Characterising Chlorogenic Acid Biosynthesis in Coffee

Daniel Knevitt

A thesis submitted to the University of East Anglia for the degree of Doctor of Philosophy

John Innes Centre
Norwich
September 2016

Abstract

Coffee is an important commodity and a major export for developing countries. There are two commercially grown species: *Robusta* and *Arabica.* The latter is more desirable, but more difficult to cultivate. It is susceptible to pests and diseases and less tolerant to environmental changes. This vulnerability is partly due to *Arabica* accumulating less chlorogenic acid (CGA) than *Robusta*. There are high levels of CGAs in the coffee beverage which contribute to the flavour and confer health benefits. In this study, I characterised enzymes involved in the biosynthesis of CGAs in coffee and investigated the function of transcription factors in controlling the phenylpropanoid pathway that lead to CGA production.

The enzyme hydroxycinnamoyl-CoA quinate hydroxycinnamoyltransferase (HQT) is entirely responsible for the synthesis of the major CGA, 5-caffeoylquinic acid (5-CQA). I also discovered two routes for the synthesis of dicaffeoylquinic acids (diCQAs) from 5-CQA utilising two different enzymes in different subcellular compartments. HQT could synthesise diCQAs in the presence of high concentrations of 5-CQA when localised to the vacuole. Hydroxycinnamoyl-CoA quinate/shikimate hydroxycinnamoyltransferase (HCT) could synthesise diCQAs through the same route but can also synthesise diCQAs at neutral pH using 5-CQA and caffeoyl CoA. Tissue distribution patterns of metabolites in developing coffee fruit confirmed the presence of these biosynthetic routes.

I cloned and characterised several *R2R3MYB* genes encoding potential regulators of CGA biosynthesis. Their analysis also led to a possible explanation for the usually high levels of CGA in coffee. Distinct MYB12-like transcription factors activated the transcription of a non-functional chalcone synthase (*CHS*) gene which is important for the synthesis of flavonols. This results in high levels of CGAs at the expenses of flavonol accumulation. Understanding CGA biosynthesis in coffee will be useful for sustainable cultivation of this important crop.

Acknowledgements

Sitting here and writing this, feels weird. Unlike writing the abstract, this feels more like I am closing a chapter in my personal journey during my PhD life. When I think about my journey towards writing-up this thesis there are so many people I would like to thank.

Firstly, my supervisor, Cathie, her support and dedication to science throughout my PhD was invaluable, especially during the last month whereby she was magical, she could read and comment on my work and give it back to me the very next morning, every time. I would also like to thank Jim McCarthy and all at Nestlé who supported and helped fund my research over these past four years.

During the last few months of my PhD, the journey train began to wobble a bit and at one point was almost derailed, Trevor Wang, Cathie Martin and Claire Domoney helped me through by continuous encouragement and reality checking my progress which was sometimes well behind schedule.

Every day in the laboratory was made an absolute pleasure by the people who I was fortunate to share it with. Since I have been a part of the Cathie Martin group many people have come and gone, all of them equally contributed to me having a good time and helping me interpret my data. Within the group, I must give special thanks to Eugenio, the only member who has been in the group throughout my whole journey, he is brilliant and passionate and a pleasure to have around, I missed him when he wasn't there. My lunchtimes were always made interesting by excellent people who always ensured I had a good respite from laboratory work through interesting debates, these helped keep me sane and meant never a lunchtime was eaten alone. People who I met in school are still friends of mine, this is unusual considering how little time I have spent with them throughout the duration of my PhD. I thank them wholeheartedly for being a significant part of my life for so long and I hope it continues for many years to come. My family have always been there for me; they are always happy and cheerful which always lifts my mood.

I have to thank my loving partner Hannah, who has stuck by me for many years now. Just her love alone was good enough to help me get through these last few months but she insisted on supporting me in many other ways. I am the luckiest man alive to have such a person who has given up so much to support me, I don't think she realises just how much help she has been.

My personal PhD journey has been one of highs and lows, (quite literally towards the end), I am sad to see it is coming to an end. I have learnt so much about myself and I feel confident now in performing fundamental skills which I have severely lacked in the past. The people who have been involved in this process are people who I will never forget.

– Now my eyes are open, I can see –

For Winnie, my first dog.
Born and passed between
starting and finishing.
Never forgotten

Contents

List of figures

List of tables

List of Abbreviations

%	Percentage
35S CaMV	35S Cauliflower Mosaic Virus
4CL	4-coumarate:CoA ligase
BHLH	Basic helix loop helix
bp	Base pairs
C3’H	p-coumaroyl quinate/shikimate 3'-hydroxylase
C4H	Cinnamate 4 hydroxylase
CaAGT	Caffeoyl CoenzymeA chlorogenate transferase
Cas	CRISPR associated
CCoAMT	Caffeoyl CoA methyl transferase
CCT	Chlorogenate:chlorogenate transferase
cDNA	Complementary DNA
CGA	Chlorogenic acid
CHS	Chalcone synthase
CoA	Coenzyme A
CQA	Caffeoylquinic acid
CRISPR	Clustered regularly interspecific short palindromic repeat
CSA	Caffeoyl shikimic acid
DBD	DNA binding domain
DESI	Desorption electrospray ionisation

diCQA	Dicaffeoylquinic acid
DNA	Deoxyribose nucleic acid
dNTP	Deoxynucleotide
dsRED	Discosoma red fluorescent protein
E.coli	Escherichia coli
eQTL	Expression quantitative trait loci
ESI	Electrospray ionisation
eYFP	Enhanced yellow fluorescent protein
FQA	Feuroylquinic acid
G	Grain
g	Grams
g	Gravity
GFP	Green fluorescent protein
GOI	Gene of interest
HCT	Hydroxycinnamoyl-Coenzyme A: quinate/shikimate hydroxycinnamoyltransferase
HPLC	High performance liquid chromatography
HQT	Hydroxycinnamoyl-Coenzyme A: quinate hydroxycinnamoyltransferase
KO	Knockout
l	Litres
LC	Liquid chromatography
LG	Large green
LG	Large green
M	Molar

m/z	Mass charge
MALDI	Matrix assisted laser desorption ionisation
MBW	MYB BHLH WD40 repeat complex
min	Minutes
ml	Millilitre
mm	Millimetre
mM	Millimolar
mRFP	Modified red fluorescent protein
MS	Mass spectroscopy
MSI	Mass spectrometry imaging
ng	Nanograms
nm	Nanometres
^{o}C	Degrees Celsius
OD_{600}	Optical density 600 nm
P	Pericarp
PAL	Phenylalanine ammonia lyase
PCR	Polymerase chain reaction
PDA	Photodiode array
R	Red
RACE	Rapid amplification of cDNA ends
RNA	Ribose nucleic acid
RPKM	Reads per kilobase of transcript per million mapped reads
rpm	Revolutions per minute

SDS	Sodium dodecyl sulphate
SG	Small green
sgRNA	Signal guide RNA
T-DNA	Transfer DNA
TIC	Total ion chromatogram
UTR	Untranslated region
UV	Ultraviolet
VSG	Very small green
VVSG	Very very small green
VVVSG	Very very very small green
WGD	Whole genome duplication
WT	Wild type
Y	Yellow
μl	Microliter
μM	Micromolar

Chapter 1: General introduction

1.1 Coffee

Coffee is the world's most widely consumed beverage and the second most traded commodity after oil. Coffee is grown and exported by developing countries that lie between the tropics, and is grown by almost 25 million small growers. Coffee is, for many countries, the most important exporting sector [1]. The coffee supply chain is very complex and involves many stakeholders which, in turn, makes it one of the most economically important products in the world.

Small coffee growers (less than 5 hectares) account for 70% of the world's production but supply only a few international coffee traders. Over the last few decades' international coffee traders have undergone internalisation and have increased their market share. The biggest three trading groups now have 45% market share and the two biggest roasting groups control 57% [1]. This has resulted in market power distribution being skewed away from the small grower. Some responsibility for the future sustainability of coffee production lies with these large trading and roasting groups.

Fair trade coffee is coffee that is certified as having been produced to 'fair trade' standards. Organisations involved actively seek to create trading partnerships that have greater equality in international trade which offer better trading conditions to coffee bean farmers. The adoption of fair trade coffee by consumers is essential to the success of the movement as currently supply of coffee that meets fair trade standards outweighs demand [2].

Two coffee species account for nearly all commercialised coffee; higher quality *Arabica (Coffea arabica)* beans are mostly produced in West African and South American countries, and poorer quality *Robusta (Coffea canephora)* beans are predominantly produced in Southeast Asia and Oceania [3]. Cultivating *Arabica* is more difficult than *Robusta*, in that it needs to be grown at an altitude higher than 1200m, and is susceptible to diseases such as coffee leaf rust and coffee berry disease [4]. In contrast, *Robusta,* grows at sea level and is resistant to many of the diseases that affect *Arabica*. Climate change threatens the sustainability of *Arabica* with yields being affected by a small variations in temperature and rainfall, for example Colombia experienced a severe 30% decline in production for 5 years as a result of excessive rainfall in 2008 [5]. Currently, *Robusta* coffee production accounts for about 30% of world production and is worth around half the price of *Arabica*. Due to lower desirable characteristics and strong flavour, *Robusta* beans are used mainly for cheap, instant coffees or as a blend with *Arabica*. Because of the hardy characteristics of *Robusta* and ease of cultivation, it is likely that strategies to improve the flavour of *Robusta* coffee will be implemented soon to increase demand and therefore, market value.

1.1.1 Genetics

Coffea is the genus to which all plants that produce coffee beans belong and consists of 103 species [4]. It is a member of the family Rubiaceae and native to tropical and southern Africa and tropical Asia. All species of coffee are diploid with the exception of *Coffea arabica,* which is tetraploid, derived from a relatively recent whole genome duplication (WGD) event that occurred between *Coffea canephora* (*Robusta*) and *Coffea eugenioides* in the plateaus of Central Ethiopia [6]. Conflicting evidence has been presented about when this WGD event took place, with a few reports suggesting that it happened 10,000 – 50,000 years ago [7, 8] and another which estimated it took place 665,000 years ago [9].

The genetic diversity of *Arabica* is small when compared to *Robusta,* due to its narrow genetic base associated with its relatively recent origin and its domestication. The world's most widely cultivated cultivars are derived from two populations known as Typica and Bourbon and these populations were derived from progeny of very few plants [10]. The low genetic diversity in populations is a limitation faced through traditional breeding programs. Analysis of wild coffee trees indicates them to have relatively high polymorphism and accessions have been introduced into breeding programs which will hopefully increase the sustainability of coffee production [6]. *Robusta* coffee is self-incompatible whereas *Arabica* is self-compatible. This imposes limitations on the applications of current technology for the improvement of *Robusta* coffee varieties.

1.1.2 History

The history of coffee consumption probably began in Ethiopia, or across the Red Sea in Yemen. It has been said that consumption started in the 9th century AD. Juicy and sweet sun-ripened red beans were consumed straight from the tree or fried in butter, creating flavoursome lactones. Six hundred years later the first 'modern' coffee was produced (roasted, ground and brewed) [11]. This beverage probably originated in the city of Mocha, Yemen, which reached its peak in the 17th century, owing to its trade in coffee [12]. Exporters of coffee in Mocha reduced the risk of losing their market share by roasting their coffee beans before exporting. This prevented competitors cultivating their *Arabica* coffee variety, Mocha [13]. The success of Mocha coffee meant that several powerful countries attacked the city of Mocha to gain control of its important and strategic port, most notably in 1820 by the British and a decade later by the Egyptians, although by this time coffee trading in Mocha had been overtaken by trading in Ethiopia [13].

Coffee consumption spread from Yemen to the rest of the Middle East and northern Africa progressively during the 16th century. When coffee appeared in Venice around 1650, intense trading through the Venetian merchant shipping routes caused coffee consumption to spread quickly throughout Europe. Around 1668 coffee was introduced to North America by immigrants, probably from England, joining the growing number of colonies. These days, the USA is the world's biggest consumer of coffee, popularised by many who rejected consumption of tea as a symbol of unjust imperial rule [11, 13]. The French were indirectly responsible for introducing coffee cultivation to South America in the 1700s through their overseas territory of Martinique [14]. Today, South America produces more coffee than any other continent with Brazil being the largest exporter.

Currently, coffee culture is booming, enjoyed by millions of people around the world. Over just a few centuries coffee has become an important part of everyday culture and one of the world's most important traded commodities.

1.1.3 Preparation

The time it takes to cultivate coffee to produce its first crop is dependent on the variety; *Arabica* will take 4-5 years while *Robusta* will take only 2-3 years [15]. The coffee seeds are the parts of the plant that are used for the preparation of the coffee beverage. Once fruit development has reached the mature red stage, the fruit is harvested either by hand or machine. Depulping or wet milling of the fruit is the first part of the processing and must take place within 24-36 hours of picking. Beyond this time, decomposition creates undesirable characteristics in the final coffee beverage [16]. The coffee is then fermented for a day or two to help soften and remove some more of the pulp. The resulting beans are then dried in the sun to reduce their water content to around 10%, and the beans are then ready for dry milling.

There are several ways in which coffee cherries are processed to make a cup of coffee, many of which are determining factors for the characteristics of the final product. Dry milling is the process that removes the last layers or 'parchment' around the dry coffee beans, cleans and sorts the beans, ready for roasting. Many processing plants still employ people to remove defective beans manually because of the high cost of complex machinery and relatively low costs of labour in communities which process coffee.

Roasting conditions are the factors that have the most influence on flavour in the final coffee beverage. There are various parameters that define the degree of roast. The most used indicator is the colour of the coffee bean [17]. The optimum degree of roast is determined by various factors

which include; the origin of the green coffee beans, the intended brewing method and personal taste.

1.1.4 Phenolic compounds in coffee

The phenolic compounds found in coffee are almost exclusively chlorogenic acids, although traces of flavan-3-ols and flavonols are present in some types from different regions, but the latter are not present in the final coffee cup because they accumulate in the bean husk which is removed in processing [18]. Coffee is a remarkably rich source of chlorogenic acids (CGA) which are important phytochemicals in the diet. Out of the two commercialised varieties of coffee, *Robusta* contains more chlorogenic acids than *Arabica.* An analysis of 38 genotypes of each variety showed that CGA in green grain ranged from 7.88 to 14.4 % dry matter (dm) for *Robusta* and 3.4 to 4.8% dm for *Arabica* [19].

CGAs are incredibly diverse in coffee, there have been 45 different compounds identified in *Arabica* and 69 in *Robusta* [15, 20]. The major CGA subgroups are quinic acid esters with one caffeic acid, the caffeoylquinic acids (CQA), or two caffeic acids, the dicaffeoylquinic acids (diCQA) or ferulic acid(s), the feruloyquinic acids (FQA), and these subgroups account for over 98% of CGAs [21]. Minor classes are *p*-coumaryolquinic acids (*p*-CoQA), caffeoylferuloylquinic acids (CFQA), diferuloylquinic acids (diFQA), di-*p*-coumaroylquinic acids (di-*p*-CoQA), dimethoxycinnamoylquinic acids and others [20, 22].

The name chlorogenic acid (CGA), used in many papers, is given to the isomer 5-caffeoylquinic acid (5-CQA), which is the most abundant CGA present in most plants. In my thesis, I will be using chlorogenic acid (CGA) to refer to the whole group of compounds, and the IUPAC nomenclature of 5-caffeoylquinic acid (5-CQA) to describe 5-CQA. Before the introduction of IUPAC nomenclature in 1976, 5-CQA was called 3-caffeoylquinic acid (3-CQA). Since then, many papers and suppliers, have continued to use the pre-IUPAC nomenclature and wrongly refer to 3-CQA as chlorogenic acid and as the main substituent of the family [23]. This is sometimes confusing for the reader when trying to distinguish between different isomers between papers.

CGA levels in the developing coffee seed change throughout maturity, and this happens differently in different cultivars of coffee, not only the changes in total CGA levels but also the relative proportions of different subgroups [24].

During coffee roasting, CGAs are degraded because of thermochemical decomposition in the absence of oxygen (pyrolysis), which generates lactones and other compounds. The main lactones

produced belong to a group called cinnamoyl-1,5-γ-quinolactones (CGL). Thus, considerably less CGA is present in coffee after the roasting process. A typical cup of coffee, consisting of 200ml will deliver anything between 15 and 325 mg of CGA, which means that regular coffee drinkers can exceed consumption of more than 1 gram per day, making coffee the biggest contributor of CGA in the US diet. Total CGA decreases as the roast grades of coffee increase [25], CGL content reaches a maximum after 7 minutes of roasting at 230°C, on average, before dropping [26].

There are many ways in which a coffee beverage can be prepared and different extraction methods produce coffee with different CGA profiles. The CGA profile is also influenced by water temperature and extraction time [27-29]. Decaffeination also changes CGA composition. There are two main methods used for commercial decaffeination of coffee, both performed prior to roasting. The most common and least costly method is use of an organic solvent such as dichloromethane or ethyl acetate to extract caffeine. The other method involves use of water only to replace the use of organic solvents. During the decaffeination process, loss of key flavour components also occurs, especially when the solvent used lacks specificity, like water. Considering that part of the 5-CQA content of coffee is known to associate with caffeine, it is highly likely that the decaffeination process affects the CGA profile. Analysis of CGA from different *Arabica* coffee varieties showed that, before roasting, decaffeinated coffee showed a 17% average increase in total CGA (mainly due to the loss of other compounds). After roasting, decaffeinated coffee had a 3-9% decrease in total CGA depending on roast grade [25, 30].

1.1.5 Chlorogenic acid: role in coffee flavour

It is well known that *Arabica* coffee produces better quality coffee than *Robusta*. *Robusta* is predominantly used in cheap instant coffees and as a filler in other coffees. There are methods that exist to improve the quality of *Robusta* coffee. One process is the steaming of *Robusta* which is used to create the specific acid taste and flavour that is found only in *Arabica* coffees. The process removes some compounds that are related to flavour such as chlorogenic acids and free diterpenes [31]. The chemistry of flavour development during coffee roasting is complex and not completely understood. Roasting of coffee would appear simple in terms of roasting conditions, but from a chemical perspective, it is not, because there are hundreds of compounds undergoing many different chemical reactions. There have been relatively few compounds identified that contribute to flavour and aroma directly, compared to the total number present. CQAs, 3-caffeoylquinic acid (3-CQA), 4-caffeoylquinic acid (4-CQA), and 5-caffeoylquinic acid (5-CQA) have all been identified as compounds that contribute to the bitterness of coffee, when roasted. Sensory analysis has revealed bitter tastes from derivatives of CGAs. These compounds are caffeoyl-γ-quinides, which

are not present in unroasted coffee but are present in roasted coffees [32]. Dicaffeoylquinic acids (diCQA) have also been identified as compounds related to bitterness when roasted, producing many different lactones, while 4,5-dicaffeoylquinic acid itself is a compound related positively to coffee flavour [22]. The abundance of CGA post roasting has been associated with having a negative impact on coffee flavour, and increasing amounts are associated with lower quality taste profiles [33].

1.1.6 Health benefits of coffee consumption

Coffee has been classified as a beverage to reduce or omit from the diet because consumption has been associated with an unhealthy lifestyle [34] and because of its caffeine content. Coffee has progressively moved to being viewed less negatively due to increasing awareness of its phytonutrient content and the positive effects CGAs have on the human body. Moderate consumption of coffee has been linked to the reduced incidence of many diseases including cardio vascular disease (CVD), coronary heart disease (CHD), strokes, diabetes, obesity, liver inflammation, Parkinson's disease, Alzheimer's disease, depression, osteoporosis, and many types of cancer [35-41]. Most health benefits have been linked to the content and diversity of the chlorogenic acids, others are realted to the caffeine [35]. Understanding the relative roles that individual compounds play in the reduction of different diseases will be important for the development of coffees which have potential therapeutic uses.

1.2 Plant secondary metabolites

Plants produce a range of compounds that are considered to be primary metabolites and are essential to photosynthesis, respiration, growth and reproduction. Examples of primary metabolites include sugars, carbohydrate polysaccharides, lipids, amino acids and nucleic acids.

Secondary metabolites have two major characteristic features: diverse chemical structures and high interspecific variation. It is because of these features that historically secondary metabolites have been viewed simply as non-essential and possibly only detoxification products. We know now that these non-essential compounds are important for the plant's ability to survive. Secondary metabolites are compounds that are not directly involved in plant growth, development or reproduction but are derived from primary metabolic processes. Many secondary metabolites are plant defence compounds, produced to help the plant survive and fight off herbivores, pests and pathogens. Others are produced in order to attract pollinators and seed dispersers, producing attractive pigments in fruit and flowers (Figure 1.1).

Research on secondary metabolism in plants arguably started in 1806 when Friedrich Wilhelm Sertürner extracted morphine ('*principium somniferum*') from opium poppy. Since then, more than 200,000 structurally different compounds have been identified (most within the last 50 years), and more are being discovered every year [42]. Research into plant secondary metabolites has proven invaluable to society because some compounds have powerful medicinal applications and, others have important roles when consumed as part of a healthy diet.

Generally, secondary metabolites are divided into three major groups, phenolics, terpenoids, and alkaloids. The terpenoid family includes examples of both primary and secondary metabolites, derived from the five-carbon precursor isopentenyl diphosphate (IPP). Alkaloids, which all contain one or more nitrogen atom, are created primarily from amino acids. Phenolic compounds are formed primarily from the shikimic acid pathway but its intermediates may also come from the malonate pathway.

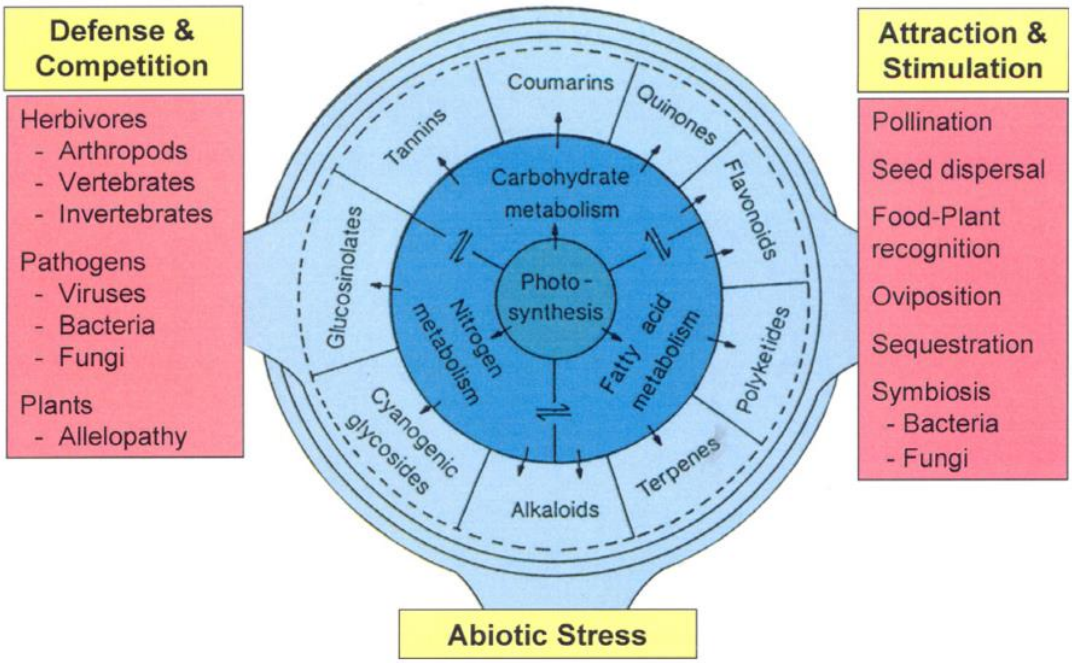


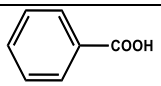
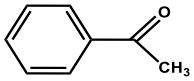
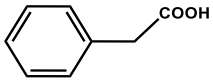
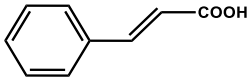
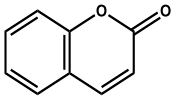
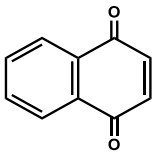
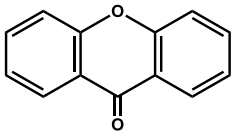
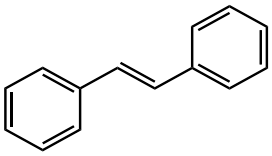
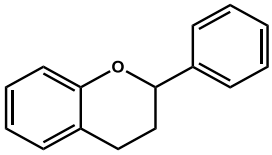
Figure 1.1 Reproduced from Hartman et al. (2007) Ecological functions of plant secondary metabolism, primary metabolites; dark blue, secondary metabolites; light blue [42].

1.2.1 Phenolic compounds

Phenolic compounds share a common feature, an aromatic ring with at least one hydroxyl subunit [43], and more than 8000 different phenolic structures have been described. These are widely dispersed throughout the plant kingdom [44]. Phenolics are diverse ranging from simple, low molecular weight, single aromatic-ring compounds to large and complex polymers such as tannins and derived polyphenols, and classification is determined by the number and arrangement of their carbon atoms (Table 1.1). These compounds are synthesised primarily from products of the shikimic acid pathway and have many important roles in plants. Some serve as defence compounds against herbivores and pathogens, such as tannins, lignans, and flavonoids [45]. Other compounds, for example flavonoids and anthocyanins, exhibit bright colours and are present to attract pollinators and seed dispersers. Lignin is another important product of phenylpropanoid metabolism, being a primary structural component and the second most abundant organic substance in plants. It mechanically strengthens and water-proofs cell walls and is a major component of xylem.

Phenolics occurring naturally in healthy plant tissue can be classified into two groups, the flavonoids and the non-flavonoids. 'Derived polyphenols' are the products of phenolic transformations and are found in processed food and beverages, such as black tea, red wine, coffee and cocoa [44]. Tannins are used to describe any large polymerised polyphenols that contain sufficient hydroxyls and other groups such as carboxyls to form complexes with various macromolecules.

Table 1.1 Basic structural skeletons of phenolic and polyphenolic compounds.

Skeleton	Classification	Basic Structure
C_6—C_1	Phenolic acids	
C_6—C_2	Acetophenones	
C_6— C_2	Phenylacetic acid	
C_6— C_3	Hydroxycinnamic acids	
C_6— C_3	Coumarins	
C_6— C_4	Naphthoquinones	
C_6— C_1— C_6	Xanthones	
C_6— C_2— C_6	Stilbenes	
C_6— C_3— C_6	Flavonoids	

1.2.2 Health benefits of phenolic compounds

Polyphenols are abundant phytonutrients in our diet, and evidence for their role in the prevention of degenerative diseases is emerging. Identification of individual compounds linked with prevention or amelioration of specific diseases is a relatively under-researched area. A few specific examples are highlighted.

Renaud and de Lorgeril,. (1993), studied the typical diets consumed across Europe in terms of calorie intake from dairy fat against incidence of mortality from coronary heart disease (CHD) present in each country. They found a linear relationship between calorie intake from dairy fat and the incidence of mortality from CHD. People in countries like Spain, Portugal and Italy who typically consumed a diet with a low proportion of calories coming from dairy fat, had less than half the mortality from CHD compared with countries with high dairy fat consumption such as Germany, the United Kingdom, and Scandinavian countries. The exception was France, with a relative calorie intake from dairy fat like the Nordic countries but a mortality rate from heart disease similar to those countries consuming Mediterranean style diets. This observation became known as the French Paradox and has been attributed to the consumption of moderate amounts of red wine by the French, particularly the health-promoting effects of the polyphenols in red wine [9].

An interesting study that demonstrated the health benefits of plant based foods was by Butelli et al. (2008), who showed that anthocyanins play a role in protection against cancer [3]. They showed that tomatoes genetically modified to accumulate increased anthocyanin levels helped to prolong the life of cancer prone mice. Food supplemented with the powder of the purple tomatoes together with food supplemented with the powder of standard, red tomatoes were fed to p53 knockout mice. p53 knockout mice develop soft tissue carcinoma at high frequency and live an average of 140 days. Mice fed a diet supplemented with powder from the purple tomatoes survived 30% longer compared to mice fed a diet supplemented with red tomatoes or on the un-supplemented standard diet [46].

Flavonols have been extensively studied because they contribute to the antioxidant properties of green vegetables, fruits, olives, soyabeans, red wine, chocolate and teas. Some flavonoids have been linked to an array of biological activities including antiviral, anti-proliferative and anti-carcinogenic properties [47]. The chlorogenic acids are linked to many health benefits, and are present in many plant-based foods. Coffee is the greatest source of CGA in the western diet but tomatoes, potatoes, eggplant, apples, carrots, plums, flax, just to name a few also contain bioavailable CGA [48, 49]. The consumption of chlorogenic acids is linked with the reduced incidence of many chronic diseases and due to high levels of consumption, has been well researched [36, 50, 51].

1.2.3 Phenylpropanoid pathway

The phenylpropanoid pathway is a secondary metabolic pathway present in all plants, it starts with the amino acid phenylalanine which is derived from the shikimate pathway. This remarkable pathway creates a vast number of secondary metabolites which are involved in a diverse range of biological functions.

The shikimate pathway uses Erythose-4-phosphate from the pentose phosphate pathway and phosphoenolpyruvate (PEP) from glycolysis in a seven step pathway [52]. Although phenylalanine is produced in the latter part of the pathway, intermediates from the pathway are used in branched pathways to create other secondary metabolites. Other amino acids produced by the shikimate pathway are the aromatic amino acids, tyrosine and tryptophan. Quinate is a product from a branch point of the shikimate pathway and shikimate is an intermediate (Figure 1.2).

The general phenylpropanoid pathway includes the initial three steps from phenylalanine catalysed by the enzymes phenylalanine ammonia lyase (PAL) (lyase, class I-like family), cinnamate 4-hydroxylase (C4H) (cytochrome P450 monoxygenase family), and 4-coumaroyl CoA ligase (4CL) (acyl CoA synthase family) the activity of which are required for the creation of most phenylpropanoids and resulting phenolic secondary metabolites. The product of the third step, *p*-coumaroyl CoA, can be classed as a central metabolite because it serves as a precursor for many different branches of the phenylpropanoid pathway (Figure 1.2). Several copies of the PAL-genes are found in most plant species, for example there are four genes present in Arabidopsis whereas in Solanaceous species an estimated 20 putative PAL-genes are encoded by the genome. In *Robusta* coffee, there are 4 copies of PAL, C4H, CHS, genes, 2 copies of CHI and C3'H and one copy of HQT and HCT [53]. Individual genes may respond differently to biotic and abiotic stresses due to the regulation of

promoter regions targeted by different stress-realted transcription factors [54]. Selected pathways using p-coumaroyl CoA as a precursor are described in the following sections.

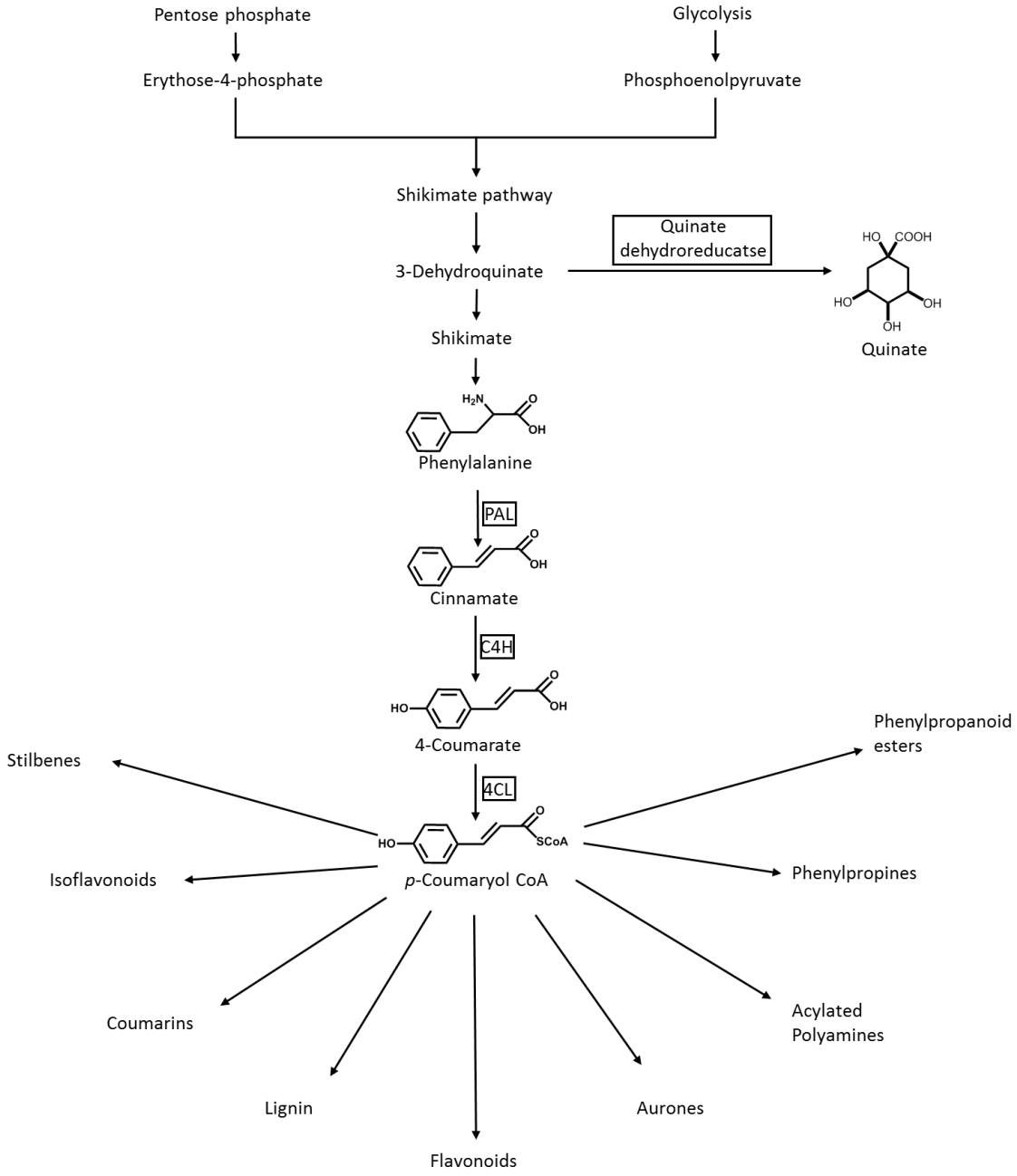


Figure 1.2 Schematic representation of the phenylpropanoid pathway showing the diversification of phenylpropanoids from the central metabolite, p-coumaroyl CoA. The pentose phosphate pathway and glycolysis feed compounds into the shikimate pathway which lead to the general phenylpropanoid pathway. The central metabolite, *p*-coumaroyl CoA is an important branch point for pathways leading to diverse phenylpropanoids. The intermediates and enzymes of the general phenylpropanoid pathway are shown: phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H), and 4-coumaroyl CoA ligase (4CL).

1.2.3.1 Flavonoids

Flavanoids as a group are complex and diverse, they all share a 15-carbon skeleton, C6-C3-C6 of which over 5000 different metabolites have been found to occur naturally in plants. Their roles within plants range from UV absorption, to signals promoting symbiotic nitrogen fixation and floral pigmentation. The synthesis of flavonols involves many enzymatic steps with many branches downstream of *p*-coumaroyl CoA to create this diverse range of polyphenols. A simplified description outlining the pathways leading to the major flavonoids is provided here.

Chalcone synthase (CHS) is ubiquitous in all higher plants and is known as a type III polyketide synthase enzyme which is able to perform a series of decarboxylation, condensation, and cyclisation reactions through the presence of two biding pockets in the enzyme [55]. CHS catalyses a range of reactions creating coumarins using *p*-courmaroyl CoA, but most commonly creates naringin chalcone through the addition of three of manoyl CoA molecules (derived from malic acid in the citric acid cycle) to *p*-courmaroyl CoA. Chalcone isomerase (CHI) then acts to isomerise naringenin chalcone to form naringenin. This reaction can also occur spontaneously, but the rate of isomerisation is enhanced by CHI activity. Naringenin then gets converted into dihydokaempferol by flavone-3-hydroxylase (F3H). Flavonoid-3'-hydroxylase (F3'H) and flavonoid-3'5'-hydroxylase (F3'5'H) convert naringenin to eriodictyol and pentahydroxyflavanone and dihydrokaempferol to dihydroquercertin and dihydromyricetin respectively. Flavonol synthase (FLS) can accommodate all three dihydroflavonols in its binding pocket to form kaempferol, quercetin and myricetin which are the major flavonols. Dihydrokaempferol, dihydroquercertin, and dihydromyricetin also serve as precursors for anthocyanin biosynthesis though the activity of a dioxygenase called anthocyanidin synthase to form pelargonidin, cyanidin and delphinidin respectively. Anthocyanins are stabilised by having a sugar added on the 3-position by a 3-glycosyl transferase. Further decoration by cytoplasmic glycosyl and acyl transferases may occur before the anthocyanins are transported to the vacuole by transporters facilitated by a glutathione S-transferase protein that facilitates transport by acting as a ligandin. Once in the vacuole, anthocyanins may be decorated further by vacuolar glycosyl and acyl transferases [56].

