, 14).

## Constructs for *Agrobacterium* transformation

To carry out overexpression studies, several vectors were used. Gateway cloning technology is a universal cloning method based on the site-specific recombination properties of bacteriophage λ Lambda (Thermo Fisher® Scientific). This technology provides a rapid and highly efficient way to move DNA sequences into a vector system for functional analysis and protein expression (15).

As a first step, the entire CDS sequence of each candidate gene was transferred into the entry donor/master vector pDONR221, which contains the M13 F and R primer recognition sites for sequencing analysis and the kanamycin resistance gene. Entry clones were generated by performing a BP recombination reaction between a pDONR 221 vector and the purified attB PCR product. *E. coli* TOP 10 cells were transformed with the BP reaction mixture.

Once each gene sequence was confirmed in the pDONR 221 vector, an LR reaction was performed with the desired destination vector (see Table 11) and later *A. tumefaciens* strain GV 3101 cells were transformed with the construct of interest.

Table 11. Plant expression vectors.

| Construct | Vector | Inserted fragment | Antibiotic resistance | Plant selection |
|---|---|---|---|---|
| Fv_3 | pK7WG2 | 2172 bp | Spectinomycin | Kanamycin |
| Fv_33 | pK7WG2 | 1962 bp | Spectinomycin | Kanamycin |
| Fv_145 | PHex2 | 1896 bp | Spectinomycin | Kanamycin |
| Fv3 | pUbi10:GW-3xHA | 2172 bp | Spectinomycin | Basta |

All the Gateway cloning-related steps were performed as described in the manufacturer's manuals (16).

## Tobacco leaf agroinfiltration

Fresh *Agrobacterium* was resuspended in 10 ml of infiltration buffer (10 mM $MgCl_2$, 0.5 M acetosyringone) to an $OD_{600}$ of 0.8, and incubated at room temperature without shaking for 3 h before infiltration. Infiltrations were performed according to the methods described by Bourras *et al*. (14): approximately 500 µl of *Agrobacterium* suspension was infiltrated at four points into a young tobacco leaf using a 1 mL syringe without needle.

## LUC/REN Assay

Promoter activity is expressed as a ratio of luciferase to renilla activity. The promoter–LUC analyzed was in the binary plasmid pGreenII 0800-LUC (7) and used for transient transformation by mixing 100 µl of *Agrobacterium* strain GV3101 transformed with the LUC reporter cassette with 450 µl of each of two other *Agrobacterium* cultures transformed with cassettes containing: 1) *M. domestica MYB10* gene fused to the 35S promoter (17) and 2) a *F. vesca* bHLH gene in either pGreenII or pHex binary vectors (7, 17).

Three days after inoculation, 3-mm leaf discs (4 technical replicates from each plant) were cut with a hole-puncher, and placed into wells of a 96-well-plate containing 100 µl of PBS (phosphate buffered saline) pH 7 in each well, and gently crushed with the help of a plastic pistil. For the Luciferase reaction, the Firefly & Renilla Dual Luciferase Assay Kit from Biotium (Fremont, United States) was used. The measurements and analysis were carried out using a Synergy 2 Multi-Mode Reader from Biotek (Winooski, United States) and the Gen5™ Reader Control and Data Analysis Software.

## Quantification of polyphenols

Polyphenol extractions and quantification from tobacco leaves were performed as described in Palmieri *et al.* (18). Briefly, the extraction was performed using 1 $cm^2$ of the leaf (approximately 100 mg of tissue) in 500 mL acetone: water (70:30 v/v), the leaf was crushed with the help of a metal pestle, and the extracts were incubated in the dark at 4°C. After 48 h, the extracts were centrifuged at 18,000 g for 15 min at 4°C, and the supernatant was collected and transferred to an HPLC glass vial that was stored at -20°C until the analysis. The analysis

was performed by an UPLC system coupled with a triple quadrupole (TQ) mass spectrometer according to Palmieri *et al.* (18).

The separation of the phenolic compounds was achieved on a Waters Acquity HSS T3 column 1.8 µm, 100 mm × 2.1 mm (Milford, MA, USA), the extract was kept at 40°C, with two solvents: A (water containing 0.1% formic acid) and B (acetonitrile containing 0.1% formic acid), the flow was set at 0.4 ml/min, and the detection was performed using multiple reaction monitoring (MRM). The calibration curves were performed previously for each compound for precise quantification, using commercially available standards.

**Floral dip *Arabidopsis* stable transformation**

For the mutant complementation assays, the mutant line *tt8* was transformed according to Bent *et al.* (19). Briefly, *Arabidopsis* plants were grown until flowering under long days in pots in soil at 24ºC. The plants that were used had many immature flower clusters, and fertilized siliques were removed. *A. tumefaciens* strain GV3101 carrying the gene of interest cloned into the pHex binary vector was grown overnight at 28ºC in LB medium with the appropriate antibiotics to select for the binary plasmid. *Agrobacterium* cells were collected by soft centrifugation and resuspended to $OD_{600}$ of approximately 0.8 in 5% sucrose solution. Silwet L-77 was added to a final concentration of 0.02% (200 µl/L). The floral parts of the *Arabidopsis* plants were dipped in *Agrobacterium* solution for 30 seconds, with gentle agitation, this step was repeated another two times, and the plants were kept horizontally and covered for 16 to 24 h to maintain high humidity. After 3 weeks, dry seeds were harvested, and transformants were selected by plating out the seeds on half MS medium using the antibiotic of selection according to the plasmid (19).

**Anthocyanin Inductive Conditions (AIC) on *Arabidopsis* seedlings**

Approximately 30 seeds of each transgenic line of *A. thaliana* plus the WT, were sterilized with ethanol 70% for 15 min, followed by another 10 min with 100% ethanol and dried under the sterile hood. Seeds were grown on a 30% sucrose solution to induce anthocyanin biosynthesis, as reported by Leipinic & Pourcel 20). The seeds were kept in cold and dark conditions for 2-3 days and later maintained at a constant temperature of 24ºC and daylight conditions of 16 h under 80 rpm agitation (20). After 6 days, the germinated seedlings were photographed and collected for further analyses.

**Total anthocyanin quantification**

*A. thaliana* seedlings were rinsed with distilled $H_2O$, dried on filter paper and between 5-10 mg of fresh tissue was collected for each one of the two replicates. Anthocyanins were extracted according to the grapes anthocyanin quantification method described by Ehrhardt *et al.* (21) by adding 500 µl 80% methanol to the tissue sample and disrupting it by plastic pestle. The tubes were shaken at 2 g for 1 h. The extraction was continued for 48 h in the dark at 4ºC. Brief strong centrifugation was performed (18,000 g for 10 min) and about 100 µl supernatant was collected in dark vials. UPLC analysis and quantification were conducted using 20 µl of the sample injected on a C18 column as described by Ehrhardt *et al.* (32) using water containing 5% v/v formic acid (A) and methanol containing 5% v/v formic acid as eluting solvents (21, 22). The commercial standard Cyanidin 3-*O*-glucoside (Sigma-Aldrich, St. Louis, Missouri, USA) was employed for the calibration curve on the total anthocyanin sample quantification.

## RESULTS AND DISCUSSION

### *Fragaria* HouseKeeping Genes (HKG)

Before measuring the gene expression of the candidate *bHLH* in *F. vesca*, first, it was necessary to identify the best housekeeping genes (HKGs) for this species, and especially the fruit tissue. The six most commonly used HKGs were taken from *A. thaliana* literature: *GADPH* (GeneBank protein accession AAF26801), *PP2A* (ABF85773), *SAND* (AEC08115.1), *UBC9* (AEE85414.1). For the gene *Actin* (AAB37098.1), two different primer sets were tested, the first one explicitly designed for *F. vesca*, and the second set reported for *Fragaria x ananassa* (13, 23).

The HKG primers were designed originally for the strawberry genome and each tissue (leaf and six fruit time points) was evaluated for consistent expression of this gene. The objective was to establish the average $C_t$ for each HKG in which the expression can be detected above background levels. As can be seen in Figure 18.1 the expression of all genes was detected between the cycles 14 and 24; values acceptable for qPCR analyses.
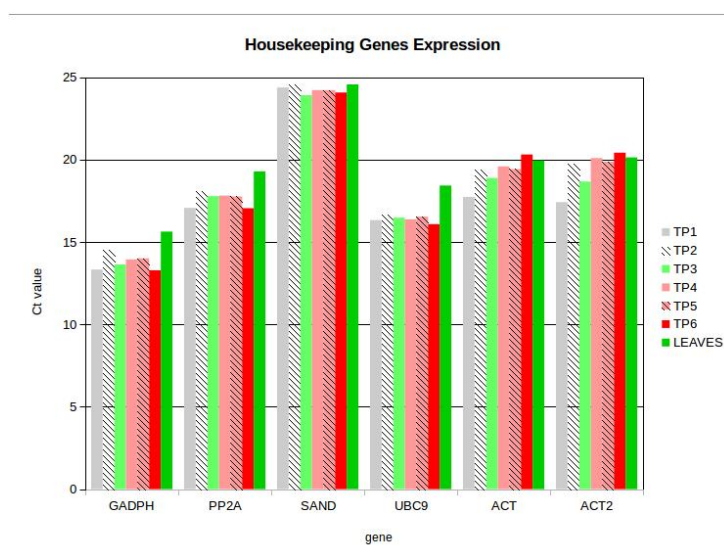


Fig 18.1: qPCR Ct values for the six HKG evaluated in *F. vesca* fruit tissues and leaf

*SAND*, *UBC*9, *PP2A*, and particularly the *Actin* primers have a relatively similar expression in the different fruit samples. Expression of *Actin* in TP1 could be detected from cycle 17, while expression in the mature fruit, or TP6, could only be detected after cycle 21. Expression of most HKGs in the leaf tissue could be detected about 2-3 cycles later, except for *Actin*, which was detected at about the same cycle number as for the fruit tissue. As each pair of primer sets for the *Actin* gene are working with proper gene expression $C_t$ values and calibration

curves (described in the next pages), we decided to continue working with the first primer set: ACT.

The gene, which presented the most constant genes expression value was *SAND*, but the expression of this gene could only be detected after about 24 cycles as seen in Table 12. For this reason, this gene was removed as HKG for the following experiments.

As the candidate HKG genes *PP2A*, ACT, GADPH, *SAND*, and *UBC9* worked well in *F. vesca*, we decided to find the homologous genes in *R. idaeus*, using the same approach as for the bHLH proteins described for *F. vesca*. In general, it can be noticed that the leaf tissue had a lower gene expression for the three HKG candidates, as shown in the summarized results of Table 12.

Table 12. Average $C_t$ values for *F. vesca* HKG qPCR analyses

| GENE | MEAN $C_T$ FRUIT | $C_T$ LEAF | MEAN $C_T$ FRUIT and LEAF |
|------|------------------|------------|---------------------------|
| GADPH | 13.79 | 15.635 | 14.05 |
| PP2A | 17.60 | 19.29 | 17.84 |
| SAND | 24.23 | 24.57 | 24.28 |
| UBC9 | 16.42 | 18.4325 | 16.71 |
| ACT | 19.23 | 19.96 | 19.33 |
| ACT2 | 19.37 | 20.1375 | 19.48 |

**Rubus HKGs**

As described for *F. vesca* above, before starting the gene expression analysis of bHLH genes during fruit development it was necessary to find good HKGs to use as controls. In the *R. idaeus* genome, seven candidate members of the *UBC* family were identified (data not shown), but no homolog of *UBC9* of *F.* vesca was found. Therefore, the *UBC9* gene is not included in our analysis. The results of the HKG evaluated are summarized in Figure 18.2.
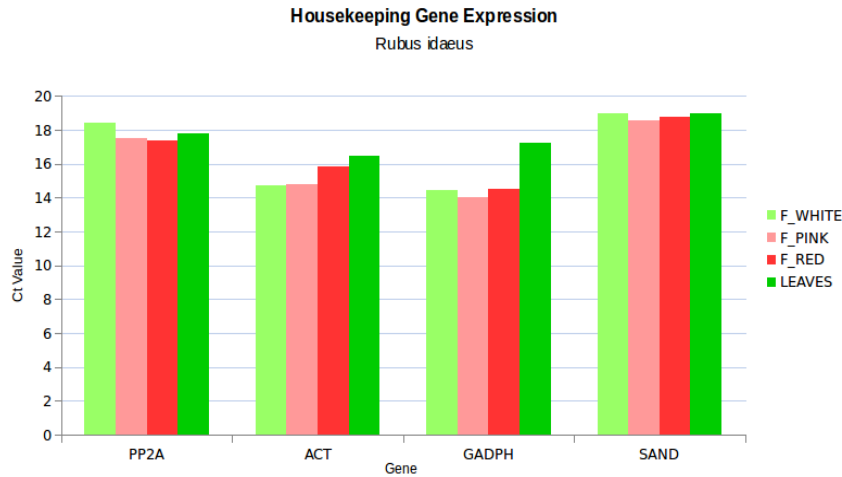
Figure 18.2: qPCR Ct values for the six HKG evaluated in *R. idaeus* fruit tissues and leaf.

HKG genes for *R. idaeus* presented good activity in the qPCR analyses, being detected after 14 and 19 cycles. Additionally, they showed a good consistency among the tissues analyzed. As seen in the *F. vesca* experiment, leaf tissue had a modestly higher gene expression value than the fruit tissues on the *ACT* and *GADPH* genes (2-3 cycles), but, interestingly, the leaf had a lower gene expression than all the fruit tissues for *PP2A* and the same values in the *ACT* HKG.

Table 13. Average $C_t$ values for *R. idaeus* HKG qPCR analyses

| GENE | MEAN $C_T$ FRUIT | $C_T$ LEAF | MEAN $C_T$ FRUIT + LEAF |
|------|------------------|------------|--------------------------|
| *PP2A* | 17.77 | 17.82 | 17.79 |
| *ACT* | 15.13 | 16.44 | 15.46 |
| *GADPH* | 14.32 | 17.265 | 15.06 |
| *SAND* | 18.76 | 19.02 | 18.83 |

*ACT* and *GADPH* were detected at the lowest $C_t$ numbers, and the *SAND* genes the same as in *F. vesca* had the higher gene expression values.

In general, the primers in *R. idaeus* behave in the same way as their equivalent primers in the homolog genes in *F. vesca*. This result helps for future reference studies on berries qPCR experiments and also helps in this study to obtain robust further analyses.

## Primer validation: specificity and standard curve

To confirm the results on the HKG genes analyzed previously, it was necessary to perform the corresponding qPCR method validation:
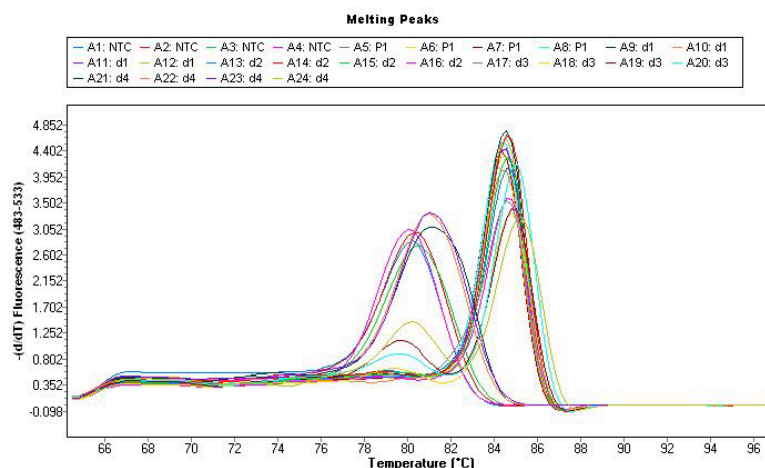


Figure 19a: Melting curves of real-time RT-PCR products from 2 HKG primer sets in *F. vesca*. Melting temperatures of *SAND* products are between 80.5°C and *PP2A* are close to 85°C.

We performed an analysis of the peaks of the melting curve of the PCR products of each of the primer combination evaluated in this study. The Figure 19a shows the result of the qPCR runs in which two different primers were analyzed (*F. vesca SAND* and *PP2A*). At the right side of the figure, a curve with a single peak around 80°C is seen; the left side of the same graph shows the melting curve of the PCR product amplified by the *PP2A* primer combination, in which some unspecific amplification is observed as a minor peak is formed at 80°C, besides the significant peak at 85°C. This small peak could be due to the presence of a second amplicon being produced during the amplification. As a result, the *PP2A* primers are excluded from our analyses.

The standard curve analyses are necessary to establish the PCR efficiency of the primers used in the reaction; a dilution series of all cDNA samples were pooled and used for this purpose, using the BioRad software the real-time PCR standard curve was made. Figure 19 b, is a graphical representation of the semi-log regression line plot of CT value vs. log of input nucleic acid, slope values of –3.32 indicate a PCR reaction with 100% efficiency (24).
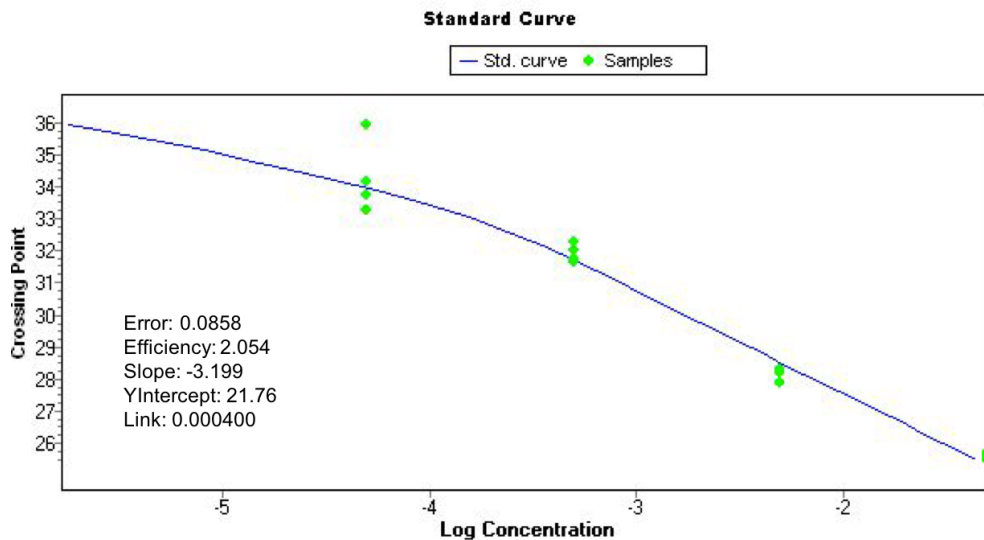
Figure 19b: Calibration curve on cDNA mix from fruit. The standard curve was generated by plotting the mean Ct values versus the Log10 of serial dilutions of RNA in water from 1:1 until 1 dilution to 1:100. Reaction equation is calculated as: -3.199 x + 21.76, based on the curve information as displayed in the graph.

As the manuals and literature report: slopes with a value exceeding –3.32 may indicate sample quality or pipetting problems, and slopes with lower values (ex. -3.9) suggest that the reaction is less than 100% efficient. The efficiency of the primers used in this study was checked and confirmed with the slope values (Figure 19 b) (24).

HKG identification in plant species is a recurrent subject in molecular biology journals, even after many years of the introduction of real-time qPCR technique. For *F. vesca* and *R. idaeus,* no specific HKG validation reports had been made so far. Studies performed in *M. domestica* on *HKG* stability are the closest reports that can be used (25).

The HKG analyses presented here, provide new reference primers that can be used for gene expression studies in *F. vesca* and *R. idaeus*, and at the same time they will provide standardized data that can be easily compared to the gene expression values of other species normalized with the same homologous genes *ACT, UBC9* and *NADPH* (25, 26, 27).

**Gene expression analysis by Real-Time PCR (qPCR) in fruit and leaf tissue**

*F. vesca bHLH* **expression patterns**

To determinate if the three bHLH candidate genes previously identified in chapter 3 of this Ph.D. project are involved in the regulation of the phenylpropanoid biosynthesis pathway, it was necessary to evaluate their gene expression patterns in the fruit tissue, and how their expression changes during the fruit maturation process. RNA extracted from *F. vesca* fruit

developmental series was employed, and after subsequent cDNA synthesis, real-time PCR was performed and the results summarized in Figure 20.

The three candidate *bHLH* genes were all found to be significantly expressed in the flower (TP1-3) and fruit time points (TP4-6). All data were normalized against the HKG *Actin* to be comparable with other studies on bHLH expression in fruit maturation (see Figure 19).
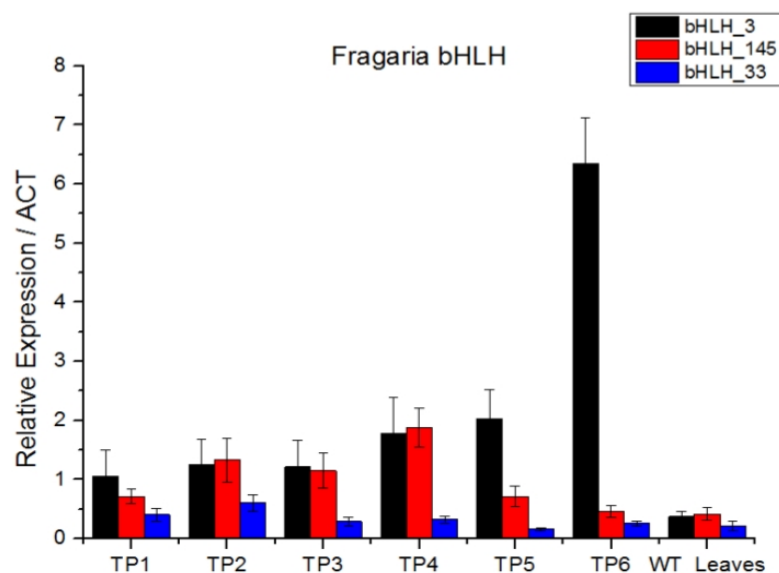


Figure 20: Bar graph showing the mRNA expression level of the three *bHLH* gene candidates of *F. vesca* during fruit development and in leaf tissue. Error bars equal to the Standard Error of four technical replicates. All data normalized to Actin.