1.2.3.2 Chlorogenic acid biosynthesis

Several pathways have been proposed for the synthesis of caffeoyl quinic acids in different plant species. It is important to note that in the most researched model in plant biology, Arabidopsis, does not accumulate chlorogenic acid in any tissue, because it lacks a functional copy of the hydoxycinnamoyl-Coenzyme A: quinate hydroxycinnamoyl transferase (HQT) gene [57]. With other evidence supporting the importance of HQT in chlorogenic acid (CGA) biosynthesis, it is therefore considered likely that a plant that accumulates chlorogenic acids should carry at least one functional

copy of this gene. 5-caffeoylquinic acid is the principle caffeoyl quinic acid, other isomers and derivatives are thought to be formed from this primary compound.

Three pathways that have been proposed for 5-CQA biosynthesis. The first is direct, and involves caffeoyl-CoA transesterification with quinic acid by hydoxycinnamoyl-Coenzyme A: quinate hydroxycinnamoyl transferase (HQT) (route 1 in Figure 1.3). However, caffeoyl CoA needs to be synthesised by the earlier activity of HCT which forms p-coumaroyl shikimate from p-coumaroyl CoA and shikimate. This ester is hydroxylated by *p*-coumaryol-3'-hydroxylase (C3'H) to form caffeoyl shikimate which is then converted to shikimate and caffeoyl CoA by the reverse activity of HCT. The second is where *p*-coumaroyl-CoA is transesterified with quinic acid to create *p*-coumaroyl quinate by HQT which then undergoes hydroxylation, catalysed by the enzyme *p*-coumaryol-3'-hydroxylase (C3'H), to form 5-CQA (route 2 in Figure 1.3). Lastly, caffeoyl-glucoside has been reported to serve as an acyl donor in a reaction with quinic acid catalysed by hydroxycinnamoyl D-glucose:quinate hydroxycinnamoyl transferase (route 3 in Figure 1.3) [58]. Understanding the relative roles that the two hydroxycinnamoyltransferases play in 5-CQA biosynthesis is a major theme of this thesis.

Hydoxycinnamoyl-Coenzyme A: quinate hydroxycinnamoyl transferase (HQT) and hydoxycinnamoyl-Coenzyme A: shikimate/quinate hydroxycinnamoyl transferase (HCT) both belong to the clade V of the BAHD family of acyltransferases. These acyltransferases catalyse reactions that use CoA thoesters for the formation of plant metabolites. This family of enzymes was named according to the first letter of each of the first four biochemically characterised enzymes of the family [59]. Enzymes that belong to the BAHD family share several conserved domains. The first is the HXXXDG domain located near the central portion of each enzyme. The HXXXDG amino acids comprise part of the active site of BAHD enzymes. The second conserved domain, a DFGQG motif, is not present in all BAHD enzymes, but is found in most. The DFGQG motif is located near the carboxyl terminus and is thought to be involved in stabilising the enzymes [59]. Mutations induced in these conserved regions result in severely reduced enzyme activity [60], particularly those that affect the HXXXDG motif in the active site of the enzymes.

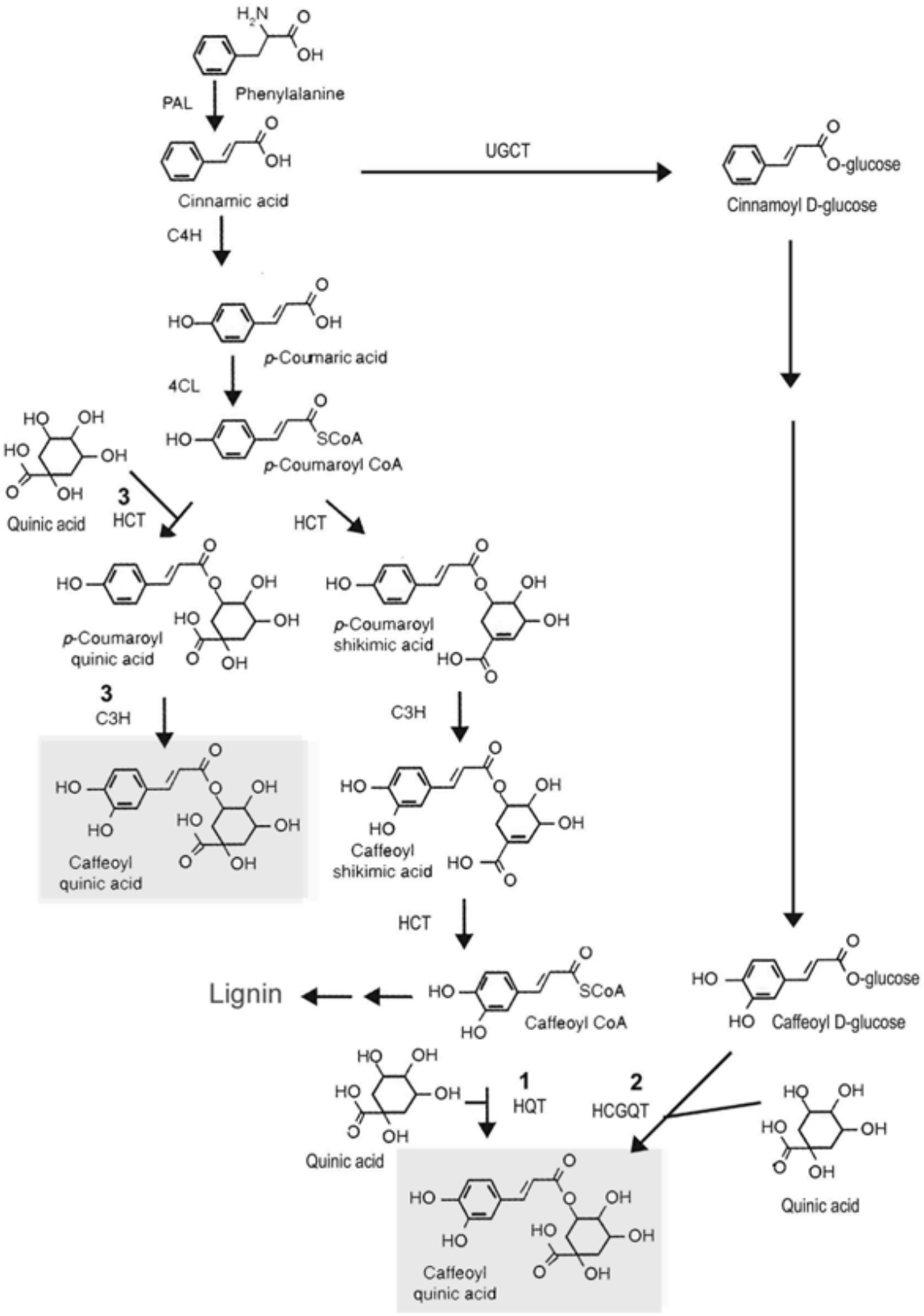


Figure 1.3 Reproduced from Niggeweg et al. (2004). Proposed pathways of chlorogenic acid biosynthesis in plants. The three different routes that have been proposed are labelled 1, 2, and 3. C4H, cinnamate 4-hydroxylase; 4CL, 4-hydroxycinnamoyl CoA ligase; PAL phenylalanine ammonia lyase; 3CH, *p*-coumarate 3'hydroxylase; HQT, hydoxycinnamoyl-Coenzyme A: quinate hydroxycinnamoyl transferase ;HCT, hydoxycinnamoyl-Coenzyme A: shikimate/quinate hydroxycinnamoyl transferase; UGCT, UDP glucose:cinnamate glycosyl transferase; HCGQT, hydroxycinnamoyl $_{D}$-glucose:quinate hydroxycinnamoyl transferase [57].

HCT and HQT are structurally very similar enzymes. Both can accept similar substrates in their active sites and they catalyse reactions involved in similar pathways; HCT in the biosynthesis of caffeoyl CoA for monolignol synthesis and HQT in the synthesis of hydroxycinnamoyl esters. HCT has evolved principally to catalyse reactions that involve shikimate as the acyl acceptor whereas HQTs use quinate as their acyl acceptor. These compounds are similar in structure; quinic acid has one extra carbon with an OH group on the carboxylic acid side chain compared to shikimate. Both HQT and HCT can use *p*-coumaroyl CoA and caffeoyl CoA as acyl donors with very similar affinities in freely reversible reactions [57, 61, 62]. *In vitro* enzyme assays have shown that HQT and HCT show substrate preferences for quinate and shikimate as acyl acceptors, respectively but each has been found to have low affinity for the other's primary acyl acceptor substrate. Manipulation of these genes *in vivo* has shown that HQT is involved predominantly in 5-CQA biosynthesis whereas HCT is involved in the formation of monolignols. A reduction in the expression of HQT in tomato affected levels of chlorogenic acid biosynthesis without affecting lignin accumulation [57]. Silencing of HCT in tobacco reduces lignin content and overall plant height, and causes the accumulation of more chlorogenic acids [63]. These reports suggest that in Solanaceous species, the main role of HQT is in chlorogenic acid biosynthesis, whereas HCT has its major role in monolignol biosynthesis. It remains to be seen if HCT has any role *in vivo* in 5-CQA synthesis or whether HQT has any role in monolignol synthesis. Within clade V BAHD acyltransferases, phylogenetic analysis shows that the enzymes HQT and HCT from different species cluster together (Figure 1.4). This shows that HCT and HCT are monophyletic and differences in their substrate specificities are likely to have been maintained over long evolutionary time scales.

The C3'H enzyme (CYP98A35) is one of two C3'H genes present in coffee but is the only allele which will hydroxylate *p*-coumaroyl quinate into 5-CQA preferentially, although both enzymes in coffee have been reported to hydroxylate *p*-coumaroyl shikimate to form caffeoyl shikimate [65]. These interactions have not been characterised as being reversible [66]. Chlorogenic acids are likely to localise in the vacuole where they are able to accumulate to high levels, because the BAHD activity of HQT is freely reversible [67]. The chlorogenic acid, 5-CQA, has been observed to form complexes with caffeine when it is in solution which is most likely to happen when both compounds are located in the vacuole [67]. The synthesis route for the other main group of chlorogenic acids, FQAs, occurs through the methylation of caffeoyl CoA to form feruloyl CoA which then donates its acyl group to quinate to form 5-feruloylquinic acid (5-FQA). The caffeoyl CoA methyltransferase (CCoAMT) methylates, and HQT probably performs the transesterification.

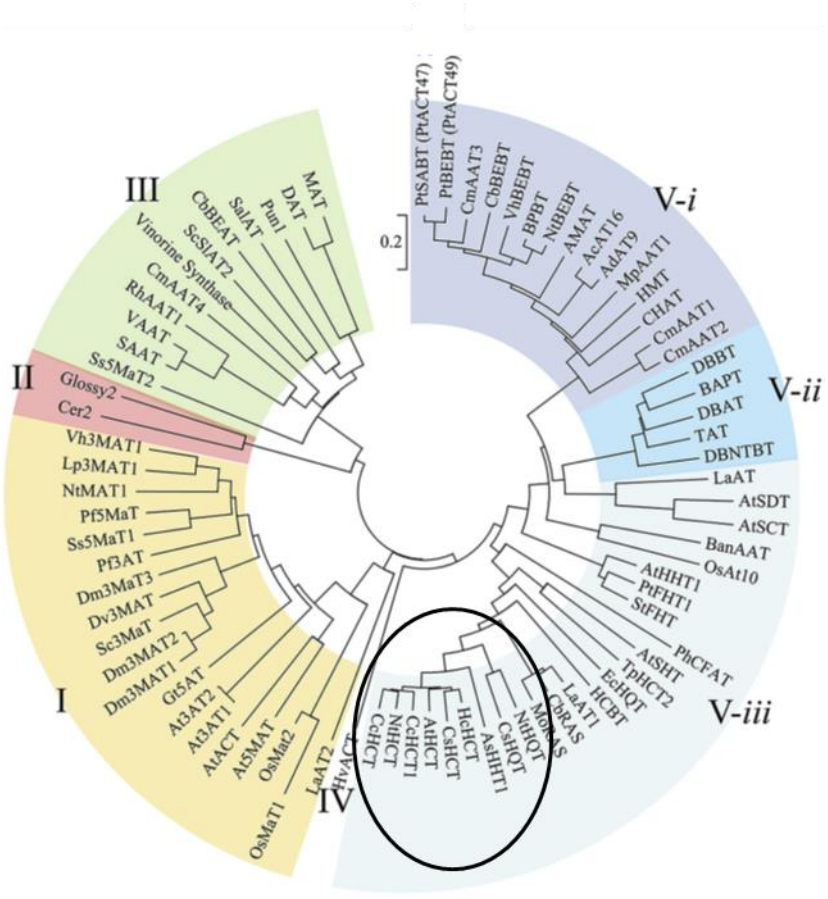


Figure 1.4 Reproduced from Chedgy et al,. (2015). Phylogenetic tree of BAHD acyltransferase family divided into different clades. The HCT and HQT enzymes from different species cluster together within clade V (circled black) [64].

1.2.3.3 Lignin

Second to cellulose, lignin is the most abundant polymer on earth. Lignin is almost completely built of phenylpropanoid units derived from the oxidative polymerisation of hydroxycinnamoyl alcohol derivatives. Due to the economic cost of removing lignin from wood for paper and biofuel formation, lignin biosynthesis and manipulation has been extensively researched [54]. Interestingly in terms of evolution, formation of lignin was unlikely to have been the primary role of phenylpropanoid metabolism. It has been suggested that the algal ancestors of early land plants evolved phenylpropanoid metabolism to provide UV-B tolerance as plants emerged onto the land. Mechanical reinforcement and waterproofing of cell walls for long distance conductance resulted from the evolution of lignin biosynthesis and allowed early land plants to grow upright and supported long distance water transportation [68].

Lignins are complex racemic aromatic heteropolymers derived mainly from three hydroxycinnamoyl alcohol monomers differing in their degree of methylation [69]. These are 4-coumaryl, coniferyl, and sinapyl alcohols. These monlignols produce *p*-hydroxyphenyl (H), guaiacyl (G) and syringyl (S) phenylpropane units that became incorporated into the lignin polymer. These units are found in different abundancies in different plant families [70]. The incorporation of these units into a lignin polymer via coupling reactions is called lignification. Compounds used in the biosynthesis of monolignols come from coumaroyl shikimate and caffeoyl shikimate hydroxycinnamic acid esters, which are structurally closely related to 5-CQA and feruloulquinic acids (Figure 1.5).

The shikimate shunt, which is catalysed by HCT, C3'H and HCT supports the formation of caffeoyl CoA (and sinapoyl CoA) and is used to synthesise G- and S- type monolignols. The shikimate shunt may have been recruited from hydroxycinnamate metabolism by duplication of a gene encoding HQT activity, to form HQT and HCT, followed by subfunctionalisation. This hypothesis would suggest that 5-CQA synthesis is an ancient pathway, which may have been lost in species such as Arabidopsis, an idea supported by the monophyletic nature of enzymes with HQT activity.

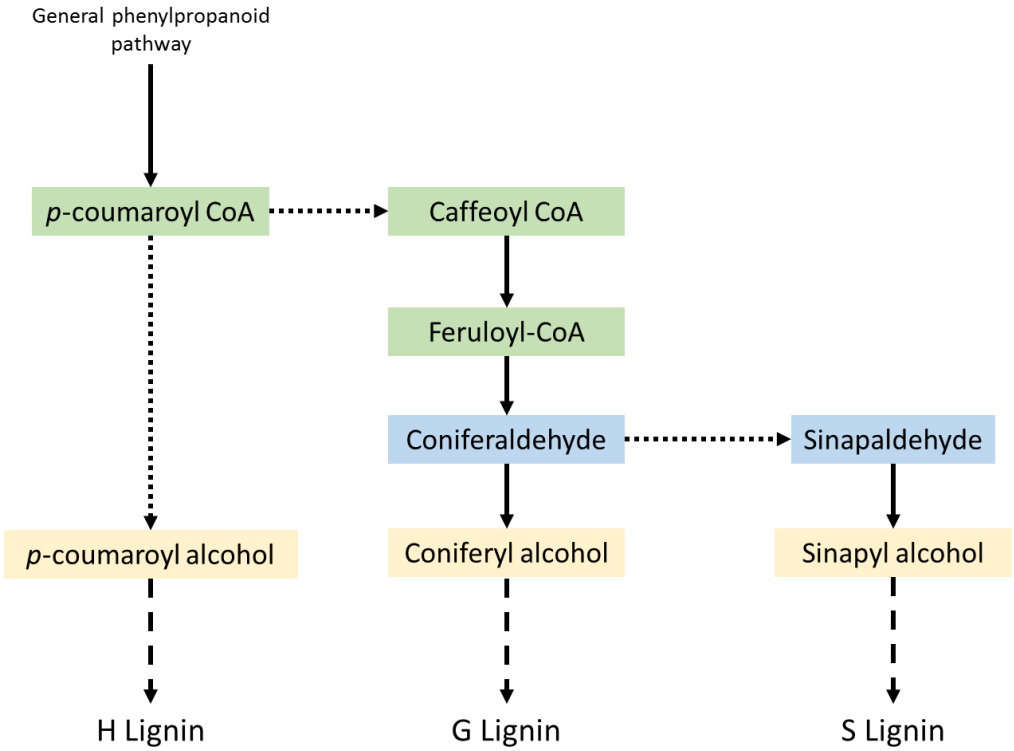


Figure 1.5 Monolignol biosynthesis pathway. The pathway highlights the precursors of H, G, and S lignin and the parts of chlorogenic acid biosynthesis from which they are derived. Dotted lines indicate pathways of multiple steps. Enzymes involved are omitted. Green represents compounds derived directly from chlorogenic and feruloylquinic acid biosynthesis. The *p*-coumaryl, coniferyl, and sinapyl alcohol precursors for lignin are shown in orange.

1.3 Transcriptional regulation

The control of gene expression underpins most of the phenotypic changes that occur during plant development and in response to environmental signals. RNA polymerase is the enzyme that produces the primary RNA transcripts, of which there are multiple types present in eukaryotes, responsible for the synthesis of different subsets of RNA. RNA polymerase I synthesises a pre-rRNA 45S (Svedberg units) which develops into 28S (25S in plants), 18S and 5.8S rRNAs, which form major components of the ribosomes in the cytoplasm. RNA polymerase II is responsible for the synthesis of mRNAs and most of the sn (small nuclear) RNA and microRNAs. RNA polymerase III synthesises tRNAs, rRNA 5S and other small RNAs which are found in the nucleus and cytosol. RNA polymerases IV and V have been found only in plants where they synthesise si (small interfering) RNA of which RNA polymerase V synthesises only siRNA involved in heterochromatin formation.

RNA polymerase II is the most studied type of RNA polymerase and due to the high level of control required for transcription of mRNA, a range of transcription factors are required for its binding and promotion of transcription. This combinatorial control allows for specific transcription factors to act upon the promoter region of a gene, recognising different regulatory elements, for the repression or activation of gene transcription.

Control of transcription of any particular nuclear gene is thought to involve modulating the rate of transcriptional initiation of that gene. RNA polymerase II (RP2) is brought to the start site of transcription by the action of general transcription factors (GTFs). The GTFs are the TATA-binding protein (TBP), TFIIB, TFIIE, TFIIF, and TFIIH. The TATA box in eukaryotes usually lies 25 bp upstream of the start of transcription, where transcription is usually initiated. An annotated flow diagram (Figure 1.6) illustrates the action of GTFs in the recruitment and promotion of gene transcription.

Proteins regulating transcription of genes are known as transcription factors (TFs). These recognise short DNA sequences known as cis elements in the genes they control. This cis-elements often lie upstream of the TATA box in the promoter of the gene, although they can also lie in the 5' Untranslated Region (5'-UTR) or in the introns. Cis-acting sequences may be bound by transcriptional activators (these are referred to as Upstream Activating Sequences; UAS) or by repressors. TFs can enhance or repress gene transcription through combinatorial control of the GTF basal machinery [71].

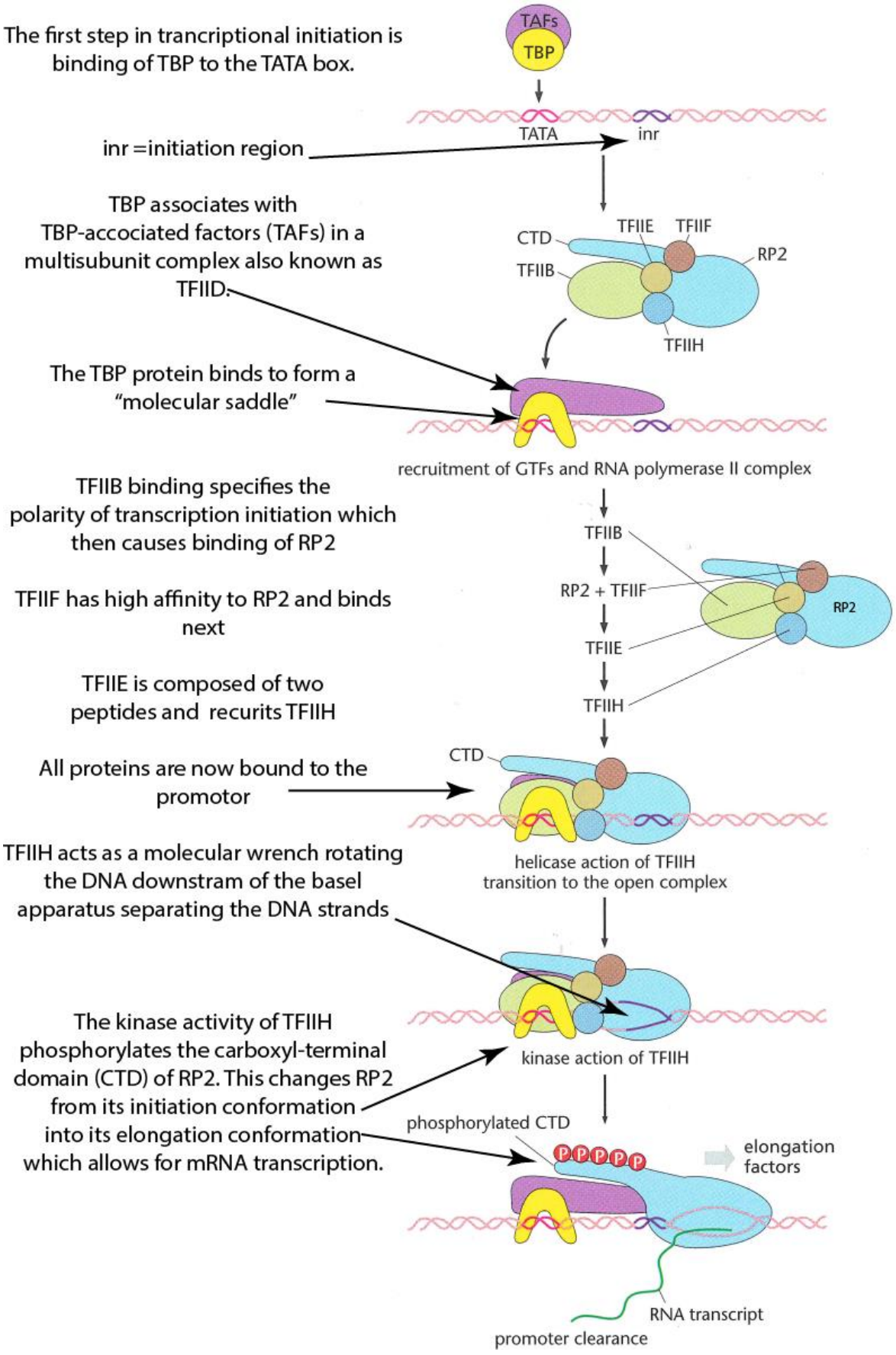


Figure 1.6. Adapted from Smith et al. (2010). Activities of the general transcription factors (GTFs). Annotated stepwise assembly of the pre-initiation complex on a promoter and transition to transcriptional elongation [71].

TFs are modular proteins and consist of a DNA binding domain and separate domains for influencing transcriptional initiation and for protein-protein interaction. DNA binding domains are highly conserved which allow TFs to be divided into families of proteins that contain structurally related DNA binding domains (Table 1.2). TFs that promote transcription often contain activation domains. These activation domains fall into categories which are influenced by their amino acid sequences, acidic domains, glutamine-rich domains and proline rich domains. These activation domains work by interacting with the basal transcriptional machinery directly, or indirectly through co-activators [72]. Transcription factors that act as repressors of gene transcription can work in a variety of ways, these are illustrated in the annotated diagram (Figure 1.7).

Generally, in eukaryotes, the specific expression pattern for any gene is defined by cell specific levels of transcriptional initiation which are controlled by the combined activity of multiple transcription factors which recognise different cis-elements within the gene's promoter region. Combinatorial control facilitates the complex regulatory networks found in higher eukaryotes [73]. The different degrees of activation and repression offered by different transcription factors recognising specific gene promoters define the level of transcriptional initiation; and are directly related to pre-mRNA transcript levels available for further processing. It is important to note that although transcriptional regulation through the production of pre-mRNA transcripts may be directly correlated to final protein levels, post transcriptional mechanisms exist that can also influence final protein levels substantially.

Table 1.2 Selectively reproduced from Smith et al. (2010). Examples of some transcription factor families in eukaryotes and their structural properties [71].

Family name	Properties of DNA binding domain	Properties of protein-protein interaction domains	Examples in plants
bHLH	Basic domain of conserved sequence	The helix-loop-helix domain interacts with other HLH domains to form symmetrical dimers composes of bindles of 4 α helices. Association allows DNA binding through the adjacent basic domain	Arabidopsis: PIF proteins in phytochrome signalling Maize: R/B proteins controlling anthocyanin biosynthesis.
bZIP	Basic domain of conserved sequence	Leucine zipper consisting of C-terminal α helix with a Leu or hydrophobic residue every 7th amino acid. α helices interact to form dimers.	Maize: Opaque2 regulator of seed storage protein synthesis.
MADS box	A 56-amino acid domain forming a pair of antiparallel coiled-coil α helices packed against an antiparallel 2 –stranded β sheet. The α helices make direct contact with DNA, and the N-terminal region of the domain contacts the DNA backbone.	Dimerization is essential for some members to bind DNA. Dimerization domains consist of β sheet and include the I and K domains of MICK-type MADS box proteins. Some MADS box proteins have C-terminal activation domains.	Majority of floral homeotic genes, including AP1, AP3, PI, AG, SEP, and FLC proteins of Arabidopsis.
MYB	Proteins containing 1 – 4 imperfect repeats of a 53/53 – amino acid motif that forms a helix-helix-turn helix structure. The 3rd helix is the recognition helix. 2- or 3- repeat MYBs (2R or 3R bind DNA as monomers. 1-repeat (1R) MYBs bind DNA as dimers and have district binding site specificities from 2R/3R MYBs.	2R and 3R MYBs may have C-terminal activation or repression domains. 1R MYBs may contain coiled-coli domains for dimerization/protein-protein interaction.	3R MYBs: regulators of cyclin gene expression. 2R MYBs: GL1, LAF1, AtMYB4, AtMYB12, PAP1, C1 of maize. 1R MYBs: Golden 2, uniform ripening in tomato (SlGLK 2) Late elongated hypocotyl (LHY) and Circadian clock associated (CCA1) from Arabidopsis.
Zinc finger	DNA binding domain contains 2 Cys and 2 His at fixed points, which together bind zinc. Two antiparallel β sheets with α helix in between bind DNA	TFs can contain several zinc finger domains which interact with proteins which effects transcription for combinatorial control.	Arabidopsis: DOF proteins involved in seed development.

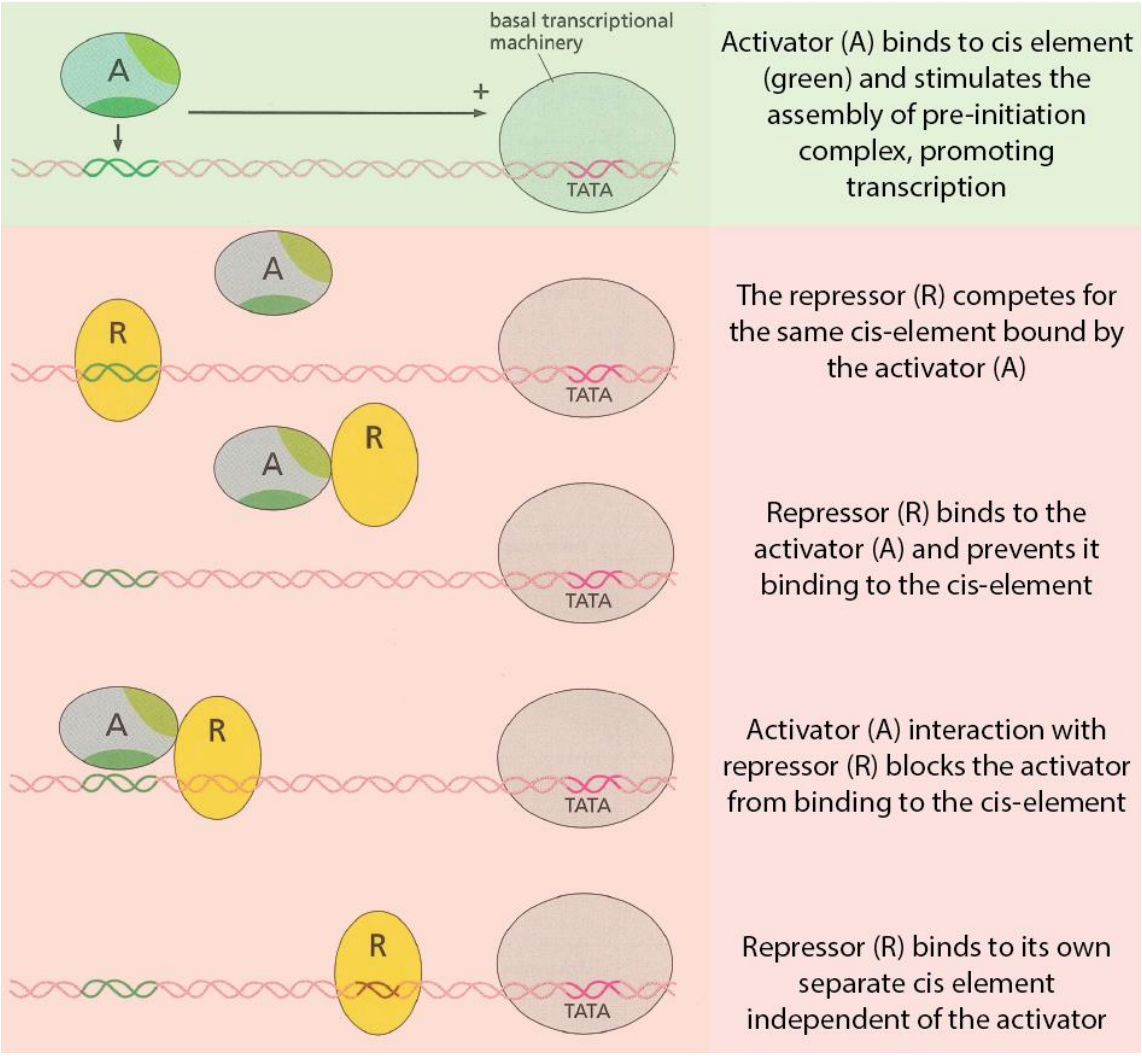


Figure 1.7. Adapted from Smith et al. (2010). Mode of action of transcriptional repressors. Area coloured green is activation of transcription. Area coloured red is repression of transcription [71].

1.3.1 MYB transcription factors

MYB transcription factors in plants play important roles in the regulation of many plant processes such as regulating branches of the phenylpropanoid pathway, tryptophan biosynthesis, control of cell fate and regulation of cell cycle. They can act on their own to promote or repress transcription or with co-activators and co-repressors to function. This transcription factor protein family is amplified in higher plants, particularly in the form of the plant-specific R2R3MYB family, which comprises 126 R2R3-type MYB genes in *Arabidopsis thaliana* [74].

MYB transcription factors exist in fungi, animals and plants and are characterised by the presence of an amino acid motif structurally and functionally related to the DNA-binding domain of the product of the retroviral oncogene v-*MYB* and its animal cellular equivalent c-*MYB* [75]. The term MYB came from the word myeloblastosis which is derived from the disease caused by the retroviral oncogene in chickens [54]. The prototypic MYB protein is c-MYB which has 3 DNA-binding repeats, called R1, R2 and R3. The MYB DNA binding domain consists of up to four imperfect amino acid sequence repeats called R; these are usually about 52 amino acids in length and form 3 α-helices [76]. The first, second and third α-helices of each repeat form a helix helix-turn-helix (HHTH) structure with three regularly spaced tryptophan (or hydrophobic) residues forming a hydrophobic core of the HHTH structure. The third helix in each repeat is the recognition helix which makes direct contact with DNA and intercalates in the major groove. There are four different classes of MYB genes characterised in plants with different numbers of adjacent R repeats; the largest family of MYB genes found in plants is called the R2R3 type and this family has two MYB repeats in its DNA binding domain. Structurally the most N-terminal repeat in R2R3MYB proteins is most similar to R2 in c-MYB, and the second repeat is most similar to R3 in c-MYB [77].

The R2R3MYB transcription factors in plants can be divided into 28 different subgroups depending on the amino acid sequences of the MYB DNA-binding domains and motifs present in their less highly conserved C terminal domains. The C-terminal domains of R2R3MYB proteins are not highly conserved in terms of primary amino acid sequences, even between functionally equivalent proteins, but may contain short conserved motifs that are linked to their function in transcriptional regulation. R2R3MYB proteins that belong to a specific subgroup, share strong similarity in their DNA-binding domains, implying that they bind the same DNA motifs. The C-terminal domain of R2R3MYB proteins may function as an activation domain, especially if regions of amphipathic alpha helix are present close to the C-terminus. Some R2R3MYB proteins have transcriptional repression domains in their C-terminal regions, such as the EAR motif in subgroup 4 R2R3MYB transcription factors [76, 78].

R2R3MYB transcription factors have evolved to bind to promoter regions independently or as a complex with a basic helix loop helix protein and WD-repeat (WDR proteins) referred as MBW complex. For those that interact with BHLH proteins, a conserved region is present in Arabidopsis MYB transcription factors which is the signature motif for interaction between MYB and BHLH proteins ([DE]Lx_2[RK]$x_3Lx_6Lx_3$R) [79]. This conserved signature lies in the R3 repeat of R2R3MYBs from positions 12 – 33 and contributes to the strength of interaction with the N-terminal region of a subgroup of BHLH proteins [79]. NMR structural analysis showed that this area is hydrophilic (solvent exposed), positioned on the surface opposite to the DNA binding site [80]. In angiosperms and probably also in gymnosperms, anthocyanin biosynthesis is initiated by the formation of the MBW complex. The R2R3MYB activator, the BHLH1 and the WDR protein form a (MBW) complex. Subgroup 6 includes R2R3MYB transcription factors that require formation of MBW complex in order to activate biosynthetic genes for anthocyanin biosynthesis. The *Anthirinum majus* R2R3MYB gene *Rosea1* requires the co-expression of the gene encoding the BHLH protein from *A.majus Delila* to induce anthocyanin biosynthesis in flowers of A.majus [81] and in tomato fruit [46]. Other R2R3MYB proteins do not need to form a complex and can activate biosynthetic genes independently, for example subgroup 7 R2R3MYB proteins. Which have differences in key residues in the motif in R3 preventing them from interacting with BHLH proteins.

Subgroup 4 contains genes that are related to repression of the phenylpropanoid pathway AtMYB4 and AtMYB32 repressed the phenylpropanoid pathway when overexpressed in Arabidopsis by binding to the promoter region of C4H causing a reduction in the accumulation of flavonols [82, 83]. This reduction in flavonols increased sensitivity to UV-B light when compared to WT. AmMYB308 and AmMYB330 overexpression caused similar phenotypes in tobacco to AtMYB4 overexpression [82, 84]. The repression of the phenylpropanoid pathway by these subgroup 4 TFs has a profound effect on the phenotype which involves shorter plants with smaller leaves with white lesions, quite distinct from WT. The overexpression of these subgroup 4 MYBs reduced the expression of cinnamate 4-hydroxylase (C4H) within the general phenylpropanoid pathway affecting all branches of phenylpropanoid biosynthesis. The tobacco plants overexpressing AtMYB4, AmMYB308 band AmMYB330 were stunted because the pathway leading to monolignol biosynthesis was repressed resulting in a reduced lignin content. The reason why white lesions appeared on the leaves was due to the lower accumulation of flavonols and chlorogenic acids for protection against UV light and the reduction in anthocyanin accumulation was responsible for the phenotype observed in the flowers [82, 84]. The strawberry FaMYB1, which is another subgroup 4 MYB protein, was overexpressed in tobacco but resulted in a different phenotype. FaMYB1 principally targets the promoters of anthocyanin synthase (ANS) and UDPglucose: glucosyl

transferase (GT) which are downstream enzymes involved in the synthesis of anthocyanins. The overexpression of FaMYB1 caused a reduction in anthocyanins and the flavonol quercetin, but the levels of kaempferol remained unchanged. The reduction in the activities of these downstream flavonol biosynthetic genes did not affect the transcription levels of the general phenylpropanoid pathway genes [85].

The subgroup 4 R2R3MYB transcription factors are all repressors which share a C2 motif (pfLNLD/ELxiG/S) with a core consensus sequences of LxLxL or DLNxxP [86, 87]. These sequences form the ethylene-response factor (ERF) – associated amphiphilic repression (EAR) motif, mutations within this domain cause reduced repression activity [87, 88]. Some MYB transcription factors can repress transcription passively, such as single-repeat R3MYBs repressors which do not contain an EAR motif but can repress TF activity by binding to BHLH proteins competitively with the R2R3MYB proteins that are required for the formation of the MBW complexes [89].

The subgroup 7 R2R3MYB Arabidopsis MYB12 protein is a transcriptional activator of phenylpropanoid biosynthesis. Mehrtens et al,. (2005) showed it to regulate expression of chalcone synthase and flavonol synthase which are genes encoding key enzymes in the synthesis of flavonols in plants [75]. Luo et al,. (2008) transformed tobacco (*Nicotiana tabacum)* with AtMYB12 controlled by the CaMV 35S promoter and tomato (*Solanum lycopersicum*) with the AtMYB12 gene under the control of the fruit-specific E8 promoter and assayed its effects on expression of different genes involved in phenylpropanoid metabolism [90]. The AtMYB12 gene functions differently in tomato to Arabidopsis. Whereas the overexpression in tomato caused AtMYB12 to bind to several PAL gene promoters, in Arabidopsis PAL transcript levels were not significantly increased following AtMYB12 overexpression [90-92]. This was likely to be the main reason why overexpression of AtMYB12 causes higher accumulation of flavonols in tomato than in Arabidopsis. When AtMYB12 was expressed in tomato fruit, nearly all the genes relating to flavonol and chlorogenic acid biosynthesis downstream of PAL were upregulated. AtMYB12 also bound to promoters of genes encoding enzymes of primary metabolism whose increased expression increased the supply of carbon to supply secondary metabolism with more aromatic amino acids [90, 92]. A range of phenolics were detected and measured in tomatoes overexpressing AtMYB12 compared to the wild-type and it was found that chlorogenic acids, monocaffeoylquinic acids (CQA), dicaffeoyl quinic acid (diCQAs), tricaffeoyl quinic acid (triCQAs), and quercetin rutinoside (QueRut), levels were significantly increased. Total antioxidant levels were measured in the transgenic tomatoes and antioxidant activity of the water-soluble fraction (containing phenolics) was shown to be increased by up to five-fold by expression of AtMYB12 in fruit. No significant differences in antioxidant capacity of the

lipophilic fraction between wild type and transgenic fruit were observed. This meant that the resulting orange colour phenotype was not due to the loss of lycopene, but due to the accumulation of flavonols [90]. Further research into AtMYB12 has identified specific promoter regions to which it binds to promote transcription of its downstream target genes. When overexpressed in tomato fruit ChIP-seq and ChIP-qPCR analyses revealed that AtMYB12 bound to the promoter regions of the two genes in intermediatory metabolism *ENO* and *DAHPS* and five promoters of genes involved in phenylpropanoid metabolism, *PAL5A, PAL5C, PAL5D, CHS-1,* and *F3H.* It is possible that AtMYB12 binds to more promoter regions of genes whose expression it enhances but sensitivity of ChIP-qPCR in tomato fruit could have limited further identification of direct targets. The ChIP-qPCR analysis also revealed the predicted binding motif of AtMYB12 (TACCTACC), which is likely to be a common recognition motif of all MYB12 homologues in other plant species [92].

R2R3MYBs which belong to the subgroup 9 regulate anthocyanin and flavonol biosynthesis and some have also been shown to influence chlorogenic acid levels when overexpressed in potato tubers. The StMtf1 R2R3 transcription factor was modified by replacing two amino acids in the serine rich N-terminal domain which increased its activity [93]. Serine residues at the N-terminus are phosphorylated which discriminates between different transcription factor responses [94]. StMf1 was mutated, from NH_2-MNSSP to NH_2-MNSTS, and both were overexpressed in potato plants. The mutated StMf1 ($StMf1^M$) caused the accumulation of anthocyanins in the stems to increase by more than 10 fold compared to overexpression of the WT version [93]. All above ground tissues showed anthocyanin pigmentation, from $StMf1^M$ overexpression, whereas the analysis of the resulting tubers showed that they did not accumulate high levels of anthocyanins, producing a mottled appearance of purple [93]. However, analysis of the other compounds which accumulated in the tubers showed that 5-CQA and 4-CQA levels were increased fourfold. The analysis of transcript levels of the HQT gene showed that it had been upregulated by $StMf1^M$ overexpression. These authors also found that the overexpression of $StMf1^M$ resulted in the upregulation of the upstream phenylpropanoid biosynthetic genes chorismate mutase and prephenate dehydratase [93]. This showed that the $StMf1^M$ which activates anthocyanin accumulation, might also influence different branches of the phenylpropanoid pathway.

The primary protein structure and biological function of most MYB proteins within a given subgroup implies their descent from a common ancestral gene. However, their action on different structural genes of different phenylpropanoid branches is sometimes diverged.

The central roles that MYB transcription factors play in the regulation of genes encoding enzymes of phenylpropanoid metabolism underline the importance of focusing research efforts on identifying the functionality of genes encoding R2R3MYB TFs when wishing to breed or engineer levels of specific phenylpropanoids in crops like coffee.

1.4 Aims of the project

I aimed to characterise the biochemical mechanisms that exist for the synthesis of different CQAs in both types of commercial coffee, *Arabica* and *Robusta*. I investigated the relative roles that HQT and HCT play in chlorogenic acid biosynthesis in tomato and coffee. Spatial localisation of phenolic compounds in developing coffee beans suggested different roles for different enzymes which localise to different sub-cellular compartments. In addition, I investigated dicaffeoylquinic acid synthesis and identified two different mechanisms for their formation. I provide evidence for why coffee does not accumulate high levels of flavonols despite the expression of transcription factors which cause high levels of flavonol accumulation in another plant species. I cloned and characterised- functionally candidate transcription factors that may regulate the phenylpropanoid pathway in both types of coffee and so identified likely regulators of 5-CQA accumulation in coffee.

Chapter 2: General materials and methods

2.1 Materials

2.1.1 Chemicals

All chemicals used were of molecular biology grade and were obtained from Thermo Scientific, Invitrogen, Sigma Aldrich, BioRad, New England Biolabs, Roche, Promega, and Qiagen. For HPLC and LCMS, column was obtained from phenomenex which was Luna 3 μm C18 100 mm x 2.0 mm and chemicals were HPLC grade and purchased from Fisher Scientific Ltd.

2.1.2 Antibiotics

For selection of bacteria and transgenic plants, ampicillin, carbenicillin, chloramphenicol, gentamycin, kanamycin, rifampicin, spectinomycin, streptomycin, and cefotaxime were used. Working concentrations are listed in Table 2.1.

Table 2.1 Working concentrations of antibiotics

Antibiotic	Solvent	Purpose	Final working concentration
Ampicillin	Water	*E.coli* selection	100 μg/μl
Carbenicillin	Water	*E.coli* selection	100 μg/μl
Chloroamphenicol	Ethanol	Gateway cloning	50 μg/μl
Gentamycin	Water	Gateway cloning	100 μg/μl
Kanamycin	Water	Transgenic plants selection, *E.coli* selection, *Agrobacterium* selection	100 μg/μl
Rifampicin	Ethanol	*Agrobacterium tumefaciencs* selection	50 μg/μl
Spectinomycin	Water	*E.coli* selection	100 μg/μl
Streptomycin	Water	*E.coli* selection	100 μg/μl
Cefotaxime	Water	Kills *Agrobacterium*	100 μg/μl

2.1.3 Plant Material

Coffea arabica and *Coffea robusta* plants were obtained from Nestlé R&D centre, Tours, France. *Solanum lycopersicum* variety MoneyMaker was edited to delete part of the HQT gene by Cas 9 by Tom Lawrenson, John Innes Centre, Norwich, for use in this thesis. *Nicotiana tabacum* variety Samsun was used for all tobacco transformations and was obtained from lab stocks.

2.1.4 Bacterial and Yeast Strains

A list of bacterial and yeast strains and their use in this thesis are listed in Table 2.2.

Table 2.2 Bacterial and yeast strains used and their function

Strain	Type	Use
LBA4404	*Agrobacterium tumefaciencs*	Stable tobacco transformation
GV3101	*Agrobacterium tumefaciencs*	Transient tomato transformation
ATTC 15834	*Agrobacterium rhizogenes*	Production of hairy roots
DH5α	*E.coli*	General *E.coli* for growth and storage of plasmids
DB3.1	*E.coli*	Growth of plasmids containing the *ccd* gene which is lethal to DH5α; used with gateway cloning.
Rosetta II	*E.coli*	Production of recombinant protein
Artic Express (DE3)	*E.coli*	Production of recombinant protein
BL21 (DE3) (pLys)	*E.coli*	Production of recombinant protein
Y182	*Saccharomyces cerevisiae*	Yeast 1 hybrid assay

2.1.5 List of plasmids used

A table of plasmids used are listed in Table 2.3.

Table 2.3 Table of plasmids used

Plasmid name	Size (bp)	Selective antibiotic	Use	Source
pGEM-T	3003	Ampicillin	T/A cloning site	Promega
pDONR207	5585	Gentamycin, Chloramphenicol	Gateway entry vector	Invitrogen
pDEST17	6354	Ampicillin	Production of recombinant protein in *E.coli*	Invitrogen
pB7RWG2	11614	Spectinomycin, Chloramphenicol	N-terminal fusion mRFP tag for localisation assays	Invitrogen
pDEST22	8930	Ampicillin	Yeast 1 hybrid transcription factor plasmid	Invitrogen
pHIS Leu 2	7200	Kanamycin	Yeast 1 hybrid promoter plasmid	Lab stocks
pBIN19	14992	Kanamycin	Gateway compatible constitutive expression (2x35S) of genes of interest	Lab stocks
pK7WG2R	13188	Spectomycin, Kanamycin	Gateway compatible 35S promoter, dsRED expression used in production of hairy roots.	Invitrogen

2.2 Methods

2.2.1 Primer design

Primers were designed using online software Primer 3 plus, (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) according to the following output parameters: the optimal length of oligo primers were from 21 to 28 bp (excluding primers designed for Gateway™ cloning). The melting temperatures (Tm) of primers were restricted to 58 – 68 °C with a GC content from 35 – 65 % with complementary sequences avoided. As G – C binding has 3 interaction points compared to the 2 between A – T, I tried to terminate with a G or C at the 3' end. qPCR primers were designed to have an amplicon ranging from 80 – 300 bp (determined by the primers used). A list of primers used can be found in Appendix 1.

2.2.2 Polymerase Chain reactions (PCR)

PCR reactions were performed using either G-Storm Thermal Cycler (Kapa Biosystems) or Mastercycler® pro (Eppendorf). The reaction mixture for taq polymerase usually consisted of 10-100 ng of template DNA, 0.5 μM of each forward and reverse primer, 100 μM of dNTPs, 1 mM Tris HCl, 50 mM KCl, and 1.5 mM and 1 x concentration of polymerase. When phusion polymerase was used the buffer composition was 25 mM TAPS-HCl (pH 9.3 @ 25°C), 50 mM KCl, 2 mM MgCl2, 1 mM β-mercaptoethanol, 10-100 ng of template DNA, 0.5 μM of each forward and reverse primer, 100 μM of dNTPs, and 1 x conventration of recombinant phusion polymerase.

A standard program was: initial denaturation, 1 minute at 95 °C, followed by 25-40 cycles of denaturation, 30 seconds at 95 °C, annealing, 30 seconds at 3 °C lower than the Tm temperature of the primer, and extension, 30 seconds at 72 °C, final extension was 5 minutes at 72 °C.

2.2.3 Purification of DNA from agarose gels

The PCR product was either used directly in a PCR clean up or it was loaded onto a 1% agarose gel in Tris Borate EDTA (89 μM Tris, 89 μM Boric acid, and 2mM EDTA at pH 8) buffer stained with ethidium bromide (0.15 mg/ml). The desired fragments were identified by running a 2 – log DNA alongside provided by NEB. Fragments were excised from the gel by visualising on a large wavelength UV transilluminator and cutting out with a clean scalpel. The gel fragment was placed into a 1.5 ml Eppendorf tube and melted into buffer QG (5.5 M guanidine thiocyanate, GuSCN, 20 mM Tris HCl pH 6.6) at 50 °C until gel has dissolved. 1 volume of isopropanol was added to the sample and then centrifuged at 17900 x g through the column membrane. The column was then

washed with buffer PB (10 mM Tris-HCl pH 7.5 and 80% ethanol) by adding 700 μl of the buffer and centrifuging at 17900 x g for a minute. The purified gel fragment was eluted from the column through the addition of 50 μl of water and centrifuging at 17900 x g for a minute to collect.

2.2.4 Purification of DNA from PCR reactions

The purification of PCR products was performed using the PCR QIAquick PCR Purification Kit (Qiagen) per the manufacturer's guidelines. 5 x the volume of buffer PB (5 M Gu-HCl, 30% isopropanol) was added to the PCR product before centrifuging through the column membrane at 17900 x g for a minute. The column was washed with buffer PB (10 mM Tris-HCl pH 7.5 and 80% ethanol) through the addition of 700 μl and centrifuging the column at 17900 x g for a minute. The purified PCR product was eluted with the addition of 50 μl of water and centrifuging at 17900 x g to collect.

2.2.5 *E.coli* Competent Cell for Heat Shock Transformation

A single colony of *E.coli* was grown in 10 ml LB at 37 ^{o}C, aerated at 220 rpm overnight. An aliquot of the overnight cultures (1 ml) was used to inoculate 100ml LB and the culture was grown until it reached a density of OD_{600} ~0.4. At this point the cells were centrifuged at 4000 x g for 5 minutes at 4 ^{o}C and washed three times in 0.1 M $CaCl_2$. The cell were placed on ice for 30 minutes before the final centrifugation. The cells were resuspended in 4 ml of 0.05 M $CaCl_2$ and 20 % glycerol, and divided into 100 μl aliquots, frozen in liquid nitrogen and stored at -80 ^{o}C. Competent *E.coli* cells can be stored at -80 ^{o}C for several months.

2.2.6 Heat shock transformation method for *E.coli* transformation

Heat shock was used to transfer desired plasmids into *E.coli*. The competent cells were thawed on ice and up to 100 ng of plasmid was added. Mixtures were kept on ice for 30 minutes and then heat shocked at 42 ^{o}C for 60 seconds. *E.coli* was placed back on ice for 1 – 2 minutes then 950 μl of room temperature SOC medium was added and the cells were placed in a shaking incubator at 37 ^{o}C for 90 minutes. The cells were spun at 4000 x g for 4 minutes and resuspended in 100 μl SOC medium and spread on LB agar plates with appropriate selective antibiotics. The plates were incubated at 37 ^{o}C overnight, and colonies usually started to appear after 16 hours.

2.2.7 Plasmid DNA isolation from *E.coli*

Plasmid DNA was isolated from 5-10 ml of *E.coli* cultures grown overnight with antibiotic selection. The QIAprep® Miniprep Kit (Qiagen) was used to isolate plasmid DNA according to manufacturer's instructions. Pelleted bacteria was resuspended in 250 μl buffer P1 (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 100 μg/ml RNaseA) and then 250 μl of buffer P2 (200 mM NaOH, 1% SDS) was added and left for a few minutes. 350 μl of buffer N3 4.2 M Gu-HCl, 0.9 M potassium acetate, pH 4.8) was then added and the mixture was centrifuged at 17900 x g for 10 minutes, the supernatant was the added to the column and then centrifuged again at 17900 x g for a minute to bind the plasmid to the membrane. The membrane was then washed with 700 μl buffer PB (10 mM Tris-HCl pH 7.5 and 80% ethanol) by centrifuging at 17900 x g and the eluted with 50 μl of water.

2.2.8 Quantification of DNA/RNA

The concentration of DNA and RNA was determined using NanoDrop 2000C UV-Vis Spectrophotometer (Thermo), following suppliers instructions. The nanodrop can measure the absorbance spectrum of 1 μl of a sample. DNA was considered pure if it had an absorbance ratio A260/280 of ~1.8. The absorbance ratio of RNA A260/A280 was considered pure if it was ~2.

2.2.9 Preparation of electrocompetent *Agrobacterium* (GV3101, ATTC 15834, AgL1, LBA4404)

Agrobacterium from a single colony was grown in TY (ATTC 15834) or LB media (GV3101, AgL1, LBA4404), supplemented with appropriate antibiotics at 28 °C until the OD_{600} was above 1.5 (24-48 hours). Overnight cultures (10 ml) were used to inoculate 100 ml of TY or LB with appropriate antibiotics at 28 °C until OD_{600} reached 0.5 – 0.8. Cells were cooled on ice and then centrifuged at 2700 x g for 10 minutes, pellets were resuspended in cold water. This wash step was repeated twice, with final resuspension in 800 μl 10% glycerol. *Agrobacterium* was divided into 40 μl aliquots, frozen in liquid nitrogen and stored at -80 °C. Electrocompetent *Agrobacterium* could be stored for many months at -80 °C.

2.2.10 Electroporetion of *Agrobacterium*

Agrobacterium was thawed on ice and 10 – 300 ng of plasmid DNA was added and was kept on ice for between 30 – 60 minutes. The mixture was transferred to a pre chilled electroporation cuvette and electrocuted using BioRad Pulser (Biorad Laboroatories) at 400 Ω, 25 μFD and 2.5kV.

Afterwards, 900 μl of SOC were added and cells were recovered at 28 °C with shaking, for 2 to 4 hours. Cells were centrifuged at 4000 x g for 4 minutes, resuspended in 100 μl SOC and plated on LB/TY agar plates with appropriate antibiotics.

2.2.11 DNA isolation from plants

DNA was isolated from plant tissue using the DNeasy plant mini kit (Qiagen) according to the manufactures protocol. Plant tissue was ground using a mortar and pestle and about 100 mg of powder was added to 400 μl of buffer AP1 (10 mM Tris-HCl pH8.0, 1 mM EDTA pH 8.0, 0.1% SDS, 0.1M NaCl, 1X PVP, 10mM DTT, RNase) and incubated at 65 °C for 10 minutes. 130 μl of buffer P3 was then added and the mixture was incubated on ice for 5 minutes. All of the mixture was then centrifuged in the shredder column for 2 minutes at 20000 x g. The supernatant was transferred to a new tube and 1.5 volumes of AW1 was added. The mixture was centrifuged through the DNA binding membrane for 1 minute at 6000 x g and washed twice with 500 μl of buffer AW2. The DNA was eluted from the column through the addition of 100 μl buffer AE, (10 mM Tris-HCl, 0.5 mM EDTA, pH 9.0) and centrifuging at 6000 x g for 1 minute.

2.2.12 RNA isolation form plant tissue

RNA was isolated from plant tissue using the RNeasy plant mini kit (Qiagen) according to the manufactures protocol. RNA quality and quantity was analysed by Nanodrop. Autoclaved ball Barings were added to an Eppendorf tube and 100 mg of plant material was added and frozen in liquid nitrogen. 100 mg of tissue was disrupted by using a liquid nitrogen chilled Qiagen tissue lyser for 2 x 1 minute cycles, tissue lyser was chilled with liquid nitrogen between cycles.

Buffer RLT (450 μl) supplemented with 10 μl of β-mercaptoethanol per 1 ml, was added to the frozen tissue then vortexed and centrifuged in the shredder tube for 2 minutes at 17900 x g. The supernatant was added to 0.5 the volume of ethanol and the liquid was centrifuged through the RNA binding column for 15 seconds at 8000 x g. The membrane was then washed with 700 μl of buffer RW1 and then twice with 500 μl buffer RPE. Total RNA was eluted by using 50 μl of water and centrifuging at 8000 x g for 1 minute.

2.2.13 First strand cDNA synthesis

First strand cDNA was created using SuperScript™ III (Invitrogen). Total RNA (2 μg) in 18 μl water was mixed with 1 μl primer mix (0.25 μg/μl oligo dT (Promega) plus 0.25 μg/μl random primer (Invitrogen)), and 1 μl 10 mM dNTPs and was incubated at 65 °C for 5 minutes then placed on ice for 5 minutes. 5X first stand reaction buffer (6 μl), 2 μl 100mM DTT, 1 μl of ribonuclease inhibitor (RNaseOUT, Invitrogen) was added to inhibit the degradation of the RNA, and 1 μl SuperScript™ III were added to the reaction mixture and incubated for 1 hour at 50 °C. The reaction was terminated by a 5 minute incubation at 85 °C to denature the enzymes, cDNA was stored at -20 °C.

2.2.14 Gateway cloning

PCR products amplified for Gateway cassettes were purified using either QIAquick Gel Purification Kit or QIAquick PCR purification kit, (QIAGEN). The DNA with attBI and attB2 sites at each end were then cloned into the pDONOR 207 entry vector using BP clonase™ (Invitrogen). pDONOR 207 (100 ng) were mixed with at least 25 ng DNA fragment and 1 μl of BP clonase™ enzyme in a total volume of 5 μl and incubated overnight at 25 °C. Proteinase K (1μl) was added and incubated for 10 minutes at 37 °C, to stop the reaction. The whole mixture was then added to competent DH5α *E.coli* cells for transformation. *E.coli* were grown on plates, then in liquid culture, using 100 μm gentamycin as selection antibiotic and then the plasmid was extracted. Inserts were confirmed using sequencing (primers found in Appendix 1).

LR clonase™ (Invitrogen) was used to inset DNA of interest into destination vectors, 100 ng of pDONOR 207 vector and 100 ng of destination vector were mixed with 0.5 μl of LR clonase™ in a total volume of 2.5 μl. Reactions were stored at 25 °C overnight and then stopped by the addition of 1μl of proteinase K solution and incubated at 37 °C for 10 minutes. The mixture was added to competent DH5α for heat shock transformation.

2.2.15 Statistics

Unless otherwise stated, paired or unpaired, two-tailed Student's t-tests were used to compare group differences throughout this thesis. p values less than 0.05 were considered significant.

Chapter 3: Mono- and di-caffeoylquinic acid biosynthesis in coffee and tomato

3.1 Introduction

Consumption of plant-based foods is important for maintaining a healthy lifestyle, and their consumption is intrinsically linked to the reduction of chronic disease [95, 96]. Research over recent decades has shown the importance of the antioxidants present in plant-based foods and linked to the consumption of different antioxidant compounds to reducing the risk of specific chronic diseases. The phenolics are the most widespread dietary antioxidants and compounds like the chlorogenic acids have important roles in promoting health. Coffee is the biggest source of chlorogenic acids in the western diet and the diversity of chlorogenic acids within the beverage links moderate coffee consumption to the reduced risk of many chronic diseases [34, 36-38, 97]. The monomer 5-caffeoylquinic acid, is the most abundant chlorogenic acid in coffee and is linked to the inhibition of body fat accumulation in mice [40]. 3,4-dicaffeoylquinic acid (3,4-diCQA) possesses antihepatotoxic activity [98] and 1,5-dicaffeoylquinic acid (1,5-diCQA) is a potent inhibitor of HIV-1 integrase [99]. Dicaffeoylquinates are much less abundant in plant foods than the monocaffeoylquinates but their antioxidant capacity is generally greater because of the larger number of hydroxyl groups present on the aromatic rings. Understanding the biosynthesis of dicaffeoylquinates could offer important opportunities for the development of plant foods with increased phytonutrient content with potential therapeutic applications.

Different studies have suggested three possible routes for 5-caffeoylquinic acid biosynthesis (5-CQA), described in the general introduction, some of which might play more predominant roles in different plant species than others [57, 100-102]. Fewer studies have investigated the biosynthesis of diCQAs, but those that have been done suggest that two similar enzymes with different roles in plants, namely hydroxycinnamoyl-CoA quinate hydroxycinnamoyltransferase (HQT) and hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase (HCT) may be responsible for diCQA synthesis. The enzymes HCT and HQT belong to the BAHD superfamily, named according to the first letter of the first four acyltransferase enzymes of the family isolated from plant species [59]. The first report of 3,5-dicaffeoylquinic acid biosynthesis (isochlorogenic acid) was made in sweet potato through an unidentified-enzyme reaction in which 5-CQA acts both as acyl donor and acceptor and this activity was named chlrogenate:chlorogenate transferase activity (CCT) [58].

The enzyme hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase (HCT) principle responsibilities are, catalysing the reaction between *p*-coumaroyl-CoenzymeA and shikimic acid to form *p*-coumaroyl shikimate, and the reaction between caffeoyl CoenzymeA and shikimic acid to form caffeoyl shikimate. The HCT enzyme has been found to catalyse the same reactions using quinic acid as a substrate but shows far greater affinity for shikimic acid than for

quinic acid *in vitro* [101, 103]. HCT activity is intrinsically linked to monolignol accumulation. In Arabidopsis, silencing of HCT results in severe dwarfing of plants [104]. *In vitro* analysis of the enzyme activity of HCT from coffee suggested that this enzyme is also responsible for 3,5-dicaffeoylquinic acid biosynthesis through activity with 5-CQA and Coenzyme A at a neutral pH. A histidine residue at position 154 which is involved in substrate binding to Coenzyme A (as shown by crystal structure docking analysis) was mutated to an asparagine, and showed an increase in catalytic activity for diCQA synthesis using the same substrates [101]. Docking studies of 3,5-diCQA to coffee HCT revealed that the substrate binding pocket is sufficiently large to accommodate this diester, therefore demonstrating coffee HCT could be the enzyme responsible for diCQA biosynthesis [101]. It was not clear from this study what the reaction mechanism involved in the synthesis of diCQAs might be. The mutation of a residue which had been shown to involve binding to Coenzyme A, strangely increased catalytic activity in the synthesis of diCQAs, where, Coenzyme A was supposedly needed as a substrate [101]. The authors did not propose a reaction mechanism for the activity observed, but, from their findings, diCQA synthesis is unlikely to occur directly through the binding of Coenzyme A and 5-CQA where, supposedly, 5-CQA acts as an acyl donor as well as an acyl acceptor, mediated by Coenzyme A interaction within the binding pocket. A more likely route for diCQA biosynthesis would be for the reverse BAHD reaction to occur first using 5-CQA and Coenzyme A, to produce caffeoyl CoA, followed by subsequent catalytic activity where caffeoyl CoA and 5-CQA can be used as substrates. This reaction pathway is plausible because both CQAs and caffeoyl CoA were present in the enzyme assays as products (Figure 3.1) [101]. In the second step of this proposed mechanism, caffeoyl CoA would act as the acyl donor and 5-CQA would act as the acyl acceptor. This would also support the fact that a mutation in a residue which is important for binding CoA increases the synthesis of diCQA because the mutation would limit the interaction of free CoA with the enzyme. Free CoA supposedly acts as a competitive inhibitor of diCQA synthesis, and the induced mutation reduces the ability of free CoA to act as a competitive inhibitor, thus making more binding pockets available to interact with caffeoyl CoA and 5-CQA in the synthesis of diCQAs [101].

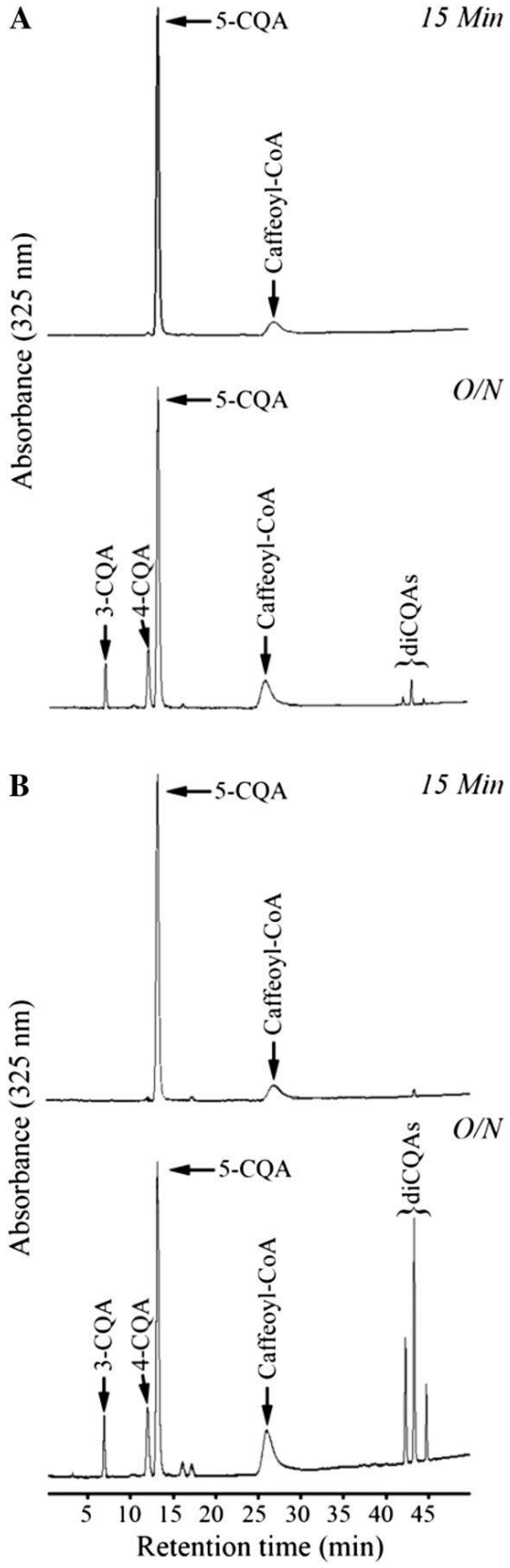


Figure 3.1. Recombinant CcHCT protein incubated with 5-CQA and CoA produce diCQA and caffeoyl CoA. Reproduced from Lallemand et al, 2012. A, native CcHCT reaction activity. B, His-154-Asn mutant CcHCT reaction activity. Assays were analysed after 15 minutes or overnight (*O/N*). DiCQAs were detected in samples analysed after an overnight incubation. His-154-Asn HCT produced more diCQAs after the overnight reaction [102].

The enzyme hydroxycinnamoyl-CoA quinate hydroxycinnamoyltransferase (HQT) has a primary role in catalysing the condensation of *p*-coumaroyl-CoenzymeA and quinic acid forming *p*-coumaroyl quinate and the condensation of caffeoyl CoenzymeA with quinic acid to form 5-caffoeylquinic acid (5-CQA). The HQT enzyme has also been found to catalyse the same reactions using shikimate as a substrate instead of quinate but shows substantially greater affinity for quinic acid *in vitro* [57, 103]. Overexpression of HQT in tomato plants resulted in increased levels of chlorogenic acids [57]. Recently, the HQT enzyme from tomato was shown to interact with 5-CQA *in vivo* where it can convert this single substrate at low pH (4-5) into dicaffeoylquinic acids (diCQA) (a chlrogenate:chlorogenate transferase reaction). To support the biological relevance of this low affinity reaction, subcellular localisation assays showed that the enzyme localises to both the cytoplasm and vacuole of onion epidermal cells, where a neutral and acidic pH, exist respectively [102]. The importance of His-276, present in HQT of Solanaceous species but not in other species, to the chlrogenate:chlorgenate transferase activity (CCT) of HQT, was demonstrated through mutating this residue to tyrosine which reduced, but did not eliminate, the catalytic activity observed. This dual catalytic activity of the HQT enzyme was suggested to allow it to moonlight in the synthesis of both mono- and dicaffeoylquinic acids when present in different cellular compartments.

In other studies, neither HQT nor HCT have been found to have activity in diCQA synthesis. The proteins encoded by the genes from globe artichoke (*Cynara cardunculus*) were assayed for enzyme activity by incubating recombinant protein with 5-CQA and caffeoyl CoA and also the hypothetical reverse reaction using diCQA with CoA at neutral pH, but no activity was detected with these substrates, suggesting globe artichoke might synthesise diCQAs, via a different enzyme [105].

I investigated the chlorogenate:chlorogenate transferase (CCT) activity of HCT and HQT from coffee *in vitro*. This activity could lead to the synthesis of diCQAs. I also investigated alternative catalysis by these enzymes. My data suggest that when localised to different sub-cellular compartments, both HQT and HCT from coffee could lead to the production of diCQAs from 5-CQA. I also characterised an alternative pathway for diCQA synthesis which I have called caffeoyl CoA chlorogenate transferase (CaAGT) activity, which could occur in the same subcellular compartment as the BAHD activity.

3.1.1 Aims

Conflicting evidence has been presented for dicaffeoylquinic acid biosynthesis in plants. I aimed to characterise the ability of the HQT and HCT enzymes from coffee to synthesise dicaffeoylquinic acids. *A priori* evidence suggested that diCQAs are derived directly from 5-CQA through a second acyl-CoA in coffee although starting from CoA and 5-CQA [101] and in tomato through CCT activity of HQT directly from 5-CQA [102]. Due to the structural similarity of these enzymes it is possible that both routes of diCQA synthesis exist in coffee. To examine this, I produced recombinant protein from both genes from coffee and tomato and assayed their activities in tandem using substrates and conditions previously described. To link *in vitro* assays to *in vivo* activity I partially resolved enzyme kinetic activity and localised both enzymes in subcellular compartments. In addition, I assessed the importance of tomato HQT to the biosynthesis of chlorogenic acid *in vivo* and drew conclusions about the functional similarity between HQT proteins from coffee and tomato.

These research objectives arouse from a publication which I was involved with at the beginning of my PhD. We reported a CCT activity of tomato HQT in the synthesis of diCQAs. A copy of the article can be found in Appendix 5 [102].

3.2 Materials and methods

To characterise different enzyme activities, I produced recombinant proteins in *E.coli* and tested catalytic activity in different environments and with different substrates. I then attempted to explore the biological relevance of the different activities of these enzymes.

3.2.1 Construction of vectors to express recombinant protein in *E.coli*

Gateway™ primers were used to amplify CcHCT, CcHQT, SlHCT, SlHQT and GFP (control), cloned in the pGEM T vector and the cDNA's were inserted into pDONR207 using the Gateway™ BP reaction. Gateway™ LR reactions were used to insert the genes into the destination vector, pDEST17. More detail about Gateway™ can be found in section 2.2.14. The pDEST17 vector contains a T7 promoter and a 6XHis N terminal tag for producing 6xHIS tagged protein in *E.coli* strains adapted for production of recombinant proteins (Figure 3.2).

3.2.2 Production of recombinant protein for assays of crude extract

Rosetta™ II (Novagen) *E.coli* is based on the BL21 strain of *E.coli* but is adapted to contain rare tRNA codons for the expression of eukaryotic genes, ensuring recombinant protein production without the need for codon optimisation.

The vectors were transformed into Rosetta II *E.coli* cells and grown in 10 ml LB media with appropriate antibiotics for selection, to an OD_{600} of 0.4 – 0.6. IPTG was added to induce protein expression to a final concentration of 1 mM and cultures were grown overnight at 25 ^{o}C. Cells were pelleted by centrifugation 2500 x g for 5 minutes and resuspended in 2 ml 1 x PBS with 1 x concentration of EDTA-free protease inhibitor cocktail (ROCHE) (Chymotrypsin 1.5 μl/ml, Thermolysin 0.8 μg/ml, Papaln 1 mg/ml, Pronase 1.5 μg/ml, Pancreatic Extract 15 μg/ml and Trypsin 0.2 μg/ml). Cultures were frozen in liquid nitrogen and thawed at 37 ^{o}C 3 times then sonicated for 5 x 30 seconds using a Soniprep 150 plus (MBE). Cell lysates were spun in a micro centrifuge at 17900 x g for 15 minutes and the supernatant was collected as crude protein extract. The crude protein could be stored at -80 ^{o}C for at least a month in 40 % glycerol.