The first candidate *FvbHLH3*, showed an increased expression as the fruit maturation process goes. This increment is subtle in the early stages and extremely high for the ripe fruit; with a 6-fold increase in expression of the ripest stage compared to the first developmental stage. This result also correlates with the color development in anthocyanin formation in the fruit tissue (see following chapter 4). The expression of *Fv3* in leaves is about half the level of that in fruit tissue.

*Fv145* showed lower values than the homolog gene *Fv3*, with an increment in the expression values along the fruit maturation process (T1 to TP5). The main differences can be seen in the TP1 tissue, where the expression levels are almost 50% lower than the *Fv3* values. Furthermore, it is necessary to mention the abrupt decrease in the expression of *Fv145* in TP5 (pink color turning fruit) and TP 6 (mature fruit), with a 60 % reduction. The expression of this candidate gene doesn't seem to correlate with the formation of color, as the candidate genes *Fv3* and 33, where their expression increases as the coloration in the fruit appears. Due to the recent identification of this IIIf bHLH, there are no literature reports of homolog genes involved in anthocyanin formation beside the paper of Rahim *et al.* (1), where Pp_bHLH 3 (homolog to Fv bHLH3) was identified as the bHLH gene responsible for the red coloration in peach fruits.

*Fv33* is lowly expressed compared to the other bHLH genes analyzed. Additional analyses were made, where other HKG were used for the gene normalization. As a result, the expression values obtained have similar results (data not shown) as before. Compared to the other two bHLH candidate genes in the same fruit tissue, the *Fv33* expression was extremely low. This suggests that FvbHLH33 is not involved as much as *Fv3* in color formation in the *F. vesca* fruit.

Interestingly, expression values for *Fv3* and *Fv33* at TP3 are similar compared to the immature tissue in TP2. These similar results can be due to technical problems in the RNA and cDNA processing and manipulation, that lead to under- or overestimation of the initial messenger RNA present in the samples. To formulate solid statements about the bHLH constant expression or not during *F. vesca* fruit formation development, it is recommended to perform another tissue sampling and cDNA synthesis. This new data will might confirm the ambiguous results obtained here.

*Fv145* seems to be important during the intermediate stages of the fruit development, based on its decrease of expression in the mature stages (TP5 and TP6). It would be informative to check this gene expression in other *Fragaria* species and varieties, to investigate this further. This result could be comparable to what has been reported in other species and for other regulatory proteins like MYBs, where it was found that one member of the protein family can have a function in early fruit development and another member during later stages. So, it is possible, that the different bHLH TFs are active during distinct developmental stages (28, 29).

**Expression of *Fragaria* flavonoid pathway genes**

After testing the expression of the bHLH candidate genes and determining their expression pattern during flower and fruit development, the next step was to determine the expression of three main biosynthetic genes involved in the anthocyanin pathway to correlate their expression with the bHLH candidate genes.

The analyzed genes were selected according to their importance for anthocyanin biosynthesis, and their previous characterization in other plant species. The chosen genes were: 1) gene chalcone synthase (*CHS*, EC 2.3.1.74), 2) the intermediate pathway gene dihydroflavonol 4-reductase (*DFR* EC 1.1.1.219) and 3) late gene hypothesized to add one possible sugar modification UDP-flavonoid glucosyltransferase (*UFGT* EC 2.4.1.91).

CHS is required for the initial step of the flavonoid biosynthetic pathway including the branches to flavones, flavan 3-ols, anthocyanins and proanthocyanins as this protein catalyzes the conversion of 4-coumaroyl-CoA and malonyl-CoA into naringenin chalcone (see also Figure 3 – pathway in Introduction chapter) (30).
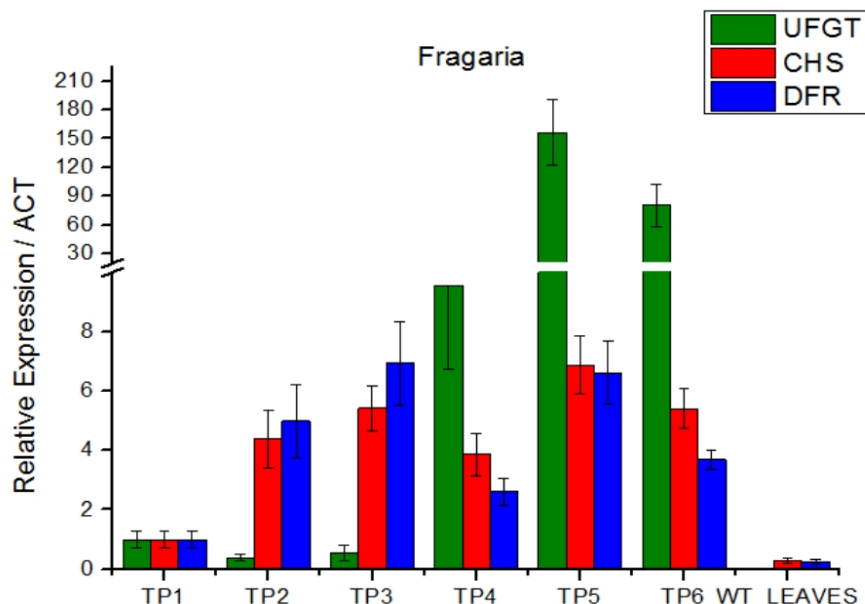


Figure 21: Bar graph showing the gene expression of three genes of the phenylpropanoid pathway in *F. vesca* in different stages of fruit development and in leaf tissue. Gene expression values normalized to *ACT* gene. Error bars equal to the Standard Error (SE) of four technical replicates.

*CHS* gene expression follows a bimodal expression pattern, with expression peaking on TP3 and TP5. *CHS* is expressed at higher levels on TP6 than on TP1 (Figure 21). It is surprising to find that *CHS* is expressed at such high levels in the later stage of fruit development because *CHS* is involved in the first step of the flavonoid biosynthetic pathway.

In Figure 21, it can be observed that the expression level of the biosynthetic pathway gene *DFR* shows a bimodal pattern, reaching the first peak at TP3 and the second peak in expression at TP5 with values that are just about the same as for TP3, and a high expression in the fruit developmental series and low expression in the leaf tissue. Expression levels of *DFR* and *CHS* are comparable. No expression was detected for the *UFGT* gene in leaf tissue. The values presented in Figure 21 are the mean of four technical replicates, and standard error (SE).

In leaf tissue, the gene expression for all three genes was low, which was expected due to the characteristics of the mature green leaf, where the principal function of this tissue is the photosynthesis process, rather than the formation of colored phenolic compounds.

In the case of red fruits, UFGTs genes are involved in the sugar modifications happening in the final steps of the anthocyanin biosynthesis. The red color in fruits appears in the middle-late stages of fruit development, same moment as UFGT, expressed in the later stages of the flavonoid biosynthesis pathway (anthocyanins and PAs). Figure 21 displays *UFGT* gene expression in the green bars, *UFGT* presents low expression values in the first three stages TP1, TP2 and TP3; with minor variations among tissues, subsequent fruit mature tissues had an 8-fold increase for TP4, until almost 100 times for the TP5 or pink tissue.

Leaves did not show any expression of the *UFGT* analyzed in this experiment. An additional qPCR was performed to confirm the absence of amplification products. Nevertheless, in the light of the results of Schulenburg *et al.* (31) in 2016, in which GT genes were identified to be involved in ellagitannin glycosylation in *F.* x *ananassa*; and comparing those *UFGT F. ananassa* sequences, with our *F. vesca* and *R. ideaus* UFGT gene models (selected for the primer design used in this experimental chapter), a high homology in their sequence could be observed, supporting the idea that our UFGT candidate genes seem to be part of the ellagitannins pathway. The fact that there is no expression of *UFGT* in leaf makes sense considering that green photosynthetic tissues are not known for the glycosylation of ellagitannin precursors (31).

*CHS* belong to a gene family with multiple members, not only in the Rosaceae family but also in the Asteraceae family. Different *CHS* genes have been reported with differences in expression patterns among different tissues and stress conditions. In *Gerbera hybrida*, the gene *GCHS3* is involved in colorless flavonoid formation. In Gerbera petals, another *CHS*, *GCHS4* is involved in the anthocyanin biosynthesis in the vegetative tissues of this species (30, 32).

Considering this and the two putative members identified in the *F. vesca* genome so far, it does not exclude the possibility that the expression patterns that were found for the fruit tissue series of *F. vesca* are in fact two different phenylpropanoid types being synthesized at different time points at the same tissue (30).

## *Rubus idaeus* bHLH

After studying the expression of the *F. vesca* bHLH candidate genes, and the gene expression of three main flavonoid pathway genes during fruit development, and following the same experimental procedures, *R. idaeus* bHLH candidate genes were analyzed to establish their role in the phenylpropanoid regulation in this species.

*R. idaeus* bHLH candidate genes were analyzed by qPCR, and as earlier mentioned in the materials section of this chapter, only three fruit time points and the leaves were examined. The HKG *SAND* was chosen in this case for normalization values, and the experimental data were confirmed by a second cDNA synthesis, due to the general low values obtained in this experimental part. Figure 21 shows the most representative information with the SE bars derived from the data of four technical replicates.

*RibHLH3* had low expression values throughout fruit development, which reached the minimum levels at 0.002 (*SAND* normalized) that can lead to false assumptions of non-existence gene expression. However, due to the same low values obtained for the homologous gene *Fv33,* it was decided to analyze these data (Figure 22).
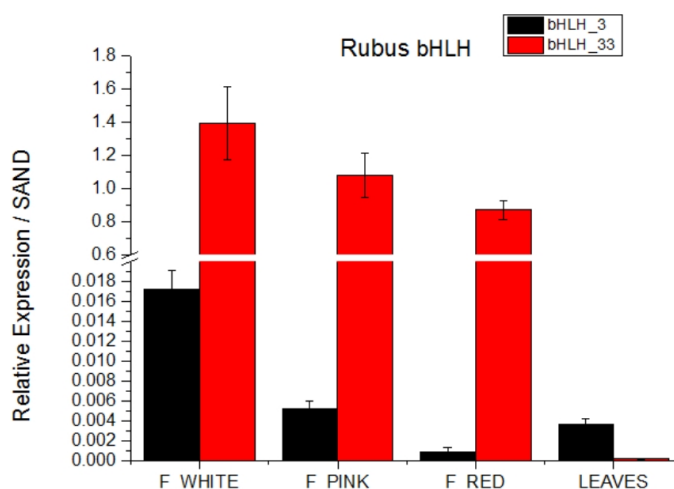


Figure 22: Bar graph showing the mRNA expression level of the three bHLH gene candidates on *R. idaeus* during fruit development time and leaf tissue. Gene expression values normalized to *SAND* gene. Error bars represent the Standard Error (SE) of four technical replicates.

The white fruits had the highest expression for *RibHLH3* among the samples (including leaves), and from this stage, until the mature red fruit, it had a decreasing tendency. Leaf sample equally had a discrete value of 0.004 of gene expression. *RibHLH33* was expressed at higher values than *RibHLH3* but displayed the same decreasing tendency as the fruit completes its maturation process, with a 35% lower value in the mature red tissue compared to the white ones.

Low values of homologous bHLH genes had been shown before in other species; for example, in *F. x ananassa* as reported by Lin-Wang *et al.* (13), and Starkevič *et al.* (28) in sweet cherries. The expression values reported for the sweet cherries were of 1x10e-2 expression value when normalized to *Actin* (13, 28).

The general expression pattern for the *R. idaeus* bHLH genes can be described as inverse to the fruit maturation and anthocyanin formation, (shown as the red color formation in the fruit, in chapter 3). The lack of a third bHLH candidate (Fv145 homolog) and the different time points of tissue samples analyzed in this project with respect to *F. vesca*, constricts our data analyses

to some extent. However, the data allow us to generate valuable information and allowed us to make conclusions on the quality of the first genome draft version, and of bHLH proteins involved in the flavonoid biosynthesis in *R. idaeus*.

As reported here for the first time (to our knowledge), there are not many reports in the literature that have shown the bHLH expression levels during the fruit maturation process. Several studies reported analysis between different genotypes and their bHLH levels in order to establish their relationships with the anthocyanin levels in crops such as peach, apple, sweet cherry and others (19, 23, 33, 34).

A similar study was made in the MYB family on Mangosteen fruit (*Garcinia mangostana* L.) where the expression of three MYB TFs was analyzed during the fruit maturation; here the correlation between the MYB expression and the anthocyanin content and fruit maturation was shown (34).

The closest related study on the Solanaceae species Russian box thorn or Goji Nero (*Lycium ruthenicum*) reported elevated transcript levels of the bHLH3 homolog AN11 during fruit maturation. These results have a similar trend as the *F. vesca* results presented here, but are contradictory to our findings in *R. idaeus* were the Ri bHLH3 expression levels decreased over time. The bHLH 33 homolog JAF13 expression in *L. ruthenicum* was low and with no statistical difference among the five-time points reported in Zeng *et al.* (35).

Based on our phylogenetic analysis, and the protein sequence data, in the previous chapter we established that *Fv3* is the gene homologous to *Ri3*. Considering that result and the gene expression patterns of *bHLH3* in *F. vesca* and other species, the contradictory data obtained in our *R. idaeus* experiments, makes necessary to confirm our *R. idaeus* bHLH gene expression analyses by analyzing a new set of samples from diverse genotypes or cultivars and perhaps new primer sets that allows: first to verify the bHLHs nucleotide sequences from both species, and second, primers designed in different intron-exon mRNA splicing from the primers used in this thesis as the initial sequence could be a non-splicing area (35).

**Expression of flavonoid pathway genes in *R. idaeus***

Due to the positive results on the three pathway genes evaluated in *F. vesca*, the same homologous genes were searched in the *R. idaeus* genome. As described before, once the genes were identified, qPCR analyses were carried out as shown in Figure 23. For all the genes, the increasing tendency along the fruit maturation is noticeable.
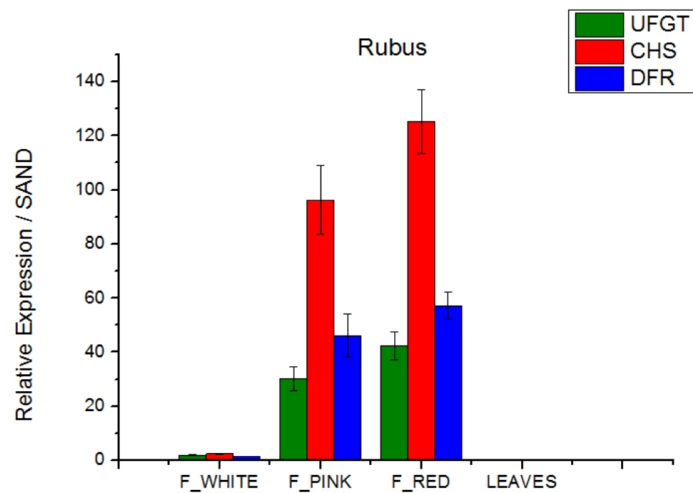
Figure 23: Graph showing the expression of three genes of the flavonoid
pathway in *R. idaeus* in different stages of fruit development and in leaf tissue. Gene expression
values are normalized to the *Actin* gene. Error bars equal to the Standard Error of four technical replicates.

*CHS* gene expression is almost undetectable in the white fruit and increases with a 90% increment in pink fruit tissue, more than 35 times the earlier tissue values. Later on, the *CHS* gene continues to increment its expression until the peak value for mature red tissues with an amount that is approximately 1.3 times the value for the previous pink tissue. Interestingly there were no CHS transcripts detectable in the green leaf tissue.

*UFGT* was the lowest transcript detected among the *R. idaeus* flavonoid pathway genes, contrary to the *F. vesca* results where *UFGT* was the most expressed gene, the *R. idaeus* result seems more accurate due to the fact that *CHS* is necessary for the construction of the backbones employed in the synthesis of not only anthocyanins but also other compounds of the phenylpropanoid pathway.

The next gene evaluated was *DFR*. This gene presented the same expression pattern as *CHS* but with a lower intensity, reaching the peak of expression value 57 in the red tissue with a 39-fold difference compared to the white fruits. The intermediate pink tissue had an expression value of 46.2 (a 15-fold increase vs. 1.5 times between pink and red fruits). This means that between the pink and red stages there was an increase of 23% in the *DFR* gene expression.

Green leaf tissue showed no expression for any of the evaluated genes. This result was once again double analyzed by running another set of qPCR reactions and the HKG previously described in this chapter.

The three pathway genes presented a bimodal distribution in *F. vesca* that was not notorious in *R. idaeus*, this could be due to the fewer time points evaluated for *R. idaeus. Rubus* "red fruit" tissue is the sample corresponding of a starting mature fruit, and cannot be compared with the full mature strawberry fruits of time point 6, specially to their difference in maturation and sugar contents (data not shown, Fondazione Edmund Mach harvest analytical studies).

The flavonoid pathway gene expression of both species, *F. vesca* and *R. idaeus*, confirms the expression pattern seen in other fruit ripening processes with a high anthocyanin formation in its tissues. The same gene expression time curve as *V. vinifera*, where the *CHS* presented an increase of the transcript, and the UFGT gene expression here evaluated in *F. vesca*, validate the previous results on fruit ripening process in *V. vinifera* of Matus *et al.* (29, 36).

Equally, the bimodal transcript expressions of CHS and DFR genes in *F. vesca* validate the *V. vinifera* findings described by Castellarin *et al.* (29), and the findings in *L. ruthenicum* by Zeng *et al.* (34, 35).

However, it is essential to consider the high number of members that the glycosyltransferase (GTs) protein family has in the plant kingdom as being involved in several plant secondary metabolite pathways. The two UFGT candidate genes evaluated here, were obtained based on a tBLASTn, using as a query the protein sequence of UDP-glycosyltransferases (UGTs) in *M. domestica* from the results of Caputi *et al.* (37). Further deep sequence analyses need to be performed to identify if the two UFGT genes (Fv UFGT and Ri UFGT) evaluated here, are homolog to reported UDP-glycosyltransferases involved in anthocyanin biosynthesis in other species (31, 36, 37).

**Functional analysis of Fv bHLH genes in *Nicotiana tabacum***

A highly extended technique for the visualization of the anthocyanin pathway activation is the transient heterologous expression of Fv bHLH genes followed by anthocyanin detection in *N. tabacum* leaves through *Agrobacterium*-mediated transient expression (38, 39). We decided to use this approach as an initial checkpoint on the anthocyanin pathway activation for the *Fragaria* bHLH candidates.

For this experimental part, the three bHLH candidate genes were cloned from cDNA synthesized from previous analyses and inserted into the vectors for plant expression pHex2 and pK7WG2. *A. tumefaciens* was transformed with the plasmids as illustrated in the materials and methods section. This fast and simple method allows the visual and chemical confirmation of the formation of colored anthocyanin compounds, without the time-consuming process of stably transformed plant regeneration.

For the proper MYB-bHLH complex formation, the candidate bHLH was combined at a 1:1 ratio with *Agrobacterium* culture containing *Fv MYB10* on the pSAK vector as reported by Lin Wang (13). The latter vector was given by the kind collaboration of the authors and facilitated by the PFR NZ and FEM partnership.

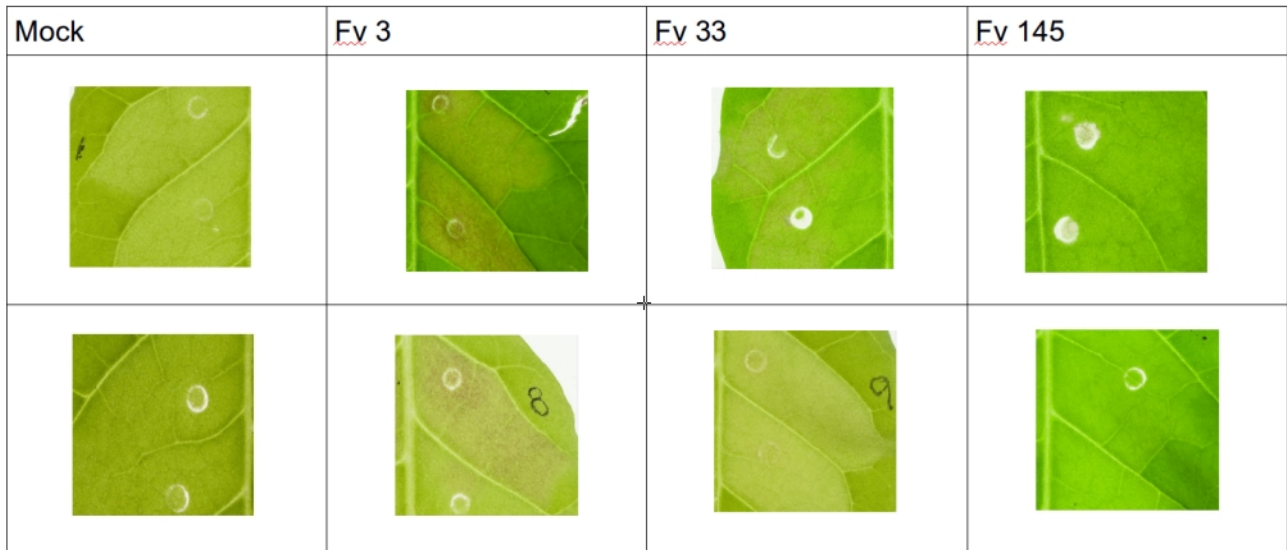| Mock | Fv 3 | Fv 33 | Fv 145 |
|------|------|-------|--------|
|      |      |       |        |
|      |      |       |        |

Figure 24: *N. tabacum* leaves 5 days after agroinfiltration with bHLH candidate genes from *F. vesca*

After one week, the red coloration could be detected in some of the leaves, following a photographic register, segments of leaf material were weighed and total polyphenols extracted using a 80% methanol solution and later measured by UPLC LC/MS as reported in Palmieri *et al.* (18).

The leaf images representative for each treatment are shown in Figure 24. Compared to mock treatment (empty vector) changes on the generally green color appearance can be seen. This change is most evident in *Fv3* treatment and followed by *Fv33*. The vector containing *Fv145* did not seem to be able to activate the anthocyanin production on tobacco leaves.

However, anthocyanins are not the only kind of phenylpropanoid compounds that can be detected in leaf tissues. To solve this problem, LC/MS-MS data can provide sufficient elements to confirm or reject the pathway activation by bHLH proteins. Triplicate experiments were performed, and the statistical analyses are discussed next; different type of compounds was present in a broad range of distribution. The extraction of total polyphenols and quantification from tobacco leaf were performed by a UPLC system coupled to a triple quadrupole (TQ) mass spectrometer according to Palmieri *et al.* (18). Later, Log10 normalization was performed to simplify the statistical analyses and data visualization of the polyphenols extracted from *N. tabacum* leaves.

The quantity of total phenylpropanoids, as shown in Figure 25 (mean and standard deviations represented in the box plots), indicate an increase or decrease value compared to the control treatment. Leaves infiltrated with *Fv3* and *Fv33* presented a higher amount of general polyphenols; a 93% and 12% more compared to mock, respectively.
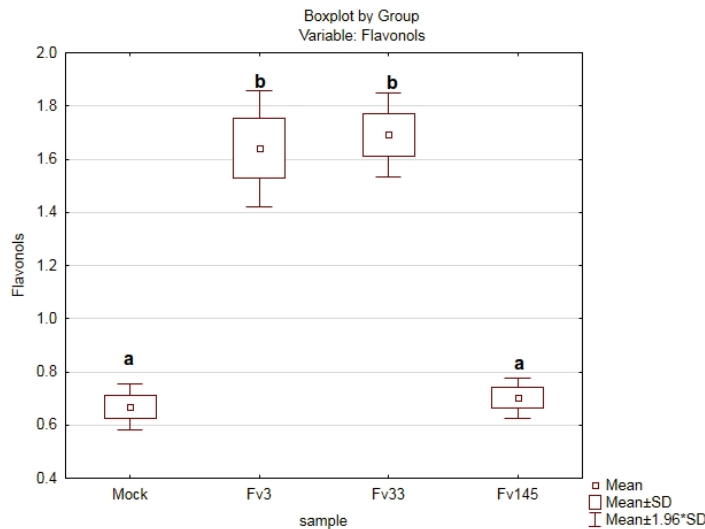
Figure 25: Box plot of total polyphenols on *N. tabacum* leaves. Values expressed in mg of polyphenols in grams of tissue. Bars sharing the same letter are not significantly different according to Tukey's HSD test. Error bars represent standard deviation.

In contrast, *Fv145* results show a 40% reduction of the total phenylpropanoids compare to mock and a 70% reduction compared to the *Fv3* treatment; equally the Tukey test (multiple comparisons of means with a $P < 0.05$) shows a significant difference among all the treatments.

As the total amount of phenylpropanoids detected on each Fv bHLH treatment was different, more exhaustive, and detailed analyses will provide information about the particular phenylpropanoid compound type causing these differences.
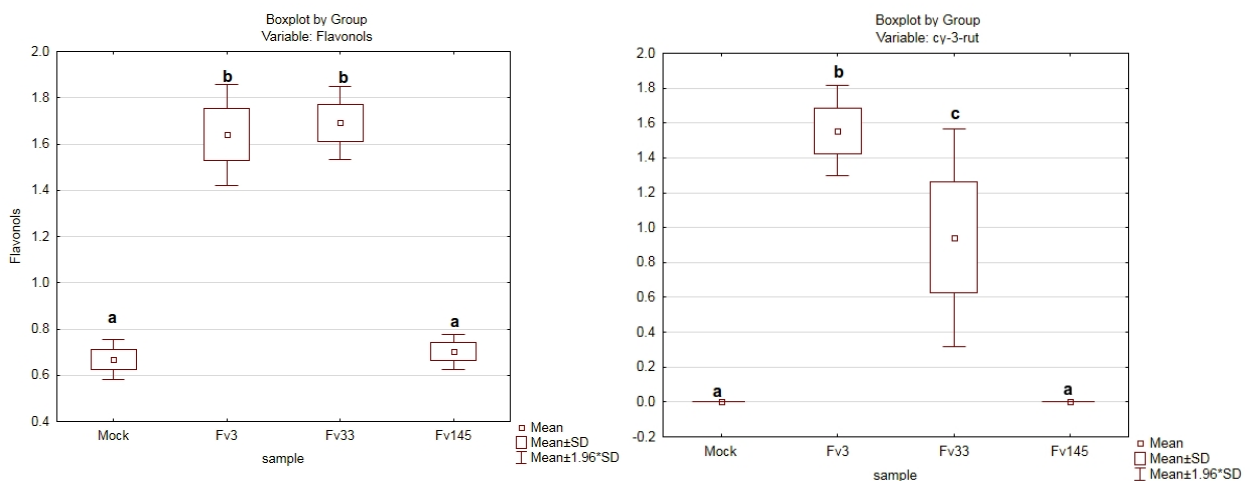


Figure 26: Box plot of the content of flavonols (left panel) and cyanidin 3-*O*-rutinoside anthocyanins (right panel) on *N. tabacum* leaves. Values expressed in mg of g of tissue. Bars sharing the same letter are not significantly different according to Tukey's HSD test.

The previous graphic (Figure 26), illustrates the significant increase of flavonols after *Fv3* and *Fv33* treatments, similar to anthocyanins in the form of Cy-3-rut, in the case of *Fv33*, leaves presented a more significant amount of anthocyanins detected. *Fv145* treated leaves did not show any significant difference compared to mock. This result does not place Fv145 in the

genetic regulation of flavonoids. Also, these low levels of flavonols do not explain yet where exactly is being directed the phenylpropanoid flux in the tobacco leaf of *Fv145* lines seen in Figure 25.

Next step in the analyses focused on the coumarins. Coumarins were detected in the form of fraxin and scopoletin compounds and later quantified due to their possible role in plant defense (40). Leaves infiltrated with *Fv145,* were the only treated leaves that reached lower levels of coumarins compared to mock treatment as it shown Figure 27 (40).
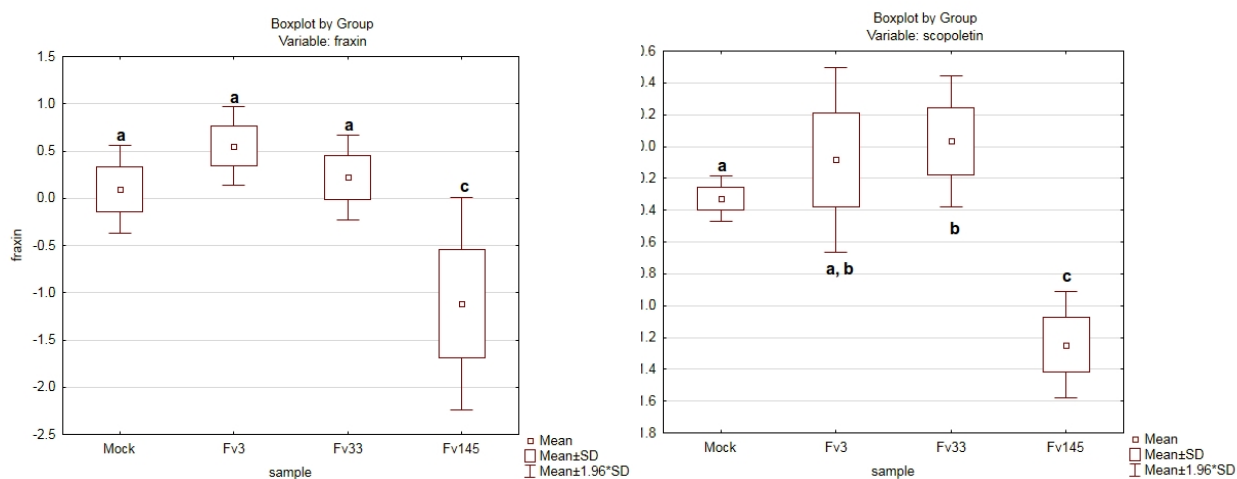


Figure 27: Box plot of the content of the coumarins fraxin (left panel) and scopoletin (right panel) in *N. tabacum* leaves. Values expressed in mg of polyphenols in g of tissue. Bars sharing the same letter are not significantly different according to Tukey's HSD test.

Other types of phenolics were detected through the UPLC LC/MS analyses; compounds like flavanones, chlorogenic acid, and dihydrochalcones, were identified and quantified without any significant difference found among mock and the three bHLH treatments (data not shown).

Similar results were reported in Matros & Mock (41), where a PY potato virus resistance in a genetically modified *N. tabacum* overexpressing TOGT (UDP-glucose glucosyltransferase involved in the glycosylation of the coumarin scopoletin) was observed, highlighting the result of the close relationship between coumarins and plant defense against pathogens. The genetic regulation of coumarin biosynthesis is an unexplored area to date, making further analyses in plants interesting/essential, especially on mutant lines where these specific phenylpropanoid compounds are being overexpressed (40, 41).

**Complementation of *Arabidopsis* mutant line *tt8* (bHLH3) by *Fv3***

To determine the effect of *Fv3* as a positive regulator of the flavonoid biosynthesis, the mutant line *tt8* complementation assay was developed in the model species *A. thaliana* using *Agrobacterium tumefaciens* as a transformation vector.

We took advantage of the plant genetic resources available from the model species *A. thaliana*, especially the *transparent testa* (*tt*) or tannin-deficient seed lines (*tds*) were transcriptional regulators and biosynthetic genes affecting phenolic compound synthesis are mutated. As an example, *tt8* lacks the effect of one TF, and the mutant line *tt3* has a depletion in the gene encoding the enzyme DFR (42).

Mutant line *tt8* reported by Nesi *et al.* (43), has truncated bHLH protein, which generates a plant with white seeds due to the absence of condensed tannins (43). Considering the gene *tt8* is one of the bHLH belonging to the IIIf group of Heim (44), and as described in the previous research chapter (Chapter 3), *Arabidopsis tt8* is a homolog gene of *Fv3* and *Ri bHLH3*.

*A. thaliana* complementation was performed as described in the material and methods section of this chapter: Floral Dip stable transformation with *Agrobacteria*. The cloned gene of *Fv3* was inserted into the vector pUbi10: GW-3xHA (unpublished vector obtained from Grefen *et al.* (15)). Following the first mature seed harvest, and antibiotic selection, a total of 23 lines were grown (T0), later seeds collected from self-pollinated T0 plants were harvested and selected again until the T2 generation was obtained.

From the 23 lines, six representative lines with differently colored seeds were selected for the analysis. The remaining 17 lines are being kept under optimal storage conditions until further studies.
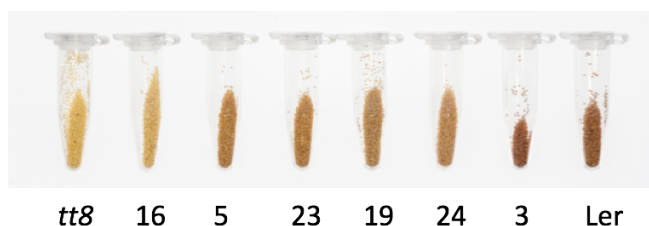
The analyzed plant lines were:



Figure 28: *A. thaliana* seeds: order from left to right side is: 1 *tt8* mutant line, 2 Line 16, 3 Line 5, 4 Line 23, 5 Line 19, 6 Line 24, 7 Line 3, 8 L*er*.

As shown in Figure 28, the complemented lines represent a wide color distribution from the white/yellow seeds of the parental *tt8* mutant line visible in line 16, until the intense brown

color from the background line for the *tt8* mutant, Landsberg *erecta* (L*er*). This later case (Line 3) represents the re-establishment of the phenolic pathways in the complemented lines as the *F. vesca* gene *bHLH 3* can supplement the regulation activity of its homologous *tt8* gene.

Seed color is considered an indication of the presence of PAs in the seed coat; as the brown coloration is due to the oxidation of the PAs (20). Considering that the compounds are not being analytically quantified, this method can only be treated as a visual indicator. Lines 5, 23, 19 and 24 have an intermediate light brown coloration. The frequency of these intermediate phenotypes was high among the total 23 lines (data not shown) compared to the contrasting white/yellow and brown seeds.

Besides the formation of PAs, another well-studied compound in *A. thaliana* plants are anthocyanins, compounds produced in specific tissues like the stem, or in the whole areal part of seedlings during stress conditions, like cold temperatures, low pH, or elevated sugar or mineral conditions (45).

As the anthocyanin extraction and quantification is a relatively straightforward procedure and it is an established method in our laboratory, it was decided to quantify this type of compounds as an initial evaluation criteria for bHLH activity on the stably transformed lines.

The following pictures show the phenotypes observed in the eight lines after one-week treatment in Anthocyanin Induction Conditions (AIC) (20).
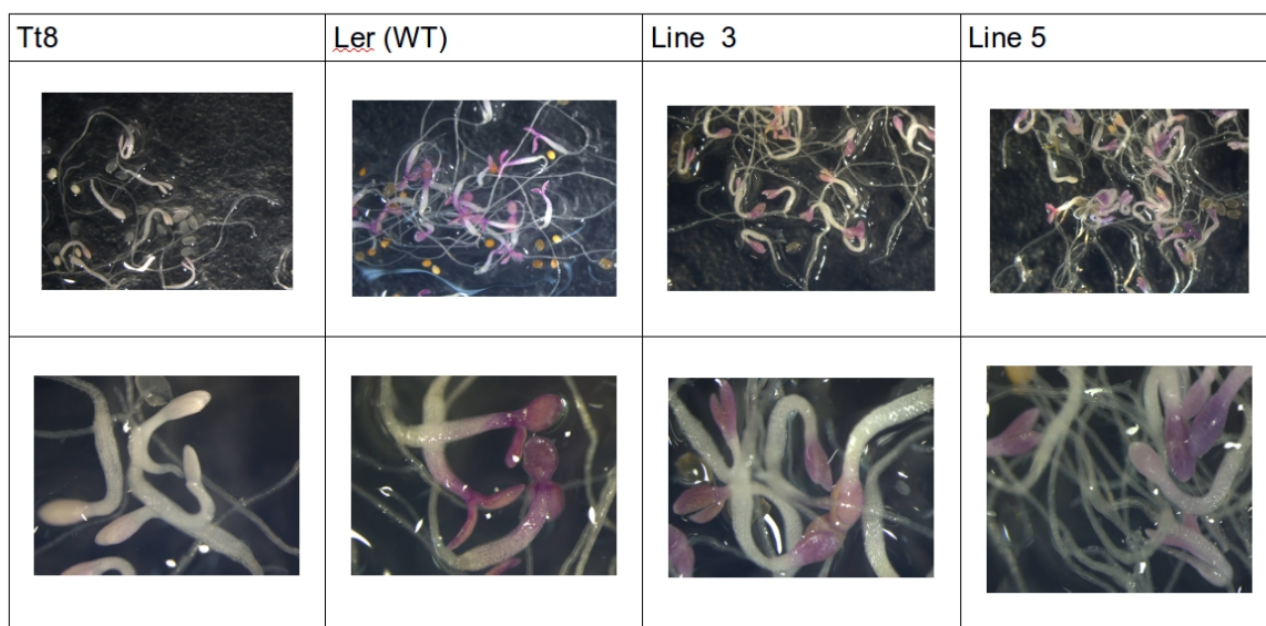


Figure 29-1: *A. thaliana* seedlings after growing 5 days on sucrose 30% (AIC). Upper panels correspond to 1X zoom, lower panels 4x zoom. Part 1

The original mutant line *tt8*, has a low germination ratio of only 40% (data not shown), and did not present any pink/violet coloration visible in the 1x lens (Figure 29-1). The opposite response is given by the parental WT line Ler, in which all seedling show an intense violet color starting on the neck base of the stem or hypocotyl hook. WT seedlings coloration was present with the same intensity in the cotyledons; the radicle and lateral roots did not present any visible coloration.

| Line 16 | Line 19 | Line 23 | Line 24 |
| --- | --- | --- | --- |
| | | | |
| | | | |
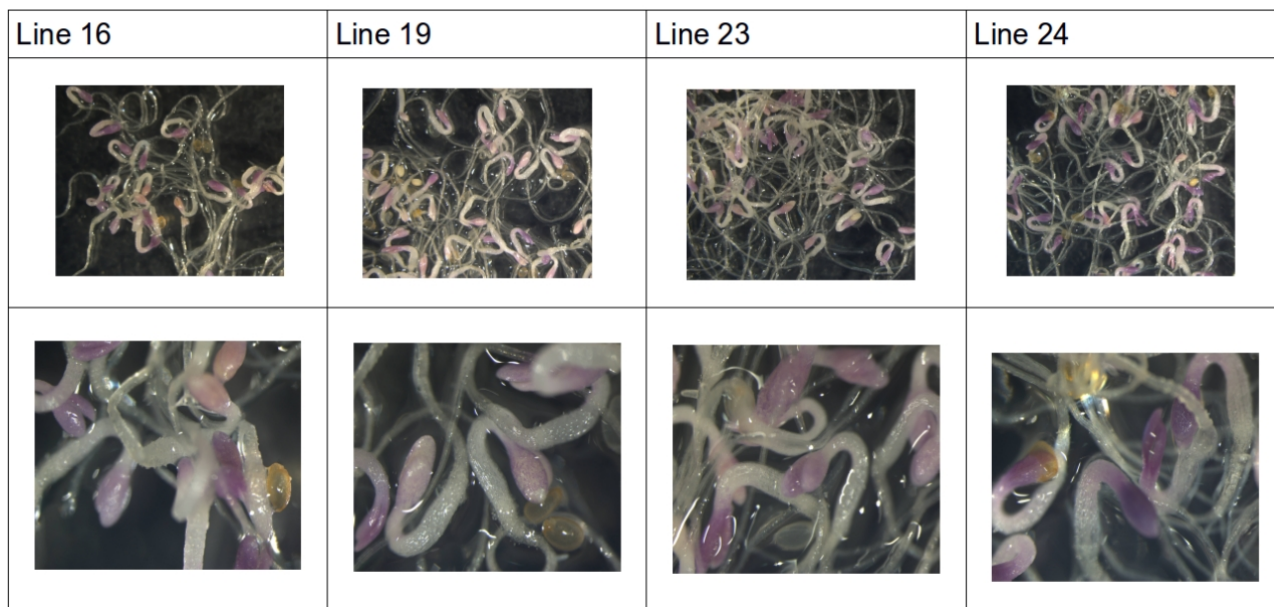
Figure 29-2: *A. thaliana* seedlings after growing 5 days on sucrose 30% (AIC). Upper panels correspond to 1X zoom, lower panels 4x zoom. Part 2

Lines 3, 23 and 24 presented a high amount of anthocyanins according to the color observed in the hypocotyl, while the lines 5, 16 and 19 showed a less intense coloration with the presence of white and green areas in the hypocotyl.

**Anthocyanin detection on complemented At *tt8*-Fv bHLH3 lines by UPLC**

To quantify the total amount of anthocyanins present in the seedlings of the complemented lines described in the previous pages, a total polyphenol methanol extraction was used on two different experimental setups. For the anthocyanin quantification, 20 µl of the methanol extract was injected into a UPLC-DAD, and run on a C18 column as described in detail in the material and methods section (21).

Cyanidin 3-*O*-glucoside commercial standard was employed for this study as the reference standard mostly used in the literature (45, 46). For the identification and quantification of the total anthocyanins present in the *A. thaliana* seedlings the UPLC-DAD results consisting of

chromatogram at 520 nm, peak retention times and UV spectra were analyzed as can be seen in Figure 30.

All samples, including the mutant line for the phenylpropanoid pathway *tt8*, presented the main peak at 17 minutes at the 520-nm wavelength DAD chromatogram (typical of anthocyanin compounds); the area of this main peak was employed for the anthocyanin quantification will be discussed later.
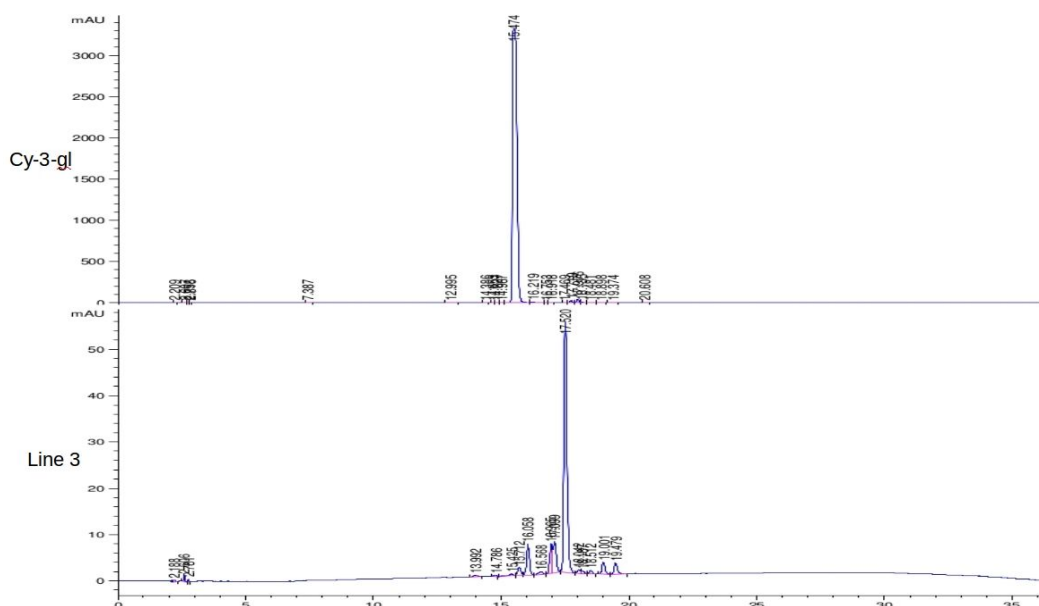


Figure 30: UPLC-DAD chromatogram at 520 nm of anthocyanins present in *A. thaliana* complemented *tt8* mutant lines. Upper part: commercial standard Cy-3-glc. Lower part: plant 3 total anthocyanin extract.