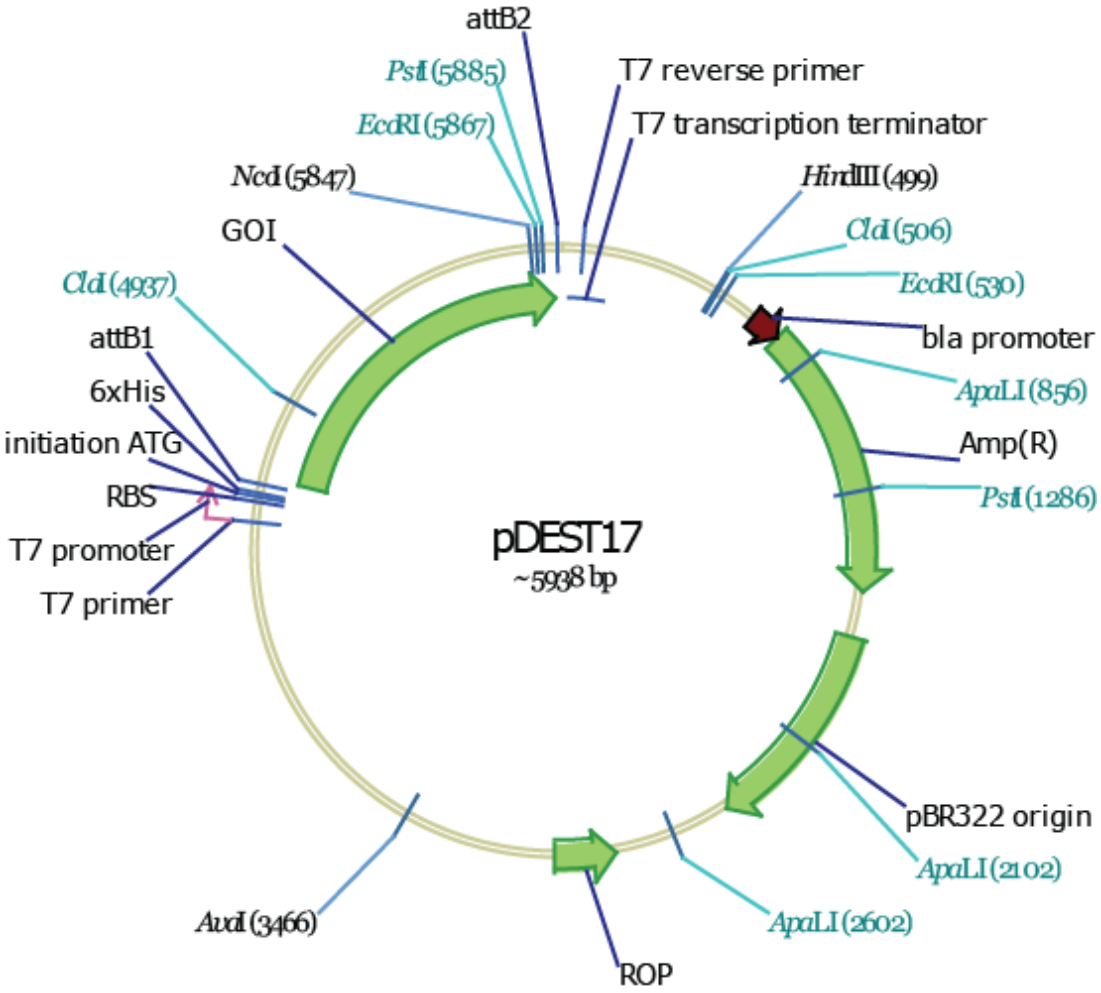


Figure 3.2. Vector map of pDEST17 from Vector NTI (Thermofisher). Vector was used to produce recombinant protein with a 6x N-terminal histidine tag. GOI, gene of interest; Amp(R), ampicillin selection gene; T7 promoter, binding site of the T7 polymerase.

3.2.2.1 *Western blot analysis for production of 6 X HIS tagged protein in crude extracts*

Crude protein extract was added to loading buffer (120 mM Tris HCl, 2% LDS, 10% glycerol, 0.51 mM EDTA, 0.22 mM SERVA Blue G, 0.175 mM Phenol Red, 8.5 pH) (Thermo) and boiled for 3 – 4 minutes and then loaded onto a NuPAGE 10 % Bis-Tris Gel (Life Technologies) together with prestained protein markers broad range (NEB). The buffer used was 1 x NuPAGE MOPS SDS running buffer (50 mM MOPS, 50 mMTris base, 0.1% SDS, 1 mM EDTA, pH 7.7) and the gel was run at 180 v until the loading buffer reached the bottom of the gel. Proteins were transferred to nitrocellulose membranes by electrophoresis, 100 v for 1 hour, in transfer buffer (25 mM Tris-HCl pH 7.5, 192 mM glycine, and 20 % methanol). To inhibit the degredation of proteins through the build-up of heat, the pack was kept cool by immersion in ice.

Successful protein transfer to the membrane was confirmed by staining with 1 x Ponceau solution (1 g/l Ponceau S, 5 % acetic acid) for 1 minute then washing in ddH_2O. The membrane was blocked with 5 % milk in TBST (10mM Tris HCl, 15mM NaCl, 0.05% Tween® 20 at pH7.5) for at least 1 hour and then incubated with the primary antibody, Anti-HIS diluted 1:1000 overnight at 4 ^oC in 5 % milk powder in TBST.

The next day, the membrane was washed 3 times in TBST for 10 mins each wash, then incubated with the secondary antibody, HRP (horseradish peroxidase) conjugated, 1:10000 dilution in 5 % milk with TBST for 1 hour. The membrane was then washed twice in TBST, for 10 minutes and antibody bound to the protein was detected with a ECL kit (SuperSignal® West Dura Trial Kit, Thermo Scientific), by mixing 500 µl of each solution together and then adding this to the membrane surface and leaving for 1 minute. An ECL (enhanced chemiluminescence) kit was used to detect the enzyme activity of the horseradish peroxidase bound to the secondary antibody and in turn, the protein of interest. The membrane was exposed to film for various lengths of time to detect the chemiluminescence and then developed.

3.2.3 Purified protein using ÄKTA purification system

After enzyme activity had been detected using crude protein extracts from *E.coli*, purified enzyme was needed to carry out enzyme kinetic analysis and to eliminate the possibility of cofactors interfering with the reaction. I therefore scaled-up the *E.coli* cultures and optimised the expression systems to increase recombinant enzyme production.

3.2.3.1 Purified CcHCT recombinant protein

A single colony of Rosetta II (Novagen) *E.coli* transformed with the pDEST17 plasmid containing the native CcHCT gene was grown in liquid LB (100 ml) with chloramphenicol (35 mg/l) and carbenicillin (100 mg/l) and grown overnight at 28 °C. The next day the starter culture was used to inoculate 4 l of LB culture pre-warmed to 37 °C supplemented with the same antibiotics. The culture was grown until the OD_{600} reached 0.6 – 1.0, at which point the culture was placed in a cold room (4 °C) to lower the temperature to below 30 °C. ITPG (1 mM final concentration) was then added to the culture to initialise the production of T7 polymerase to transcribe the gene of interest. The flasks were kept in a shaking incubator at 180 rpm at 30 °C overnight. The *E.coli* culture was then centrifuged at 6000 x g for 15 minutes and the supernatant was discarded, the *E.coli* pellet was frozen in liquid nitrogen and stored at -80 °C until further use.

3.2.3.2 Purified CcHQT recombinant protein

Artic Express *E.coli* were used to express recombinant CcHQT protein because this *E.coli* strain has been engineered to produce protein at low temperatures and synthesises cold-adapted chaperone proteins (Cpn20 and Cpn60) to assist with the folding of proteins that form insoluble inclusion bodies when produced in other strains of *E.coli*. This *E.coli* strain was utilised for the production of recombinant CcHQT because the protein was difficult to express in other *E.coli* expression strains.

A single colony of Artic Express (Aligent) *E.coli* was transformed with the pDEST17 plasmid containing the native CcHQT gene and grown overnight at 37 °C in liquid LB (200 ml) with chloramphenicol (35 mg/l) and carbenicillin (100 mg/l) placed in a shaking incubator at 220 rpm. The next morning, the culture (200 ml) was used to inoculate LB medium (10 l) prewarmed to 30 °C with no antibiotics in a shaking incubator at 220 rpm. After 3 hours of shaking, the culture was placed into the cold room (4 °C) to bring the temperature down to below 10 °C at which point IPTG (1 mM final concentration) was added. The culture was placed in a shaking incubator at 10 °C for 24 hours then the culture was centrifuged at 6000 x g for 15 minutes and the supernatant discarded. The *E.coli* pellet was frozen in liquid nitrogen and stored at -80°C until further use.

3.2.3.3 ÄKTAxpress purification of proteins

E.coli pellets containing recombinant protein were resuspended in 100 ml buffer A1 (50 mM Tris-HCl pH 8.0, 50 mM glycine, 5 % (v/v) glycerol, 500 mM sodium chloride, 20 mM imidazole, protease inhibitor cocktail (Roche)). The solution was sonicated using vibra-cell 750 Watt ultrasonic processor (SONICS) using 40 % intensity with 1 second on, 2 seconds off sonication pulses for 5 minutes of sonication 'on' time. The suspension was placed on ice to prevent the suspension from

heating up. The suspension was then centrifuged at 20,000 x g for 30 minutes at 4 ^{o}C to pellet the cell debris. The centrifugation step was repeated if the lysate was not clear of cell debris.

Clarified lysate was loaded onto a 5 ml Nickel (II) (Ni^{2+}) charged HisTrap FF column (GE Healthcare) which was pre-equilibrated with buffer A1 using the automated ÄKTAxpress (GE Healthcare) purification system. The system washed the column with 10 volumes of buffer A1 to elute unbound proteins. B1 buffer (50 mM Tris-HCl pH 8.0, 50 mM glycine, 5 % (v/v) glycerol, 500 mM sodium chloride, 500 mM imidazole) was then added to elute the protein of interest bound to the Nickel column (flow rate 3 ml/min). Eluate was loaded immediately onto a Hi-Load 26/60 Superdex 75 prep grade gel filtration column (GE Healthcare) to purify the protein of interest by size exclusion chromatography. The Superdex gel filtration column had been pre-equilibrated with buffer A4 (50 mM HEPES, 150 mM sodium chloride, pH 7.5) and the flow rate of the column was 3 ml/min. The absorbance at 280 nm was continuously monitored and fractions spanning peaks were analysed by SDS-PAGE for the presence of the purified protein of interest. The fractions that contained the purified enzyme of interest (after visualising on SDS-PAGE, Figure 3.3) were pooled and concentrated using a 30,000 kDa cut off concentration spin tube (ThermoFisher) by centrifuging at 4000 x g for 5 – 10 minutes until the protein was concentrated to the desired volume. Glycerol was added to the purified protein (final concentration 20%) and they were frozen in liquid nitrogen and stored at -80^{o}C until further use. The concentration of the protein was measured using a Peirce^{TM} BCA protein assay Kit (ThermoFisher) per the manufacturer's guidelines. The absorbance (562 nm) of the standards and the samples was measured using a nanodrop (described in general materials and methods, 2.2.8). For each sample the average of 5 reads was used to increase accuracy.

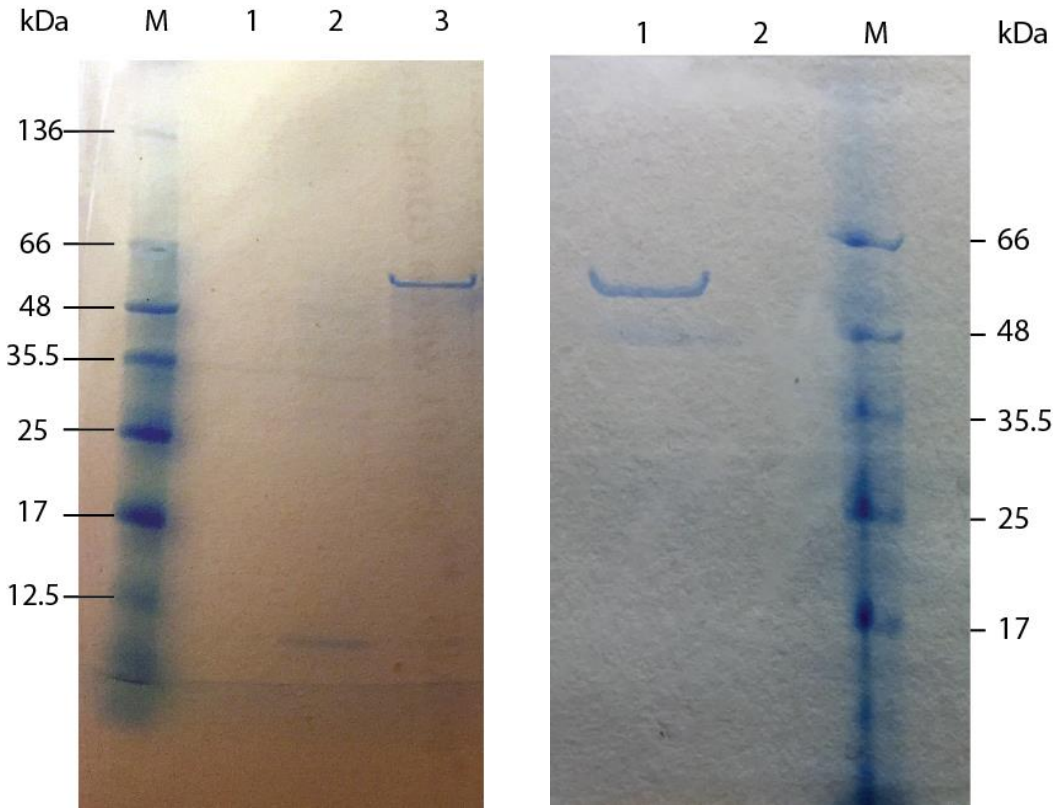


Figure 3.3. SDS PAGE gels of purified recombinant CcHCT and CcHQT enzymes. HQT Left. 1 and 2 are fractions before purified protein, present in 3, from gel filtration. **HCT Right**. 1 showed purified protein, 2, fraction eluted after purified protein. Both gels showed a single band of about 54 kDa which corresponded to CcHQT and CcHCT each with a 6 x His tag used for purification.

3.2.4 Enzyme assays

The reaction time for the enzyme assays was 90 minutes, the reaction was stopped by passing the solution through a microcentrifuge filter tube (0.22 μm) at 8000 x g for 1 minute to remove the protein. The sample was then transferred to a sampler vial and analysed immediately by LC MS. The CcHCT recombinant purified protein was at a concentration of 295 μg/ml and the CcHQT purified protein was at a concentration of 567 μg/ml. Plasmids for expressing mutated versions of CcHCT and CcHQT were received from Andrew McCarthy of EMBL Grenoble, France which were over expressed in *E.coli*, the crude protein extracts were used for the same reactions as the native proteins. These results can be found in Appendix 3.

3.2.4.1 Caffeoyl CoA chloroGenate Transferase (CaAGT) activity enzyme assay (pH 7)

Crude protein for initial assays and purified enzyme for subsequent assays were used to measure enzyme activities. The assays were undertaken in a total volume of 100 μl and contained sodium phosphate buffer 100 mM pH 7.0 and EDTA 1 mM, caffeoyl CoA 60 μM, 5-caffeoyl quinic acid 120 μM. Crude protein extract or purified enzyme was added to start the reactions (40 μl).

3.2.4.2 Chlorogenate:chlorogenate transferase (CCT) enzyme assay (pH 5)

Crude protein for initial assays and purified enzyme for subsequent assays were used for enzyme activity measurements. The assays were done in a total of 100 μl and contained sodium phosphate buffer 100 mM pH 5, DTT 1 mM, $MgCl_2$ 2.5 mM, EDTA 1 mM and varying concentrations of 5-caffeoylquinic acid from 1 mM to 20 mM. Crude protein extract or purified enzyme were added to start the reactions (40 μl).

3.2.4.3 Enzyme kinetics of CCT activity

I used purified enzyme of CcHQT and CcHCT to analyse the enzyme kinetic activity for the CCT activity. An initial time course was performed using the previously described conditions for the CCT enzyme assay where the single substrate 5-CQA (5 mM) was used to explore when the reaction rate was at its fastest. As it had been reported previously that diCQA production through CCT activity is slow, samples were taken every 20 minutes [102]. I discovered that the initial reaction rate of both enzymes slowed after 120 minutes.

I then varied the substrate concentration (1 – 25 mM) and used 60 μm/ml concentration of purified enzyme and stopped all the reactions at 120 minutes and measured the concentration of diCQA. I fitted the values obtained using the Michaelis Menten equation to work out the enzymes respective

K_m and V_{max}. Due to time constraints, only one replicate was performed for each CCT activity, future studies will need to include more replicates for improved reliability.

3.2.5 Liquid chromatography mass spectroscopy identification of dichlorogenic acids

For the identification of phenolic compounds liquid chromatography mass spectrometry (LC MS) was used because of its increased sensitivity compared to high performance liquid chromatography (HPLC). A Shimadzu IT-ToF (time of flight) system was chosen for the analysis because the IT-ToF mass spectrometer can detect the fragmentation of compounds insuring accurate identification.

For the quantification of diCQA in the enzyme assays the mobile phase consisted of formic acid (0.1 %, solution A) running against increasing concentrations of acetonitrile (100 %, Solution B) in a binary gradient. The run method for initial identification assays for each sample was 25 minutes and the total mass flow was 300 µl/min (details of which can be found in Table 1.1). A Luna 3 µm C18 100 mm x 2.0 mm (Phenomenex) column was used for this run method. For enzyme assays for quantification of diCQA for enzyme kinetics, the run method was made shorter and the column was changed to a Kinetex 2.6 µm EVO C18 100 Å 50 x 2.1 mm (Phenomenex) column. The flow rate was 600 µl/min in a binary gradient lasting 10.3 minutes, other parameters were kept the same. The run method can be found in Table 1.2.

Table 3.1 Run method for the initial enzyme assays for accurate detection of compounds

Time (mins)	**Solution A (%)**	**Solution B (%)**	**Column oven temperature (°C)**
0	97	3	40
5	97	3	40
15	20	80	40
17	20	80	40
18	97	3	40
25	0 (Stop)	0 (Stop)	40

Table 3.2 Run method used for the study of the enzyme kinetic activity

Time (mins)	Solution A (%)	Solution B (%)	Column oven temperature (°C)
0	97	3	40
2.50	97	3	40
7.50	20	80	40
8.50	20	80	40
8.60	97	3	40
10.30	0 (Stop)	0 (Stop)	40

Every sample (5μl) was injected at the start of the run and the standards were dissolved in a solution that was not too dissimilar to the sample solution to replicate run characteristics, i.e stock solutions of standards in methanol (80 %) were diluted with water opposed to methanol. For the creation of standard curves for quantification, 6 dilutions were used which were predicted to span the range of unknown concentrations.

Compounds were measured by their absorbance in the UV and visible wavelengths spanning from 200 nm to 600 nm using a photo diode array (PDA) detector model PDS-M20A sampling at 12.5 Hz. The UV spectra of peaks of interest were checked against standards and data were extracted from 325 nm which is the maximal absorbance of chlorogenic acids. Mass spectrometry was performed in positive mode [ESI^+] and negative mode [ESI^-] with a CDL temperature of 250 °C, nebulising gas flow (nitrogen) of 1.5 l/min and the heat block at 300 °C. MS^1 and MS^2 of the mass spec ([ESI^+] and [ESI^-] respectively) acquired data for the full run time of the program at m/z (mass charge) 200 to 2000, every 125 msec. Before each run, mass spec data were calibrated, measuring the m/z of a standard sample solution and correcting the data accordingly through the creation of a tuning file for each dataset.

A run of standard samples showed that chlorogenic acids ionise better when the MS is running in negative mode [ESI^-], small quantities of the sample were not always detectable when the MS was running in positive mode [ESI^+], and therefore the detection of synthesised compounds in the enzyme assays might only be detected in [ESI^-]. The details of standard compounds used for enzyme assays and their fragmentation in [ESI^+] and [ESI^-] when run in the MS IT-ToF can be found in Table 3.3.

Table 3.3 Fragments detected of standard compounds in different modes of mass spectrometry. Brackets represent heavy isotope peaks for fragments detected.

Standard compound	ESI^+	ESI^-
5-caffoeyl quinic acid (5-CQA)	377.083 (378.086) (379.088)	353.091 (354.094) (355.096)
	215.050	215.057
3,5-dicaffoeyl quinic acid (3,5-diCQA)	499.123 (499.316) (499.447)	499.236
	539.114 (540.117) (541.118)	515.124 (516.126) (517.128)
	377.083 (378.086) (379.088)	353.091 (354.094) (355.096)
	215.050	215.057
caffeoyl Coenzyme A	309.133 (310.136) (311.139)	530.673 (530.875) (531.008)
	291.118 (293.136) (295.134)	718.552 (719.559)
	203.139 (204.143)	299.150 (300.160)

3.2.6 Subcellular localisation of enzymes

Coffee HQT and HCT cDNAs were amplified using primers for Gateway™ recombination, (a list of primers used can be found in Appendix 1) and recombined into Entry vector pDONR207. The resulting clones were recombined into pB7RWG2.0, an expression vector with a 35S promoter and a modified red fluorescent protein (mRFP) C-terminal fusion tag. The control vector was cut with *Eco*RV and *Spe*I to remove the Gateway destination cassette, filled with the Klenow fragment of DNA polymerase and self-ligated to yield 35S:mRFP. mRFP was used as the fluorescent tag because its stability at lower pH makes it suitable for subcellular localisation studies in vacuoles.

The plasmids were introduced into epidermal membrane peels of onion skin using biolistics. Bombardment was conducted using a particle inflow helium gun. Onion epidermal tissue was isolated from the inner tissue of fresh, large onions and placed on 0.8 % MS agar plates with the lower side facing up. Plasmid DNA (5 μg) was precipitated onto 2 mg gold particles (1.0 μm diameter) through the addition of 50 μl of 2.5 M $CaCl_2$ and 20 μl of 100 mM spermidine and left on ice for 15 minutes. After the precipitation, all liquid was removed and gold particles were washed twice in ethanol and finally resuspended in ethanol (50 μl). This was enough for 5 separate bombardment assays. The Petri dish containing the onion epidermal tissue was placed in the vacuum chamber 160 mm away from the particle gun. The gold suspension (10 μl) was bombarded using a 50-ms burst of helium at a pressure of 7580 kPa within a vacuum of -98 kPa. After the bombardment, the tissue was stored on MS medium in the dark at 20 °C for 24 or 96 hours. Bombarded tissue was imaged using a Leica SP5 confocal microscope. For mRFP, a 561-nm laser

line was used and emission at 570 to 615 nm was imaged. For EYFP, a 488-nm laser line was used and emission at 500 to 550 nm was imaged using a x25, 0.7 multimersion objective.

3.2.7 CRISPR/Cas9 HQT tomato plants

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (Cas) genes were used to edit the genome of tomato (MoneyMaker) by Tom Lawrenson, John Innes Centre Norwich. Two single guide RNA sequences were designed to enable the CRISPR Cas9 complex to splice two sites of genomic DNA corresponding to sequences in the 5'UTR and 1st exon of the hydroxycinnamoyl-CoA quinate hydroxycinnamoyltransferase (HQT) gene (Solyc07g005760). The sgRNA sequences used can be found in Appendix 1.

Tomato plants were transformed with *A.tumefaciens* to insert the T-DNA containing the sequences to express the sgRNAs, Cas9 and nptII proteins. The production of Cas 9 proteins was controlled by the cauliflower mosaic virus 35S promoter. Plant shoots were regenerated from the resulting callus and rooted in MS media containing kanamycin. Rooted plants were potted up and placed in the glasshouse. Dr. Fabio D'Orso of John Innes Centre, Norwich analysed the T0 generation for genomic changes in the HQT gene. He genotyped 65 lines of the T0 generation plants by PCR on the genomic DNA and found 6 lines with the expected deletion, one of which carried a homozygous deletion of the target sequence, line 12. This homozygous knockout of HQT is the line I used for my analysis.

3.2.8 Production of transgenic hairy roots from tomato leaves using *A.rhizogenes*

For the overexpression of HQT and HCT from coffee and tomato in HQT KO tomato plants, a hairy root assay system was used. The pBR7WG2R Gateway^{TM} compatible vector was recombined with HQT or HCT from coffee or tomato, or a eGFP protein as a control. This vector contains a gene encoding the dsRED protein between the left and right boarders which can help identify roots that are transformed and expressing proteins of interest under the control of the CaMV 35S promoter (Figure 3.4 top). The vectors were inserted into electrocompetent *A. rhizogenes* ATTC15834 and grown on TY media for a minimum of 48 hours. The methods for the creation of electrocompetent cells and electroporation of Agrobacterium can be found in the General Materials and Methods sections, 2.2.9 and 2.2.10 respectively.

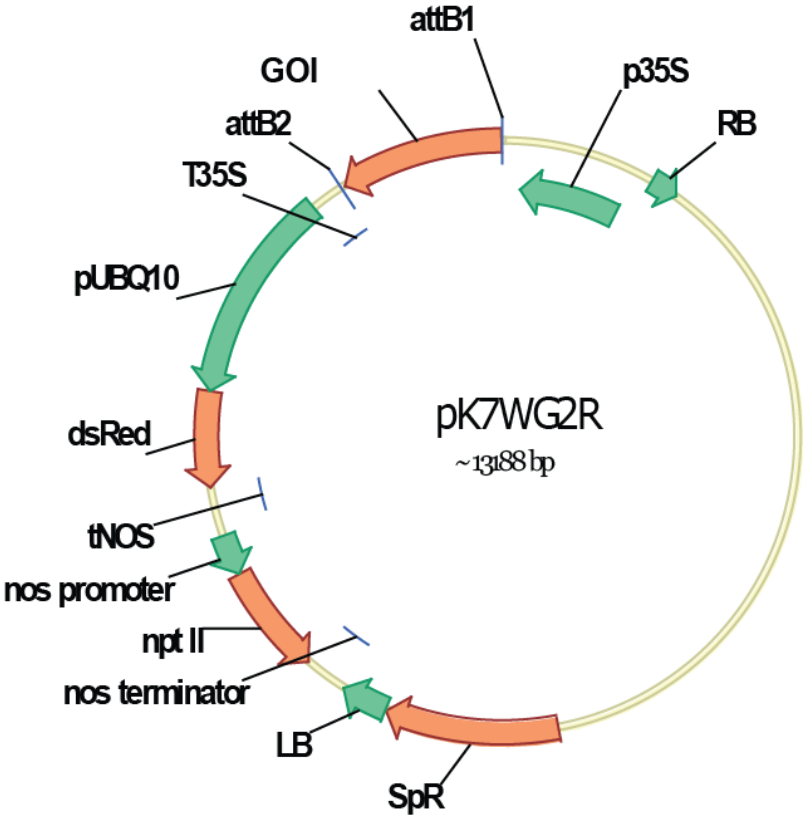
A large colony of Agrobacterium was selected and grown in liquid TY medium (10 ml) with selective antibiotics overnight at 28 ^{o}C, shaking at 220 rpm. The next day the cultures were grown to an OD_{600}

of 0.2 – 0.4 and the cells were resuspended in MS liquid (10 ml) following centrifugation at 3000 x g for 5 minutes and removing the supernatant.

Line 12 of the HQT KO plants was screened for the insertion of T-DNA backbone with a nptII gene conferring kanamycin resistance by PCR because the T1 generation of this line did not contain a functional HQT gene. This was because analysis of the T0 plant of this line showed a homozygous deletion of 410 bp of coding region and 5'UTR of the HQT gene (Solyc07g005760). Analysis showed that line 12 had a 1 in 4 chance of not inheriting the T-DNA carrying the Cas9 machinery Figure 3.4 bottom. This observation meant that line 12 from the T0 generation was a single insert of the T-DNA. Tomato plants with no T-DNA in the T1 generation were selected for hairy root transformation to ensure that any future selection with kanamycin conferred plants with an nptII gene from the pK7WG2R plasmid.

Young leaves were sterilised in 10 % bleach (Domestos) for 10 minutes and then rinsed 3 times in sterile ddH_2O. The leaves were cut into 1 cm^2 pieces and submerged into *A.rhizogenes* solution for 20 – 30 minutes. Explants were placed on sterile filter paper to remove excess liquid for 5 minutes and then placed on 0.8 % MS agar (no antibiotics). The co-cultivation plates were kept in the dark for 3 days at 23 ^oC.

Explants were then placed onto 0.8 % MS agar with kanamycin (100 mg/l) and cefotaxime (200 mg/l) bottom side up, ensuring the edges of the tissue were slightly under the surface of the medium. The plates were incubated in the dark at 23 ^oC and subcloned every 2 weeks. Callus and roots started appearing after 2 weeks of selection. Roots were visualised under the microscope for dsRED production and eGFP for the control (Figure 3.5). Roots which were positively expressing dsRED were placed into liquid MS medium with kanamycin (100 mg/l) at 23 ^oC until they had grown enough to be sampled. Two lines for every overexpression construct were established, individual lines were declared from roots arising from different individual calli.



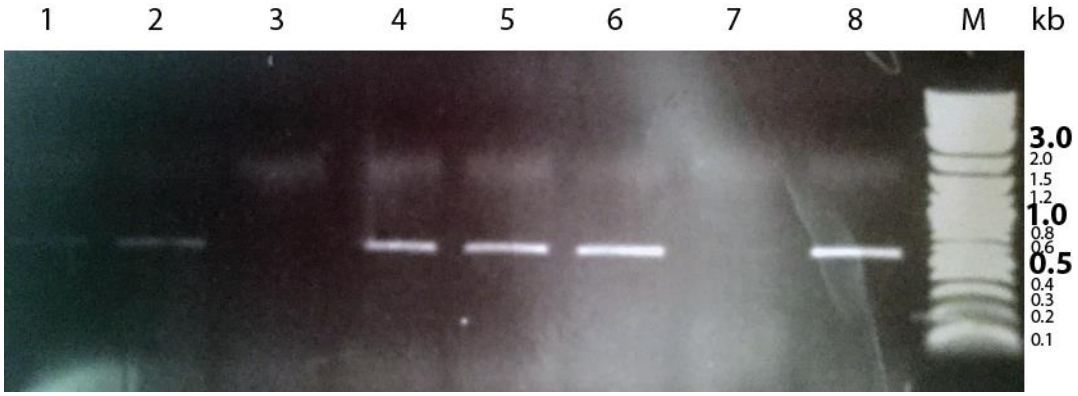


Figure 3.4. Vector used to produce hairy roots and the analysis of T-DNA in #12 HQT KO plants. Top, pK7WG2R vector imaged from Vector NTI (Thermofisher) used for overexpression of BAHD acyltransferases. GOI (gene of interest) is the gene to be overexpressed. The vector has the spectinomycin gene, SpR, for resistance in bacteria and the nptII gene for kanamycin resistance in plants. LB/RB left and right borders of T-DNA, attB1, attB2, Gateway™ sequences for cloning, pUBQ10, ubiquitin-10 promoter, tNOS, nos terminator. Bottom, bands observed from amplifying T-DNA of genomic DNA from #12 HQT KO tomato plants. A DNA band was not observed in plants 3 and 7. Ubiquitin was amplified as a control. A list of all primers used can be found in Appendix 1.

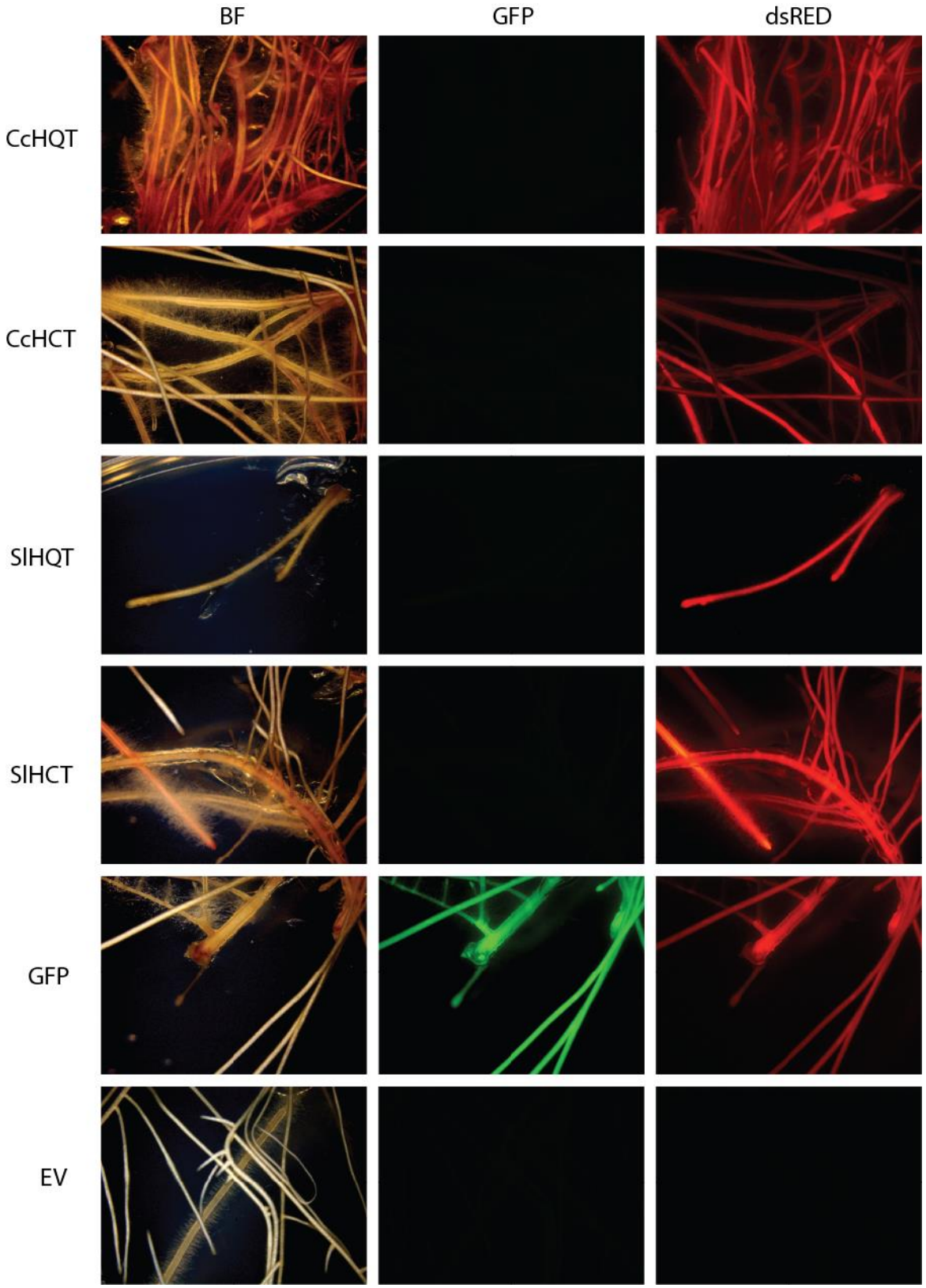


Figure 3.5. Fluorescence observed from hairy root transformations of HQT KO tomato plants. All the hairy root transformations resulted in the expression of dsRED and GFP in the control. Empty vector (EV) control did not contain dsRED or GFP genes. No fluorescence was observed for the EV control roots. dsRED was observed at 575 nm and GFP is visualised at 480 nm.

3.3 Results

I assayed two different activities for the HQT and HCT enzymes. I assayed whether HCT and HQT from coffee and tomato could synthesise diCQA's using the substrates caffeoyl CoA and 5-CQA. Here, caffeoyl CoA would act as the acyl donor and 5-CQA, the acyl acceptor. I named this reaction **Ca**ffeoyl Co**A** chloro**g**enate **T**ransferase activity (CaAGT). I also assayed whether HQT and HCT from coffee were able to synthesise diCQAs *in vitro* using the previously characterised CCT (**C**hlorogenate:**C**hlorogenate **T**ransferase) activity [58, 102]. The proposed reactions for CaAGT and CCT activities, along with the well characterised BAHD activity, in the synthesis of diCQAs and 5-CQA respectively are outlined (Figure 3.6).