Due to time limitations, and the lack of new *tt8* seeds, it was decided to analyze the complemented lines in this preliminary study, taking subsequent precautions from the molecular characterization point of view on the further investigations of the *F. vesca* and *R. idaeus* bHLH. However, new stock seeds need to be ordered as the integrity of the *tt8* mutant line seems to be compromised, and the complementation transformation repeated to confirm our results.

However, I would like to address the fact that the presence of anthocyanin compounds in the *Arabidopsis* mutant line *tt8*, this finding is not exclusive to my research according to other researchers experience a complete depletion of anthocyanins can only be achieved with double mutations of *tt8* and *GL3*. Unfortunately, no scientific reports can be found in the literature. The presence of small amounts of anthocyanins could represent a form of redundancy role from another bHLH regulator in *Arabidopsis*.

**Total anthocyanin quantification on complemented *At tt8-Fv bHLH3* lines**

Once we detected the presence of anthocyanin pigments in the total polyphenol methanol extraction, a calibration curve of five points of Cy 3-*O*-glucoside (Cy 3-glu) was developed for the quantification measurements, from the highest concentration of 1000 ng/µl until 0.1 ng/µl (data not shown). *A. thaliana* seedlings anthocyanin quantification values are expressed in µg of Cy 3-glu equivalents in mg of fresh weight (µg/mg FW) according to literature (20, 43).

As can be seen in Figure 30, in all the analyzed lines anthocyanin quantities were detected, and as it was mentioned before, the mutant line *tt8* contained a low level of anthocyanins, 0.25 µg/mg FW. This means not an absolute absence of anthocyanins, but represents a 66% reduction of the anthocyanins compared to the WT L*er*. The 0.25 µg/mg FW, was used as a reference point for comparison with the complemented lines.

Despite the coloration observed on the stereomicroscope images, Line 16 presented a low amount of Cy 3-glu: 0.33 µg/mg FW, representing only a 35% increase compared to the 241% increase in anthocyanins detected in Line 3 (0.83 µg/mg FW).
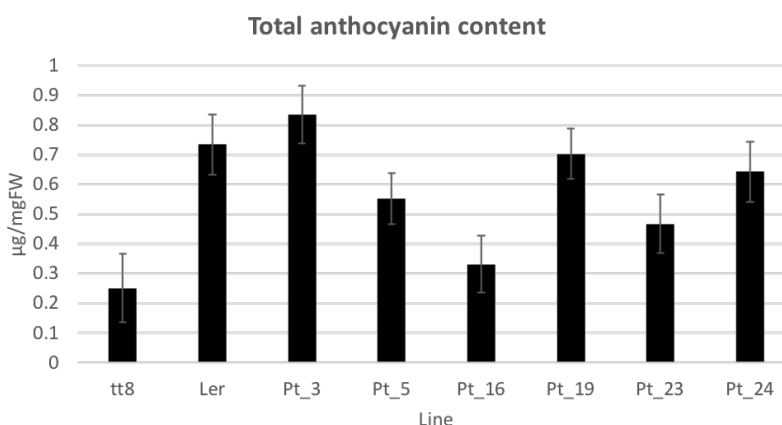


Figure 31: Total anthocyanin content in *A. thaliana* seedlings of complemented lines Fv_bHLH 3 from *tt8* growth on AIC. Values expressed in µg/mg FW

The broad range of anthocyanin levels detected in the six complemented lines (see Figure 31), plus the seed color variation previously discussed, represent the difference in *tt8* complementation obtained from the flower dip transformation experiments. As reported by Ghedira *et al.* (47) high transformation variability from plant to plant is a commonly obtained result and as mentioned earlier additional analyses to estimate the copy number of the T-DNA insertion in transformed lines need to be performed (47).

However, lines 3, 19 and 24 successfully exceeded the WT L*er* total anthocyanin levels by more than 85%, making these lines the candidates for more detailed studies about the

complementation itself: number of copies inserted, homozygosity of the insertions, etc. The analysis that unfortunately for time restrictions could not be performed in this thesis project.

### *M. domestica* CHS promoter activation through Luc/Ren assay

One important function to characterize a TF gene, is its ability to activate a promoter gene (7). The most common way to perform this analysis, is to clone the promoter of interest with a reporter gene. It is important that the promoter contains specific binding element regions, in the case for the bHLH an E-box.

To see the effect of the bHLH proteins in the activation of a reporter gene, the CHS Promoter Activation through Luc/Ren Assay was performed. This experimental part is made using a commercial Luciferase/Renilla assay system from Biotum (Fremont, United States). Due to technical and time constraints for this initial study on the TF involved in *F. vesca* and *R. idaeus* fruit development, the published sequence 2kb upstream promoter of CHS gene from the close species *M. domestica* was employed; this decision was made after several unsuccessful attempts at cloning the promoter regions of both species into the p-Green 800 vector (7).

In general, the transient transformation of *N. benthamiana* leaves is a widely used technique for reporter gene assays. In this study for the Luciferase activity assay, the *Luc* gene is always accompanied by another bioluminescence gene, in this case, the *Renilla* luciferase gene under the control of the 35S promoter.

Double gene transformation is necessary for normalization purposes due to the high variation that this technique has. Inner variation between repetitions is so high that literature and scientific forums even recommended avoiding data comparison among experiments performed on different days and different gene sets and always included negative controls for each gene promoter activation to test (48).

Initially, it was assumed that difficulties in cloning the promoters from *F. vesca* and *R. idaeus* were caused by the poor genomic DNA extracted from immature fruit tissues. However subsequent trials using leaf tissue lead to the same negative results. To confirm this hypothesis the commercial vector pCR®4-TOPO (Thermo Fisher Scientific) was used to clone the promoter region of *Fv3*. The result of this cloning was successful, and the sequence analysis of this promoter region was correctly aligned to the *F. vesca* genome, confirming the integrity of the DNA samples.

Additional CHS promoter sequence comparison was made between *M. domestica, F. vesca* and *R. idaeus*, to establish the *cis*-regulatory element similarities among them, as expected for genes regulated by the MYB-bHLH-WD40 complex. E-boxes and ARE regions (anthocyanin regulatory elements) were identified in the 600 bp upstream region from the ORF (open reading frame). This sequence similarities and literature reports in which homolog gene promoter regions are used for the LUC/REN assays allow us to use with confidence the MdCHS promoter in the following analyses (2, 4).

For the bHLH activation assay the presence of the other two proteins involved in the complex formation is necessary; MYB and WD40. For this reason, the N. *benthamiana* leaf infiltration was made with a culture mixture of three or more vectors: 1. Luc/Ren pGreen vector (ratio 1:3), 2. bHLH candidate gene, and 3. Fv_MYB10 on pSAK vector previously described. Additionally, in the negative controls for each experiment Fv_MYB10 was replaced with the pSAK vector containing the GUS reporter gene.

The absence of additional vectors containing WD40 proteins is because the endogenous *N. benthamiana* WD40 proteins can interact with the bHLH-MYB complex through their highly conserved binding domains (2, 33, 38).
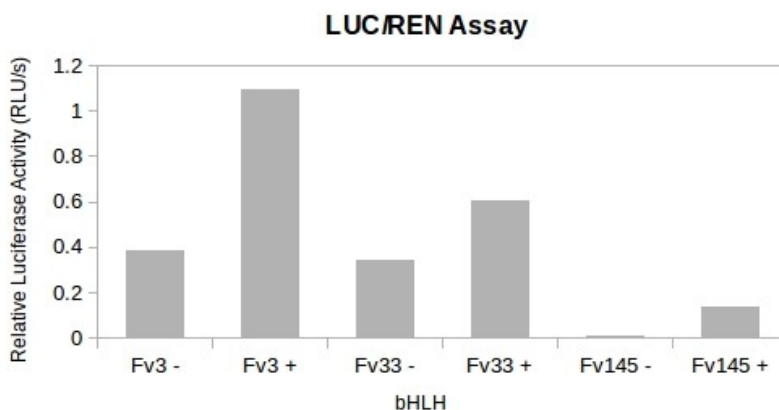


Figure 32: Renilla luciferase reporter assays (LUC/REN) *F. vesca* bHLH activation on MdCHS promoter.

Figure 32 represents the data obtained for *Fv3, Fv33* and *Fv145* promoter activation experiments*,* and it is the result of three different experimental sets put together for visualization purposes. On the x-axis, the candidate genes are compared with their respective negative controls, and the y-axis represents the activation results on RLU (relative light units, in this case, Luciferase ones). All the values have been normalized to the Renilla control.

*F. vesca* bHLHs genes behave on a broad range of the *MdCHS* promoter activation results; *Fv3* presented almost three-fold activity compared to the negative control Fv3.

*Fv33* presented 1,7-fold activity compared to the respective control, and at least *Fv145* showed a lower promoter activation, with only 0,134 RLU units, a value that is 76% lower if compared to the results observed for *Fv3* and *Fv33*.

Lin-Wang *et al.* (2) reported similar results for *A. thaliana DFR* promoter activation with the *F. ananassa* MYB10 plus *M. domestica* bHLH protein sequences, where *Md_bHLH3* had a more significant activation than *Md_bHLH33*. Later, the same author (13) reported the *Fv_bHLH33* activation of several promoters including *F. vesca DFR* and *UFGT*. However, the results discussed here need to be confirmed in the presence of their native original promoter from *F. vesca* (2, 13).

## CONCLUSIONS AND REMARKS

With the results presented in this chapter, we established the initial steps in the study of the bHLH genes of two of the most important berries of the European market: wild strawberry and raspberry, *F. vesca* and *R. idaeus*, respectively. Also, we performed an identification of HKGs for both species and created a set of validated primers for quantitative PCR analyses. Later these genes were used in the data normalization of an initial gene expression analysis during fruit development.

Our *F. vesca* data confirm the results reported by Schaart *et al.* (3), where the bHLH homologous genes from the commercial strawberry (*F. x ananassa*) are found to be highly expressed during fruit development, however the *R. idaeus* data was contradictory as both bHLH candidate genes were expressed to higher values in immature fruit tissue and decreased in the mature fruit.  I strongly recommend further studies and consider to increase the number of fruit time points to be analyzed in *R. idaeus* from 3 to 5 to obtain a dataset that is more comparable to the one from *F. vesca.*

### *Fv3* characterization

From the three *F. vesca* bHLH candidate genes, only *Fv bHLH_3* seems to be mainly responsible for the flavonoid positive regulation according to the gene expression analyses. This result corroborates the previously reported data on other Rosaceae species as *M. domestica* and *P. persica*, where the bHLH 3 homolog is the bHLH involved in the anthocyanin (phenylpropanoid) synthesis together with MYB 10 (1, 2, 3).

Thus, the gene *Fv3* was selected for *in vivo* characterization experiments using the heterologous plant system *N. tabacum* for the anthocyanin detection in leaves, and *A. thaliana* for the complementation of *tt8* mutants. Both experiments generated data enough for confirming the activation role, or positive regulation of *Fv3* in the phenylpropanoid biosynthesis, and especially in the anthocyanin formation.

These results contribute not only to the state of art of the wild species *F. vesca*, but also to the commercial *F. x ananassa* field, and can help the community of strawberry breeders to develop new strategies on the color management of this crop.

However, more analyses and studies still need to be done, as protein interaction experiments like Y2H (yeast two-hybrid) assays, to confirm the interaction of *Fv3* and the Fv MYB (MYB10), as it is required for the formation of the Myb-bHLH-WD40 protein complex. Likewise, similar experiments as the ones presented in this chapter, are still needed for the characterization of the remaining two homologs bHLH genes found on *R. idaeus*, in addition to the *Fv145* gene whose role is still unclear.

**REFERENCES**

1. Rahim, M. A., Busatto, N., and Trainotti, L. (2014). Regulation of anthocyanin biosynthesis in peach fruits. *Planta* 204, 913–929.

2. 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.

3. Schaart, J. G., Dubos, C., Romero De La Fuente, I., van Houwelingen, A. M. M. L., de Vos, … Bovy, A. G. (2013), Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidins biosynthesis in strawberry (*Fragaria × ananass*a) fruits. *New Phytologist*, 197: 454–467.

4. Meier, I., & Gruissem, W. (1994). Novel conserved sequence motifs in plant G-box binding proteins and implications for interactive domains. *Nucleic Acids Research*, *22*(3), 470–478.

5. Yang, Y., Li, R. and Qi, M. (2000), In vivo analysis of plant promoters and transcription factors by agroinfiltration of tobacco leaves. *The Plant Journal,* 22: 543–551.

6. Potenza, C., Aleman, L., & Sengupta-Gopalan, C. (2004). Targeting transgene expression in research, agricultural, and environmental applications: Promoters used in plant transformation. In Vitro Cellular & Developmental Biology - *Plant*, 40(1).

7. Hellens R.P., Allan A.C., Friel E.N., Bolitho K., Grafton K., Templeton M.D., … Laing W.A. (2005). Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants*. Plant Methods* 1: 13-16.

8. Zupan, J. R., & Zambryski, P. (1995). Transfer of T-DNA from *Agrobacterium* to the plant cell. *Plant Physiology,* 107(4), 1041–1047.

9. Smale, S. Luciferase assay. *Cold Spring Harbor Protocols*. 2010 May.

10. Jiwaji, M., Daly, R., Pansare, K., McLean, P., Yang, J., Kolch, W., & Pitt, A. R. (2010). The Renilla luciferase gene as a reference gene for normalization of gene expression in transiently transfected cells. *BMC Molecular Biology,* 11.

11. McNabb, D. S., Reed, R., & Marciniak, R. A. (2005). Dual Luciferase Assay System for Rapid Assessment of Gene Expression in Saccharomyces cerevisiae. *Eukaryotic Cell*, 4(9), 1539–1549. 29.

12. Manual https://biotium.com/wp-content/uploads/2016/08/PI-30081.pdf

13. Lin-Wang, K., McGhie, T. K., Wang, M., Liu, Y., Warren, B., Storey, R., ... Allan, A. C. (2014). Engineering the anthocyanin regulatory complex of strawberry (Fragaria vesca). Frontiers in Plant Science, 5, 651

14. Bourras, S., Rouxel, T., & Meyer, M. (2015). *Agrobacterium* tumefaciens Gene Transfer: How a Plant Pathogen Hacks the Nuclei of Plant and Nonplant Organisms. *Phytopathology*, 105(10), 1288–1301.

15. Karimi, M., Inzé, D., Depicker, A., (2002) Gateway vectors for Agrobacterium-mediated plant transformation. *Trends in Plant Science*. 2002 May;7(5): 193-195.

16. Invitrogen Manual Gateway. Gateway® Technology A universal technology to clone DNA sequences for functional analysis and expression in multiple systems. Publication Number MAN0000282

17. Espley, R. V., Hellens, R. P., Putterill, J., Stevenson, D. E., Kutty-Amma, S., & Allan, A. C. (2007). Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *The Plant Journal,* 49(3), 414–427.

18. Palmieri, L., Masuero, D., Martinatti, P., Baratto, G., Martens, S. and Vrhovsek, U. (2017), Genotype-by-environment effect on bioactive compounds in strawberry (*Fragaria x ananassa Duch.*). *Journal of Science Food Agriculture*. Version of record online: 31 March 2017.

19. Bent, A. (2006). *Arabidopsis thaliana* Floral Dip Transformation Method. In K. Wang (Ed.), *Agrobacterium Protocols*. Totowa, NJ: Humana Press 87–104.

20. Pourcel, L., Routaboul, J.M., Kerhoas, L., Caboche, M., Lepiniec, L., & Debeaujon, I. (2005). *TRANSPARENT TESTA10* Encodes a Laccase-Like Enzyme Involved in Oxidative Polymerization of Flavonoids in *Arabidopsis* Seed Coat. *The Plant Cell*, *17*(11), 2966–2980.

21. Ehrhardt, C., Arapitsas, P., Stefanini, M., Flick, G. and Mattivi, F. (2014), Analysis of the phenolic composition of fungus-resistant grape varieties cultivated in Italy and Germany using UHPLC-MS/MS. *Journal of Mass Spectrometry*, 49: 860–869.

23. Kozera, B., & Rapacz, M. (2013). Reference genes in real-time PCR. *Journal of Applied Genetics*, *54*(4), 391–406.

24. Schmittgen T.D., Livak K.J. (2008) Analyzing real-time PCR data by the comparative CT method. Nature *Protocols,* 3:1101–1108.

25. Storch, T. T., Pegoraro, C., Finatto, T., Quecini, V., Rombaldi, C. V., & Girardi, C. L. (2015). Identification of a Novel Reference Gene for Apple Transcriptional Profiling under Postharvest Conditions. PLoS ONE, 10(3), e0120599.

26. Borges, A. F., Fonseca, C., Ferreira, R. B., Lourenço, A. M., & Monteiro, S. (2014). Reference Gene Validation for Quantitative RT-PCR during Biotic and Abiotic Stresses in *Vitis vinifera. PLoS ONE*, 9(10), e111399.

27. Katayama-Ikegami, A., Katayama, T., Takai, M., & Sakamoto, T. (2017). Reference gene validation for gene expression studies using quantitative RT-PCR during berry development of 'Aki Queen' grapes. VITIS. *Journal of Grapevine Research*, 55(4), 157-160.

28. Starkevič, P., Paukštytė, J., Kazanavičiūtė, V., Denkovskienė, E., Stanys, V., Bendokas, V., ...Ražanskas, R. (2015). Expression and Anthocyanin Biosynthesis-Modulating Potential of Sweet Cherry (Prunus avium L.) MYB10 and bHLH Genes. *PLoS ONE*, 10(5).

29. Castellarin, S. D., Pfeiffer, A., Sivilotti, P., Degan, M., Peterlunger, E. Daspero, G. (2007), Transcriptional regulation of anthocyanin biosynthesis in ripening fruits of grapevine under seasonal water deficit. *Plant, Cell & Environment*, 30: 1381–1399.

30. Deng, X., Bashandy, H., Ainasoja, M., Kontturi, J., Pietiäinen, M., Laitinen, R. A. E., Albert, V. A., Valkonen, J. P. T., Elomaa, P. and Teeri, T. H. (2014), Functional diversification of duplicated chalcone synthase genes in anthocyanin biosynthesis of *Gerbera hybrida*. *New Phytologist*, 201: 1469–1483.

31. Schulenburg, K., Feller, A., Hoffmann, T., Schecker, J. H., Martens, S., & Schwab, W. (2016). Formationof β glucogallin, the precursor of acid in strawberry and raspberry. *Journal of Experimental Botany*, 67(8), 2299–2308.

32. Martens, S., & Mithfer, A. (2005). Flavones and flavone synthases. *Phytochemistry*, 66(20), 2399–407.

33. Ramsay, N. A., & Glover, B. J. (2017). MYB2013-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends in Plant Science,* 10(2), 63–70

34. Palapol, Y., Ketsa, S., Lin-Wang, K., Ferguson, I. B., & Allan, A. C. (2009). A MYB transcription factor regulates anthocyanin biosynthesis in mangosteen (*Garcinia mangostana L.*) fruit during ripening. *Planta*, 229(6), 1323–1334.

35. Zeng, S., Wu, M., Zou, C., Liu, X., Shen, X., Hayward, A., … Wang, Y. (2014), Comparative analysis of anthocyanin biosynthesis during fruit development in two *Lycium* species. *Physiology Plantarum*, 150: 505–516.

36. Matus, J. T., Poupin, M. J., Cañón, P., Bordeu, E., Alcalde, J. A., & Arce-Johnson, P. (2010). Isolation of WDR and bHLH genes related to flavonoid synthesis in grapevine (Vitis vinifera L.). *Plant Molecular Biology*, 72(6), 607–620.

37. Caputi, L., Malnoy, M., Goremykin, V., Nikiforova, S. and Martens, S. (2012), A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land. *The Plant Journal*, 69: 1030–1042.

38. Franzmayr, B. K., Rasmussen, S., Fraser, K. M., & Jameson, P. E. (2012). Expression and functional characterization of a white clover isoflavone synthase in tobacco. *Annals of Botany*, *110*(6), 1291–1301.

39. Liu, Y., Lin-Wang, K., Espley, R. V., Wang, L., Yang, H., … 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(8), 2159–2176.


40. Matern, U. (1991). Coumarins and other phenylpropanoid compounds in the defense response of plant cells. *Planta Med*, Oct 57 15-20.

41. Matros, A., & Mock, H. (2004). Yctopic expression of Y glucosyltransferase leads to increased resistance of transgenic tobacco plants against infection with potato virus. *Plant Cell Physiology*, 45(9): 1185–1193.

42. Buer, C. S., & Djordjevic, M. A. (2009). Architectural phenotypes in the *transparent testa* mutants of *Arabidopsis thaliana*. *Journal of Experimental Botany*, *60*(3), 751–763.

43. Nesi, N., Jond, C., Debeaujon, I., Caboche, M., and Lepiniec, L. (2001). The *Arabidopsis* TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. *Plant Cell* 13, 2099–2114.

44. Heim, M., Jakoby, M., Werber, M., Martin, C., Weisshaar, B., Bailey, P. (2003). The basic helix–loop–helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. Molecular Biology and Evolution. 20 (5), 735-747.

45. Kovinich, N., Kayanja, G., Chanoca, A., Riedl, K., Otegui, M. S., & Grotewold, E. (2014). Not all anthocyanins are born equal: distinct patterns induced by stress in *Arabidopsis*. *Planta*, 240(5), 931–940.

46. Saito, K., Yonekura-Sakakibara, K., Nakabayashia, R., Higashi, Y., Yamazaki, M., Tohge, T., Fernie, A. (2013). The flavonoid biosynthetic pathway in *Arabidopsis*: Structural and genetic diversity. *Plant Physiology and Biochemistry*. 72, 21–34

47. Ghedira, R., De Buck, S., Nolf, J., & Depicker, A. (2013). The efficiency of *Arabidopsis thaliana* floral dip transformation is determined not only by the *agrobacterium* strain used but also by the physiology and the ecotype of the dipped plant. *Molecular Plant-Microbe Interactions*, 26(7), 823–832.

48. Schagat, T., Paguio, A., Kopish, K. (2007). Normalizing genetic reporter assays: approaches and considerations for increasing consistency and statistical significance. *Cell notes* Issue 17.

# Chapter 5

Effect of bHLH down-regulation during the maturation of *Fragaria x ananassa* fruits

# INTRODUCTION

## Strawberry Fruit Metabolites

Recently, the nutritional and health features of strawberry and wild strawberry have been studied. These benefits can be associated with the high presence of polyphenols in these fruits. The nutritional value of strawberries is further correlated with the presence of soluble sugars, organic acids, amino acids, vitamins, besides the mentioned important secondary metabolites such as polyphenols (1, 2).

As it is well known, polyphenols, mainly flavonoids, form an anti-oxidative potential and protect against chronic diseases such as tumors or heart disorders. Cultivated strawberries (*Fragaria* x *ananassa*) accumulate large quantities of polyphenols, which have been reported to have a positive impact on human health due to their antioxidant capacity (1, 3, 4)
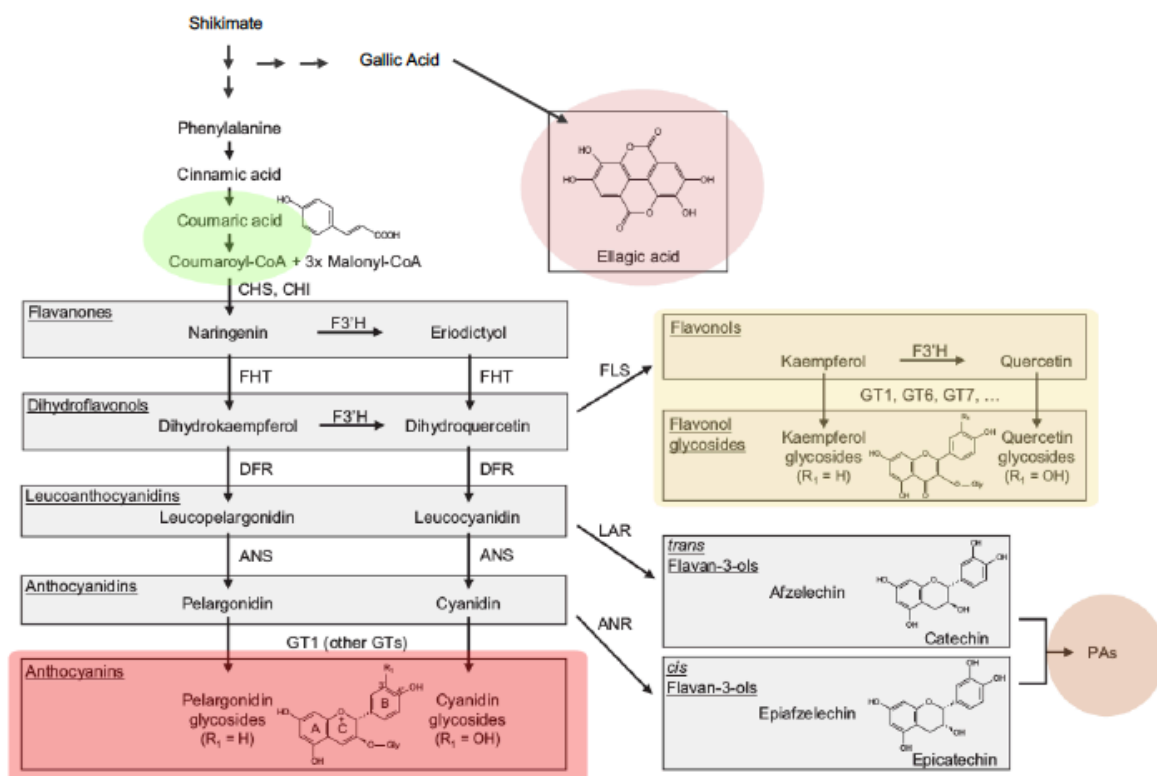


Figure 33: Simplified representation of the phenylpropanoid and flavonoid pathways in strawberry. Adapted from Winkel-Shirley (5)

Among the polyphenols present in strawberries are proanthocyanidins, anthocyanins, flavonols, phenolic acids, and ellagitannins. The anthocyanins are present in the flesh, expanded flower and receptacle of the strawberry also referred to as the fruit. The most

abundant anthocyanins are 4′-hydroxylated pelargonidin-type, and the most abundant proanthocyanidins are 3′,4′-hydroxylated (catechin- and epicatechin-derived) (2, 3).

Ellagitannins are another kind of polyphenols present in *Fragaria* fruits. They are derived from gallic acid, that later is transformed to 1,2,3,4,6-pentagalloylglucose. Gallic acid itself, is formed from an intermediate compound of the upstream reactions of the shikimate pathway. Interestingly, shikimate it also a precursor of phenylalanine, the compound that later is modified to coumaroyl-CoA (precursor of anthocyanins, flavonoids, and proanthocyanidins) (6).

However, the highest amount of ellagitannins are found not in the flesh but in the nuts or the real seed, and being agrimoniin, the main ellagitannin found in *Fragaria* green fruits, specifically in the achenes. Figure 33 represents a general overview of the phenylpropanoid pathway, and the colored areas highlight the compounds present in the strawberry fruits mentioned before (3, 6).

**RNA silencing in Plants**

RNA silencing in plants, also known as post-transcriptional gene silencing (PTGS), is the remarkable process where foreign RNA molecules are recognized and degraded by the enzyme Dicer in a sequence-specific manner in the cytoplasm of the cells. This process occurs in a wide variety of organisms, including plants, animals, and fungi (8, 9).

A critical early step in RNA silencing is the formation of double-stranded (ds) RNA. In the case of most plant viruses, the PTGS process involves recognition of a target RNA, and the initiation of a sequence-specific RNA degradation pathway in the cytoplasm. Targets for PTGS may be recognized because of the presence of large double-stranded RNA (dsRNA) structure or because of an unusual feature of the RNA. A graphic representation of the process is shown in Figure 34 (9, 10).

dsRNA is formed during the intermediate steps of virus genome replication, and this may explain why viruses are often potent inducers of RNA silencing (19). Small RNAs of 21–23 nucleotides, corresponding to both sense and antisense strands of the target, are consistently associated with PTGS, and it was proposed that these short RNAs provide specificity for target RNA degradation through association with an RNase III-like enzyme.
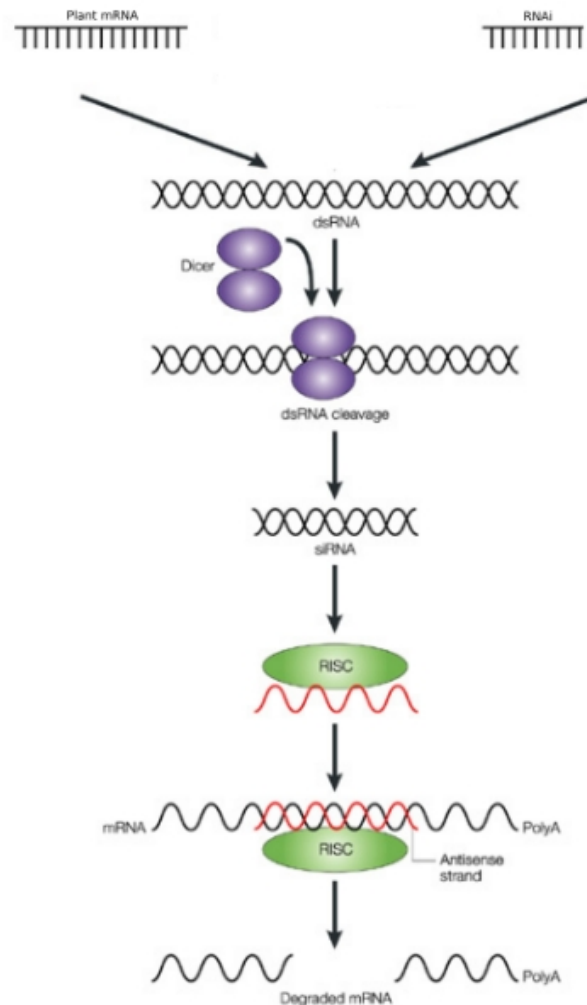
Figure 34: The current model of RNA-mediated gene silencing in plants. Adapted from:" The current model of RNA-mediated gene silencing in plants". Waterhouse & Helliwell (11).

In plants, PTGS has been widely used to accelerate the identification of the biological function of genes considering the evidence that can be collected on the phenotype of organisms that contain mutations in X gene. One of the most famous and early genes studied was the theobromine synthase of the coffee plant that was knocked down with the hairpin construct of the transgene, leading to the production of decaffeinated coffee plants (9, 12).

**Transient transformation of fruits**

As described in Chapter 4, *Agrobacterium* cells can be infiltrated into the intercellular spaces of plant cells, to transfer T-DNA into the plant cell nucleus and express the gene of interest. The most popular host plant for agroinfiltration is *N. benthamiana*. Over time, the power of the technique has been improved, and other species like *Medicago sativum*, tomato (*Solanum lycopersicum*), and *Arabidopsis* among others have been successfully used (ref 8 from chapter 4).

The efficiency of agroinfiltration varies from host to host. Some species and tissues seem more challenging to use than others; the reasons for these differences in efficiency are still not well understood, but some factors like the compactness of the tissue, innervations pattern, and bacteria-host compatibility are probably the main responsible features (13 - 16).

Protoplasts from leaf tissue are the easiest and most accessible kind of cells to use for transient expression analyses not only of model species such as *Arabidopsis* and tobacco but also of maize, *Petunia* and others; considering the relative easiness of the protoplast isolation technique and little requirements for their isolation. However, some problems can arise when trying to obtain protoplasts from the specific tissue where the studied gene is expressed, as in the case of ripe fruit tissue (15, 16, 17).

Many transient transformation studies on flesh fruits (as apple, pear, tomato, peach, and strawberry) have been performed in the last 10 years, focusing on reverse genetics in order to characterize gene function of the candidate genes. The majority of those studies were finalized to understand the role of genes involved in ripening processes as shown in Table 14 (15, 16, 18). These transient transformations mostly use a 35S promoter fused to the GUS-intron as a reporter gene to evaluate the effectiveness of the agroinfiltration (18).

Table 14. Genes functionally characterized from the flavonoid pathway through transient transformation in strawberry (*Fragaria x ananassa*). Modified from Guiradelli & Baraldi (18).

| Gene | Putative Function | Reference |
| --- | --- | --- |
| Chalcone synthase (*FaCHS*) | Pigment formation, flavonoid biosynthesis. | Hoffmann *et al*. (19), Miyawaki *et al*. (20), Ring *et al*. (21) |
| FaMYB10 transcription factor | Regulation of phenylpropanoid/flavonoid pathways | Medina-Puche *et al*. (22) |
| Flavanone 3-Hydroxylase (*F3H*) | Pigment formation, flavonoid biosynthesis | Jiang *et al*. (23) |
| Dihydroflavonol 4-reductase (*FaDFR*) | Pigment formation, flavonoid biosynthesis | Lin *et al*. (24) |
| Glycosyltransferase (*FaGT1*) | Pigment formation, flavonoid biosynthesis | Griesser *et al*. (25) |
| FaSHP transcription factor | Regulation of ripening time | Daminato *et al*. (26) |

# MATERIALS AND METHODS

## Plant material

Strawberry plants from the commercial species *F. x ananassa* var. "Elsanta" were grown in a greenhouse under conditions of a controlled temperature of 25ºC and a 16h photoperiod. The "Elsanta" variety was chosen due their continuous flowering period and fruit production performance.

The entire experimental procedure was performed at the facilities of the Group "Biotechnologie der Naturstoffe" (Biotechnology of Natural Products) at the Freising campus of the Technische Universität München, Germany, under the direction of Prof. Wilfried Schwab. This stay was possible thanks to the COST action grant for Short Term Scientific Mission (STSM) under the COST Action *FA1306*.

## Plasmid construction for RNAi silencing

Considering the close homology of the three bHLH genes, it was essential to determine the appropriate region of the gene sequence to be used for the amplification of the specific fragments to silence with the RNA interference approach. It was also a critical point to control the presence of restriction site sequences that could interfere with the following cloning steps.

Table 15. Information on PCR products used for the RNAi plasmid construction

| Constructs | Length of RNAi amplicon | Restriction Site added to primer sequence |
|---|---|---|
| Ri26_RNAi | 202 | HindIII + BamHI |
| Ri36_RNAi | 205 | HindIII + BamHI |
| Fv3_RNAi | 186 | HindIII + BamHI |
| Fv33_RNAi | 150 | HindIII + BamHI |
| Fv145_RNAi | 161 | HindIII + BamHI |

Once these primers were designed, the corresponding PCR amplification reactions were done and the digestion, with the specific restriction enzymes, was performed in order to produce the insert of interest to be introduced into the general vector of the RNAi technique. After the ligation and transformation of *E. coli* cells, the corresponding selection for the positive colonies

was performed followed by DNA sequencing, to confirm the presence of the correct sequences.

The vector p9U10–RNAi (DNA-Cloning Services e.K., Hamburg, Germany) was the binary vector employed. p9U10 is an RNAi-based gene silencing vector, widely used in research. This vector works by inducing the transcription of large inserts directed by two flanking 35S promoters in opposite orientation (27).

## Fruit infiltration

Once the five constructs previously mentioned were obtained and the bHLH sequences confirmed, the binary plasmid was transferred to *A. tumefaciens* cells AGL0 strain, a strain which is able to perform the infection on the plant tissue to be analyzed. A fresh *Agrobacterium* liquid culture was grown overnight at 28ºC, and after 16 h ($OD_{600}$ of 0.8) the cells were collected by soft centrifugation at 2 g. Later the cells were resuspended in the Infiltration Solution Media (ISM) containing 200 µM acetosyringone as reported by Hoffmann *et al.* (19). The identification of the correct developmental stage of the fruits is essential. In general, the appropriate fruits can be described as immature fruits with a subtle green color.

The *Agrobacterium* infiltration procedure was performed according to previously described methods published by the team of Prof. Schwab (19). The infiltration was done with extreme caution to preserve the fruit integrity. *Agrobacterium* suspension was injected into the compact tissue using a hypodermic needle as seen in Figure 35; a slow and constant pressure was applied to the plunger, and a maximum of three injection points were made. After the infiltration, the surface of the fruit was cleaned with a paper towel impregnated with ethanol 70% to avoid the contamination of the fresh wounds.

A minimal number of ten fruits was used per treatment due to the low success rate of this technique. On estimate, only 40% of the cells will accept the injected vector according to literature (15, 18, 19).

Figure 35: Detail of the agroinfiltration process of the *F*. x *ananassa* fruit.

**Fruit collection**

14 days after infiltration was performed, the fruits were collected and analyzed for phenotypic changes (*i.e.*, color), changes in gene expression and changes in flavonoid levels. Fruits infiltrated with the pBi *Gus* construct were used as a negative control. The fruits were individually labeled and kept on ice before freezing with liquid nitrogen and storage at -80°C until use. Each fruit was cut in half, and the interior was photographed to compare changes between the treatments and the controls.

**RNA isolation and cDNA synthesis**

The RNA isolation and cDNA synthesis were performed in Italy at the Fondazione Edmund Mach, and the procedure used was the same as in chapter 4.

**qPCR**

cDNA was diluted 1:20 and qPCR was performed under the same conditions as previously described (see chapter 4). Primers for *bHLH 145* and the HKGs were the same as described in chapter 4, primers employed for CHS were described in Hoffmann (19), primer sequences for *F*. x *ananassa bHLH 3* and *33* were obtained from Schaart *et al.* (Reference 42, chapter 4). Primers for *Fa GT2* were used according to Schulenburg *et al.* (28).

**Metabolomic analysis**

The fruits were freeze-dried for 48 h using a lyophilizer, and later the achenes were separated from the flesh, as this latter tissue was under the direct effect of the *Agrobacterium* infiltration.

Once the flesh was cleaned, it was ground with the help of a cryomill, and the obtained powder was kept in the dark at -20ºC until further analyses.

An approximated 10 mg of tissue was extracted in 2500 µL 80% methanol for 48 h at 4ºC, and the supernatant was filtered through a 0.22 µm filter. The extract was analyzed by UPLC MS/MS method as described by Palmieri *et al.* (29). A total of 39 phenolic compounds were identified and quantified.

**Statistical analysis**

For the statistical analysis of both genetic and metabolomic data, the software STATISTICA version 13 (data analysis software system) was used (Dell Inc. (2016)). ANOVA followed by parametric Tukey, and non-parametric Kruskal-Wallis post hoc test was applied to make multiple comparisons of mean values to evaluate if there were significant differences between treatments. P-values of less than 0.01 were considered statistically significant.

## RESULTS AND DISCUSSION

In the previous chapters of this thesis, we identified the bHLH genes involved in genetic regulation of the flavonoid pathway in the wild strawberry, the diploid *Fragaria vesca*. Due to the small fruit size characteristic of *F. vesca*, an agroinfiltration procedure was not successfully performed, and the result was in most of the cases the detachment of the immature fruit. Because of this reason, and after having discussed this experimental problem with Dr. Schwab (TUM, Freising, Germany) we decided to use the big fruit of the commercial strawberries var. Elsanta, for the RNAi agroinfiltration experiments in *Fragaria* fruits.

A total of nine different infiltration treatments were done (Table 16). Two treatments consisted of: a positive control consisting of a reporter construct, the RNA interference sequence for the *F. x ananassa* gene *CHS* (19) and negative control to check the infiltration success, which is a pBi vector with the reporter gene *GUS* inserted.

Table 16. Summary of the constructs infiltrated in strawberry fruits

| Treatment | Constructs | Number infiltrated fruits |
|---|---|---|
| Gus (- Control) | pBI_*Gus*_intron | 7 |
| CHS_i (+ Control) | pBI_*CHS*_i | 10 |
| Fv3_i | *Fv3*_RNAi | 10 |
| Fv33_i | *Fv3*_RNAi | 10 |
| Fv145_i | *Fv145*_RNAi | 10 |
| Mix 3 | *Fv 3*i + *Fv145*i | 10 |
| Mix 4 | *Fv 33*i + *Fv145*i | 10 |
| Mix 5 | *Fv 3*i + *Fv 33*i + Fv145i | 10 |

The following diagram (Figure 36) shows the sample processing done per each fruit, for each treatment. Fruit samples were divided into two, each half to be used on metabolomic analyses (through methanol extraction) or gene expression analyses (after RNA extraction):

**METABOLOMIC ANALYSIS**

All samples

Freeze dried, ground

MetOH 80% extraction 48 hrs

Quantitative Analysis UHPLC-MS/MS

Total of 39 phenolics identified (cal curves)

Classification in 12 main groups

Statistical analysis by KW, and Fisher

**GENE EXPRESSION ANALYSIS**

5 samples per treatment

Tissue ground by Mortar (N2)

RNA extraction cDNA preparation

qPCR

HKG

CHS

DFR

Fv GT2

Figure 36: Simplified scheme for the analysis of the *F*. x *ananassa* fruits treated for RNAi silencing experiment.

**CHS control as evaluation for the agroinfiltration assay**

Before analyzing the possible changes in either the metabolites or gene expression levels of each *Fv bHLH* RNA_i treatment, it was necessary to evaluate first the successful rate of the agroinfiltrated fruits with the positive control *Fa CHS*_i.

As mentioned in the material and methods section, specific primers *for F. x ananassa* were available for the gene *CHS*. These primers were used not only to quantify the positive control gene expression but the *CHS* gene expression of all considered treatments. Data normalization was done using a different approach than the one reported on chapter 4; in this case, the best two HKGs analyzed for *F. vesca* in Chapter 4 *SAND* and *Actin* were combined

by the geNorm mathematical model to generate a normalization factor calculated for each sample (30).

In Figure 37, it can be seen how the relative expression of *CHS* gene (positive control) was strongly down-regulated by the pB9i vector compared to the GUS control by 81%. Considering the low success rate of the fruit infiltration technique reported, it can be considered that *CHS* RNAi silencing was a successful transient transformation experiment (19, 31).
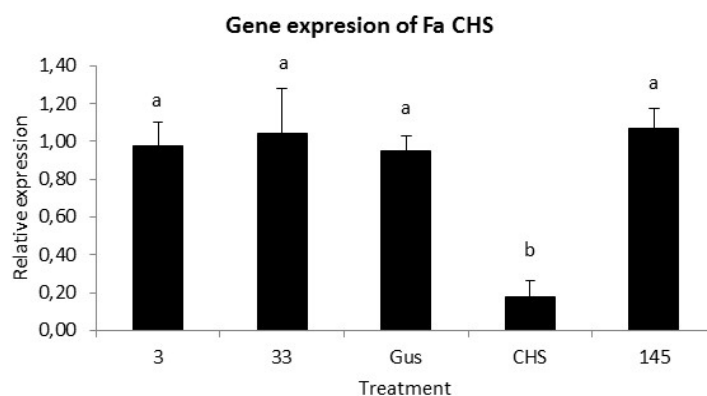


Figure 37: Gene expression values of *Fa CHS* in flesh tissue of *F. x ananassa*. Y-axis represents the gene expression value normalized with the combined HKG data. Bars sharing the same letter are not significantly different according to Tukey's HSD test. Error bars representing SD (standard deviation) of technical replicates. Letters on top of each bar represent statistically significant differences.

On the contrary, the *Fv3_i*, *Fv33_i*, and *Fv145_i* infiltrated fruits did not present any significant difference in the *CHS* expression values either on the t-test values (data not showed) or in the Tukey's honest significance test, that also confirms that there is no significant difference among the control *GUS* and the bHLH treatments.

This result was to some extent unexpected; as our initial hypothesis was to consider bHLH genes as positive regulators of the whole phenylpropanoid biosynthesis pathway, *CHS* being included in this genetic regulation as it has been demonstrated by Ling *et al*. (32). The lack of effect on the *Fv-bHLH* treated fruits can be due to: a) infectivity of the RNAi constructs designed using the *F. vesca* genome, and in this experiment, we used in the octoploid commercial strawberry *F. x ananassa* or, b) where the bHLH are controlling the flavonoid pathway at a more specific level, like the final compounds of the pathway, this last hypothesis will be analyzed and described in more detail in the following results.