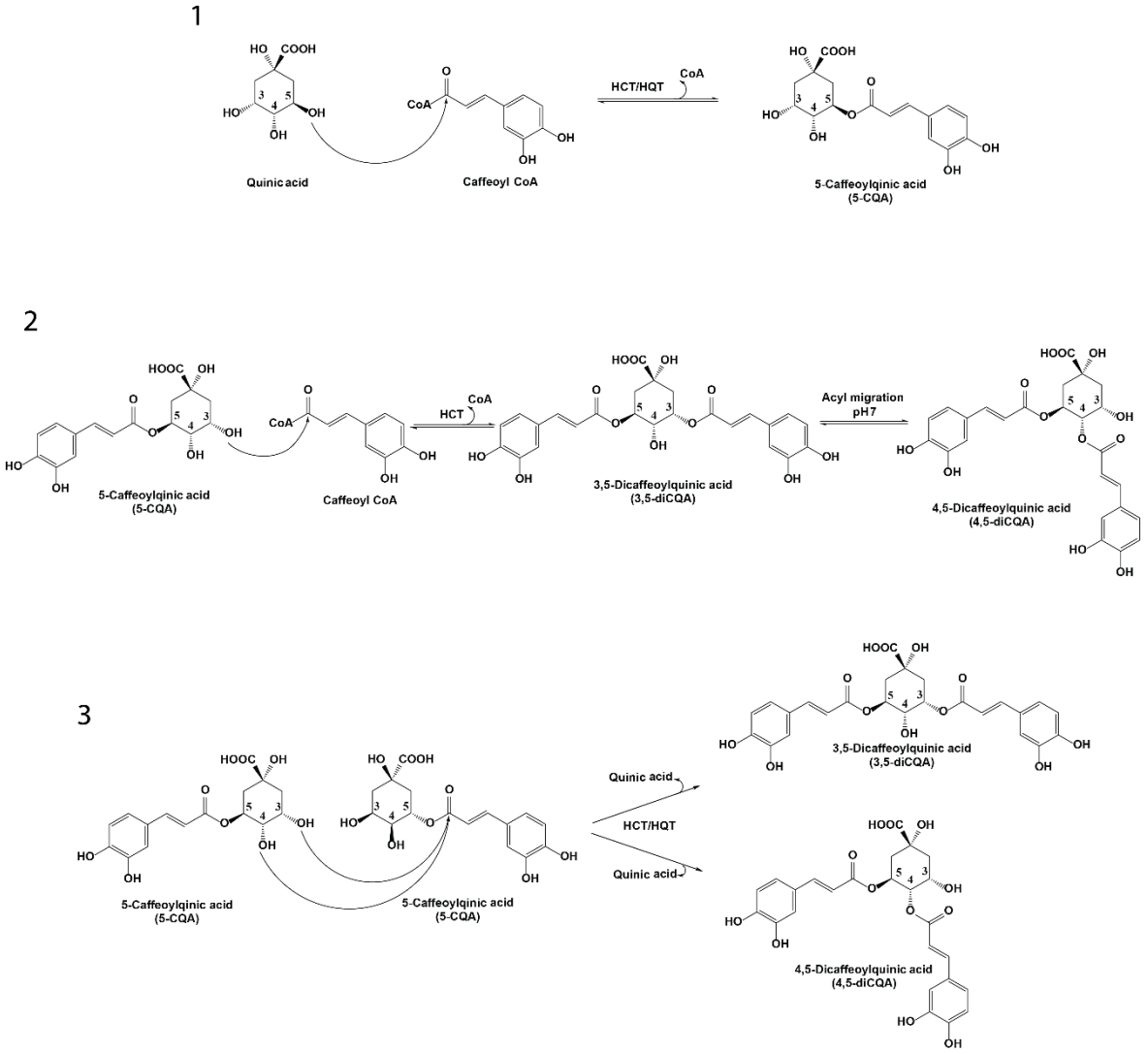


Figure 3.6. Proposed reaction schemes for CcHQT and CcHCT. 1. BAHD activity of HCT and HQT in the synthesis of chlorogenic acid using caffeoyl CoA as the acyl donor and quinate as the acyl acceptor. 2. CaAGT activity for the synthesis of diCQAs. 5-CQA serves as the acyl acceptor and caffeoyl CoA as the acyl donor. 3. CCT activity where 5-CQA acts as both the acyl acceptor and the acyl donor in the synthesis of diCQAs.

3.3.1 Expression and assay of recombinant proteins from crude and purified extracts of *E.coli*

I expressed cDNAs encoding HCT and HQT from coffee (CcHCT and CcHQT) and tomato (SlHCT and SlHQT) in *E.coli* for characterising their catalytic activities in vitro. Western blot analysis of crude *E.coli* extracts overexpressing CcHCT, CcHQT, SlHCT, SlHQT and eGFP showed that all proteins were expressed in crude extracts had the expected molecular weight. Bands that corresponded to CcHQT and SlHQT were less visible compared to the bands corresponding to CcHCT, SlHCT and GFP (Figure 3.7). Therefore, *E.coli* expressing HQT enzymes were less efficient at recombinant protein production then *E.coli* expressing HCT enzymes and GFP. This means a large proportion of the protein forms in inclusion bodies when expressed in *E.coli.*

I performed previously-characterised forward reactions of the purified and crude extracts (Figure 3.8) [61, 101] to test their abilities to form caffeoyl quinates. Extracts of *E.coli* overexpressing HQT and HCT from coffee and tomato all showed the production of 5-CQA after an incubation with caffeoyl CoA and quinate for 20 minutes (Figure 3.8). *E.coli* overexpressing eGFP (enhanced green fluorescent protein) did not produce any 5-CQA. There were large differences observed for the enzyme activity of crude *E.coli* extracts where the relative area of the peaks which corresponded to 5-CQA differed between samples. HQT enzymes from both coffee and tomato produced more 5-CQA than HCT enzymes (Figure 3.8). Notably, despite HQT being present in *E.coli* extracts in much smaller quantities than the HCT enzymes, they produced far more 5-CQA.

The differences observed between HQT and HCT enzyme activity in the forward reaction, with quinic acid and caffeoyl CoA, were because HQT enzymes have a greater affinity for quinate than do HCT enzymes which have a greater affinity for shikimate, although both enzymes have been found to accept both quinate and shikimate as substrates *in vitro* [57, 61, 101]. These assays confirmed that the HIS tags did not impair the activity of either HQT or HCT.

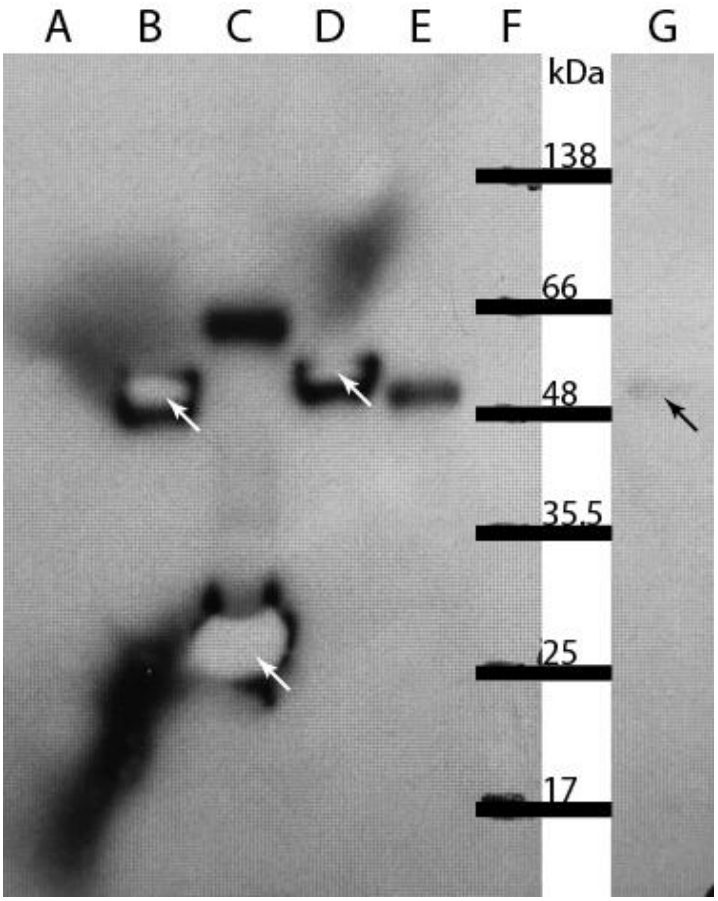


Figure 3.7. Western blot of crude protein extract of recombinant protein production in *E.coli*. Membranes were probed with ECL kit and exposed to medical X-Ray film for 10 seconds and then developed. Black bands indicate His-tagged proteins. White arrows indicate areas of concentrated HRP secondary antibody which are washed away with the ECL kit. Black arrow indicates a faint band observed at 50 kDa. A, empty vector; B, CcHCT; C, GFP; D, SlHCT; E, SlHQT; F, Molecular masses of prestained protein marker broad range (NEB); G, CcHQT.

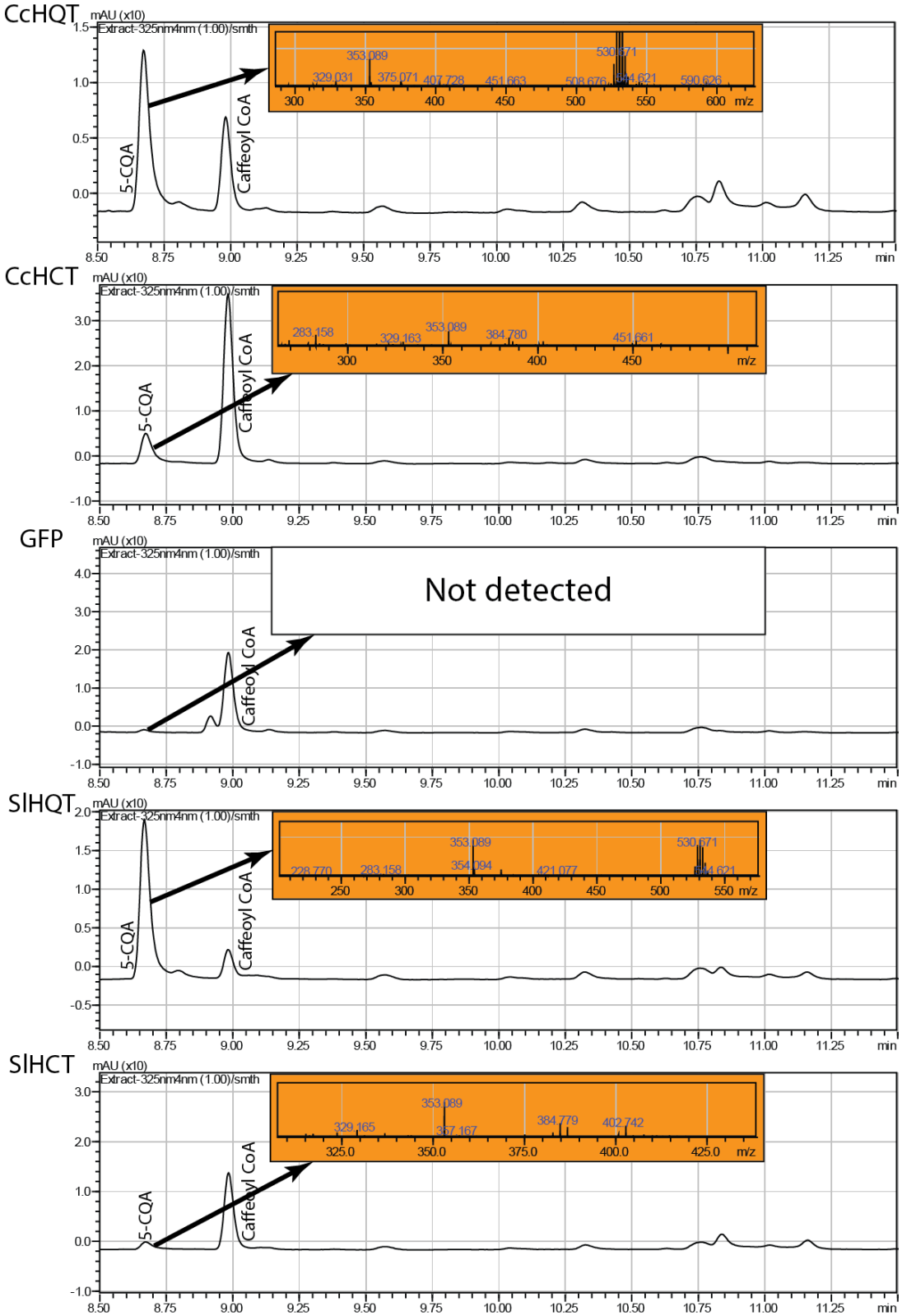


Figure 3.8. Control forward reactions of HCT and HQT coffee and tomato. Caffeoyl CoA and quinic acid was used for all reactions. CcHQT, CcHCT, SlHQT and SlHCT enzyme assays resulted in formation of 5-CQA. Quinic acid was not observed with the run method. The data shown were extracted at 325 nm from PDA data. Orange boxes represent the mass spectrum observed for 5-CQA. 5-CQA was not detected in the control reaction with GFP. HQT has a higher affinity for quinate than HCT and therefore produced more 5-CQA.

3.3.2 Coffee and tomato HCT can synthesise dicaffeoylquinic acid using caffeoyl CoA and chlorogenic acid

I next tested whether HCT or HQT from coffee or tomato could catalyse the synthesis of diCQAs from caffeoyl CoA and 5-CQA. I incubated crude extracts of *E.coli* which overexpressed HCT or HQT from *C.canephora* and *S.lycopersicum* with caffeoyl CoA and 5-caffeoylquinic acid (5-CQA) at a range of pH's and I detected the formation of a new product from the crude extracts of HCT but not from HQT from either coffee or tomato (Figure 3.9). This activity occurred only at pH 7. I identified the product to be 3,5-diCQA by liquid chromatography (LC)-photodiode array detector (PDA)-tandem mass spectrometry (MS/MS) analysis (Figure 3.9). The peak of the 3,5-diCQA standard had a retention time of 10.25 minutes and had a mass-to-charge ratio (m/z) of 539.1160 in positive electrospray ionisation mode [ESI^+] and 515.1220 (m/z) in negative electrospray ionisation mode [ESI^-]. 3,5-diCQA also fragmented in negative mode to 353.0870 and, in positive mode, to 377.0920 which was also the m/z of 5-CQA which had a retention time of 8.75 minutes with the run method used. The observed fragmentation pattern of 3,5-diCQA was expected because of its structure which comprises caffeoylquinic acid with an additional caffeoyl group. 5-CQA and 3,5-diCQA exhibited best ionisation in negative electrospray ionisation mode [ESI^-] (Figure 3.9). This is **Ca**ffeoyl Co**A** chloro**g**enate **T**ransferase activity (CaAGT) from a BAHD acyltransferase enzyme. HQT from tomato and coffee did not show CaAGT activity using the substrates caffeoyl CoA and 5-CQA in the 90-minute reaction (Figure 3.9) or even when they were incubated overnight with the same substrates (data not shown). The activity of crude extracts of CcHCT expressed in *E.coli* was confirmed by repeating these assays with native CcHCT protein purified by gel filtration. Assays with the purified protein showed detectable levels of 3,5-diCQA after a reaction time of 90 minutes.

I showed that after a relatively short incubation time of 90 minutes, it was possible to detect 3,5-diCQA in these assays using the native version of the HCT protein. I also used a mutated version of the HCT in which Lys-210 and Lys-217 were converted to Ala (which is reportedly more resistant to proteolysis) which showed similar catalytic activity to the native protein (Appendix 3) [101].

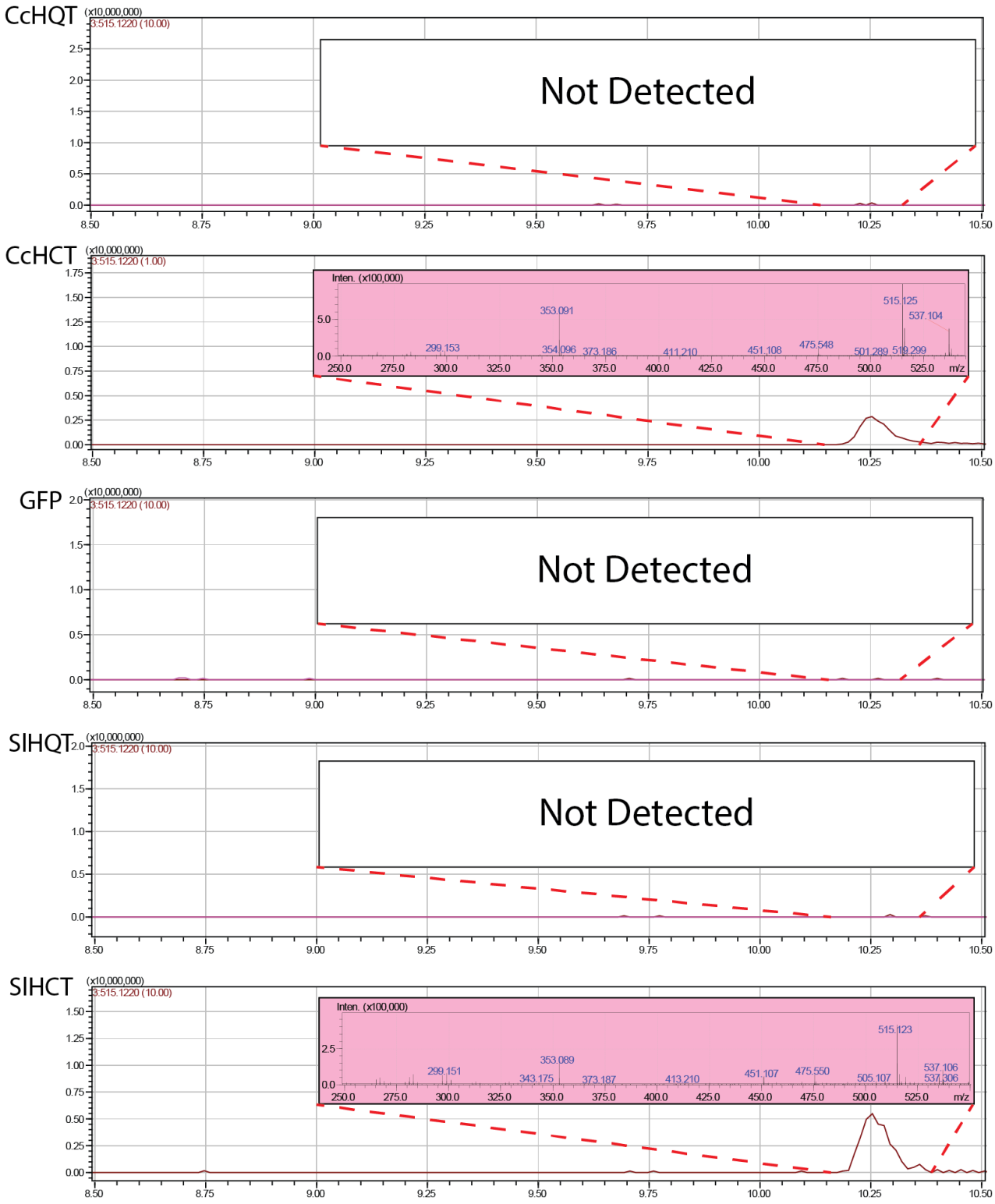


Figure 3.9. CaAGT activity of HCT from coffee and tomato in the synthesis of diCQAs. MS data are shown isolated for ESI$^-$ m/z 515. Peaks identified as 3,5-diCQA are present in samples labelled CcHCT and SlHCT. Data represent one dataset from three biological replicates. Pink boxes represent the average mass spectrum of the indicated area where diCQA fragments were detected. Fragments of standard diCQA are found in Table 3.3 materials and methods.

3.3.3 HCT from coffee and tomato can synthesise diCQA using Coenzyme A and 5-CQA

It had been reported previously that native CcHCT can create 3,5-diCQA through overnight incubation with 5-CQA and CoA [101]. The native CcHCT showed the production of only minimal amounts of diCQA following a long overnight incubation and a mutated version of CcHCT also showed detectable activity only after a long overnight incubation. It is not clear whether the diCQA created from the native version of CcHCT in the assay was biologically relevant as levels were only detectable after an overnight incubation.

To further examine differences between the BAHD enzymes I incubated the *E.coli* protein extracts overexpressing HQT and HCT from coffee and tomato with 5-CQA and CoA for 90 minutes. This was the reaction reported earlier by Lallemand et al,. (2012) to result in the formation of caffeoyl CoA and diCQAs through the catalytic activity of CcHCT after an overnight incubation [101].

By repeating this enzyme assay with HQT and HCT from coffee and tomato, I was able to detect 3,5-diCQA synthesis after an incubation of 90 minutes in reactions containing the HCT enzyme from coffee and tomato (Figure 3.10). This assay also produced detectable amounts of caffeoyl CoA, which is the product of the reverse reaction of both these BAHD acyltransferases as previously described [57, 61, 101], which was not present in the control. No diCQA was detected in samples containing *E.coli* extracts overexpressing HQT from coffee or tomato (Figure 3.10).

The presence of caffeoyl CoA as a product of this enzyme assay was an indication that CaAGT activity through HCT is actually dependent on production of caffeoyl CoA to serve as a substrate. The characterised reverse reaction occurs first (using CoA and 5-CQA to produce caffeoyl CoA and quinate) before 5-CQA and caffeoyl CoA can be used as substrates in the synthesis of diCQA via the CaAGT activity of HCT. This assay also provided further confirmation that HQTs from coffee and tomato do not possess CaAGT activity.

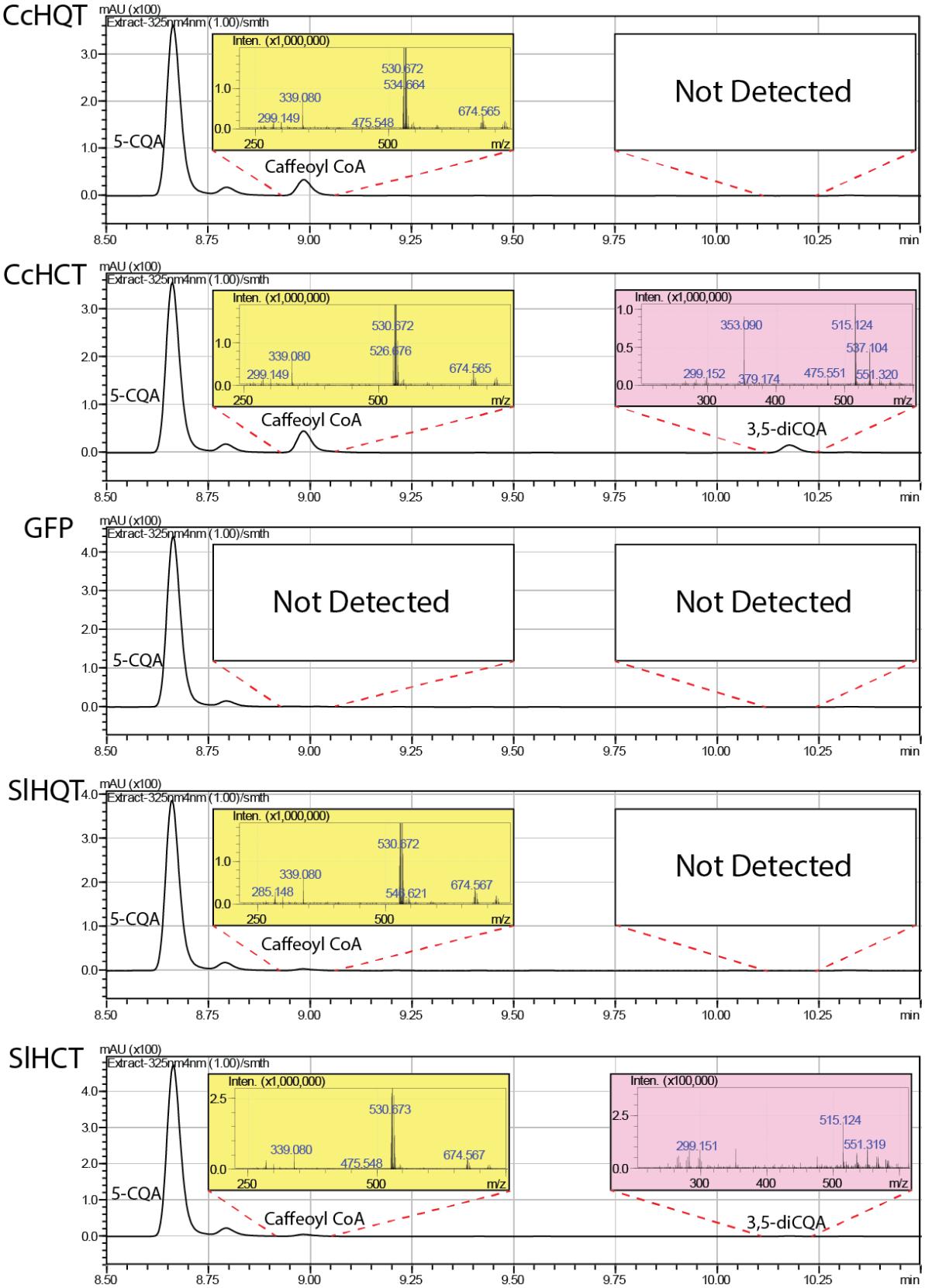


Figure 3.10. Assay of crude *E.coli* protein extracts overexpressing HQT and HCT from coffee and tomato incubated with CoA and 5-CQA. Samples were analysed by LC MS. Protein extracts containing HCT enzymes produced 3,5-diCQA. Mass spectra are in pink boxes. HQT and HCT from coffee and tomato produced caffeoyl CoA, mass spectra in yellow boxes. Caffeoyl CoA was not detected in the control GFP. DiCQAs were not detected in samples containing CcHQT, SlHQT and GFP. Fragmentation of standards for the detected compounds are found in Table 3.3, materials and methods.

3.3.4 Both coffee HQT and HCT can synthesise dicaffeoylquinic acid from chlorogenic acid by CCT activity

As previously described, HQT from tomato has the ability to synthesise 3,5-diCQA *in vitro*, with high concentrations of 5-CQA at a relatively low pH [102]. I incubated crude extracts of *E.coli* overexpressing HQT and HCT from tomato and coffee with different concentrations of 5-CQA at a range of pHs and I detected the formation of diCQA. HQT and HCT from both tomato and coffee were able to synthesise diCQA from 5-CQA at pH 5 compared to the control reactions after incubation for 90 minutes (Figure 3.11). 3,5-diCQA was the most abundant compound present in all samples and therefore it was presumed to be the initial product of the reaction. 3,5-diCQA may then undergo spontaneous acyl migration to form other diCQA isomers [101, 102]. CcHQT and CcHCT purified by gel filtration were able to form diCQA under the same reaction conditions. This chlorogenate:chlorogenate transferase (CCT) activity has not previously been reported for HCT.

Enzyme kinetic analysis was performed on the CCT activity of CcHQT and CcHCT. First a time course at a fixed substrate concentration was analysed to determine at what time the reaction occurred at its quickest. The time course revealed that the CCT activity from CcHQT and CcHCT was slow with the initial rate of reaction not declining until after 120 minutes of incubation. To determine the enzyme kinetics using the Michaelis-Menten equation, reactions were stopped at 120 minutes to allow for maximal diCQA accumulation to aid quantification, as smaller amounts of product were difficult to quantify. The CCT activity of CcHCT showed a Km of 8.297 ± 3.2961 mM CGA and a *V*max of 0.2623 ± 0.0371 μmol/min. The CCT activity of CcHQT had a Km of 3.8942 ± 1.012 mM CGA and a *V*max of 0.1482 ± 0.0102 μmol/min (Figure 3.12). The Km values for the two enzymes were not significantly different because of the standard errors of the curves. To determine whether the Km (and Vmax) values did show any significant differences between the two enzymes, the assays would need to be repeated at least another 2 times as 1 replicate is not enough for reliable conclusions. The Km values for CCT activities were very similar to those reported for HQT from tomato [102].

The analysis of the amino acid sequences showed that CcHQT, CcHCT, and SlHCT all contain the tyrosine (Y) residue at equivalent His-276 (H) present in tomato HQT (Figure 3.13). This residue was considered important for the CCT activity from SlHQT previously described [102]. The tyrosine residue present in CcHCT, CcHQT and SlHCT could mean that CCT activity is less important in these enzymes in the synthesis of diCQAs. Because of their similarity in structure, function and amino acid alignment, presumably 5-CQA interacts in the same way for CCT activity in these BAHD acyltransferases as previously described [102, 106].

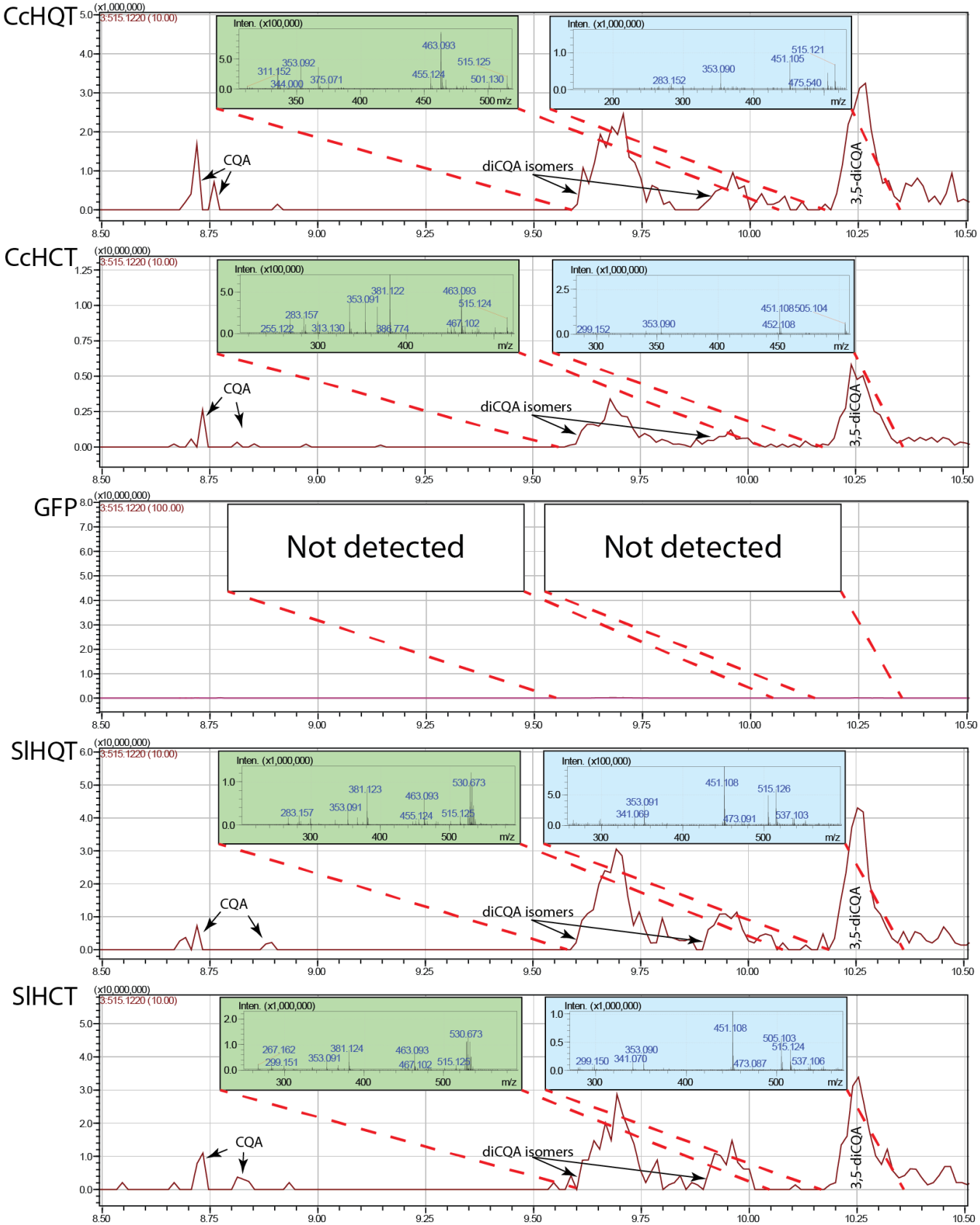
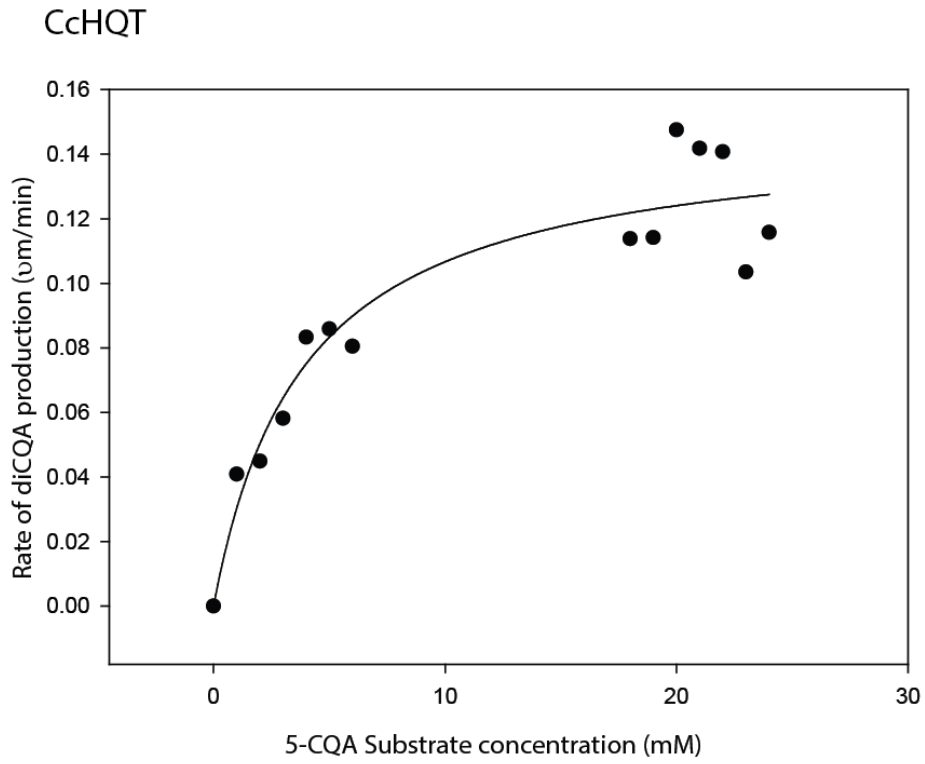


Figure 3.11. Assays of CCT activity using recombinant HQT and HCT from coffee and tomato incubated with 5 mM 5-CQA. The reactions were analysed by LC MS. MS data is shown isolated for ESI$^-$ 515.122. The formation of a new products is observed for CcHQT, CcHCT, SlHQT and SlHCT which was identified as 3,5diCQA and other isomers. The average mass spectra of 3,5-diCQA is shown in the blue boxes. The average mass spectra for other diCQA isomers is shown in the green boxes. DiCQA was not detected in samples containing GFP. Fragmentation of diCQA standard in MS are found, Table 3.3, materials and methods.