However, we could establish the successful inhibition of gene *CHS* in the control treatment *CHS*_i, with a gene expression decrease of almost 80% compared to the control *GUS*.

**Effect on *Fa bHLH* gene expression levels**

Taking advantage of the study of Schaart *et al.* (31) in which *Fa bHLH 3* and *Fa bHLH 33* genes were characterized as the bHLH involved in the proanthocyanidin genetic regulation, we evaluated the gene expression of these two genes in the transiently transformed fruits.
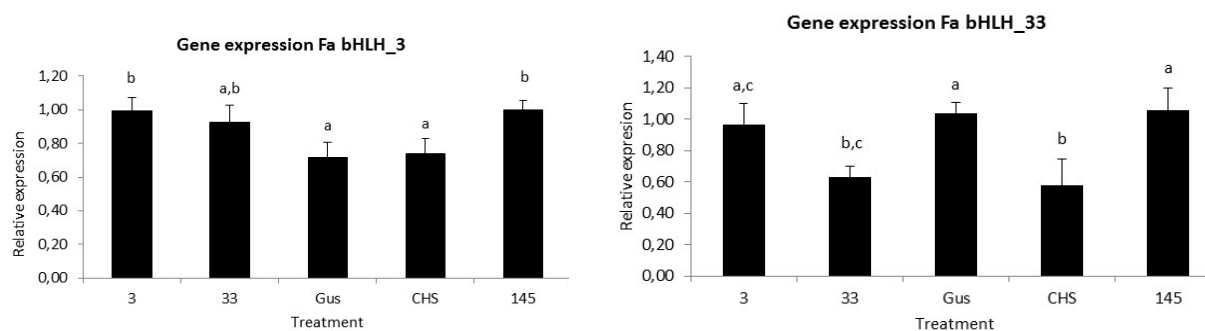


Figure 38: Gene expression analyses of *Fa bHLH* candidate genes in the flesh tissue of *F. x ananassa*. Y-axis represents the gene expression value normalized with the combined HKG data, X-axis representation of the different treatments. Error bars representing SD of the technical replicates. Letters on top of each bar represent statistically significant differences.

As shown in Figure 38, in all the treated fruits the relative expression of *Fa bHLH 3* seems to be affected positively, with an average of 25% higher gene expression compared to both controls (positive and negative). This increase of *Fa bHLH3* expression was seen in the three treatments where the fruits were infiltrated with *F. vesca* bHLHs. According to t-test results, the treatment Fv_145 has an average expression of Fa_3 significantly different from GUS control (data not shown). However, this increase in the gene expression cannot be considered an up-regulation. I strongly recommend an experiment, to compare the Fv_145 gene expression data with the results of an experiment where *Fa bHLH3* it is over-expressed.

On the other hand, the gene expression of *Fa bHLH 33* presented a diverse trend among all the treatments, where the *Fa_33*i fruits presented gene expression reduction of about 44% compared to the Gus control. A similar result was detected in the positive control CHS. This significant difference was confirmed by the p-value obtained on the t-test for this gene (0.004). The expression levels for both *Fv3_i* and Fv145i were the same as the Gus control.

**RNAi silencing effect of bHLHs on phenylpropanoid pathway genes**

Expression levels of the genes involved in the phenylpropanoid pathway were quantified by detected transcripts in qPCR. Here, we present the obtained results to clarify the regulatory

events occurring in the fruit tissue due the effect of the RNA interference vector pBi9 used, also to look to identify new possible relationships for the Fv bHLH candidate genes.

The gene expression analysis was performed on the first five fruit treatments from table 3; the treatments containing only one Fv bHLH RNA interference construct; this decision was made due the non-significant results in the metabolic profiling of the four mix treatments (see next section: Metabolite analyses), along with the elevated reagent costs if the additional samples were included.
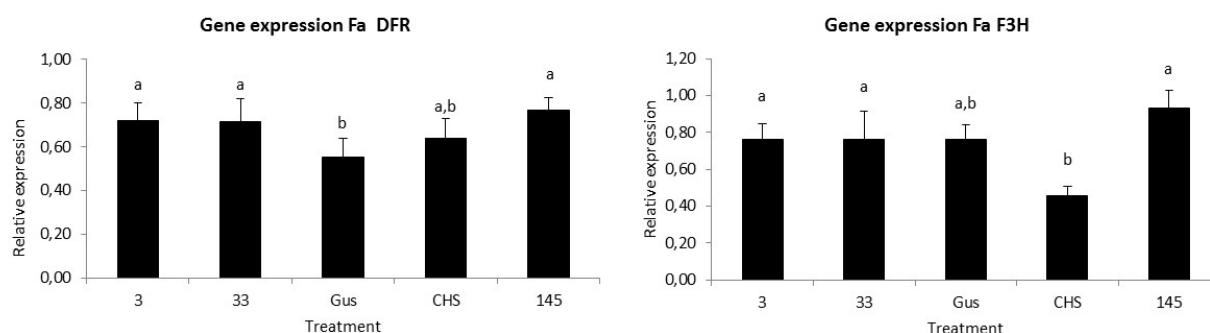


Figure 39: Gene expression values of flavonoid pathway genes in flesh tissue of *F. ananassa*. X-axis represents the different treatments. Bars sharing the same letter are not significantly different according to Tukey's HSD test. Error bars representing SD of technical replicates. Letters on top of each bar represent statistically significant differences

The expression of the pathway gene *DFR* seems to be not affected by any of the treatments evaluated (Figure 39), not even by the positive control CHS. DFR is considered as a protein typically present in more than one isoform. This is in agreement to the reported papers on *M. domestica* and *Pyrus* by Fischer *et al.* (33), and the annotation of at least two DFR putative proteins in the genomes of *F. vesca* and *R. idaeus*. It is possible that in this experiment, the set of primers being used are detecting the expression levels of the *DFR* transcript that is not involved in this branch of the anthocyanin biosynthesis but instead can be responsible for the lignin or PAs regulation (33).

**Fruit color variations**

*F. ananassa* cv. "Elsanta" fruits have an intense red color in the flesh. Color changes on infiltrated fruits were recorded on the day of the fruit collection. A representative picture of the different treatments can be seen in the following Figure 40.

GUS control berries are brilliant red, very similar to non-infiltrated berries. On the contrary, the fruits of positive control CHS lack red color formation and show white areas on the receptacle,

an indication of the lack of anthocyanin production. This result confirms the previous report from Hoffmann *et al.* (19) where the fruit infiltration was first described.
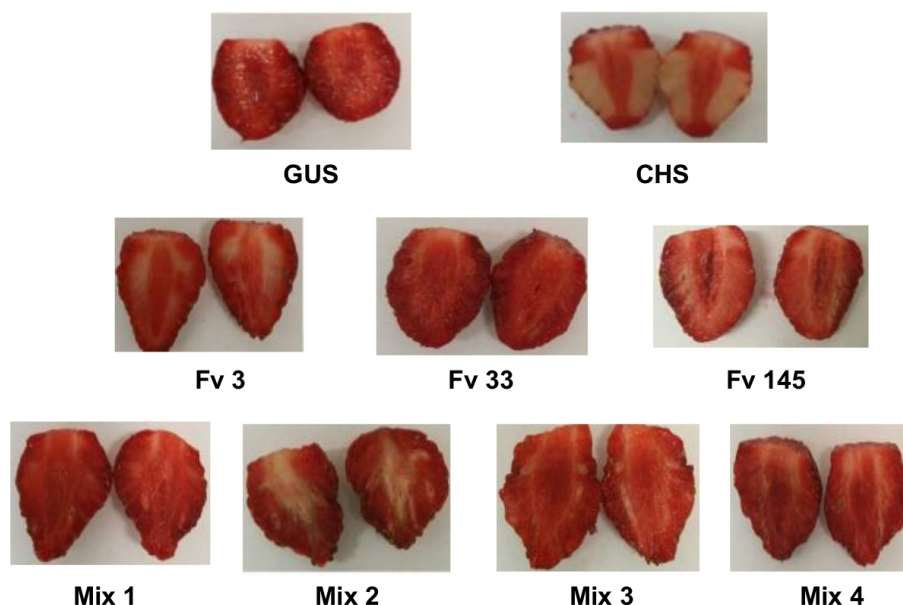


Figure 40: Photographic record of most representative fruits for each treatment. Controls are in the top level.

In all treated fruits, some small white areas can be seen on some fruits from the other treatments, including the negative controls, especially around the inner core of the fruit receptacle; this lack of coloration zones can also be seen on fruits grown for commercial purposes. Due to this random variation, it is also essential to include a higher number of samples and more precise methods of evaluation like metabolomic analyses that will be discussed later (34, 35).

As discussed before, final statements cannot be made based only on visual criteria, as it is possible that other kinds of phenolic compounds besides the colored anthocyanins are affected: therefore, in the next section, the results from the metabolite analysis will be presented and discussed.

The fruits treated with the bHLH candidate genes from *F. vesca* can be seen in the middle part of Figure 40: for all the three genes, the color observed is the same as fruit color in the negative control GUS. On the *Fv bHLH3*_i, treated fruits some small white areas with blurred edges can be seen. However, those white areas differ from the consistent white patches seen on fruits from CHS-control. The corresponding fruits for the mixed treatments (combination of two or the three *F. vesca* bHLHs), also showed an intense red coloration.

Strawberry fruits had been reported to be successfully transiently transformed in several studies. The limitation of this technique has also been pointed out. As an example, in the work

of Spolaore *et al.* (30), the reported *GUS* gene expression in the infiltrated tissue is described as "uneven and patchy." Another problem mentioned is the loss of samples due to contamination and poor tissue integrity (15, 18).

In this study, the quality of the tissue to analyze was increased due to the initial measures on stage identification and surface cleaning after the agroinfiltration. However, the non-transformed "patches" or regions can still be found according to the visual examination (e.g., fruits treated with CHS_i). Further analyses of the reporter gene *GUS* activity on the control fruits need to be done in order to assure the area percentage successfully transformed.

Additionally, for the evaluation of the success rate of our bHLH RNAi treatments, it is necessary to examine the presence of a reporter gene or protein tag in our RNA_i vector as the pAUL1-20 vector systems (15, 18, 36). As a recommendation for further experiments, it can be considered to divide each fruit to analyze in three pieces to have an additional tissue for the reporter gene evaluation either by histological assays in the case of GUS or by Western blot for protein tag.

## Metabolite analysis

Once the agroinfiltration of our positive control treatment CHS_i was confirmed, and the bHLH gene expression of the bHLH involved in the flavonoid regulation was analyzed, we continued with the second part of our analyses (as described in Figure 35), that is the analyses of the phenolic compounds obtained by methanol extraction.

A total of 37 different compounds were detected by the UPLC/MS-MS. Due to practical and analytical reasons, we decided to group these metabolites into 12 main groups as shown in Table 18:

Table 18: Phenylpropanoids detected by UHPLC/MS general classification

| Classification | Compounds |
|---|---|
| Benzoic acid derivatives | anthranilic acid, *p*-hydroxybenzoic acid, vanillin, vanillic acid |
| Phenylpropanoids | *p*-coumaric acid, ferulic acid, chlorogenic acid, cinnamic acid |
| Stilbenes | *trans*-piceide, *cis*-piceide |
| Dihydrochalcones | phloretin, phloridzin, trilobatin |
| Flavones | luteolin |
| Flavanones | hesperidin |
| Flavan–3-ols | catechin, procyanidin B1, procyanidin B3 (as B1) |
| Kaempferol derivatives | kaempferol, kaempferol-3-*O*-glucoside, kaempferol-3-*O*-rutinoside, kaempferol-3-*O*-glucuronide |
| Quercetin derivatives | quercetin-3-*O*-glucoside+quercetin-3-*O*-galactoside (as que-3-glc), quercetin-3-*O*-glucuronide, quercetin-3-*O*-glucoside acetyl |

| Other flavonols | rutin, isorhamnetin-3-O-rutinoside, taxifolin, isorhamnetin-3-O-glucoside |
|---|---|
| Anthocyanins | Pg 3 glucoside, Cy 3-O glucoside + Cy 3-O-galactoside (as cy-3-glc), Cy 3-O-rutinoside, Pg 3-O-rutinoside |
| Ellagitanins | casuarictin + sanguiin H-6 (as casuarictrin), agrimoniin, ellagic acid, methyl ellagic acid rhamnoside |

The following analysis was done per group of compounds, and not per individually compound due to the large sampling size (10 berries for each construct, times 7 constructs = 70 samples to analyze) and the time-consuming process that would require the analysis of every single one of the 37 compounds in the 170 samples.

Considering that the scope of this investigation is to elucidate the effect of bHLH TF on the biosynthesis of main phenolic compounds or even classes and some of their intermediates and not the bHLH effect in the final steps of the flavonoid pathway, the idea to analyze the bHLH TF effect on the products of specific enzymes that are involved in final modifications can only be seen as the next logical research projects in the field.

Table 19 presents the statistical mean with their corresponding standard deviation (SD) for each treatment. All the data are expressed in mg/kg of fresh flesh.

Table 19: Average values of polyphenols quantified on the flesh tissue of agroinfiltrated fruits. Showing the results for both controls and the single bHLH treatments. Data expressed in mg/kg of fresh flesh.

| | Gus (contol -) | CHS_i | Fv 3i | Fv-33i | Fv-145i |
|---|---|---|---|---|---|
| Benzoic Acids | 7.66 ± 1.23 | 10.46 ± 2.04 | 3.94 ± 0.73 | 3.24 ± 0.51 | 7.10 ± 1.42 |
| Phenylpropanoids | 22.62 ± 0.64 | 24.68 ± 1.16 | 23.17 ± 0.50 | 30.44 ± 3.15 | 21.47 ± 0.33 |
| Stilbenes | 0.95 ± 0.05 | 0.60 ± 0.11 | 1.22 ± 0.05 | 1.43 ± 0.12 | 1.17 ± 0.05 |
| Dihydrochalcones | 40.56 ± 4.32 | 7.75 ± 1.26 | 74.14 ± 5.76 | 73.44 ± 10.11 | 43.16 ± 4.48 |
| Flavones | 2.84 ± 0.64 | 1.06 ± 0.19 | 14.26 ± 2.90 | 15.34 ± 2.67 | 1.42 ± 0.22 |
| Flavanones | 0.12 ± 0.05 | 0.20 ± 0.04 | 0.13 ± 0.03 | 0.13 ± 0.04 | 0.11 ± 0.03 |
| Flavan-3-ols | 990.69 ± 169.76 | 714.92 ± 256.16 | 2,029.74 ± 273.64 | 2,739.21 ± 344.80 | 1,352.90 ± 158.52 |
| Keampferol Deriv | 46.44 ± 1.64 | 17.98 ± 2.29 | 60.94 ± 3.83 | 60.30 ± 6.69 | 40.74 ± 2.55 |
| Quercerin Deriv | 31.03 ± 1.99 | 19.76 t 2.00 | 38.96 ± 3.37 | 51.20 ± 8.15 | 22.56 ± 1.92 |
| Other Flavonols | 4.42 ± 0.37 | 3.02 ± 0.33 | 5.27 ± 0.31 | 6.95 ± 0.83 | 3.97 ± 0.21 |
| Anthocyanins | 6,966.97 ± 252.87 | 3,182.63 ± 370.89 | 7,713.50 ± 251.09 | 8,144.86 ± 585.28 | 6,589.80 ± 298.32 |
| Ellagitanins | 827.42 ± 167.20 | 656.72 ± 107.03 | 1,887.35 ± 253.14 | 2,276.97 ± 260.68 | 873.39 ± 70.22 |

The negative control treatment GUS was considered as reference for all the infiltrated fruits including the positive control CHS_i. In the positive control CHS_i, eight compounds of the 12 groups presented show a reduced value, especially the dihydrochalcones with the highest difference of 80.6%. The anthocyanin levels did also decrease by 54%, and the stilbenes by 34%. Flavan 3-ols were reduced by 27.8% and ellagitannins by 20%.

Some precursors of the pathway were also quantified in the forms of benzoic acids, and phenylpropanoids "building blocks" as *p*-coumaric. It is interesting to see how the levels of these compounds increased due to the blocked flux generated by the CHS RNA interference vector. There was a 10% increase for the "phenylpropanoids", and 36.5% for the benzoic acids. This result in addition to the previously described reductions on dihydrochalcones, serve as a confirmation of the effectiveness of the positive control treatment.

The results obtained for both bHLH transient infiltrations Fv3_i and Fv33_i can be described together in all the compound groups, only with slight increases or decrease. The most particular change was seen in the ellagitannins level that surprisingly increased compared to both controls. The flesh of fruits treated with the *Fv145* interference vector didn't seem to differ statistically from the negative control GUS. The metabolite quantification data will be discussed in detail by the proper statistical analyses and tests in order to determine the associations between all the bHLH-silenced treatments and the values of mg/kg of the 12 groups of compounds mentioned before.
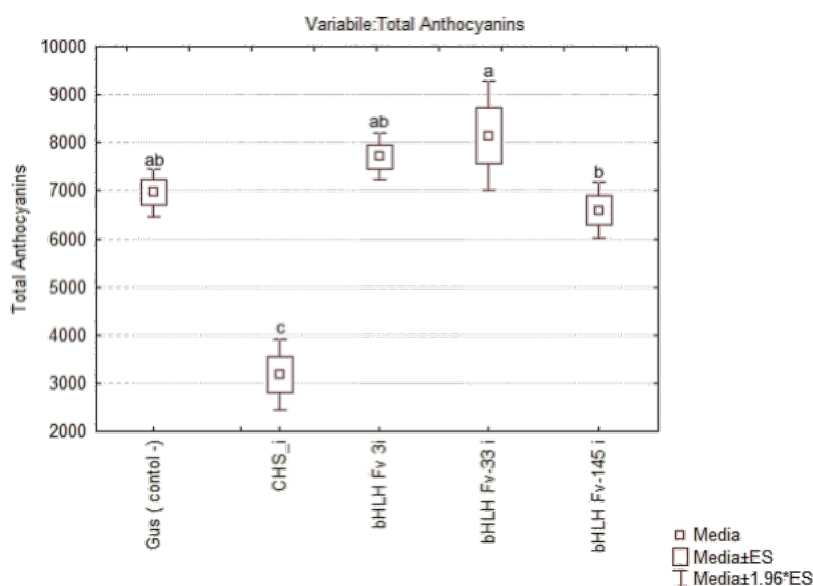
**Anthocyanins**



Figure 41: Mean anthocyanin compound content in *F. x ananassa* fruits expressed in mg/kg fresh fruit. Error bars refer to the standard deviation of technical replicates. Different letters were significantly different ($p < 0.1$) according to Tukeys's test.

As expected, the positive control of infiltrated fruits (CHS_i in Figure 41) presented a 54% reduced level of anthocyanins compared to the control GUS. Figure 41 shows that the anthocyanin levels in the *F. vesca* bHLH-silenced treatments were similar to the levels detected in the negative control GUS, including the *Fv3_i* with a 10% increase in the

anthocyanins levels. However, this increase in *Fv3_i* did not present a significance according to the Tukey test results (or Tukey's HSD: Honest Significant Difference).
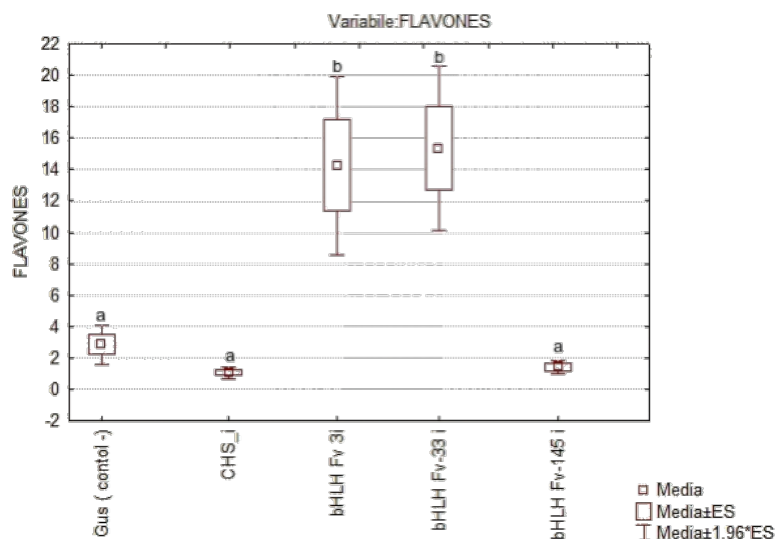
**Flavones**



Figure 42: Boxplot representing mean flavone compound content in *F. x ananassa* fruits expressed in mg/kg. Error bars refer to the standard deviation of technical replicates. Different letters were significantly different (p < 0.1) according to Tukeys's test.

The content of the flavone luteolin was increased more than 5.1 and 5.4 times for Fv3_i and Fv33_i, respectively. This increase is also confirmed by the significance groups of the Tukey test (Figure 42). This increment of the flavone content is only seen in "single bHLH treatments," but no effects on the level of flavones can be detected in the mix 2 (Fv3_i + Fv33_i) or mix 3 (Data not shown).

Possible confirmation of the specific activation of this branch of the pathway is to analyze the expression levels of Flavone Synthase (*FNS*) since FNS is one of the key responsible enzymes involved in flavone biosynthesis (14). As *Fragaria* is not known for the presence of high levels of flavones in their tissues, the *FNS* gene was not considered initially in the gene expression analyses. Further experiments will need to focus on the possible effect of bHLH on the flavone regulation in Rosaceae species, as it has been described in other species such as *Z. mays* (37).

Luteolin is a type of flavone reported in some plant innate response to pathogen attacks and is believed to have a positive effect on human health in terms of anti-inflammatory and antioxidant properties (13, 38). There are no reports about the luteolin content in *Fragaria* sp. fruits, or other fresh fruits according to the data provided by the website Phenol Explorer (http://phenol-explorer.eu/contents/polyphenol/229), making the data found in this study an interesting aspect of the roles of bHLH TFs in fruit development.
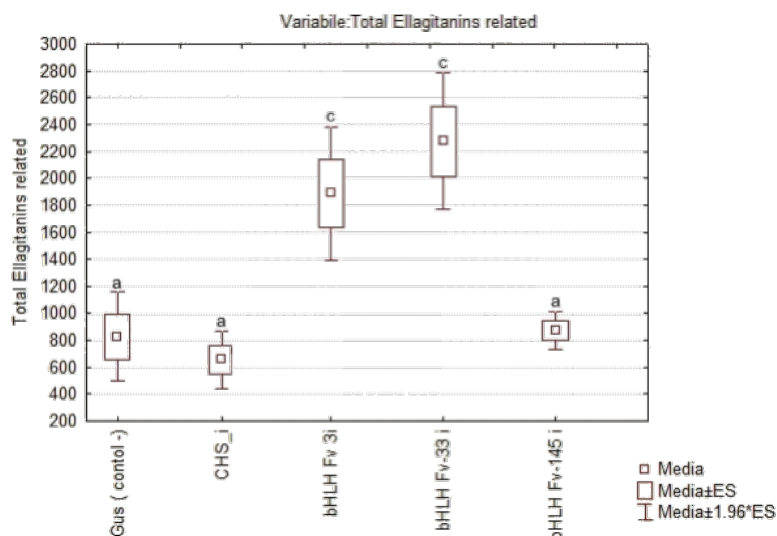
**Ellagitannins**



Figure 43: Boxplot representing mean ellagitannin compound content in *F. x ananassa* fruits expressed in mg/kg. Error bars refer to the standard deviation of technical replicates. Different letters were significantly different (p < 0.1) according to Tukeys's test.

The hydrolysable tannins or ellagitannins are polyphenols commonly known to occur mainly in the strawberry achenes, and, as mentioned before, ellagitannin quantities are usually low in ripe fruit tissue (39). Figure 43 represents the levels of ellagitannins in all the infiltrated fruits, and it can be seen clearly that the ellagitannin levels in Fv3_i and Fv3_i are 2.3 and 2.7-fold times higher, respectively, than both controls. Fruits infiltrated with Fv145_i did not undergo any change in the ellagitannin level.

In general, the values for the phenylpropanoid compounds evaluated in this study are higher than those reported in other commercial varieties like cv. "Camarosa" in which proanthocyanidins are between 137–179 mg/kg of fresh weight compared to the 990.69 mg/kg of fresh weight detected in the GUS control (Table 19). However, *p*-coumaroyl esters or phenylpropanoid building blocks as called in Table 18 presented a consistent value in our control samples compared to the 23 mg/kg of fresh weight reported in the same cv. "Camarosa" study (40).

Compared to the values reported in Gasperotti *et al*. (39), where the cv. "Elsanta" variety total ellagitannins were quantified between 262.5 mg/kg and 256.2 mg/kg in fresh and overripe fruits, respectively, in our results, the average ellagitannin content in the GUS control fruits is 827.42 mg/kg, a value that is more than 3.1 times higher than the previously mentioned cv. "Elsanta" quantification (6).

Nevertheless, it is important to consider the effect of the environmental growing conditions on the ellagitannin formation in the strawberry fruit and the additional stress that the agroinfiltration technique generates on the tissue, and the fact that, in this study, the achenes were removed from all the samples prior the extraction of metabolites. The ellagitannin values reported here are of fruits subjected to the agroinfiltration process, so new ellagitannin analyses of non-agro infiltrated fruits need to be done. Other strawberry varieties like cv. "Camarosa" have been described to contain 87-117 mg/kg of fresh weight of ellagitannins, this is 7.1 times less than the amount found in this study, but closer to other values reported in literature (3, 6, 40).

**Ellagitannin-related gene expression**

The ellagitannin content in Rosaceae species as *R. idaeus* and *F. vesca* has been previously reported and extensively studied due to the low number of fruits and vegetable that are rich in these kinds of compounds and their associations with reducing the risks of cardiovascular diseases and some cancers (3, 27, 41). A recent study published in 2015 reported and characterized two of the glucosyltransferases involved in the later steps of ellagic acid (and ellagitannin) formation in *F. vesca* and *F. x ananassa* (27). Based on the results of the mentioned study and our metabolomic data, the gene expression of the gene *FaGT2* was examined.
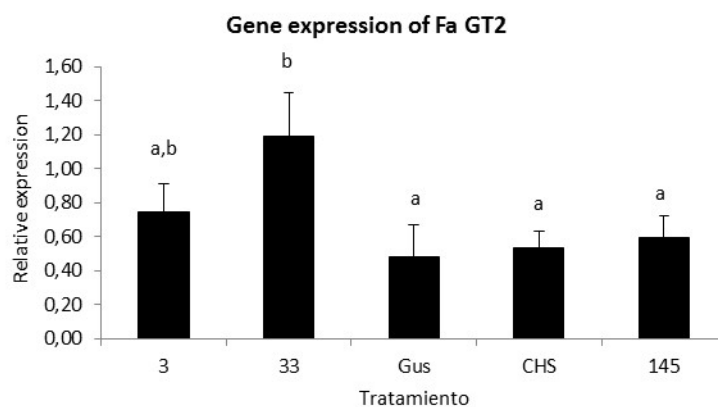


Figure 44: Gene expression values of *Fa GT2* in flesh tissue of *F x ananassa*. X-axis represents the different treatments. Bars sharing the same letter are not significantly different according to Tukey's HSD test. Error bars refer to the standard deviation of technical replicates. Letters on top of each bar represent statistically significant differences

Figure 44 represents the gene expression values for the gene *Fa GT2*, and it can be observed the increment in the expression of this gene for the treatments *Fv3_i* and *Fv33_i* compared to both controls and *Fv145_*i. *Fv33_i* treated fruits show an expression of 2.4 fold higher than the controls, followed by a significant 1.5 fold increase of *Fv3_i.*

## CONCLUSIONS AND REMARKS

In this experimental chapter, we evaluated the effect of specific bHLH RNA-interference vectors on the flavonoid pathway in the commercial species *F. x ananassa*, using information retrieved from the wild *F. vesca* as a template. The results presented here indicate that the RNAi constructs for *Fv 3*, *Fv 33* and *Fv145* (bHLH members of the sub-family IIIf, involved in the flavonoid pathway regulation), did not alter the color formation in the fruits, or change the total anthocyanins detected.

In general, the effect of *Fv bHLH3* and *Fv bHLH33* (see Table 18), was very similar in most of the metabolite groups analyzed here besides the ellagitannin content, where the content was elevated more than double compared to the controls. This result is a discovery in the bHLH TFs regulation of the phenylpropanoid pathway as it represents the first report of bHLH genes being involved in the ellagitannin regulation. This remark needs to be confirmed by the standard methodologies and approaches of gene regulation studies as same species promoter sequence analysis and activation studies, and protein interaction assays among others.

Further analyses in the change of ellagitannin composition in the fruits, where the bHLH RNA transcripts had been altered, need to be done to understand better this new part of the regulatory effect on the genes of the ellagitannin pathway.

Considering the genetic differences among the species investigated, *F. vesca* (2n) and the commercial *F. x ananassa* (8n), the numbers of genes present in *F. x ananassa* is 4 times bigger than its wild counterpart *F. vesca*. This was a preliminary study, so the fact that the whole phenylpropanoid and more specifically the anthocyanin production could not be stopped or down-regulated by the RNA interference vectors designed for the diploid *F. vesca*, employed here has to be investigated in more detail.

Our results suggest either a high rate of the change in the nucleotide sequence in *F. x ananassa* or a redundant role of the other bHLH present in their genomes. About the first hypothesis, it can be said that, nevertheless, the predicted models from *F. vesca* found in this study match more than 90% the amino acid sequence of the reported *F. x ananassa* by Schaart *et al*. (31). This hypothesis should be tested in a future study, by sequencing and comparing the gDNA sequences of both species. Unfortunately, due to time constrictions, this experimental part could not be included in this Ph.D. thesis.

The second hypothesis, bHLH redundancy, has been initially suggested by Heim *et al.* (22 Chaper 1), where genome duplication events may have led to gene redundancy. This phenomenon was supported by the results of Zhao *et al.* (42) where *Arabidopsis* bHLH TFs *AtMYC1* and *GL3* and *EGL3*, all members of the IIIf subfamily, have partial redundancy even though they have such a different function in the plant. Likewise, the reports of Morohashi and Grotewold (43) demonstrate beside the partial functional redundancy among *GL3* and *EGL3*, a higher complex level of genetic regulation between these two genes, a feed-forward regulation loop. Having said this, it is is possible that a similar mechanism is involved in the regulation of *Fv3*, *Fv33* and *Fv145*.

Further experiments in this redundancy subject need to be done in order to evaluate this phenomenon, for example through Y2H (yeast to hybrid), and double mutant *Arabidopsis* complementation, or even by using both RNAi vectors for the genes *Fv 3* and *Fv 33* together to evaluate their combined effect in the phenolic compounds biosynthesis in *Fragaria* during fruit formation.

## REFERENCES

1.  Najda A., & Dyduch, M. (2009). Chemical diversity within strawberry (*Fragaria vesca* L.) species. *Herba Polonica* Vol 55 No 3, 140-145.

2. Carbone, F., Preuss, A., De Vos, R. C. H., D'Amico, E., Perrotta, G., Bovy, A. G., … Rosati, C. (2009). Developmental, genetic and environmental factors affect the expression of flavonoid genes, and metabolites in strawberry fruits. *Plant, Cell & Environmen*t, 32(8), 1117–1131.

3. Vrhovsek, U., Guella, G., Gasperotti, M., Pojer, E., Zancato, M., Mattivi, F. (2012). Clarifying the identity of the main ellagitannin in the fruit of the strawberry, *Fragaria vesca* and *Fragaria ananassa* Duch. *Journal of Agricultural and Food Chemistry* 60, 2507–2516.

4. Ibrahim, D. S., & Abd El-Maksoud, M. A. E. (2015). Effect of strawberry (*Fragaria × ananassa*) leaf extract on diabetic nephropathy in rats. *International Journal of Experimental Pathology*, 96(2), 87–93.

5. Winkel-Shirley B. (2001): Flavonoid biosynthesis. A colorfulmodel for genetics, biochemistry, cell biology, and biotechnology. Plant Physiology 126(2):485–493.

6. Gasperotti, M., Masuero, D., Guella, G., Palmieri, L., Martinatti, P., Pojer, E., … Vrhovsek, U. (2013). Evolution of ellagitannin content and profile during fruit ripening in Fragaria spp. *Journal of Agricultural and Food Chemistry*, 61(36), 8597–8607.

7. Johansen, L. K. and Carrington, J. C. (2001) Silencing on the spot. Induction and suppression of RNA silencing in the Agrobacterium-mediated transient expression system. *Plant Physiology*. 126, 930–938.

9. Llave, C., Kasschau, K. D., & Carrington, J. C. (2000). Virus-encoded suppressor of posttranscriptional gene silencing targets a maintenance step in the silencing pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 97(24), 13401–13406.

10. Meins, F. (2000). RNA degradation and models for post-transcriptional gene silencing. In M. A. Matzke & A. J. M. Matzke (Eds.), *Plant Gene Silencing*. pp. 141–153.

11. Waterhouse, P. M., & Helliwell, C. A. (2003). Exploring plant genomes by RNA-induced gene silencing. *Nature Reviews. Genetics*, 4(1), 29–38.

12. Agrawal, N., Dasaradhi, P. V. N., Mohmmed, A., Malhotra, P., Bhatnagar, R. K., & Mukherjee, S. K. (2003). RNA Interference: Biology, Mechanism, and Applications. *Microbiology and Molecular Biology Reviews*, *67*(4), 657–685.

13. Liu, C.-W., & Murray, J. D. (2016). The Role of Flavonoids in Nodulation Host-Range Specificity: An Update. *Plants*, *5*(3), 33.

14. Martens, S., & Mithöfer, A. (2005). Flavones and flavone synthases. *Phytochemistry*, 66(20), 2399–407.

15. Spolaore, S., Trainotti, L., Casadoro, G.; A simple protocol for transient gene expression in ripe fleshy fruit mediated by *Agrobacterium. Journal of Experimental Botany* 2001; 52 (357): 845-850.

16. Orzaez, D., Mirabel, S., Wieland, W. H., & Granell, A. (2006). Agroinjection of Tomato Fruits. A Tool for Rapid Functional Analysis of Transgenes Directly in Fruit. *Plant Physiology*, *140*(1), 3–11.

17. Carvalho, R. F., Carvalho, S. D., O'Grady, K., & Folta, K. M. (2016). Agroinfiltration of Strawberry Fruit a Powerful Transient Expression System for Gene Validation. *Current Plant Biology*, 6, 19–37.

18. Guidarelli, M., & Baraldi, E. (2015). Transient transformation meets gene function discovery: the strawberry fruit case. *Frontiers in Plant Science*, 6, 444.

19. Hoffmann, T., Kalinowski, G. & Schwab, W. (2006). RNAi-induced silencing of gene expression in strawberry fruit (*Fragaria × ananassa*) by agroinfiltration: a rapid assay for gene function analysis. *Plant Journal*. 48, 818–826.

20. Miyawaki, K., Fukuoka, S., Kadomura, Y., Hamaoka, H., Mito, T., Ohuchi, H., et al. (2012). Establishment of a novel system to elucidate the mechanisms underlying light-induced ripening of strawberry fruit with an *Agrobacterium*-mediated RNAi technique. *Plant Biotechnology*. 29, 271–277.

21. Ring, L., Yeh, S. Y., Hu_cherig, S., Hoffmann, T., Blanco-Portales, R., Fouche, M., *et al.* (2013). Metabolic interaction between anthocyanin and lignin biosynthesis is associated with peroxidase FaPRX27 in strawberry fruit. *Plant Physiology*. 163, 43–60.

22. Medina-Puche, L., Molina-Hidalgo, F. J., Boersma, M. R., Schuurink, R. C., López-Vidriero, I., Solano, R., *et al.* (2015). A R2R3-MYB transcription factor (FaEOBII) regulates eugenol production in ripe strawberry (*Fragaria x ananassa*) fruit receptacles. *Plant Physiology.*

23. Jiang, F., Wang, J. Y., Jia, H. F., Jia, W. S., Wang, H. Q., and Xiao, M. (2013). RNAi-mediated silencing of the flavanone 3-hydroxylase gene and its effect on flavonoid biosynthesis in strawberry fruit. *Journal of Plant Growth Regulation*. 32, 182–190.

24. Lin, X., Xiao, M., Luo, Y., Wang, J., and Wang, H., (2013). The effect of RNAi-induced silencing of FaDFR on anthocyanin metabolism in strawberry (*Fragaria x ananassa*) fruit. *Science Horticulture*.

25. Griesser, M., Hoffmann, T., Bellido, M. L., Rosati, C., Fink, B., Kurtzer, R., *et al.* (2008). Redirection of flavonoid biosynthesis through the down-regulation of an anthocyanidin glucosyltransferase in ripening strawberry fruit. *Plant Physiology*. 146, 1528–1539.

26. Daminato, M., Guzzo, F., and Casadoro, G. (2013). A SHATTERPROOF-like gene controls ripening in non-climacteric strawberries, and auxin and abscisic acid antagonistically affect its expression. *Journal of Experimental Botany*. 64, 3775–3786.

27. Schulenburg, K., Feller, A., Hoffmann, T., Schecker, J. H., Martens, S., & Schwab, W. (2016). Formation of β glucogallin, the precursor of acid in strawberry and raspberry. *Journal of Experimental Botany*, 67(8), 2299–2308.

28. Zhang, J., Wang, X., Yu, O., Tang, J., Gu, X., … Fang, C. (2011). Metabolic profiling of strawberry (*Fragaria x ananassa* Duch.) during fruit development and maturation. *Journal of Experimental Botany*, 62(3), 1103–1118.

29. Palmieri, L., Masuero, D., Martinatti, P., Baratto, G., Martens, S. and Vrhovsek, U. (2017), Genotype-by-environment effect on bioactive compounds in strawberry (*Fragaria x ananassa* Duch.). Journal of Science Food Agriculture. Version of record online: 31 March 2017.

30. Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., … Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, 3(7).

31. Schaart, J. G., Dubos, C., Romero De La Fuente, I., van Houwelingen, A. M. M. L., … Bovy, A. G. (2013), Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (*Fragaria × ananassa)* fruits. *New Phytologist*, 197: 454–467.

32. Lin-Wang, K., McGhie, T. K., Wang, M., Liu, Y., Warren, B., Storey, R., … Allan, A. C. (2014). Engineering the anthocyanin regulatory complex of strawberry (*Fragaria vesca*). *Frontiers in Plant Science*, 5, 651.

33. Fischer, T. C., Halbwirth, H., Meisel, B., Stich, K., & Forkmann, G. (2003). Molecular cloning, substrate specificity of the functionally expressed dihydroflavonol 4-reductases from Malus domestica and Pyrus communis cultivars and the consequences for flavonoid metabolism. *Plant Journal*, 412, 223–230.

34. Le Mière, P., Hadley, P., Darby, J., & Battey, N. H. (1998). The effect of thermal environment, planting date and crown size on growth, and yield of *Fragaria × ananassa* Duch. cv. Elsanta. *The Journal of Horticultural Science and Biotechnology,* 73(6), 786–795.

35. http://www.beeren-plantproducts.com/en/aardbeienras-elsanta

36. Lyska, D., Engelmann, K., Meierhoff, K., & Westhoff, P. (2013). pAUL: A Gateway-Based Vector System for Adaptive Expression and Flexible Tagging of Proteins in *Arabidopsis. PLoS ONE*, *8*(1), e53787.

37. Falcone Ferreyra, M. L., Rius, S., Emiliani, J., Pourcel, L., Feller, A., … Grotewold, E. (2010), Cloning and characterization of a UV-B-inducible maize flavonol synthase. *The Plant Journal*, 62: 77–

38. Lopez-Lazaro, M. (2009). Distribution and Biological Activities of the Flavonoid Luteolin. Mini-Reviews in *Medicinal Chemistry*.

39. Gasperotti, M., Masuero, D., Guella, G., Palmieri, L., Martinatti, P., … Vrhovsek, U. (2013). Evolution of ellagitannin content and profile during fruit ripening in Fragaria spp. *Journal of Agricultural and Food Chemistry*, 61(36), 8597–8607.

40. Cerezo, A. B., Cuevas, E., Winterhalter, P., Garcia-Parrilla, M. C., & Troncoso, A. M. (2010). Isolation, identification, and antioxidant activity of anthocyanin compounds in Camarosa strawberry. *Food Chemistry*, 123(3), 574–582.

41. Landete, J. M. (2011). Ellagitannins, ellagic acid and their derived metabolites: a review about source, metabolism, functions and health. *Food Research International*, 44, 1150-1160.

42. Zhao, H., Wang, X., Zhu, D., Cui, S., Li, X., Cao, Y., & Ma, L. (2012). A Single Amino Acid Substitution in IIIf Subfamily of Basic Helix-Loop-Helix Transcription Factor AtMYC1 Leads to Trichome and Root Hair Patterning Defects by Abolishing Its Interaction with Partner Proteins in *Arabidopsis*. *The Journal of Biological Chemistry*, 287(17), 14109–14121.

43. Morohashi, K., and Grotewold, E. (2009) A systems approach reveals regulatory circuitry for *Arabidopsis* trichome initiation by the GL3 and GL1 selectors. *PLoS Genetic*, 5.

# Chapter 6

**General Conclusions**

# GENERAL CONCLUSIONS

The phenylpropanoid biosynthetic pathway is one of the most studied and well-known plant metabolic pathways, not only in model species as *A. thaliana* and *M. truncatula* but also in economically important crops such as *Z. mays*, *V. vinifera* and *M. domestica*. This pathway generates several kinds of secondary metabolites, which, besides conferring beneficial properties to plants like the defense-related flavonoids and colored anthocyanins, they also provide benefits for human health, due to their anti-oxidative stress capacity.

Genetic regulation of phenylpropanoid biosynthesis is mediated by the MYB-bHLH-WD40 complex, which recognizes specific motifs in promoter regions through the DNA-binding domains of the TF proteins. The presence and proper assembly of the protein complex is necessary for the gene expression of each one of the biosynthetic structural genes involved in the multiple steps of the pathway, from the early pathway gene *CHS* until the later genes encoding enzymes that catalyze chemical modifications such as glycosylation like the UDP-Glucose: flavonoid 3-O-glucosyltransferases (*3GT*).

In strawberry and raspberry, the red coloration, typical for the mature fruits, is due to the presence of anthocyanins. Even though both commercial species present higher anthocyanins contents compared to the wild varieties, the market requirements for better quality fruits regarding health benefits are a constant pressure for breeders around the world. Thanks to the antioxidant and anti-inflammatory properties of the anthocyanins, and the relatively easy detection by the red color intensity, these compounds are the principal trait to improve. In an important crop as apple, a member of the Rosaceae family that also includes *F. vesca* and *R. idaeus*, it has been shown that the red coloration in the flesh of fruits can be due to the effect of overexpression of TFs of particular MYB type (Espley *et al.* 2007 (Ref. 40, Chap 1)). Other crops as tomatoes, potatoes, and rice have also been studied in the recent years in order to identify the genes involved in the color formation of wild-type cultivars and to bring this trait to new commercial varieties.

In this doctoral thesis, the basic helix-loop-helix or bHLH proteins involved in the regulation of the phenylpropanoids in two berries were studied: *F. vesca* or woodland strawberry and *R. idaeus,* raspberry. Both species belong to the Rosaceae family, which also contains other fruits such as apple, peach, pear, and cherries. The results described here are aimed to develop a better knowledge of the regulation of the phenylpropanoid pathway in both species and their activation mechanism on the fruit color formation.

The bHLH proteins represent a large family of TF regulators, not only in plants but also in animals and yeast. There are so far 150 members identified in the model species *A. thaliana* and more than 240 in apple (*M. domestica*). In plants, bHLH proteins are involved in the regulation of several processes such as phytochrome signaling, gynoecium development, trichome development, cold tolerance response, tryptophan, and phenylpropanoid biosynthesis among other processes, indicating the participation of the bHLH family during all plant cycle stages. Members of this large family can be identified and classified by the secondary motifs or domains present in their amino acid sequence. The classification made by Heim *et al.* (Ref. 22, Chap 1) is still used nowadays in the study of this family.