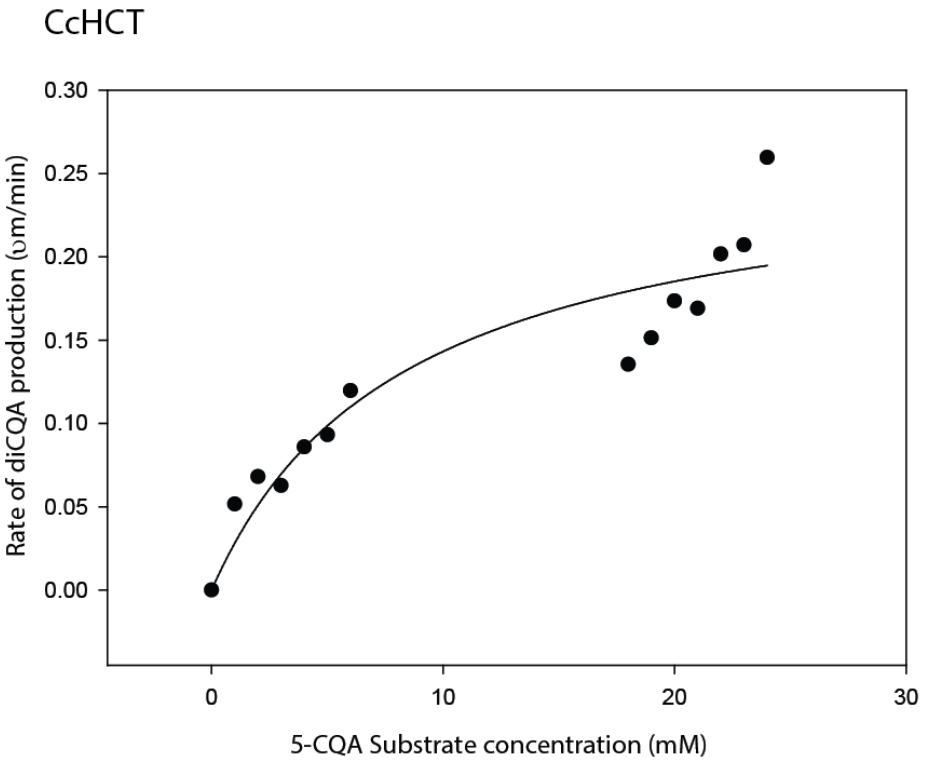


Figure 3.12. Michaelis Menten plot of enzyme CCT activity from CcHQT and CcHCT. The curve represents the best fit of the Michaelis-Menten equation to the data using SigmaPlot (Systat Software) using a dataset obtained from only one experiment.

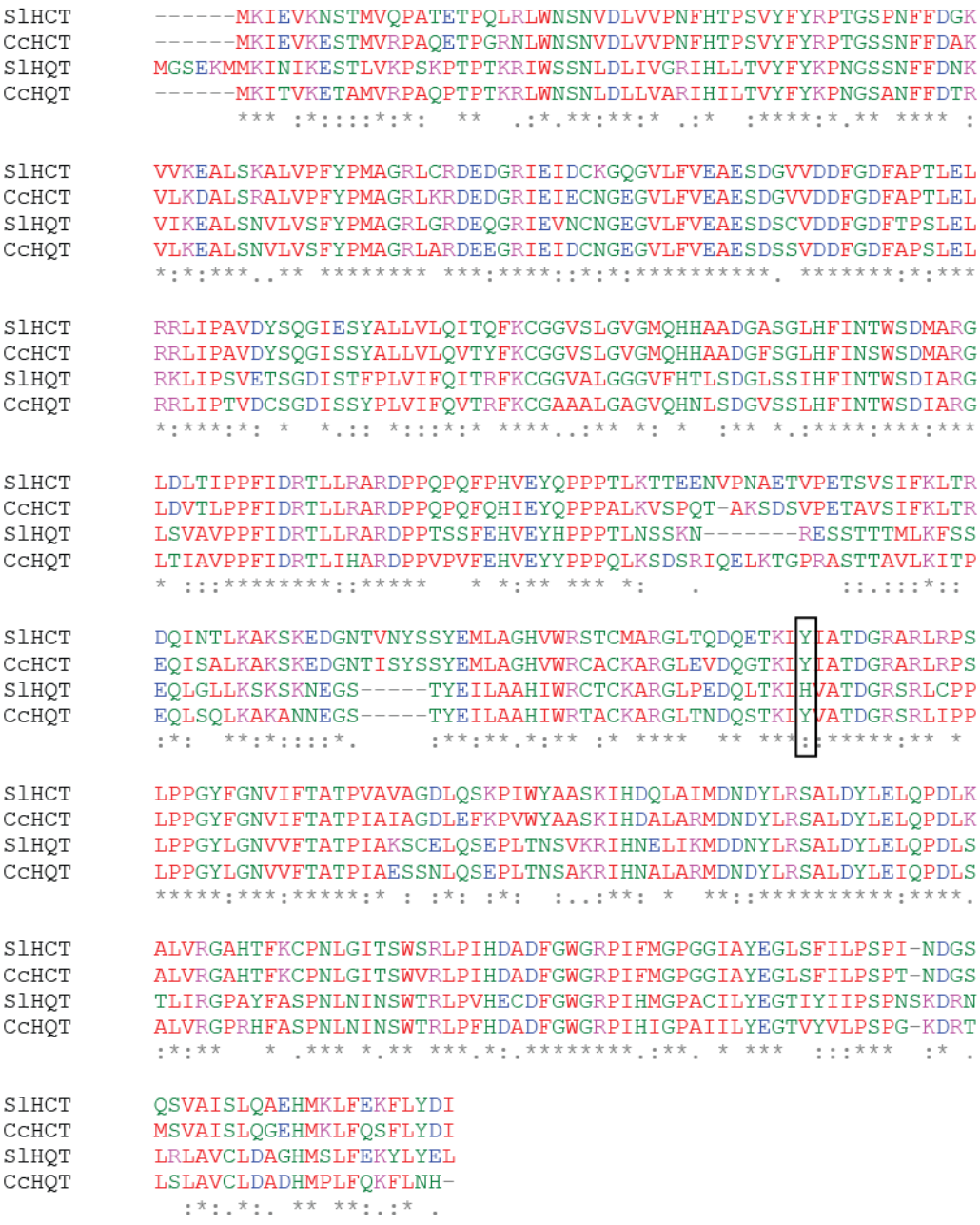


Figure 3.13. Sequence alignment of HQT and HCT from tomato and coffee. The black box indicates His 276 present in SlHQT which is considered key for CCT activity and the equivalent residues in other proteins that show CCT activity [102].

3.3.5 Coffee HCT and HQT can localise to the vacuole

I investigated the subcellular localisation of CcHCT and CcHQT using a vector expressing each enzyme, fused C-terminally to monomeric red fluorescent protein (mRFP). This technique is the same as I used in Moglia et al., 2014 [102]. Plasmid DNA was bombarded into onion (*Allium cepa*) epidermal tissue as these cells have no chloroplasts and chlorophyll can cause interference with mRFP as it fluoresces red. As a control, I used a vector for expression of mRFP alone under the control of the 35S promoter which I also bombarded into onion epidermis.

Both 35S:CcHQT:mRFP and 35S:CcHCT:mRFP transformed cells, imaged after 24 hours, showed clear cytoplasmic localisation, indicated by strong fluorescence at the periphery of the cells and in the cytoplasmic strands (Figure 3.14). Fluorescence could also be observed in the nucleus of transformed cells. As previously described, SlHQT could be observed localised to the vacuole following incubation in the dark for 96 hours [102], and therefore it was possible that similar enzymes from different plants could show similar localisation characteristics. I therefore repeated the experiment but this time imaged the onion epidermis assays after leaving them in the dark for 96 hours after bombardment. Again, the predominant fluorescence was cytoplasmic, as evidenced by the strong fluorescence from the zone delimiting the inner edge of the cells and in the occasional cytoplasmic strands. The nucleus also fluoresced strongly, like 35S:mRFP. However, fluorescence was also clearly visible in the vacuolar region of onion epidermis bombarded with both 35S:CcHQT:mRFP and 35S:CcHCT:mRFP compared to cells bombarded with 35S:mRFP (Figure 3.14).

To demonstrate that the localisation observed was in the vacuole I performed a cobombardment with a vector expressing enhanced yellow fluorescent protein (EYFP) under the control of the cauliflower mosaic virus 35S promoter along with the vectors with the C-terminal mRFP tag of CcHQT and CcHCT. I imaged the onion epidermis after incubating in the dark for 96 hours. Cells expressing both plasmids showed that both CcHQT:mRFP and CcHCT:mRFP were in the vacuole as well as in the cytoplasm while EYPF was completely cytoplasmic (Figure 3.15). An overlay of the two images obtained for each fluorescent protein showed that the periphery of the cells was yellow, where red and green signals had merged, but the central vacuole was red showing specific localisation of CcHQT and CcHCT in the vacuole. These results demonstrated that CcHQT and CcHCT localised to the vacuole as well to the cytoplasm of plant cells.

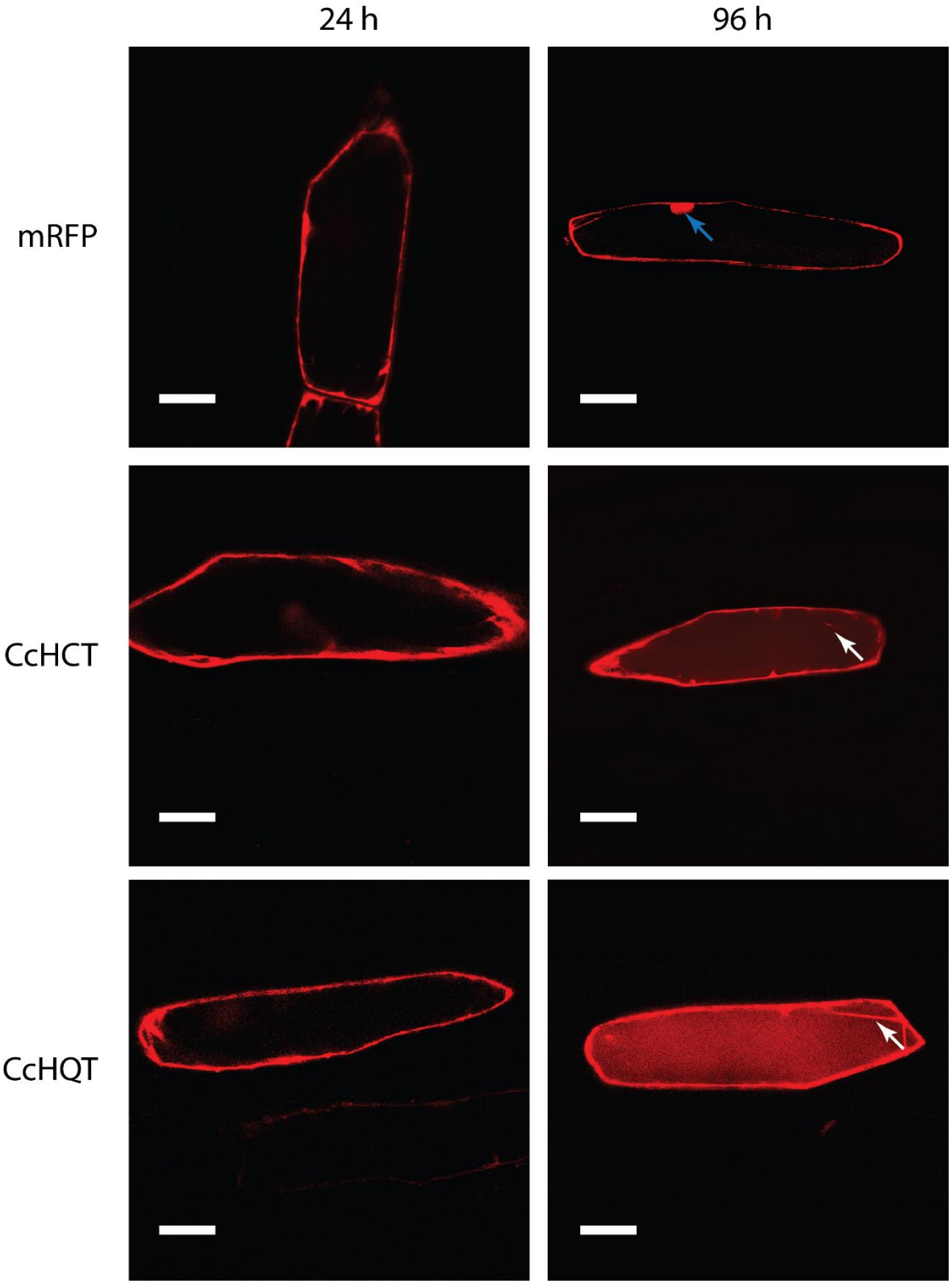


Figure 3.14. Subcellular localisation of CcHCT and CcHQT. Onion epidermis was bombarded with 35S:CcHCT and 35S:CcHQT fussed to mRFP or 35S:mRFP as controls for cytoplasmic localisation. After 24 hours of incubation in the dark, fluorescence was clearly observed in the cytoplasm for CcHCT and CcHQT and the mRFP control, a slightly higher than background fluorescence was just detectable in the vacuole. After 96 hours of incubation in the dark fluorescence was visible in the vacuole for CcHCT and CcHQT which was not observed for the mRFP control localised only in the cytoplasm. White arrows indicate cytoplasmic strands, Blue arrows indicate fluorescence observed in the nuclei. Bars = 50 μm.

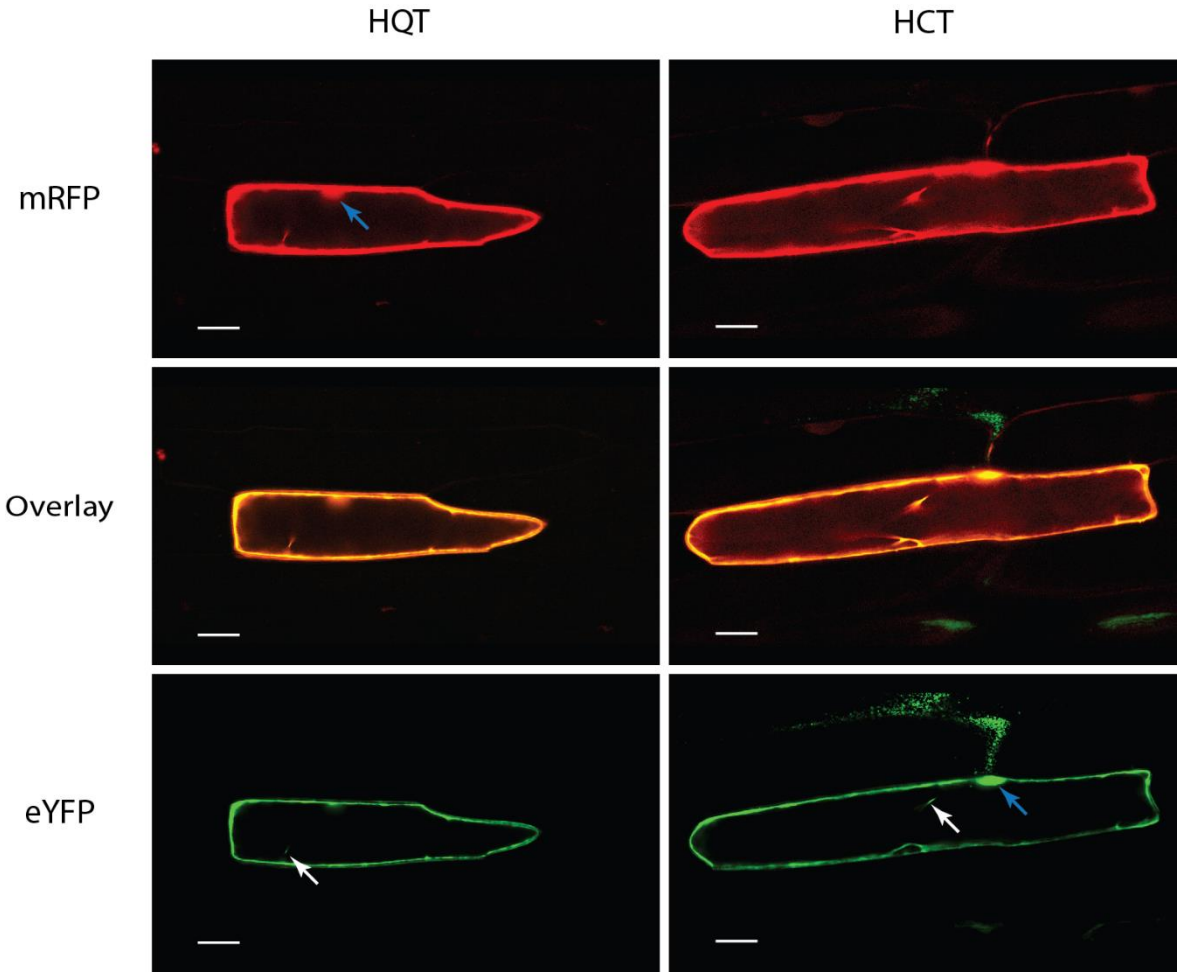


Figure 3.15. Subcellular localisation of CcHQT and CcHCT cobombardment with EYFP. Onion epidermis was bombarded with 35S:CcHCT and 35S:CcHQT fused to mRFP and 35S:EYFP was used as a control and incubated in the dark for 96 hours. CcHQT and CcHCT showed localisation to the vacuole with their fusion to mRFP while EYFP showed localisation to the cytoplasm only when the same cell was imaged. The merged imaged shows fluorescence from CcHQT and CcHCT in the vacuole as well as the cytoplasm while the EYFP remained entirely cytoplasmic. White arrows indicate cytoplasmic strands, Blue arrows indicate fluorescence observed in the nucleus. Bars = 50 μm.

3.3.6 3,5-dicaffeoylquinic acid undergoes spontaneous acyl migration at neutral pH

It had been suggested previously that 3,5-dicaffeoylquinic acid is the main product of both CCT and CaAGT activity which then undergoes spontaneous acyl migration to form other isomers of diCQA [101]. I tested the ability of 3,5-diCQA to undergo spontaneous acyl migration by incubating 3,5-diCQA at different pHs. I did not detect any isomers of 3,5-diCQA stored in buffer at pH 5 after 24 hours of incubation at 30 ^{o}C (Figure 3.16 top). 3,5-diCQA stored at a pH 7 showed spontaneous acyl migration into at least two other diCQA isomers (Figure 3.16 bottom) (probably 4,5-diCQA and 3,4-diCQA, the next most abundant diCQA compounds found in coffee [24].) The spontaneous acyl migration of 3,5-diCQA was dependent on pH such that non-enzymatic dependent isomers were formed only when 3,5-diCQA was at a neutral pH. I also tested 5-CQA pH stability in the same way which showed the same characteristics, this has been previously reported [107].

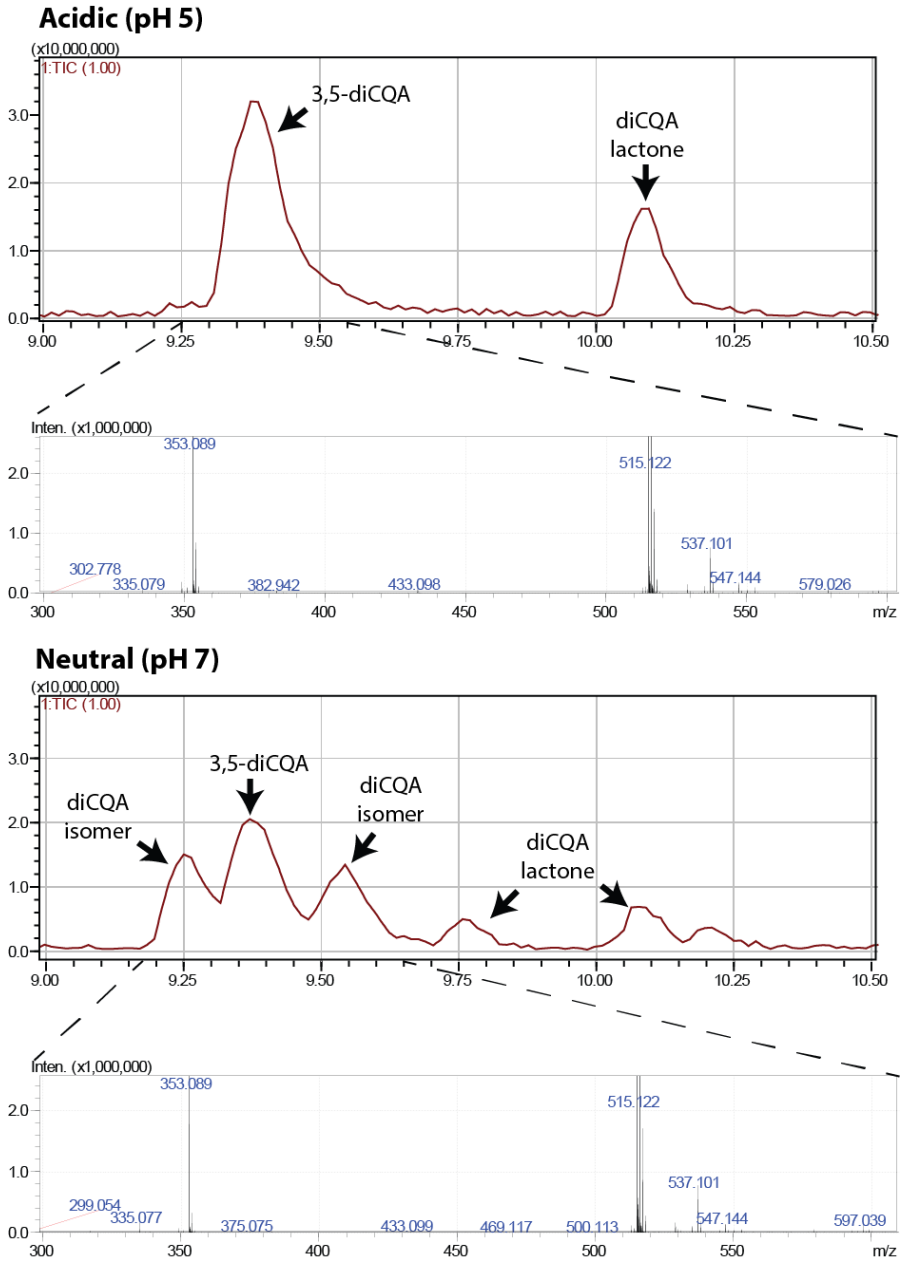


Figure 3.16. 3,5-diCQA stored in acidic conditions remain stable whereas diCQA stored neutral pH shows spontaneous acyl migration. A standard sample of 3,5-diCQA was stored in neutral (sodium phosphate buffer 100 mM pH 7.0) and acidic (sodium phosphate buffer 100 mM pH 5.0) conditions at 30 °C for 24 hours and then analysed by LC MS. 3,5-diCQA stored in an acidic pH showed the presence of an extra peak which was identified as a 3,5-diCQA lactone which was also present in the standard (data not shown). 3,5-diCQA stored in a neutral pH showed the presence of peaks which were identified as other diCQA isomers by MS along with other peaks which were identified as lactones. The average mass spectra are shown for areas identified as diCQA.

3.3.7 Tomato fruit lacking functional HQT do not produce caffeoyl quinic acids

Solanum lycospersicum was transformed with a construct driving the expression of Clustered Regulatory Interspaced Short Palindromic Repeats (CRISPR) and CRISPR associated (Cas) genes expressing two single guide RNAs (sgRNA). This construct was designed to delete a region of genomic DNA that exists between the 5'UTR and the first exon of the coding region of hydroxycinnamoyl-CoA quinate hydroxycinnamoyltransferase (HQT) (Solyc07g005760) (Figure 3.17 top). This deletion resulted in a functional knock out of the enzyme HQT which catalyses the condensation of caffeoyl CoA and quinic acid to make 5-caffeoylquinic acid (5-CQA). The functional knock out of HQT in tomato resulted in no detectable chlorogenic acids in tomato fruit (breaker + 10 days) (Figure 3.17). This suggested that in tomato fruit at least, chlorogenic acid biosynthesis is wholly dependent on the presence of a functional HQT enzyme and that there are no other pathways for chlorogenic acid biosynthesis in tomato.

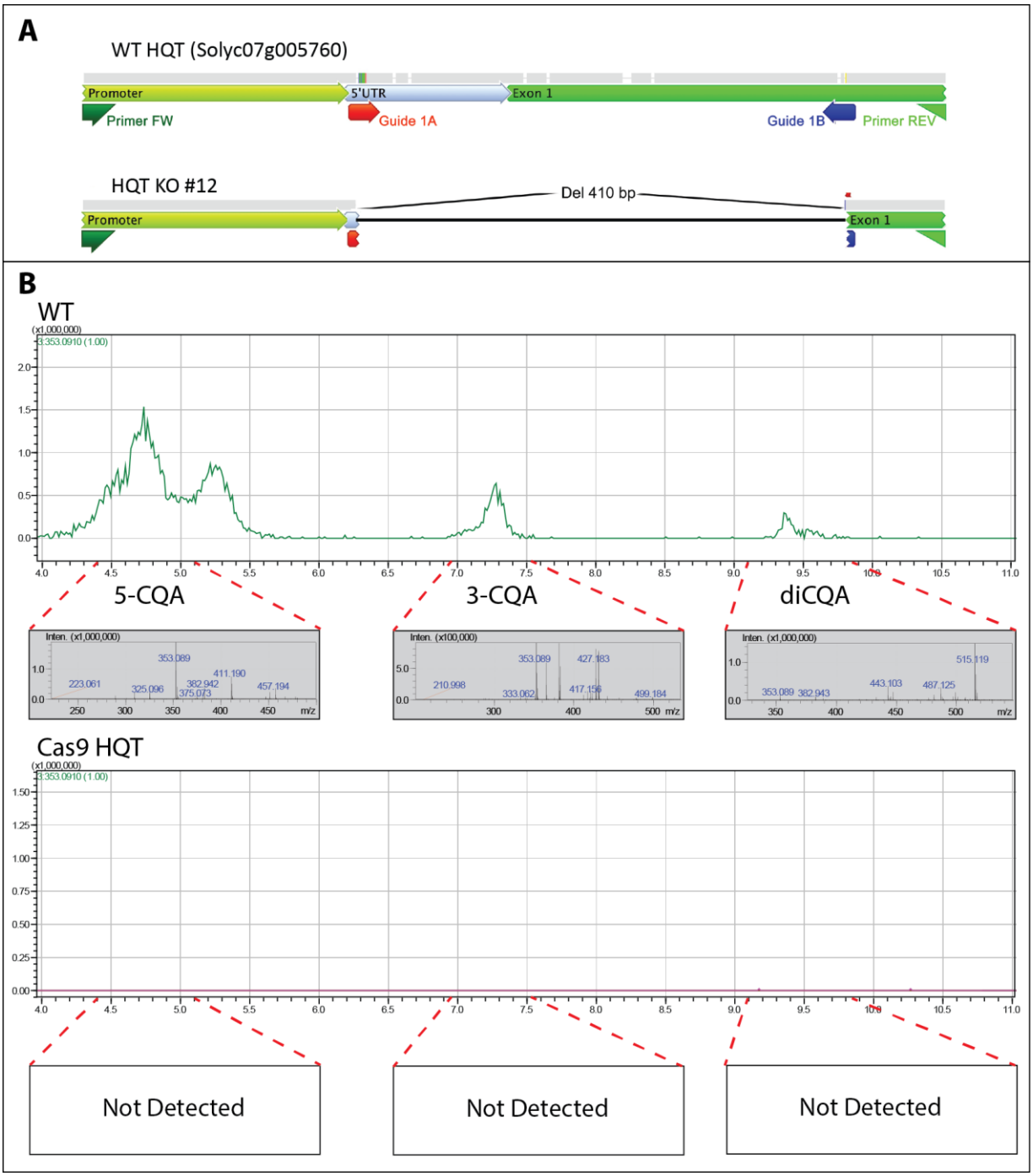


Figure 3.17. Analysis of line 12 tomato HQT KO plants. A. Schematic of genomic DNA sequence edited in line 12 HQT KO plants. A 410 bp deletion in the 5'UTR and Exon 1 of the HQT gene caused a functional knockout the gene. Guide 1A and 1B represent where sgRNA oligos were designed. Primer FW and REV were for analysis of the knock plants. B. Methanol extracts of tomato fruit analysed by LC MS. MS data is isolated for ESI^- m/z 353.091. Peaks identified as 5-CQA, 3-CQA and diCQA are present in the WT fruit whereas there are no peaks present in HQT KO fruit which identifies the presence of chlorogenic acids.

3.3.8 Coffee HQT reintroduces chlorogenic acid biosynthesis when overexpressed in tomato roots lacking functional HQT

Plasmids containing HQT and HCT from coffee and tomato under the control of the 35S promoter were transformed into the HQT KO tomato leaves, lacking the T-DNA insert conferring kanamycin resistance, using *A.rhizogenes* to create hairy root transformants. The analysis of hairy root transformants showed traces of 5-CQA that were detected in all of the samples, including the control GFP which was unexpected (Figure 3.18). I confirmed that trace levels of 5-CQA in the GFP roots were not caused by contamination, firstly by checking the blank samples of the LC MS experiment for 5-CQA traces and secondly by repeating the LC MS analysis with another HQT KO line expressing GFP. This meant that, a functional HQT KO did not result in the complete loss of chlorogenic acid in tomato as traces of 5-CQA were present in HQT KO roots. As previously shown *in vitro*, HCT is able to form 5-CQA from caffeoyl CoA and quinate with low efficiency in vitro. *In silico* analysis of transcript levels of HCT (Solyc03g117600) of RNA seq reads from different tomato tissues revealed that HCT is expressed most highly in tomato roots and much lower in tomato fruit, a difference of almost 4 in data represented as Log 2 ratio difference in RPKM values (Figure 3.19).

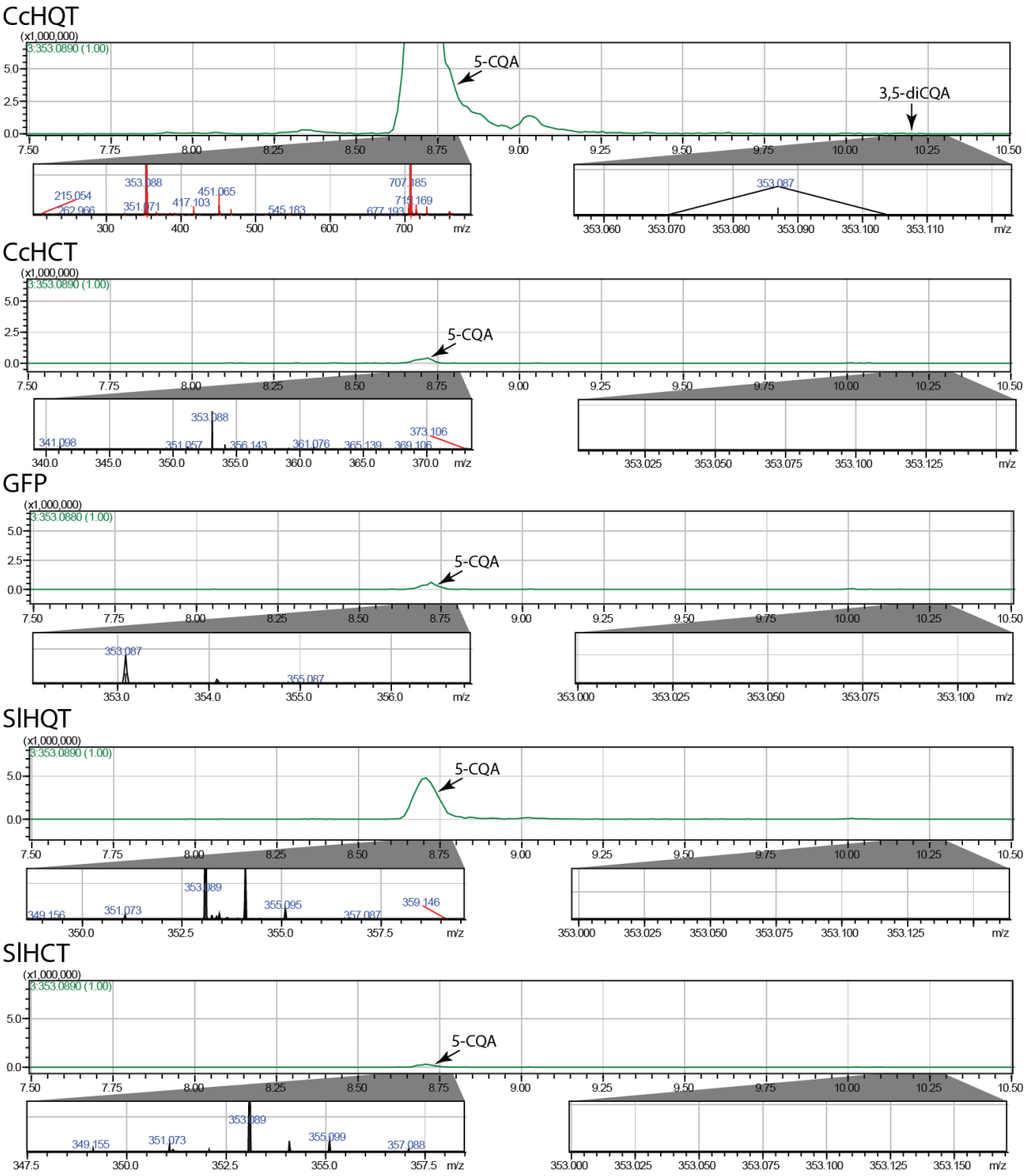
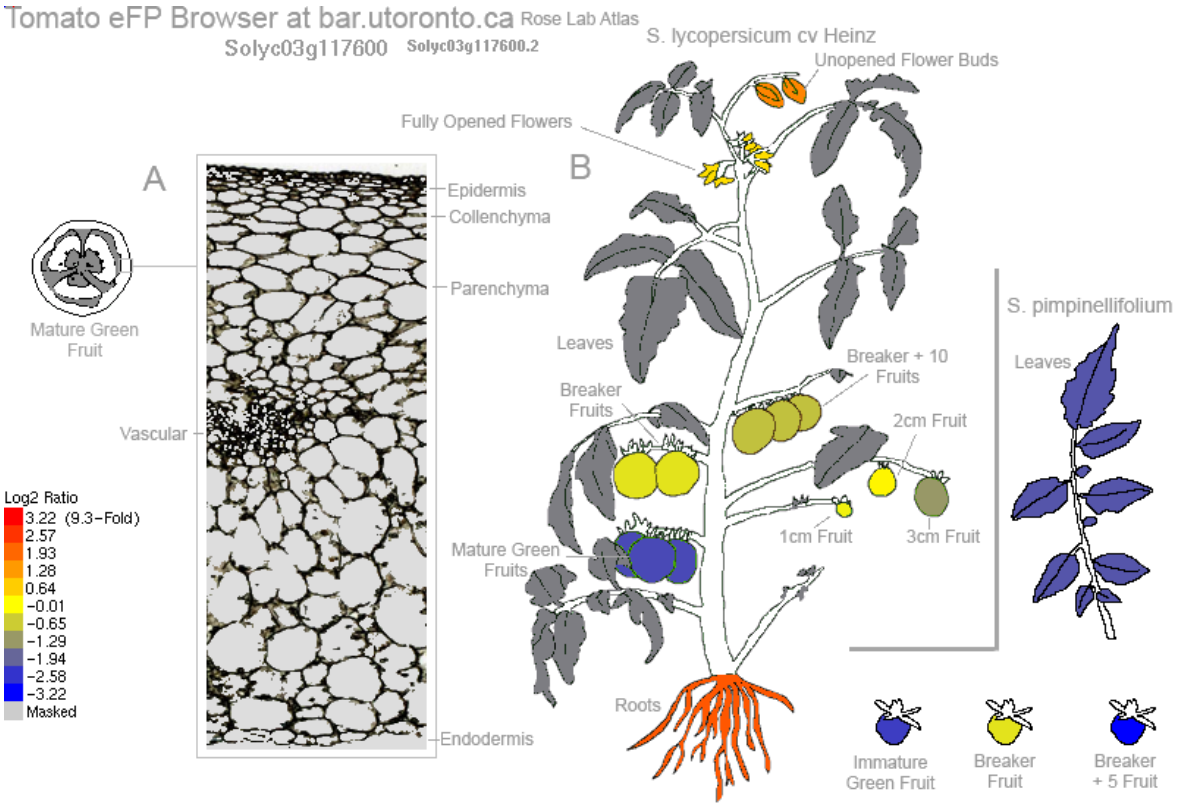


Figure 3.18. Analysis of HQT KO hairy roots overexpressing HQT and HCT from coffee and tomato. Methanol extractions of hairy roots were analysed by LC MS. MS data is isolated for ESI^- m/z 353.089. Analysis showed chlorogenic acid is present in all samples. Samples overexpressing CcHCT, SlHCT and GFP showed very small traces of 5-CQA, whereas roots overexpressing CcHQT and SlHQT showed larger peaks. A trance of diCQA can be detected in the hairy root sample overexpressing CcHQT, indicated by the fragment 353.087 m/z. No diCQA was detected in hairy roots overexpressing CcHCT, GFP, SlHCT or SlHQT.