Based on the *A. thaliana* bHLH classification and phylogenetic studies reported, the genomes of *F. vesca* and *R. idaeus* were screened, and putative gene candidates were found for both species in this thesis. Posterior sequence analyses based on protein length and motif conservation were performed on the initial gene candidates models, and a total of 98 bHLH protein-coding sequences were found in *F. vesca* genome v1.0 and 90 protein-coding sequences in the unreleased draft version of the *R. idaeus* genome.

Phylogenetic analyses for each one of the species were performed and compared to characterized bHLH proteins in closely related (*M. domestica*) and model species (*A. thaliana*). All protein sequences were aligned and the pairwise distance calculated between them, followed by Maximum likelihood (ML) analyses followed by 1000 bootstrap calculations to obtain a consensus tree. Groups of proteins such as ICE, PIFs, and BEE were grouped similarly to previous reports. A significant result was the identification of the bHLH members of the IIIf group in *F. vesca* and *R. idaeus*, recognized as the proteins involved in the flavonoid regulation in *A. thaliana*: homologs of *AtMYC1*, *GL3*, *EGL3*, and *TT8*. The *in-silico* results obtained in chapter 4 provide three and two gene candidates for the woodland strawberry and raspberry, respectively: *Fv3-FV2G25270*, *Fv33-FV7G08120*, *Fv145-FV5G02910*, *Ri3 gene36602* and *Ri3-gene26116*.

Previous reports of bHLHs in the closely related Rosaceae species *P. persica* (peach) did also reveal three bHLH genes belonging to the clade of *Arabidopsis TT8*, *GL3*, and *MYC1* (Rahim *et al.* (Ref. 11, Chap 4), expressed in the peel of red skinned varieties. In the case of apple, so far only two bHLHs have been identified to be involved in the anthocyanin-flavonoid pathway regulation.

The expression of candidate genes involved in the flavonoid pathway was analyzed during fruit developmental stages to establish a first correlation between the color formation (as an indication of anthocyanin biosynthesis) of this specific tissue and the regulatory effect of the

different bHLH genes. As both species are not yet widely studied in molecular genetic aspects, specific HKG were designed and tested for future reference studies on berries qPCR based gene expression experiments. *F. vesca* and *R. idaeus* homologous HKGs behave in the same way, having as a result that genes *PP2A*, *ACT*, *SAND*, and *UBC9* were suitable for gene normalization purposes.

In *F. vesca*, the three candidate bHLH genes were all significantly increased in their expression in the flower/fruit ripening process, and, in general, the leaf tissue had a lower gene expression. Interestingly, *FvbHLH145* seems to be less important during the later stages of fruit development (time points 5 and 6), based on its decreased expression. For *R. idaeus* the gene expression of the two bHLH candidate genes was inverse to *F. vesca* results and contrary to the fruit maturation and anthocyanin formation. The evaluation of the function of a new gene requires a series of experiments. Here the activation of the flavonoid pathway was evaluated using heterologous systems as *A. thaliana* and tobacco plants, and specific gene promoter activation using reporter-gene assays - either GUS activity or Luciferase bioluminescence assay.

Using the *Agrobacterium*-mediated transient transformation system, *N. tabacum* leaves were used as an initial checkpoint of anthocyanin pathway activation. Reported results so far show that transient expression of bHLH is not enough to activate the whole biosynthetic pathway of anthocyanins in *N. benthamiana* leaves. For this experimental part, the *F. vesca* bHLH candidates were selected. The corresponding DNA was cloned from the cDNA synthesized from previous analyses and inserted into vectors for plant expression pHex2 and pK7WG2. As a result, infiltrated *N. tabacum* leaves with FvbHLH3 and FvbHLH33 presented a higher amount of the general phenylpropanoids; a 93% and 12% more compared to the control treatment respectively, but the result obtained with the gene *FvbHLH145* was the opposite: a 40% reduction of the total polyphenols compared to mock control and 70% to the FvbHLH3 treatment.

Several types of phenylpropanoids were detected by UPLC LC/MS in the *N. tabacum* leaf extracts. Compounds like kaempferol derivatives, chlorogenic acid, dihydrochalcones, and quercetin, were identified and quantified without any significant difference found among mock and the three bHLH treatments. However, a significant increase of flavonoids after FvbHLH3 and FvbHLH33 treatments, in the form of Cy-3-rut, was observed. Additionally, coumarins in the form of fraxin and scopoletin were increased in the previous treatment mentioned compared to the mock treatment. The HPLC LC/MS results observed in the FvbHLH145 treatment failed to demonstrate a direct regulation of the anthocyanin biosynthesis. Interestingly *Fv bHLH145* in our phylogenetic analyses groups with *Arabidopsis* GL3, a gene

involved in leaf trichome formation, further experiments in this subject need be considered and performed.

The results presented here, help to elucidate the regulatory role of the *F. vesca* bHLH genes *3* and *33* in the flavonoid pathway biosynthesis as anthocyanin specific regulators. To further study the function of FvbHLH3, *Arabidopsis* mutant line *tt8* (*FvbHLH3* homolog) complementation assay was also performed. The analyzed transgenic *Arabidopsis* lines are from the T1 generation, in which heterozygosis segregation was still undergoing, making the results inconclusive. Unfortunately, due to time constrictions, it was necessary to perform the analysis on these T1 lines. However, I highly recommend analyzing the T2 homozygous stable transformed lines.

The six T1 complemented mutant lines were analyzed by the Anthocyanin Induction Conditions (AIC) for the anthocyanin pathway reactivation. After five days in the AIC treatment, the original mutant line *tt8* did not present any visible coloration (pink/violet), contrary to the response of the parental WT line L*er*. Three out of the six complemented lines (lines 3, 23 and 24) showed a high level of anthocyanins according to the color observed in the hypocotyl. The remaining three lines (lines 5, 16 and 19) showed a less intense coloration with the presence of white and green areas in the hypocotyl. A total anthocyanin quantification was performed for all lines, using a cyanidin 3-*O*-glucoside standard curve: chromatogram at 520 nm, peak retention times and UV spectra were analyzed.

The HPLC detection analyses show that the mutant line *tt8* contained a low level of anthocyanins, 0.25 µg/mg FW representing a 66% reduction of the anthocyanins detected in the WT. This value was used as a reference point for anthocyanin quantification and comparison with the complemented lines. In personal communications with Dr. Antje Feller (former researcher on anthocyanin regulation in *Zea mays*), this anthocyanin formation in *tt8* lines was discussed, and similar results were detected previously.

Detected values from the lines presented a broad range from line 16 = 0.33 µg/mgFW equivalent to a 35% increase of the anthocyanins of the mutant line until the 241% increase in anthocyanins detected in line 3 (0.83 µg/mg FW). This broad range of anthocyanins detected is not a surprising result, as this represents the intrinsic variation of the complementation by T-DNA insertion, from the flower dip transformation procedure. Additional analyses as copy number determination of the T-DNA insertion in transformed complemented *tt8* lines need to be performed.

Overall, the mutant complementation of *Arabidopsis tt8* lines allowed us to confirm the phenylpropanoid regulator role of Fv bHLH 3, as the stable complemented lines obtained and analyzed were able to produce anthocyanins in the proper conditions (sugar osmotic stress). Likewise, the anthocyanin detection by HPLC in methanol extracts proved to be a powerful and sensitive detection method for this type of studies.

The promoter of the phenylpropanoid pathway gene *CHS* from *M. domestica* was used for the analysis of the capacity of activation of *F. vesca* bHLH genes through the LUC/REN assay, which was performed in *N. benthamiana* leaves by coinfiltration with MdMYB10. *F. vesca* bHLH genes presented a broad activation range of the *MdCHS* promoter. The promotor of *FvbHLH3* presented almost 4 times more activity compared to the negative control. *FvbHLH33* presented 1.7 times more activation than their respective control and *FvbHLH145* showed a 76% lower activation compared to the results of *FvbHLH3* and *FvbHLH33*. These results confirm the previous reports on *Malus CHS* promoter where *Md_bHLH3* seems to activate *CHS* much better than *Md_bHLH33*.

After having characterized the regulatory properties of the *F. vesca* bHLH genes, the next experimental part was designed to evaluate the possible roles of the bHLHs in plants. For this, the immature fruits of the commercial octaploid strawberry cultivar "Elsanta" were transiently transformed with agrobacteria in order to evaluate the effect of bHLH on the maturation of *F. x ananassa* fruits. RNAi-silencing vectors were designed for each *F. vesca* bHLH gene. As a positive control, a reporter construct containing an RNA interference sequence for the *F. x ananassa* gene *CHS* in the vector p9U10–RNAi was used. In total, nine different infiltration treatments were performed, two treatments consisted of the positive control and negative control to check the infiltration success, which is a pBi vector with the reporter gene *GUS* inserted.

After 10 days of the treatment, the mature fruits infiltrated with the positive control CHS lack red color formation and showed white areas in the receptacle, an indication of the lack of anthocyanin production. This result confirms the previous report from Hoffman *et al.* (Ref. 19, Chap 5) where the fruit infiltration was first described. The fruits treated with the bHLH candidate genes from *F. vesca* did not show a visible color change in their fruit receptacle. Additional metabolites analyses were made by HPLC/MS. FvbHLH3_i and FvbHLH33_i treated fruits, did not present a particular change in anthocyanins, stilbenes, flavanones and kaempferol derivatives. However, it was seen that the flavone and ellagitannin levels were increased compared to both controls.

Compared to the values reported in Gasperotti *et al.* (Ref. 6, Chap 5), where the total ellagitannins of "Elsanta" variety were quantified between 262.5 mg/kg and 256.2 mg/kg in fresh and overripe fruits respectively, in the results obtained here the average ellagitannin content in the GUS control fruits was 827.42 mg/kg, a value more than 3.1 times higher than the previously mentioned "Elsanta" quantification. The gene *FaGT2* presented an increment in transcript levels detected by qPCR analyses in the cDNA from fruits infiltrated with Fv bHLH 3_i and Fv bHLH 33_i, resulting in the first reported relationship of the regulation of ellagitannins by bHLHs.

The lack of reduced anthocyanins in the infiltrated berries might be due to redundancy among the FvbHLHs. Redundancies have been shown in trichome formation in *Arabidopsis* in Zhao *et al.* (Ref. 42, Chap 5) and the studies of Morohashi and Grotewold 2009 (Ref. 43, Chap 5). Our results also showed that in *tt8* mutant lines, where the main bHLH phenylpropanoid regulator is not functional, it is still possible to detect the presence of anthocyanin formed under stress conditions.

To date, ellagitannin formation and especially its genetic regulation, have been a new subject of study for the plant biochemistry community, among other reasons because they are mostly present in a non-model organism like Rosaceae. Looking at the ellagitannin biosynthetic pathway and main precursors the close relationship that this type of compounds (may) have with other better-studied compounds like anthocyanins and proanthocyanidin it is evident. Considering Fv bHLH 3 and Fv bHLH 33 as regulators in the repressor form, for ellagitannins, it is a crucial step for possible future scenarios to improve ellagitannin content in berries

Overall, the results presented in this thesis contribute to a deeper understanding of gene regulatory mechanisms in non-model plants *F. vesca* and *R. idaeus* by identifying the bHLH TFs involved in the flavonoid pathway, and their expression during fruit developing process. Additionally, *F. vesca* bHLH genes were partially characterized and their effect on the biosynthesis of phenylpropanoid compounds evaluated.

To conclude, the results presented in this thesis contribute to a deeper understanding of gene regulatory mechanisms in the non-model plants *F. vesca* and *R. idaeus* by identifying and characterizing the role of bHLH3 and 33 in the regulation of flavonoid and ellagitannin biosynthesis.

# Statutory Declaration

Ich, Andrea Lorena Herrerera Valderrama, versichere, dass ich meine Dissertation:

„Identification and characterization of regulatory proteins involved in anthocyanin biosynthesis in *F. vesca* and *R. idaeus* "

selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe. Alle vollständig oder sinngemäß übernommenen Zitate sind als solche gekennzeichnet.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den..........................

.........................................

Andrea Lorena Herrera V.

# ACKNOWLEDGMENTS

Il bar Aquila Nera, sulla rotunda di San Michele, anche a voi, tanti apperitivi doppo lunge ore di laboratorio.

To my loved family, in Colombia and to my Dutch family, thanks for their comprehension and support during this journey. For understanding that I like to ask questions about plants. Jaap Wolters, dank je wel for the amazing dinners, academic talks, micro-gardening and fun times.

A mi familia y amigos en Colombia (y el resto del mundo), por creer siempre en mi y darme su amor, comprension y paciencia. And last but not least, I would like to acknowledge my strength and dedication.

# ANNEX 1

Table 1: bHLH protein list from *Fragaria vesca*

| bHLH | Fvesca geneme ab initio v1-1 Protein Accesion | Fvesca EST database accesion number | aa length |
|---|---|---|---|
| FvbHLH1 | gene00160 | 1509361 | 196 |
| FvbHLH2 | gene00194 | 1509412 | 405 |
| FvbHLH3 | gene00549 | 1509760 | 411 |
| FvbHLH4 | gene00557 | 1509769 | 411 |
| FvbHLH5 | gene00635 | 1509846 | 464 |
| FvbHLH6 | gene00687 | 1509897 | 185 |
| FvbHLH7 | gene00689 | 1509899 | 426 |
| FvbHLH8 | gene00756 | 1509967 | 277 |
| FvbHLH9 | gene00989 | 1510204 | 418 |
| FvbHLH10 | gene01065 | 1510283 | 189 |
| FvbHLH11 | gene01886 | 1511108 | 438 |
| FvbHLH12 | gene02195 | 1511413 | 1038 |
| FvbHLH13 | gene02373 | 1511595 | 344 |
| FvbHLH14 | gene02549 | 1511774 | 377 |
| FvbHLH15 | gene02616 | 1511843 | 620 |
| FvbHLH16 | gene03243 | 1512385 | 448 |
| FvbHLH17 | gene03286 | 1512429 | 231 |
| FvbHLH18 | gene03658 | 1512804 | 432 |
| FvbHLH19 | gene03693 | 1512839 | 235 |
| FvbHLH20 | gene03860 | 1513005 | 371 |
| FvbHLH22 | gene04195 | 1513341 | 226 |
| FvbHLH24 | gene04485 | 1513639 | 573 |
| FvbHLH25 | gene04505 | 1513660 | 368 |
| FvbHLH26 | gene04818 | 1513980 | 188 |
| FvbHLH27 | gene05031 | 1514190 | 228 |
| FvbHLH28 | gene05531 | 1514685 | 366 |
| FvbHLH29 | gene05580 | 1514735 | 376 |
| FvbHLH30 | gene05582 | 1514737 | 431 |
| FvbHLH31 | gene06060 | 1515218 | 369 |
| FvbHLH32 | gene06396 | 1515560 | 535 |
| FvbHLH33 | gene06409 | 1515573 | 366 |
| FvbHLH34 | gene06701 | 1515862 | 238 |
| FvbHLH35 | gene07234 | 1516402 | 238 |
| FvbHLH36 | gene07275 | 1516444 | 171 |
| FvbHLH37 | gene07310 | 1516480 | 332 |
| FvbHLH38 | gene07345 | 1516516 | 325 |
| FvbHLH39 | gene08016 | 1517172 | 711 |
| FvbHLH40 | gene08157 | 1517312 | 292 |
| FvbHLH41 | gene08199 | 1517355 | 504 |
| FvbHLH42 | gene08374 | 1517535 | 692 |
| FvbHLH43 | gene08899 | 1518060 | 192 |
| FvbHLH44 | gene09076 | 1518239 | 244 |
| FvbHLH45 | gene09188 | 1518346 | 450 |
| FvbHLH46 | gene09188 | 1518668 | 274 |
| FvbHLH47 | gene09705 | 1518859 | 200 |

| | | | |
|---|---|---|---|
| FvbHLH48 | gene10027 | 1519185 | 459 |
| FvbHLH49 | gene10170 | 1519326 | 168 |
| FvbHLH50 | gene10468 | 1519623 | 682 |
| FvbHLH51 | gene11392 | 1520557 | 231 |
| FvbHLH52 | gene11527 | 1520693 | 685 |
| FvbHLH53 | gene11799 | 1520968 | 371 |
| FvbHLH54 | gene11976 | 1521151 | 340 |
| FvbHLH55 | gene12069 | 1521238 | 324 |
| FvbHLH56 | gene12143 | 1521312 | 339 |
| FvbHLH57 | gene12607 | 1521779 | 609 |
| FvbHLH58 | gene12857 | 1522026 | 256 |
| FvbHLH59 | gene12909 | 1522079 | 479 |
| FvbHLH60 | gene13060 | 1522229 | 625 |
| FvbHLH61 | gene13318 | 1522485 | 235 |
| FvbHLH62 | gene13427 | 1522597 | 757 |
| FvbHLH63 | gene13430 | 1522600 | 199 |
| FvbHLH64 | gene13777 | 1522825 | 353 |
| FvbHLH66 | gene14381 | 1523545 | 281 |
| FvbHLH67 | gene14560 | 1523724 | 321 |
| FvbHLH68 | gene14775 | 1523930 | 262 |
| FvbHLH69 | gene16010 | 1525145 | 811 |
| FvbHLH70 | gene15196 | 1525332 | 335 |
| FvbHLH71 | gene15535 | 1525675 | 377, |
| FvbHLH72 | gene16835 | 1525981 | 550 |
| FvbHLH73 | gene17240 | 1526382 | 286 |
| FvbHLH74 | gene17402 | 1526539 | 544 |
| FvbHLH75 | gene17419 | 1526557 | 178 |
| FvbHLH76 | gene17651 | 1526788 | 698 |
| FvbHLH77 | gene17695 | 1526832 | 228 |
| FvbHLH78 | gene17848 | 1526990 | 296 |
| FvbHLH79 | gene17902 | 1527046 | 298 |
| FvbHLH80 | gene18322 | 1527463 | 737 |
| FvbHLH81 | gene18509 | 1527651 | 275 |
| FvbHLH82 | gene18575 | 1527717 | 482 |
| FvbHLH83 | gene18617 | 1527760 | 491 |
| FvbHLH84 | gene18854 | 1527997 | 411 |
| FvbHLH85 | gene19111 | 1528252 | 616 |
| FvbHLH86 | gene19288 | 1528430 | 653 |
| FvbHLH87 | gene20109 | 1529255 | 286 |
| FvbHLH89 | gene20375 | 1529520 | 675 |
| FvbHLH90 | gene20523 | 1529667 | 211 |
| FvbHLH91 | gene20584 | 1529730 | 431 |
| FvbHLH93 | gene20758 | 1529905 | 214 |
| FvbHLH94 | gene20775 | 1529923 | 1216 |
| FvbHLH95 | gene20959 | 1530108 | 261 |
| FvbHLH96 | gene21255 | 1530410 | 157 |
| FvbHLH97 | gene21391 | 1530548 | 273 |
| FvbHLH98 | gene21933 | 1531096 | 319 |
| FvbHLH100 | gene22228 | 1531395 | 342 |
| FvbHLH101 | gene22424 | 1531596 | 316 |
| FvbHLH102 | gene22799 | 1531965 | 272 |

| FvbHLH103 | gene22835 | 1532002 | 361 |
|-----------|-----------|---------|-----|
| FvbHLH104 | gene22900 | 1532068 | 768 |
| FvbHLH105 | gene23542 | 1532818 | 583 |
| FvbHLH106 | gene23543 | 1532819 | 621 |
| FvbHLH107 | gene23749 | 1532922 | 948 |
| FvbHLH108 | gene23750 | 1532933 | 250 |
| FvbHLH109 | gene23815 | 1532987 | 391 |
| FvbHLH111 | gene24504 | 1533677 | 365 |
| FvbHLH112 | gene24534 | 1533707 | 327 |
| FvbHLH113 | gene24535 | 1533708 | 302 |
| FvbHLH114 | gene24582 | 1533754 | 245 |
| FvbHLH115 | gene25078 | 1534249 | 345 |
| FvbHLH116 | gene25199 | 1534371 | 332 |
| FvbHLH117 | gene25479 | 1534644 | 435 |
| FvbHLH118 | gene25488 | 1534652 | 269 |
| FvbHLH119 | gene25489 | 1534653 | 407 |
| FvbHLH120 | gene25750 | 1534919 | 258 |
| FvbHLH121 | gene25903 | 1535071 | 182 |
| FvbHLH123 | gene25914 | 1535082 | 778 |
| FvbHLH124 | gene26004 | 1535171 | 597 |
| FvbHLH125 | gene26068 | 1535237 | 582 |
| FvbHLH126 | gene25811 | 1535977 | 224 |
| FvbHLH127 | gene26826 | 1535992 | 354 |
| FvbHLH128 | gene25898 | 1536064 | 429 |
| FvbHLH129 | gene25970 | 1536136 | 1122 |
| FvbHLH130 | gene27354 | 1536519 | 333 |
| FvbHLH131 | gene27351 | 1536526 | 906 |
| FvbHLH132 | gene27760 | 1536924 | 723 |
| FvbHLH133 | gene27771 | 1536937 | 262 |
| FvbHLH134 | gene28247 | 1537417 | 518 |
| FvbHLH136 | gene28702 | 1537882 | 1035 |
| FvbHLH138 | gene29524 | 1538799 | 168 |
| FvbHLH139 | gene29747 | 1538924 | 262 |
| FvbHLH140 | gene30387 | 1539573 | 577 |
| FvbHLH141 | gene30575 | 1539760 | 484 |
| FvbHLH142 | gene30578 | 1539762 | 263 |
| FvbHLH143 | gene32054 | 1541266 | 187 |
| FvbHLH144 | gene32240 | 1541441 | 299 |
| FvbHLH145 | gene32388 | 1541589 | 631 |
| FvbHLH146 | gene32430 | 1541631 | 567 |