Overexpression of CcHQT induced much higher levels of 5-CQA in the roots from HQT KO plants (Figure 3.18). This showed that CcHQT shows functional identity *in vivo* to SlHQT as the overexpression of native SlHQT in HQT KO roots also caused high production of 5-CQA (Figure 3.18). I also detected small amounts of 3,5-diCQA present in the hairy root samples overexpressing CcHQT which were not present in any of the other samples (Figure 3.18). An analysis of WT roots of tomato revealed that diCQAs are synthesised in the roots of WT tomato plants (data not shown). The reason why SlHQT overexpression in HQT KO roots did not show diCQA synthesis might be due to the low expression of SlHQT in the hairy root lines. Analysis of the intensity of dsRED fluorescence of hairy roots overexpressing CcHQT and SlHQT showed a more intense fluorescence from CcHQT overexpression (Figure 3.20). Although dsRED and Cc/SlHQT are not under the control of the same promoter in the hairy roots, (ubiqutin-10 gene promoter (pUBQ10), and cauliflower mosaic virus 35S (CaMV35S) respectively), differences observed between fluorescence intensity maybe directly correlated with Cc/SlHQT expression, particularly if higher dsRED fluorescence indicates multiple T-DNA inserts in these lines. *In silico* analysis of RPKM values of SlHQT (Solyc07g005760) expression showed that SlHQT is relatively highly expressed in tomato roots compared to other tissues (data not shown). All these data suggested that SlHQT is capable of synthesising diCQA in tomato roots but this activity is directly correlated to its expression levels. The HQT enzyme from coffee probably caused a higher accumulation of CQA, and consequent diCQA production through the CCT activity of HQT which will occur only in the presence of high concentrations of 5-CQA. However, these data do not exclude the possibility that diCQAs were formed by the CaAGT activity of SlHCT, following 5-CQA synthesis by CcHQT.



eFP by R. Patel. Drawings for experiment "B" by R. Patel and also based on images provided by Zhangjun Fei and Jocelyn Rose at Cornell University for experiment "A". Data for "A" from Tissue- and Cell-Type Specific Transcriptome Profiling of Expanding Tomato Fruit Provides Insights into Metabolic and Regulatory Specialization and Cuticle Formation: Matas, A.J., Yeats, T.H., Buda, G. J., et al. Plant Cell 2011; 23(11):3893 - 3910. Data for "A" is 454-derived and FPKM-Normalized. Data for "B" from The tomato genome sequence provides insights into fleshy fruit evolution: Tomato Genome Consortium. Nature 2012 485 (7400): 635 - 641. Data for "B" is Illumina-derived and RPKM-normalized.

Figure 3.19. *In silico* analysis of HCT transript levels (Solyc03g117600) in tomato. Data obtained from eFP browser. (An electronic Fluorescent Pictograph browser for exploring and analysing large scale data sets). RPKM, (Reads Per Kilobase of transcript per Million mapped reads) are shown as Log2 ratio changes. A red colour indicates high and blue, low RPKM reads. HCT shows highest RPKM in the roots of *S.lycopersicum* and low RPKM in reads of toamto fruit. (http://bar.utoronto.ca/efp_tomato/cgi-bin/efpWeb.cgi) [108].

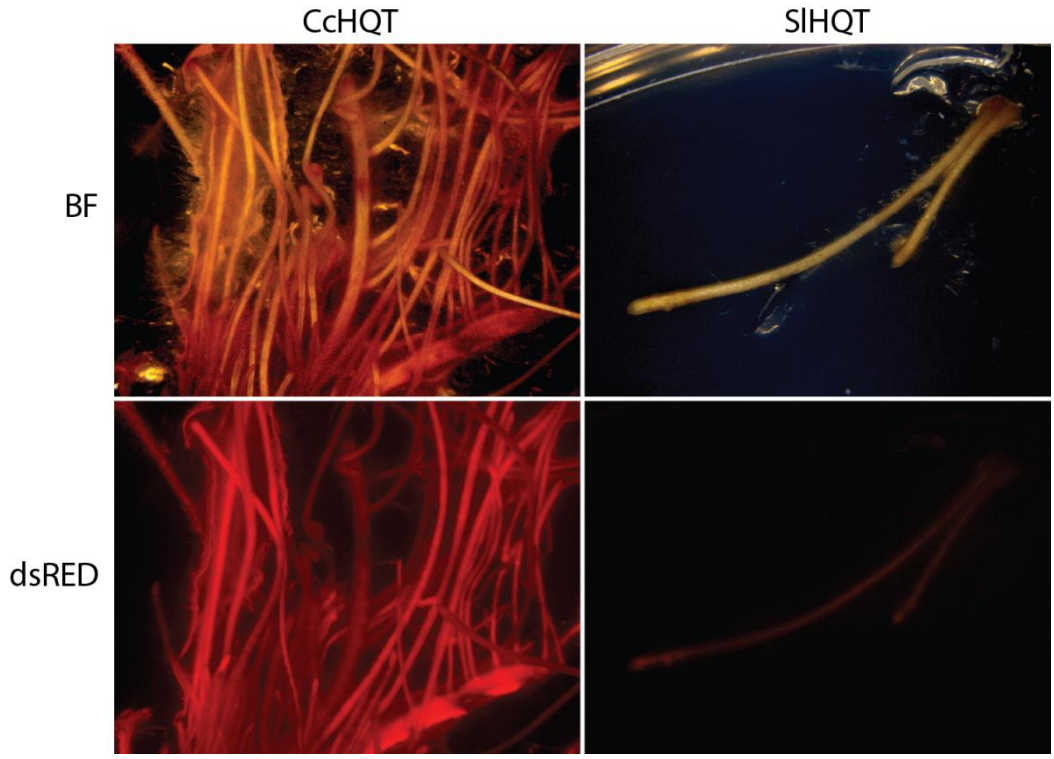


Figure 3.20. Difference observed in the intensity of dsRED expression in hairy root transformations of CcHQT and SlHQT in HQT KO tomato. Roots expressing CcHQT showed a more intense red colour compared to roots expressing SlHQT when observed in bright field mode (BF), and at 575 nm (dsRED).

When CcHCT and SlHCT were overexpressed in the HQT KO roots, the amounts of chlorogenic acid measured in the hairy roots did not change. Although 5-CQA was not quantified accurately in the hairy roots, it would be very difficult to calculate small changes in the presence of 5-CQA due to sensitivity problems using mass spectra to quantify trace amounts of metabolites.

In addition to overexpressing HQT and HCT from tomato and coffee in hairy roots of tomato HQT KO plants, I also transiently overexpressed the enzymes in the leaves of the HQT KO plants (supplementary figures). The overexpression of CcHQT and SlHQT caused the production of detectable levels of all 3 isomers of mono-caffeoylquinic acid (3/4/5-CQA). I also observed a peak which might have corresponded to diCQA, although this peak was also present in the HQT KO control. This means that it probably was not diCQA. The reason why it was not possible to formally identify whether this peak was 3,5-diCQA was because data were acquired by HPLC, not LC MS, and therefore retention time and PDA data at 325 nm were used for the analysis. If 3,5-diCQA was indeed present following CcHQT and SlHQT overexpression in HQT KO leaves, coupled with the fact that all 3 isomers of CQA were present and more was detected than in the hairy root transformations, these data would support the hypothesis that diCQA biosynthesis is directly correlated to CQA levels, and that diCQA synthesis was likely predominantly associated with the CCT activity of HQT in tomato. Determination of the kinetic parameters for the CaAGT activity of HCT and comparison to those for the CCT activity of HQT and HCT might resolve which enzyme activity is primarily responsible for diCQA biosynthesis, although unfortunately time constraints did not allow these experiments to be completed.

3.4 Discussion

A functional knock-out of HQT in tomato caused the complete loss of chlorogenic acid accumulation in tomato fruit and the almost complete loss of chlorogenic acid accumulation in the roots, where only traces of 5-CQA were observed. The traces of 5-CQA observed in the HQT KO roots were likely correlated to the high levels of HCT expression in WT tomato roots. As HCT has been shown to have low affinity for the BAHD reaction of quinic acid with coumaroyl CoA/caffeoyl CoA *in vitro,* the trace levels of 5-CQA observed in the HQT KO hairy roots were likely to be caused by high levels of native HCT.

Using this HQT KO line I showed that HQT is almost entirely responsible for chlorogenic acid biosynthesis because 5-CQA was undetectable in nearly all tissues of the KO line. This supports earlier reports where silencing of HQT in tomato caused an almost complete loss of CGA accumulation [57]. Although 5-CQA levels were not quantified in the HQT KO roots, the role of HCT in the biosynthesis of CGA *in vivo* is small, and detectable levels of 5-CQA were present in roots of the KO line only where HCT expression was high. Expression of SlHQT in HQT KO tomato roots restored high levels of 5-CQA, whereas expression of SlHCT did not, confirming the predominant role of HQT in 5-CQA synthesis in vivo. The overexpression of CcHQT in tomato HQT KO introduced significant synthesis of 5-CQA. This showed that CcHQT has functional identity to SlHQT in the synthesis of 5-CQA in tomato plants. I also detected trace levels of diCQA synthesis following the overexpression of CcHQT, which were not present in the roots of the HQT KO plants. This showed that the introduction of CcHQT into tomato plants lacking native functional HQT activity reintroduced the synthesis of diCQA. It is not clear at this point the mechanism(s) responsible for this reintroduction of diCQA synthesis as native HCT is expressed highly in tomato roots. Therefore, it is possible that diCQA synthesis through CcHQT overexpression was caused by the CCT activity of the overexpressed HQT, or caused by the native HCT, through its CaAGT activity.

Assays of crude extracts of *E.coli* overexpressing the BAHD acyltransferases HCT and HQT from coffee and tomato confirmed earlier reports that diCQA synthesis can occur through different enzymatic mechanisms. Chlorgenate:chlorgenate transferase (CCT) activity occurs under acidic conditions using high concentrations of 5-CQA as a substrate where it acts as both the acyl acceptor and acyl donor and is associated with both HCT and HQT enzymes. With the newly-identified **Ca**ffeoyl co**A** chlor**G**enate **T**ransferase (CaAGT) activity, caffeoyl CoA acts as the acyl donor and 5-CQA as the acyl acceptor, directly. Purification of the enzymes CcHCT and CcHQT and repetition of the enzyme assays showed that diCQA biosynthesis through these mechanisms occurs independently of other cofactors.

I have shown that CCT activity for the synthesis of diCQAs is not limited to SlHQT, but is also a property of SlHCT, CcHQT and CcHCT. The CCT activities of CcHQT and CcHCT had very similar catalytic properties. CcHCT showed a lower affinity for 5-CQA at low pH for CCT activity (K_m, 8.297 ± 3.2961 mM 5-CQA) than CcHQT (K_m, 3.8942 ± 1.012 mM 5-CQA). However, the standard errors indicate that the differences observed are unlikely to be significant. Future repetition of the assays will confirm whether any differences are significant. While the relatively low affinity of these enzymes for 5-CQA when acting as a CCT could exclude it from having *in vivo* activity in the cytoplasm, the CCT activity could be favoured when the when the 5-CQA levels in the vacuole are high. This CCT activity is likely to be biologically relevant only when the enzymes are localised in the vacuole of tomato and coffee cells where an acidic pH prevails. Vacuolar localisation of CCT activity would favour diCQA biosynthesis if 5-CQA levels are high in the vacuole (mM range). I showed that CcHQT and CcHCT could localise to the vacuole as well as to the cytoplasm where acidic and neutral pH prevail respectively. The duel subcellular compartmentalisation of CcHQT and CcHCT indicates that CCT activity can occur *in vivo*.

HQT and HCT in coffee could act as BAHD acytransferases at a neutral pH when localised in the cytoplasm in the presence of aromatic acyl-CoA donors. By contrast, they could have CCT activity in the vacuole at an acidic pH in the absence of aromatic acyl-CoA donors and the presence of high concentrations of 5-CQA. After high concentrations of 5-CQA have accumulated through BAHD activity it is possible that CcHQT and CGA get sequestered into the vacuole to carry out diCQA synthesis via their CCT activity (Figure 3.21). This possibility is supported by the close association between levels of CQAs and levels of diCQAs in different coffee varieties at different stages of coffee grain development [24]. However, the newly identified CaAGT activity of HCT *in vitro*, indicates that this enzyme could synthesise diCQAs in the cytoplasm. The low V_{max} observed for CCT activity of HQT and HCT could limit the biological relevance of CCT activity in some organisms but in coffee fruit, which develops over a period of months, this level of activity could be significant. These data all suggest, as has been suggested for SlHQT, that CcHQT is able to perform two different functions in separate subcellular compartments with contrasting environments. The dependency of CCT activity on high vacuolar 5-CQA levels suggests that HCT is unlikely to contribute significantly to diCQA biosynthesis through its CCT activity, except in tissues where it is highly co-expressed with HQT.

Through the incubation of recombinant HCT from tomato and coffee with caffeoyl CoA and 5-CQA, I could identify a new activity for diCQA biosynthesis at neutral pH. This catalytic activity has been given the name **C**affeoyl Co**A** chloro**g**enate **T**ransferase (CaAGT) activity where 5-CQA acts as the

acyl acceptor and caffeoyl CoA the acyl donor. This contrasts to the standard BAHD activity where quinic acid acts as the acyl acceptor and caffeoyl CoA is the acyl donor (as proposed in Figure 3.6). It should be noted that the primary role of HCT is to form caffeoyl CoA in the shikimate shunt of monolignol biosynthesis using shikimic acid as the acyl acceptor and coumaroyl CoA/caffeoyl CoA as the acyl donors. At this point it is not clear whether the CaAGT reaction mechanism is biologically relevant as there was not time to perform kinetic analysis to determine the substrate binding affinity or the catalytic efficiency of the CaAGT activity of HCT.

Previously reported in coffee, the synthesis of diCQAs has been proposed to occur via an uncharacterised activity of HCT on 5-CQA and CoA [101]. By incubating CcHCT with caffeoyl CoA and 5-CQA, I could show that CcHCT has CaAGT activity which is likely to occur using caffeoyl CoA as the acyl donor as opposed to CoA and 5-CQA. Caffeoyl CoA and 5-CQA could both fit into the binding pocket of HCT where presumably 5-CQA acts as the acyl acceptor as quininc acid does in BAHD activity. The catalytic activity observed by HCT where CoA and 5-CQA are used as substrates to synthesise diCQA would have to undergo the reverse BAHD reaction first to synthesise caffeoyl CoA to be used as the acyl donor substrate for CaAGT activity.

Enzymes which exhibited CCT activity also caused the accumulation of isomers of diCQA, which was likely not to be due to spontaneous acyl migration as 3,5-diCQA remained stable in acidic conditions. These data suggest that, in addition to HCT and HQT being able to synthesise 3,5-diCQA through CCT activity, they may be able to synthesise other diCQA isomers directly. This activity either happens directly from forward CCT activity, which would be able to synthesise two different isomers from 5-CQA, or by further activity from the enzyme where it can cause further acyl migration for the diversification of diCQAs. Although the reverse reaction was not investigated here, it has previously been reported that CCT activity through SlHQT is not reversible [102]. As the reaction is therefore likely irreversible, further isomerisation might occur directly through the interaction of diCQA with the enzyme as opposed to the reverse reaction occurring to create isomers of CQA.

I have found that BAHD acyltransferases, HQT and HCT, can both synthesise diCQAs *in vitro*. I confirmed that CCT activity, which has previously been reported for SlHQT [102], and an unidentified enzyme [58], is also exhibited by CcHQT and CcHCT. This activity is likely to occur when the enzymes are localised to the vacuole, in the presence of high concentrations of 5-CQA, where an acidic pH prevails. In addition to this, I also observed the synthesis of diCQAs through an activity that has not been directly reported before which I have called **Ca**ffeoyl coenzyme **A** chloro**G**enate **T**ransferase (CaAGT) activity. This CaAGT activity is the property of SlHCT and CcHCT, with the

substrates, caffeoyl CoA and 5-CQA at a neutral pH, when the enzymes are localised to the cytoplasm (Figure 3.21). This study identified two plausible pathways for diCQA synthesis in coffee and suggests mechanisms by which the diversification of diCQAs may be achieved.

I observed diCQA production in the overexpression of CcHQT in the hairy root HQT KO. I could not identify peaks which corresponded to diCQAs when I overexpressed CcHQT in HQT KO leaves. Due to there being two different pathways responsible for the synthesis of diCQAs, CCT activity occurs primarily through CcHQT and SlHQT, and CaAGT activity occurs through CcHCT and SlHCT. In the HQT KO roots, where high levels of native HCT is expressed, diCQA synthesis is more likely to occur through the CaAGT activity native HCT, rather than CCT activity through overexpressed CcHQT. This observation is supported by the fact I could not identify diCQAs in the HQT KO leaf overexpression where much lower levels of native HCT is transcribed.

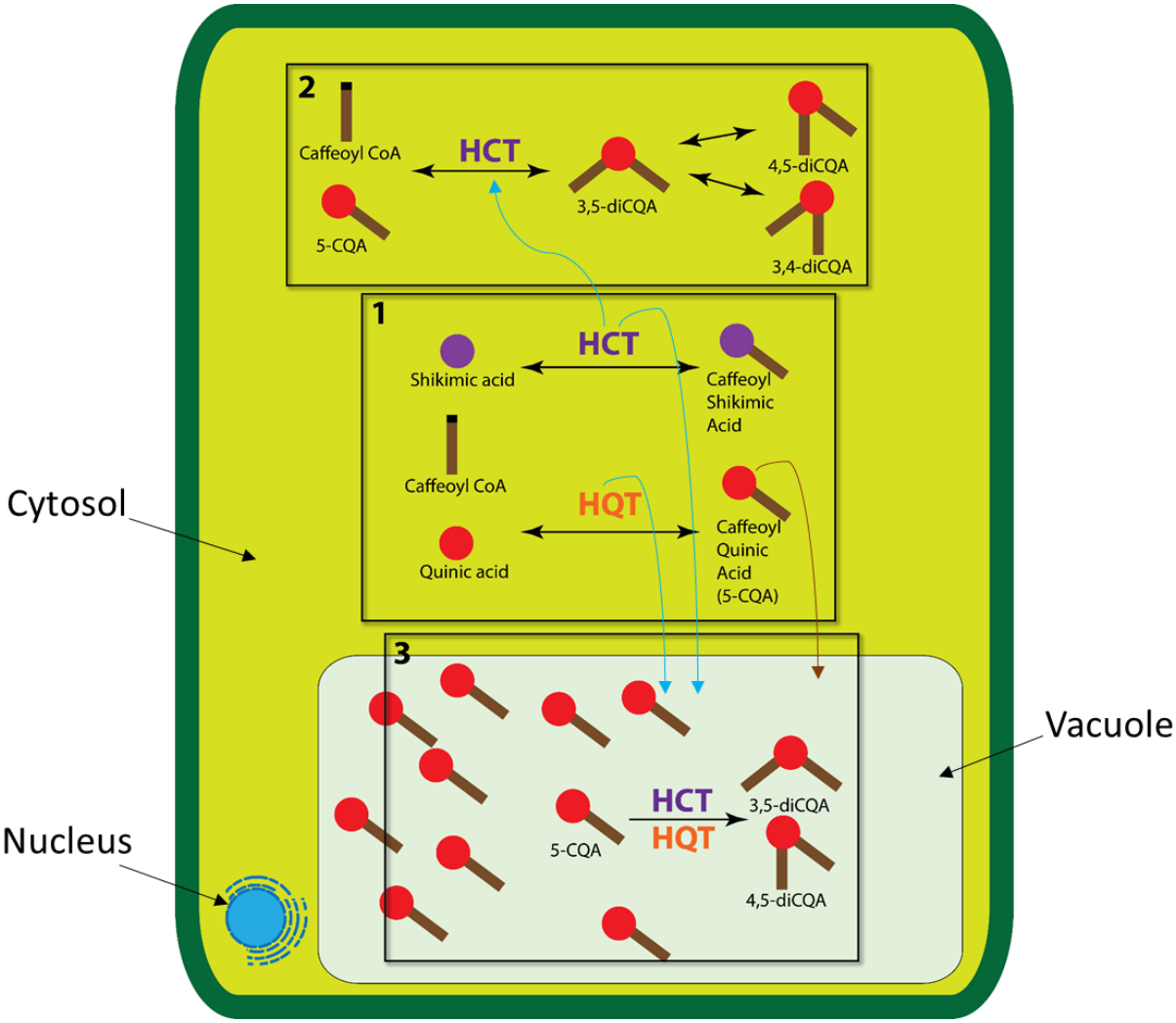


Figure 3.21. Proposed reactions schemes of BAHD, CaAGT and CCT activity in different subcellular compartments with proposed movement of selected enzymes and compounds where HCT and HQT are expressed in the same cell. 1. BAHD activity of HCT and HQT catalysing the formation of caffeoyl shikimate and 5-CQA respectively. HQT and HCT and 5-CQA sequester into the vacuole over time (blue and brown arrows). 2. CaAGT activity of HCT for diCQA synthesis in the cytosol using caffeoyl CoA and 5-CQA. 3,5-diCQA can undergo spontaneous acyl migration to form isomers. 3. CCT activity of HCT and HQT when localised to the vacuole when high concentrations of 5-CQA exist in the synthesis of diCQAs.

Chapter 4: Spatial localisation of chlorogenic acids in developing coffee

4.1 Introduction

Coffee beans are elliptical, plano-convex in shape and consist mostly of endosperm tissue which makes up around 99% of the mature seed mass. The endosperm tissue is a major storage tissue and is composed of living cells which undergo programmed cell death once their reserves have been mobilised during seedling growth. The relative contribution of the endosperm to the final mass of the mature seed varies greatly depending on the species. Some like soybean and rape do not have endosperm, others like Arabidopsis have a small layer of endosperm surrounding the embryo. Other seeds like coffee, castor bean, and palm oil, have endosperms which make up most of the seed volume [109]. Triploid coffee endosperms like *Arabica* mainly consist of cell wall polysaccharides (48 – 60% dry matter), lipids (13 – 17% dry matter), proteins (11 - 15% dry matter), sucrose (7 – 11% dry matter) and chlorogenic acids (5 – 8% dry matter) [109]. There is no other known plant species that has been reported to accumulate such high levels of chlorogenic acid during development of its seed [109].

The chlorogenic acid content in coffee beans is directly related to the bitterness of the beverage after roasting [32, 110]. The main classes of chlorogenic acids are, caffeoylquinic acids (CQAs), dicaffeoylquinic acids (diCQAs), feruloylquinic acids (FQAs), and *p*-coumaroylquinic acids. Chlorogenic acids are diverse within coffee, over 45 have been discovered in *Arabica* and 69 in *Robusta* [15, 20]. The biological roles of chlorogenic acid are to protect the plant from UV damage, defence against bacterial and fungal infections, antioxidant radical scavenger, and damage from herbivores, it may also be an intermediate in lignin biosynthesis [57, 111-113].

Histochemical localisation of chlorogenic acids in developing *Robusta* leaves has suggested that CGAs have different functions due to differences in their accumulation patterns during leaf development. In young leaves, CGAs associate around the chloroplasts, suggesting a protective role against UV damage, whereas in older leaves, CGAs localise to the vascular tissue. Mondolot et al., (2006) concluded that CGAs may be acting as precursors for lignin formation and that they are transported from young leaves after they have performed their protective roles [113]. They used *Neu's* reagent staining combined with microscopy. *Neu's* reagent will stain all polyphenols present, and therefore cannot be used as definitive evidence that CGAs are mobilised for lignin biosynthesis [113, 114]. The primary evolutionary function of chlorogenic acids, was likely to protect the plant from UV light exposure and pathogen infections, although a second evolutionary role as a storage compound for lignin biosynthesis cannot be excluded [68, 109, 113].

Ongoing research suggests that 5-CQA is not an intermediate for monolignols. In the previous chapter, we characterised a knockout mutant of the enzyme HQT in tomato which resulted in almost the complete loss of chlorogenic acid accumulation. It has been observed that the HQT KO plants did not show the stunted growth which is characteristic of a reduction in lignin synthesis or loss of HCT activity (D'Orso, *personal communication*)[63, 84].

Recently, mass spectral imaging (MSI) showed the relative accumulation of caffeoylquinic acid (CQA) in processed green coffee beans, i.e the beans were picked and dried using different methods. They showed that different post-harvest processing techniques influenced CQA localisation and that CQA localised to areas which had been attacked by the coffee berry borer [115]. These authors concluded that the unusually high levels of CGA in coffee beans were probably related to the protection of the seed during fruit development and dormancy. The spatial localisation of different compounds during fruit development in coffee is a relatively unexplored field, given the content and diversity of chlorogenic acid compounds.

Previously, I showed that HQT is the gene responsible for 5-CQA biosynthesis. Once high levels of 5-CQA have built up in the vacuole, HQT also in the vacuole, would be able to synthesise diCQAs through its CCT activity. HCT was also able to perform the same CCT activity to synthesise diCQAs. CCT activity relies on high concentrations of 5-CQA. Therefore, the synthesis of diCQAs by CCT activity is most likely to occur via HQT when 5-CQA levels are high because of the BAHD activity of HQT. Where 5-CQA levels are lower or declining (implying lower HQT activity) diCQAs are more likely to be formed by the CaAGT activity of HCT. HCT can synthesise diCQAs via another pathway, at neutral pH, presumably in the cytoplasm using 5-CQA and caffeoyl CoA to form diCQAs. Although my results for this activity are preliminary, they do, however, suggest another biological route for the synthesis of diCQAs, which might not rely on the high concentrations of 5-CQA. Understanding the spatial localisation of different chlorogenic acid groups may help to clarify further, the biosynthetic pathways for different chlorogenic acids.

4.1.1 Aims

I aimed to understand the spatial localisation of different chlorogenic acids and other hydroxycinnamic acids related to the phenylpropanoid pathway in developing coffee beans. To do this, I optimised a method whereby I could cut 15 μm transverse sections of coffee beans and fruit at different developmental stages which I fixed to a slide. Together with Dr Gerhard Saalbach, we optimised a method using mass spectrometry imaging (MSI) to resolve the spatial distribution of metabolites and their relative abundance in the transverse sections of developing coffee beans. The data were mined for hydroxycinnamic acids and polyphenols. Images were generated for targeted compounds, which revealed different tissue localisations for different compounds. The dicaffeoylquinic acids showed distribution patterns distinct from those of mono-caffeoylquinic acids and feuroylquinic acids showed further distinct distribution patterns. These data indicated different ratios for synthesis of different hydroxycinnamic acid compounds in developing coffee seeds that can be explained by differential use of HCT and HQT enzymes in synthesis. Interestingly in addition, I could not detect masses which corresponded to the m/z of flavonols, indicating that flavonols may not be present in *Robusta* coffee during fruit development.

4.2 Materials and methods

4.2.1 Cryosectioning of developing coffee beans

Coffea canephora (Robusta) fruit at five different developmental stages were received from Victoria Berry from Nestlé R&D centre, Tours, France. The samples were frozen in liquid nitrogen after picking from the trees and stored at -80°C before use. Fruits were stored whole to reduce the chances of cracking the inner seed at low temperatures. These different stages were named; very small green (VSG), small green (SG), large green (LG), yellow (Y), and red stage (R). Beans at different stages of fruit development were processed differently according to the developmental stage. Due to the delicate nature of VSG and SG beans, these samples were sectioned whole, whereas for the fruit at later stages, LG, Y and R the pericarp was removed before sectioning. The fruit were taken from storage and slowly brought up to a temperature suitable for processing. During the removal of the pericarp, the samples were kept as cool as possible (from 0 – 4°C), before they were mounted onto the cryosectioning stage. I tried to identify the end of the bean which did not contain the embryo and I cut a flat surface into it to create a greater surface area to bind to the stage.

The cryosectioning stage was placed on dry ice for 10 seconds and a small droplet of optimum cutting temperature (O.C.T) compound (Sakura®) was placed in the middle of the stage and the flat surface of the coffee bean was pressed into the rapidly freezing O.C.T compound. O.C.T compound fixes the sample to the stage and it has a freezing point of -10°C. I took care to ensure the frozen O.C.T compound was limited to the base of the bean as this can interfere with downstream applications of sectioned samples. From this point forward, beans mounted on the cryosection stage spoiled if the temperature rose above -10°C, so the samples were stored at -20°C until further processing.

A CryoStar™ NX70 Cryostat (ThermoFisher) was used to section the coffee beans. The mounted cryostat stage was secured onto the cryostat specimen head (Figure 4.1). I initially trimmed the coffee bean sections (100 μm), until I reached the area surrounding the embryo. If the embryo was not present (due to the orientation of the sample), an area was chosen where all of the folds of the endosperm were visible. Optimisation of sample sectioning was required because samples which were sectioned when they were too cold would splinter and samples which were too warm would roll, and flat sections were required for mounting onto the slide. Samples were cut (15 μm) and required different conditions, to ensure flat section for mounting, depending on the coffee sample sectioned. VSG samples were sectioned when the specimen head and the blade were at -15°C. The SG samples were sectioned at -15°C on the stage when the blade was cooler, at -20°C. I used the

same sectioning conditions for the final stages of development, LG, Y, and R, which were -15°C for the sample temperature, and -25°C for the blade temperature.

When the samples were cut by the cryostat, they could be pushed underneath the anti-roll plate (Figure 4.1). Mounting slides, made of indium tin oxide (ITO) for MALDI purposes, or a standard microscope slide for DESI, were brought close to the sample, and the difference in temperature between the slide and the sample caused the sample to stick to the slide. The samples were then pressed flat onto the slide and thaw mounted and dried on a 50°C heat block. Up to five different developmental stages were placed onto one slide. To check the samples were sufficiently fixed, the slides were placed in a vacuum chamber, which also removed any excess moisture.

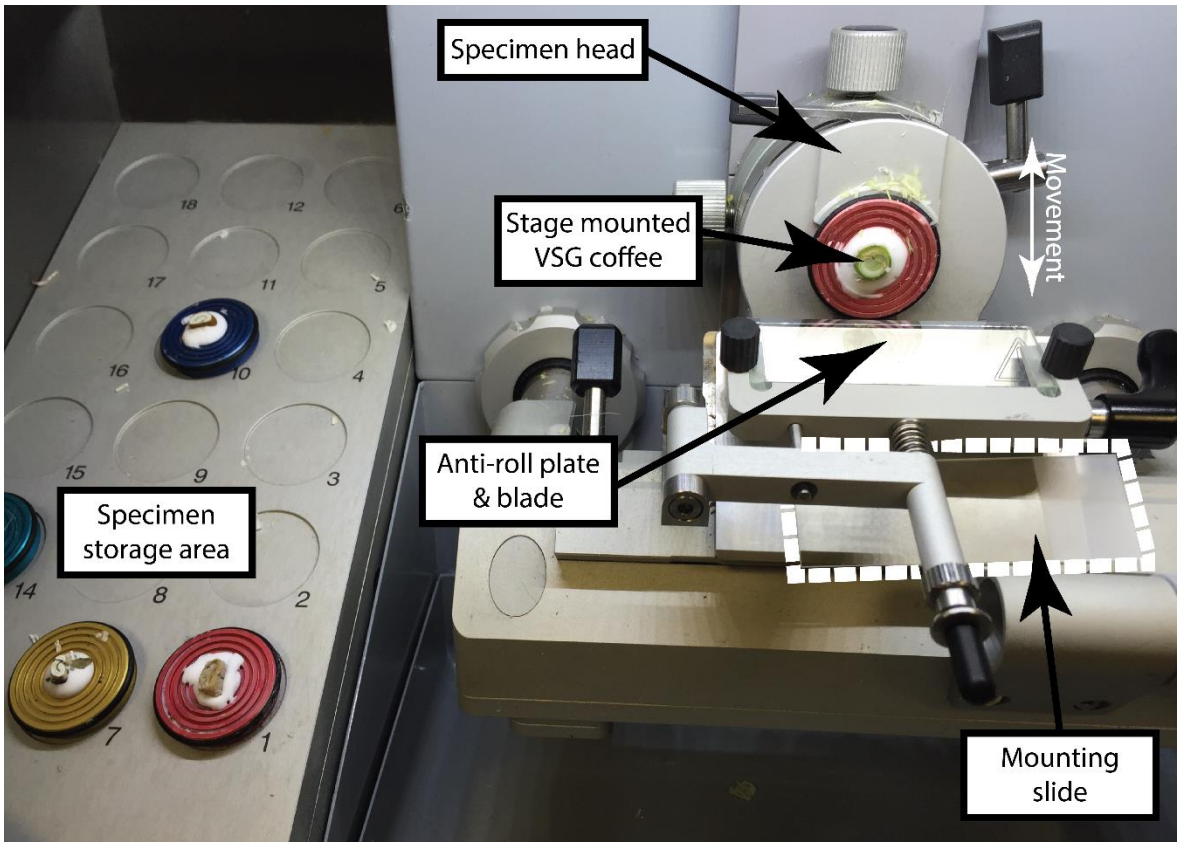


Figure 4.1 Annotated image of cryostat chamber for cutting cross-sections of developing coffee beans. Important components for the cross-sectioning of samples are labelled. The linear movement of the specimen head caused the stage mounted sample to be sectioned by the blade, sectioned samples are collected underneath the anti-roll plate. The mounting slide was brought close to the sectioned sample, the temperature difference caused the sample to stick to the mounting slide. The slide was brought up to room temperature and dried.

4.2.2 Mass spectrometry imaging (MSI)

Mass spectrometry based imaging (MSI) enables the visualisation of the spatial distribution of intact compounds across the surface area of tissue sections based on their molecular mass [116]. Two techniques of MSI imaging were used, matrix-assisted laser desorption/ionisation (MALDI) and desorption electron spray ionisation (DESI). We tried to optimise MALDI imaging first because of its increased lateral resolution (20 μm) but we came across problems with the matrix because we could not identify one that efficiently crystallised with a standard sample of 5-CQA. Failure of crystallisation resulted in little or no detection when the laser beam caused desorption of the matrix crystals. Desorption electron spray ionisation (DESI) was then used because it was easier to optimise than MALDI. This was because it did not require crystallisation of compounds within a matrix. However, it did have a disadvantage because the spatial resolution of DESI was lower than for MALDI. In addition, initial experiments with MALDI using 2,5-Dihydroxybenzoic acid (DHB) and α-cyano-4-hydroxycinnamic acid (α-CHCA) as matrices showed that the compounds of interest could not be detected using either matrix. With α-CHCA alone (no sample) a peak at m/z 353 (negative mode) was present which would interfere with CQA.

Therefore, experiments were performed with DESI. I assisted Dr Gerhard Saalbach of the John Innes Centre, Norwich, in the optimisation and acquisition of data using desorption electron spray ionisation (DESI). A Synapt G2Si mass spectrometer (Waters, Manchester, UK) was used for both MALDI and DESI imaging. For DESI imaging, a DESI-2D-Source (Model OS-6205) from Prosolia Inc. (Indianapolis, USA) was used. Peaks corresponding to CQA and many other compounds could be detected easily in negative mode. The Masslynx 4.1 and HDImaging v1.4 software (Waters) controlled the DESI acquisition. Methanol was used as the spray solvent and delivered through the auxiliary solvent manager of a nano Acquity UPLC (Waters) at 1.5 μl min^{-1} with a nebuliser gas pressure of 7 bar. The solvent contained LeuEnk peptide (1ng μl^{-1}) as a lock mass which was used for the calibration of mass spectra. The spray capillary was adjusted to an angle of 70°, a distance of 5 mm to the DESI capillary, and 3 mm to the surface of the sample. The pixel size was set to 100 μm and the stage was moved at a speed of 800 μm s^{-1} resulting in a scan time of 0.11 s. The Synapt G2Si instrument was calibrated with sodium formate and was run in negative TOF MS sensitivity mode over the mass range m/z 50-1200. The capillary voltage was set to 3.5 kV, the sampling cone to 50 V, and the source temperature to 150 °C. The acquired data were processed with HDImaging v1.4 software (Waters). Intensities were normalised by total ion chromatogram (TIC).

4.3 Results

I used the MSI technique DESI to capture spatial data for coffee fruit development using transverse cross sections. These data allowed me to visualise the relative abundance of metabolites, predominantly chlorogenic acids, detected by MSI in negative mode, localised to different cells and tissues of coffee fruit.

Five stages of coffee fruit development were mounted on cryosection stages. Due to the delicate nature of the VSG and SG fruit, the samples were prepared simply by cutting down the middle to split the embryo sacs, and the pericarp was left mostly intact. The pericarp, mesocarp and parchment were removed from LG, Y and R stages and silver skin remains were left on. Transverse sections (15 μm) were cut and placed onto a microscope slide in order of increasing maturity. Some of the acquired sections were from an area of the seed where the endosperm surrounded the embryo (Figure 4.2). If the embryo was not observed in the first quarter of the fruit, an area that represented the complete folding of the endosperm was sampled. The VSG section was taken from an area that represented the inner perisperm tissue and expanding endosperm tissue. The fruit pericarp tissue from VSG and SG samples appeared on final sections for analysis (Figure 4.2). The later development stages, LG, Y and R all showed remains of the silver skin on the outside of the folded endosperm (blue arrows Figure 4.2).

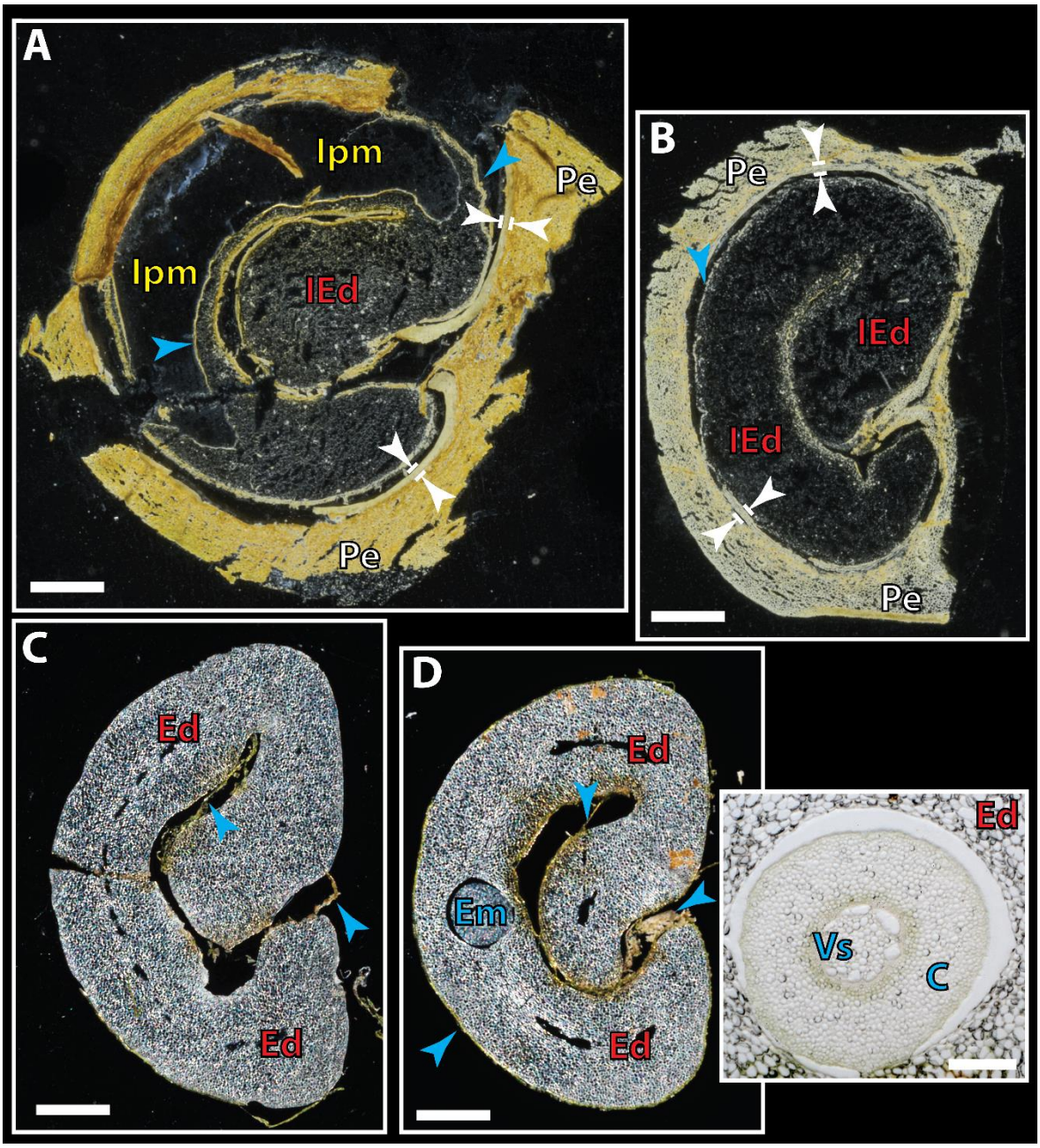


Figure 4.2 Transverse section of coffee beans at different stages of fruit development. A, very small green (VSG); B, small green (SG); C and D are representitive of stages late green (LG), yellow (Y), and red (R), example of a section without and with the embryo present repectively. Samples VSG and SG were sectioned with the pericarp (Pe) surrounding the imature endosperm (IEd) The white arrows in these samples represent the developing parchment layer and blue arrows point to the silver skin. The VSG stage also contained an area which represented the immature perisperm (Ipm). C showed that nearly everything present represents mature endosperm (Ed) with a few pieces of sliver skin present (blue arrows) present on the outside. D contains and area which was the embryo, an enlargened version showed that the vacular tissue (Vs) and the coretex (C) are also present. Bars: whole bean – 1000 μm, enlarged embryo – 250 μm.

I analysed 500 of the most intense peaks detected by MSI in negative mode from the surface of coffee fruit development (Figure 4.3). The most intense peaks corresponded to different chlorogenic acid groups and quinic acid, a central component in the synthesis of all chlorogenic acids (Figure 4.3). When analysed by mass spectroscopy in negative mode [ESI^-], quinic acid has an m/z of 191.0559, CQAs have an m/z of 353.0881, FQA have an m/z of 367.1034, and diCQAs have an intact mass of 515.1204. DESI MSI imaging is not able to distinguish between isomers of the different compounds, therefore these fragments corresponded to the totals for each chlorogenic acid group.

CQAs were present in all the developmental stages analysed. The youngest coffee fruit, VSG, showed a strong localisation to the inner perisperm, and CQA was not detected in the endosperm (Figure 4.4 A; arrows, white and black respectively). In the SG stage, CQAs were seen concentrated to the pericarp, and only traces of CQAs were present in the endosperm. When the developing coffee fruit reached LG, Y and R stages, CQAs localised strongly to the endosperm. CQAs were observed in the embryo at the LG stage, whereas they were less concentrated in the embryo at the R stage (Figure 4.4 A, green arrows).

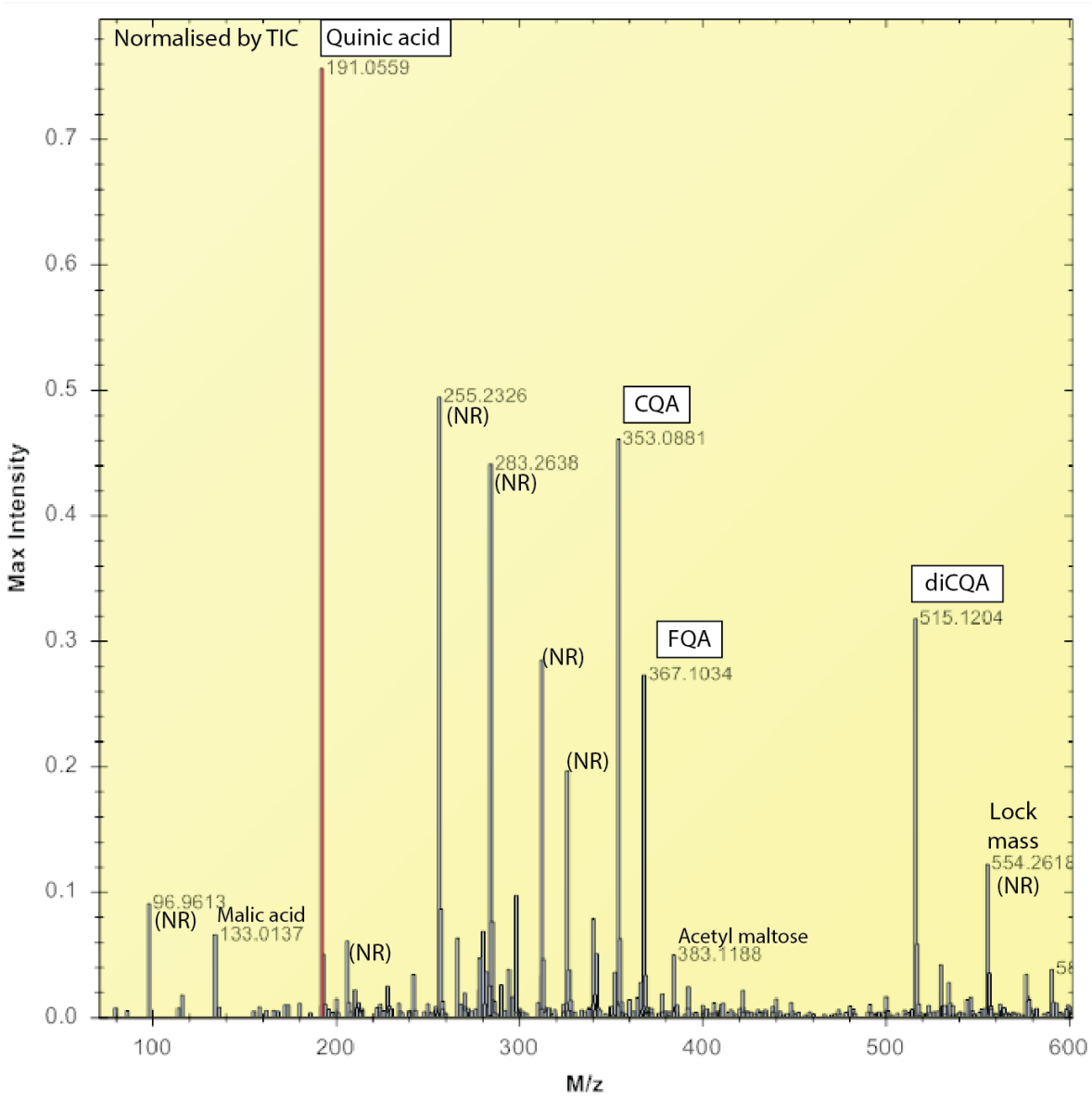


Figure 4.3 Mass spectrum representing the total ion chromatogram (TIC) of all detected fragments on the surface of coffee sections of fruit development. The most abundant peaks identified are labelled, NR represents fragments that are not relevant and they are either lock mass fragments or contaminants on the surface of the DESI stage, which were not on the microscope slide.

CQAs are esters of quinic acid and caffeic acid. Routes to their synthesis include the hydroxylation of *p*-coumaroyl quinate by the enzyme coumaroylshikimate 3' hydroxylase quinate (C3'H). A BAHD acyltransferase, hydroxycinnamoyl-CoA quinate hydroxycinnamoyltransferase (HQT) forms *p*-coumaroyl quinate using *p*-coumaroyl CoA as its acyl donor and quinic acid is an acyl acceptor. The other is where HQT will form caffeoylquinic acid directly, through the transesterification of caffeoyl CoA with quinic acid. I searched the MSI data for intermediates in 5-CQA synthesis by either route. I could detect fragments of the acylated precursors; there were no data for the intact compounds, which meant that they are either not present in the samples, or more likely, not detectable. I did not use the detected fragments for analysis because the fragments could belong to other compounds as all CGA components show similar fragmentation patterns whereby they lose caffeic groups attached to the central quinate readily when analysed by mass spectral analysis. Caffeic acid showed the same pattern of localisation as CQA compounds. CQAs were the most abundant compound detected and therefore most likely to show more fragments then other compounds. Consequently, the caffeic acid detected was likely derived from CQA.

I discovered that apart from the main chlorogenic acids and quinic acids, most other compounds related to CQA biosynthesis were present in very low levels. I detected caffeoyl shikimic acid (CSA), which was present in the perisperm that later forms the silver skin but none were detected in any area that accumulated CQA (Figure 4.4, C). This suggested strongly that HCT expression is restricted to the perisperm and that HQT expression is concentrated in the endosperm of late stage coffee beans. These ideas could be confirmed or refuted by *in situ* hybridisation of the HCT and HQT transcripts. The precursor of CSA, *p*-coumaroyl shikimate, was not detected in any of the samples, however the precursor to CQA, *p*-coumaroyl quinate was. It showed that during the first two stages of development, it co-localised with CQA whereas in the last three stages of development, L, Y, and R, it localised to areas where CQA was less abundant. At stage L, coumaroyl quinate showed stronger localisation to the embryo, whereas CQA was evenly distributed throughout the surface (Figure 4.4, A and B). At stage Y, coumaroyl quinate localised to the hard-outer region of endosperm while CQA localisation did not change. Stage R, coumaroyl quinate localised very strongly to the embryo, and CQA showed less localisation there compared to the endosperm. This change in coumaroyl quinate levels could be due to changes in the enzyme activity of C3'H.

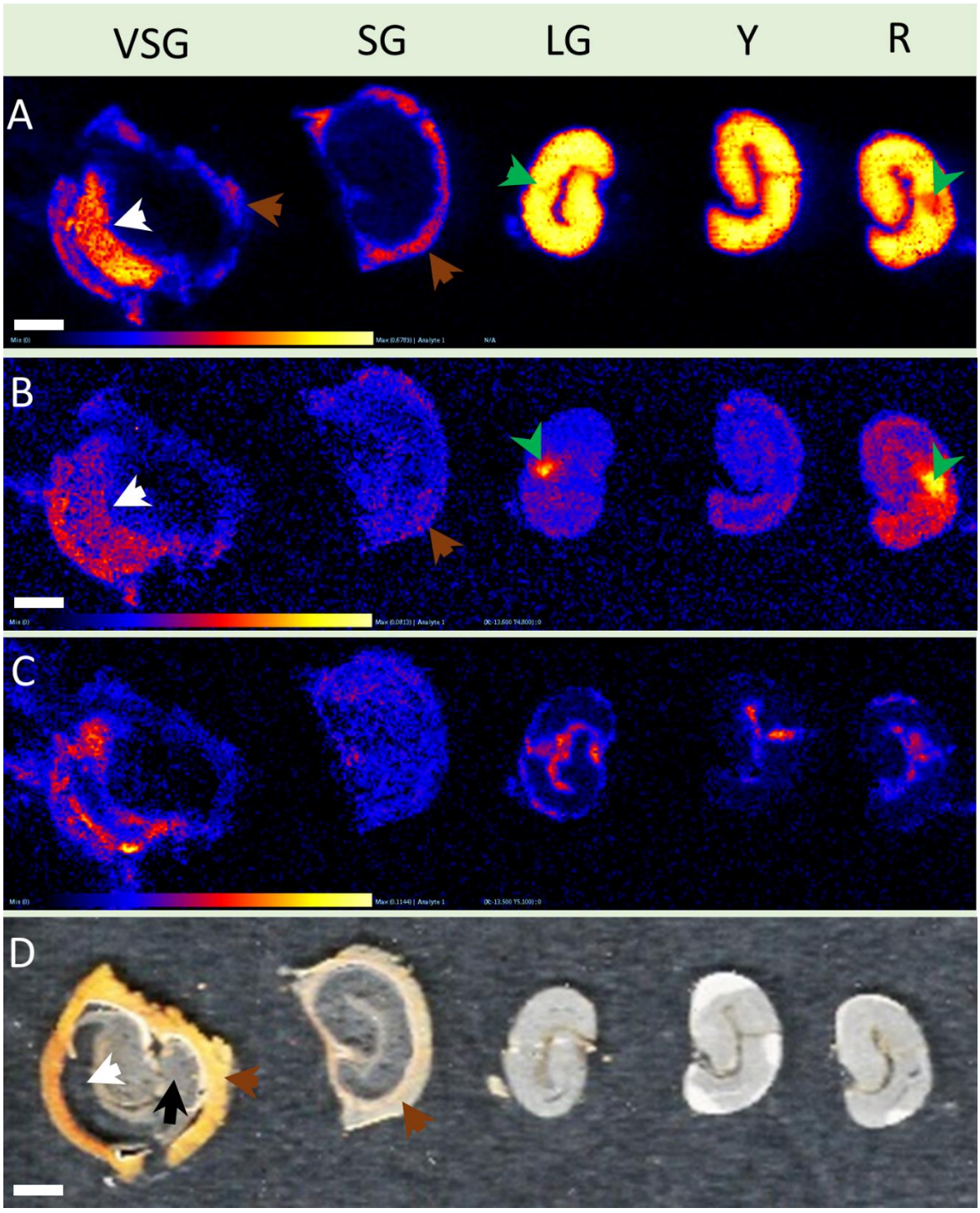


Figure 4.4 Caffeoylquinic acid, coumaroyl quinate and caffeoyl shikimate showed different localisation patterns in late coffee fruit development. A. MSI data for CQA (m/z 353.0881) B. MSI data for coumaroyl quinate (m/z 337.0924) C. MSI data for CSA (m/z 352.0793) D. Transverse section of beans used for MSI data. Bean sections are arranged per the stages of development (left to right) VSG, very small green; SG, small green; LG, large green; Y, yellow; R, red. In the early stages of development, CQA and coumaroyl quinate are proportionally localised to the same regions while in the last three development stages, they showed a different pattern of accumulation. CQA is predominantly localised to the endosperm at LG, Y, and R stages whereas coumaroyl quinate localises more to the embryo in LG, outer edge of endosperm at Y stage and the embryo at R stage. Caffeoyl shikimate (CSA) showed localisation to the inner perisperm in VSG and later in LG, Y, and R to the outer endosperm or silver skin (see Figure 4.2 for anatomy). Inner perisperm tissue (white arrows). Endosperm tissue (black arrow). Fruit pericarp tissue (brown arrows). Data was normalised by total ion chromatogram (TIC). Bars = 2 mm

Data for the relative abundance of quinic acid were also observed on a heat map. They showed that quinic acid levels generally drop throughout development (Figure 4.5 A). These observations confirmed the expression patterns previously described by Lepelley et al., (2007) where they found quinic acid levels dropped from 15.32% dry mater in SG stage to 0.73% in the R stage [24]. I compared quinic acid distribution in these samples to CQAs. They showed a clear inverse association in levels (Figure 4.5 A). Where quinic acid was high, chlorogenic acid was at low, the highest levels of quinic acid were in the endosperm of the VSG stage, where CQAs were not detectable (Figure 4.5). Although quinic acid is involved in many other plant metabolic pathways, the direct inverse associations between the two compounds suggested strongly that the quinic acid accumulating in the early stages of fruit development is used for the synthesis of CGAs at the LG stage.

I was also able to detect intact mass ions corresponding to shikimic acid (m/z 173.0452) which showed that shikimic acid levels rose throughout fruit development, an opposite pattern to quinic acid. Similar to quinic acid, shikimic acid was evenly distributed throughout individual stages and its high accumulation towards the end of fruit ripening could be required as a storage component to act a precursor for monlignol production during germination. Although shikimic acid is also used in a range of other metabolic processes, its pattern of accumulation may not be quantitatively connected to these hydroxycinnamic acid ester metabolic pathways.

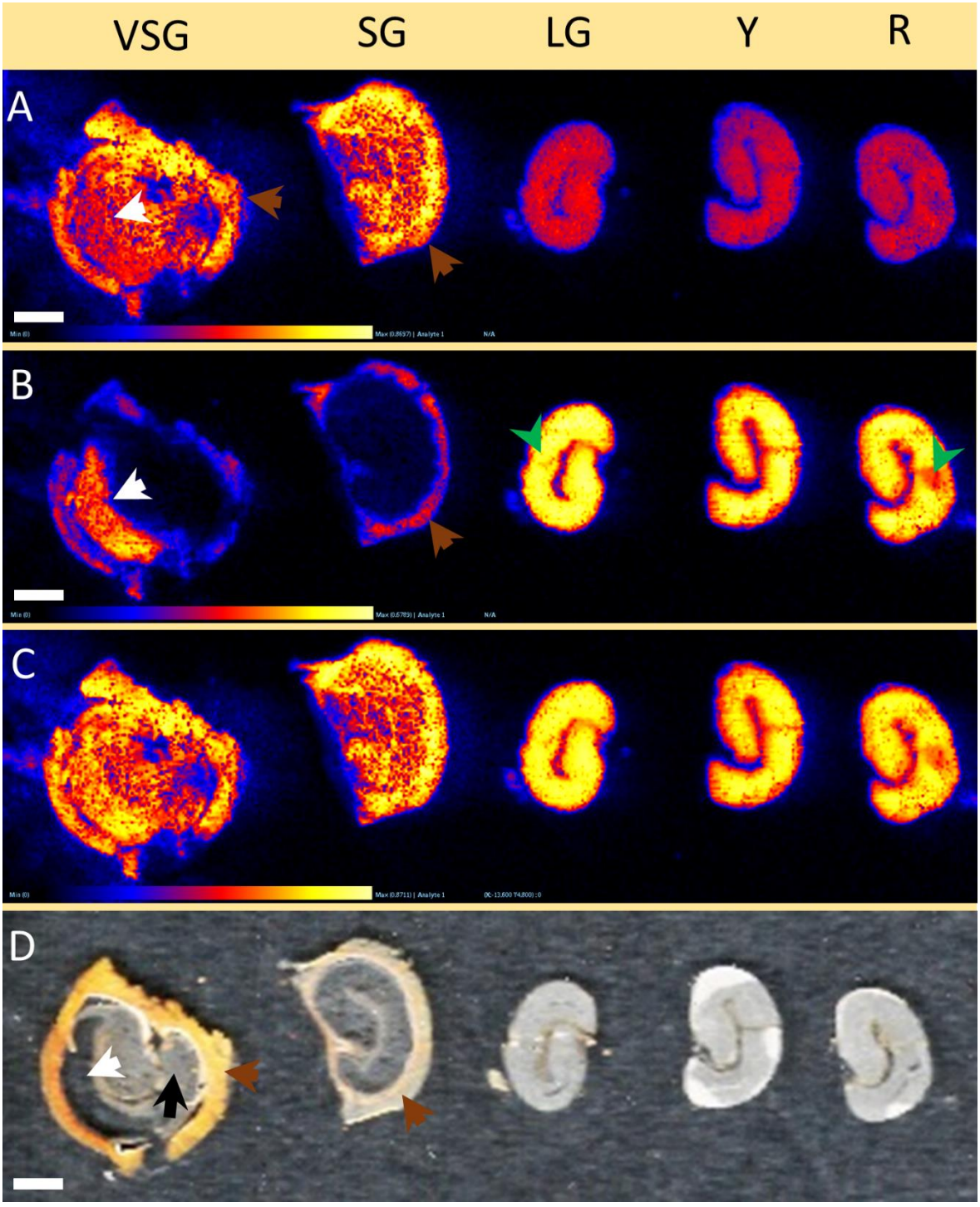


Figure 4.5 Localisation of quinic acid and CQA in coffee during fruit development analysed by MSI. A. MSI data for quinic acid (m/z 191.0559) B. MSI data for CQA (m/z 353.0881) C. Overlay of the two compounds normalised by the square root of intensity D. Cross section of beans used for MSI data. Bean sections are arranged per the stages of development (left to right) VSG, very small green; SG, small green; LG, large green; Y, yellow; R, red. In the early stages of development, quininc acid was present in VSG and SG stages, CQAs were present at trace levels in the endosperm. In LG, Y and R stages quinic acid levels drop and CGA levels are strongly elevated. In VSG, CQA localises to the inner perisperm (white arrow) whereas the quinic acid in the same region is not as concentrated. Where high levels of CQA are present, there are lower levels of quininc acid. Endosperm tissue (black arrow). CQA also localises to the fruit pericarp tissue (brown arrows) CQAs localise to the embryo in LG and then shows reduced localisation in R (green arrows). Data was normalised by total ion chromatogram (TIC). Bars = 2 mm

The dicaffeoylquinic acids were another abundant fragment in the MSI data. Their localisation was directly proportional to CQA localisation during the first three stages of fruit development (Figure 4.6, B vs A respectively). DiCQAs were present in the pericarp and perisperm in relatively high levels, whereas no intact ions were detected in the central area of the immature endosperm. The last two stages of development, showed some surprising results. DiCQAs showed stronger localisation in compartments to which CQAs were least concentrated. In fruit development stage Y, diCQAs localised to the harder outer endosperm region, whereas CQAs showed greater abundance in the soft centralised endosperm. Stage R, diCQAs showed a very strong signal in the area which represents the embryo, conversely CQAs localised least strongly to this area (Figure 4.6, B). DiCQAs were also highly located in the embryo at the Y stage, the samples analysed did not include the region where the endosperm surrounded the embryo. DiCQA shares the same localisation patterns as coumaroyl quinate (Figure 4.4, B). These data indicate that diCQA synthesis was not necessarily directly proportional to CQA concentration and suggests different enzymes might be contributing to diCQA biosynthesis in the latter stages of coffee fruit development compared to early stages. There is a clear switch in the metabolic relationship between the LG and Y stages, which could be due to changes gene expression or the metabolites used in different ways for the last stages of development.

Feruloylquinic acid (FQA) is another type of chlorogenic acid (CGA) which is found in coffee, FQAs generally make up the third largest group of CGAs in coffee. The biosynthesis of FQAs within coffee occurs through the caffeoyl CoA methyltransferase (CCoAMT), which is an important enzyme in lignin biosynthesis, and the conversion of caffeoyl CoA to feruloyl CoA leads directly to the synthesis of S and G lignin units [24, 117]. Feruloyl CoA acts as an acyl donor to quinic acid catalysed by the enzyme HQT which forms FQA, from there, other FQA related compounds are synthesised such as diferuloylquinic acids (diFQAs). The spatial localisation of FQA and diFQA revealed that they do not follow the same accumulation patterns as either CQA or diCQA during the late stages of development (Figure 4.6). Once again, the pattern of accumulation for the first two stages of fruit development in coffee was the same as for CQA and diCQA. FQA started to show differential accumulation in the LG stage where it showed relatively stronger localisation in the outer endosperm region. Then in Y and R stages, FQA and diFQA strongly localised to the outer endosperm (Figure 4.6, C and D). This was completely different to the pattern of accumulation observed for either CQA or diCQA. DiCQA did localise to the outer endosperm but did show strong localisation to the embryo at R stage whereas FQA did not localise strongly to this region at any stage. DiFQA was not abundant, trances of diFQA localised to the same region as FQA in the later stages of development but in much lower levels. CQA localised throughout the whole endosperm

and did not show preferential localisation towards any part of it (Figure 4.6, A). These differential localisation patterns of different chlorogenic acid groups have not been observed before in developing coffee fruit. Many different pathways and enzymes exist for the synthesis of these compounds, and changes in gene expression can cause differential accumulation. In addition, compounds could be transported to different compartments.

A closer inspection of the cells showing differential accumulation of chlorogenic acids showed that localisation of diCQA and FQA could be linked to the size of cells in the endosperm and the embryo. I observed a higher concentration of diCQAs and FQAs in the larger cells towards the outside of the endosperm. The cells which were near the outside of the endosperm tissue, were larger than the cells which were on the inside of endosperm, cells of the embryo were smaller still (Figure 4.7). In addition to this, in the final stages of fruit development, diCQAs were localised strongly to the embryo. The concentration of diCQAs in the embryo in stage R were even higher than the concentration of CQA in the embryo. This observation was not quantifiable, however, due to the different factors which limit the ability to quantify concentrations of different compounds using this technique.

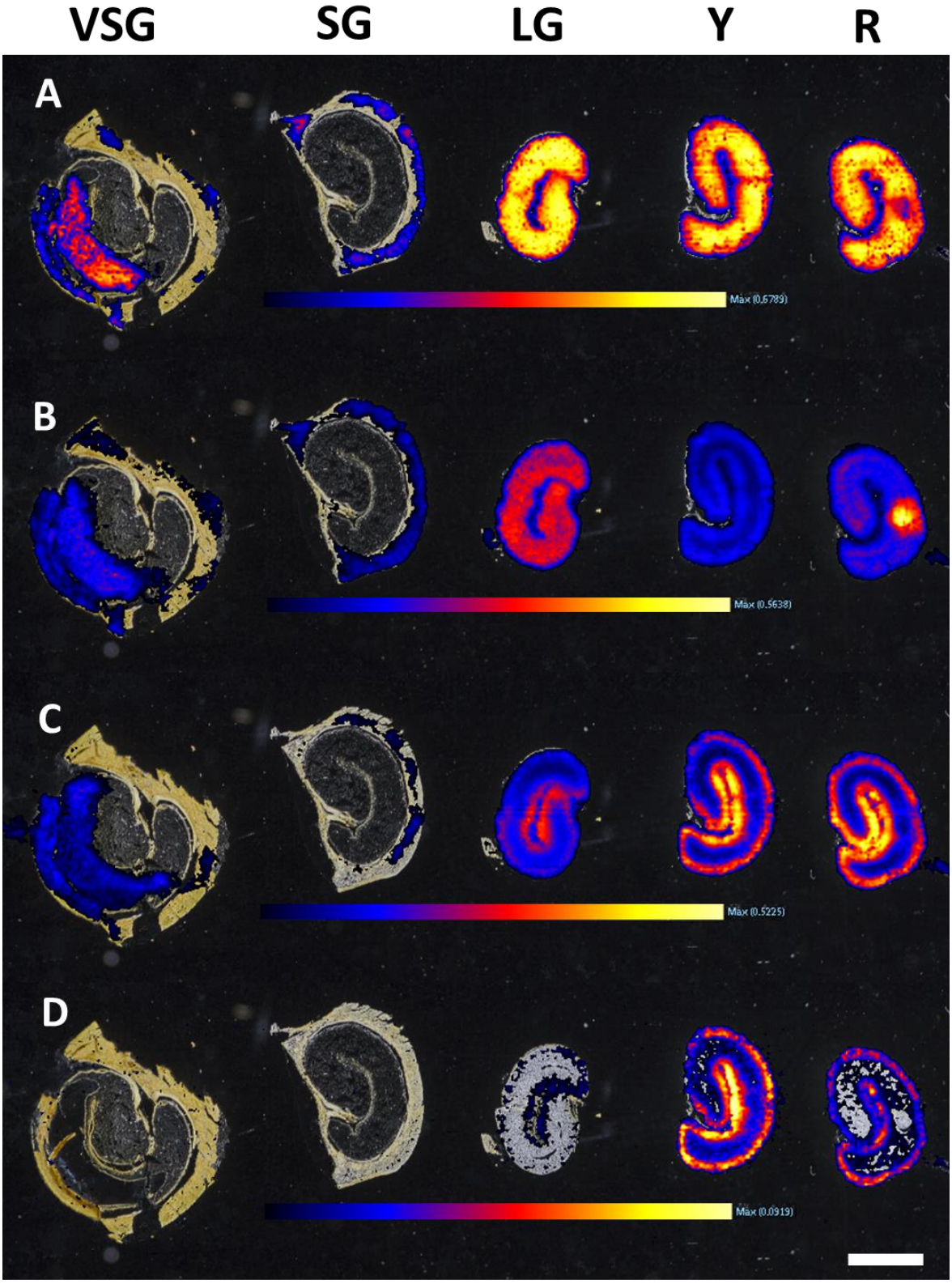


Figure 4.6. The major groups of chlorogenic acid present in coffee all show differential localisation patterns during late stages of fruit development. A. MSI data for CQA (m/z 353.0881) B. MSI data for diCQA (m/z 515.1204). C. MSI data for feruloylquinic acid (FQA) (m/z 367.1034). D. MSI data for diferuloylquinic acid (diFQA) (m/z 543.1508). Bean sections are arranged per the stages of development (left to right) VSG, very small green; SG, small green; LG, large green; Y, yellow; R, red. CQA, FQA and diCQA all showed similar localisation patterns in VSG and SG stages of fruit development whereas in LG, Y and R stages, CQA consistently localises predominantly to the endosperm. FQA and diFQA localised to the outer endosperm, and diCQA showed even localisation in LG, outer embryo towards the silver skin in Y and predominant embryo in R. Data was normalised by total ion chromatogram (TIC). Bar = 3 mm

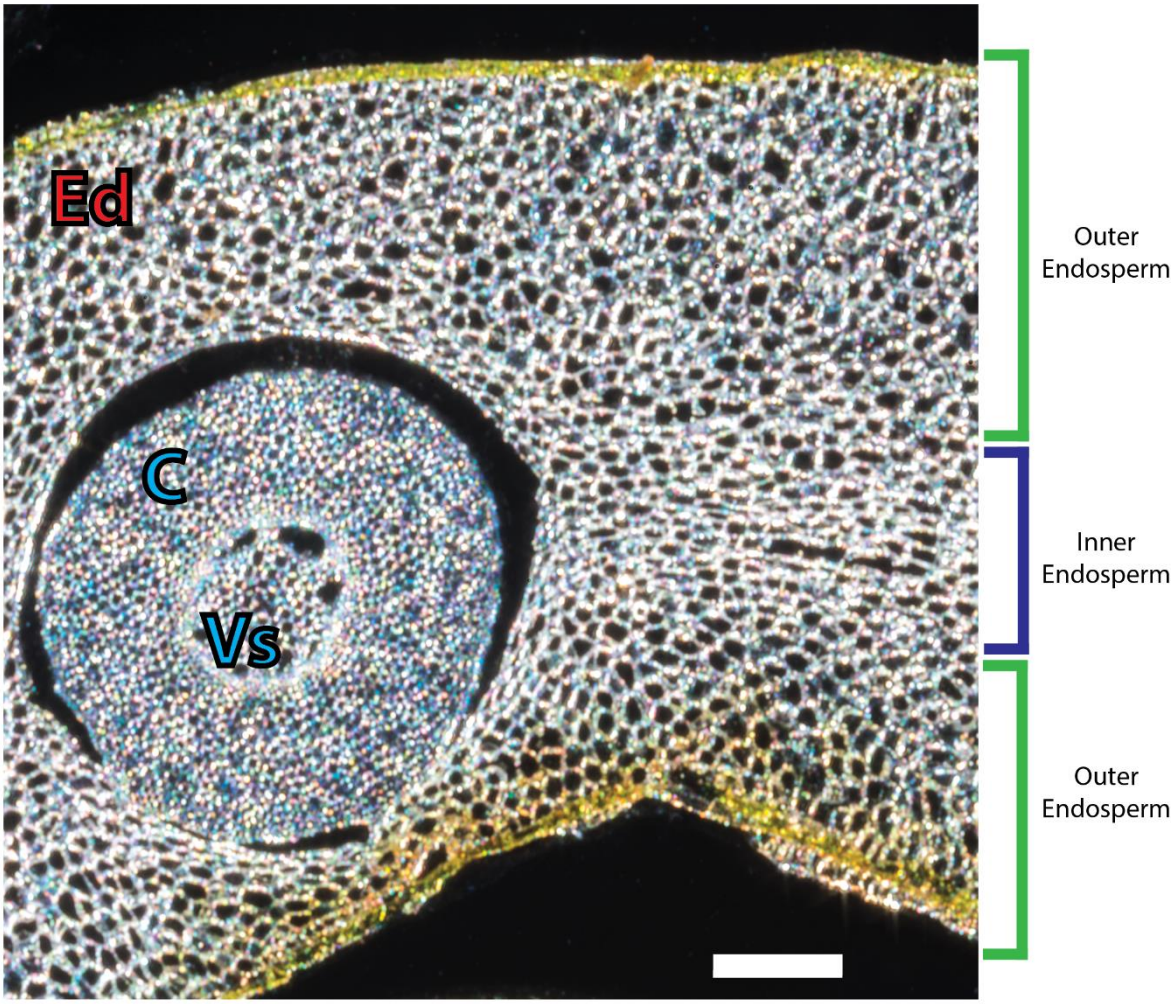


Figure 4.7 Cells representing the inner endoperm, outer endosperm and embryo showed different sizes in later stages of development. Representitive of cross sections from LG, Y and R stages of fruit development in coffee. The outer endosperm cell size is the largest, followed by the inner endosperm cells. The smallest cells are present in the embryo, the cortex (C) and the vascular tissue (Vs) are labbelled within the embryo. Bar = 250 um.

In addition to searching for different chlorogenic acid groups and their precursors, I mined the data for the presence of flavonols which have been reported to be absent or at very low levels in coffee [18]. The masses, which I selected to search for different flavonols, were chosen by identifying the m/z of different flavonol compounds detected in [ESI^-] from different articles [18, 93, 104, 118]. I could not detect any flavonol compounds present on the surface of coffee during fruit development. Although others have reported being able to detect traces of glycosylated quercetin present in the fruit husk and pericarp, I could not detect any quercetin or quercetin derivatives in these tissues, which were present in the two earliest developmental stages I sampled.

The absence of flavonols in developing coffee beans suggested a block on flavonoid biosynthesis. However, other flavonoids were detected in the pericarp tissue of young beans. Signals for, epicatechin and flavan 3-ol derivatives were detected in this tissue suggesting that the flavonoid pathway was active there. A catechin was detected in the embryo at late developmental stages, but none were detected in endosperm at any stage (Figure 4.8).

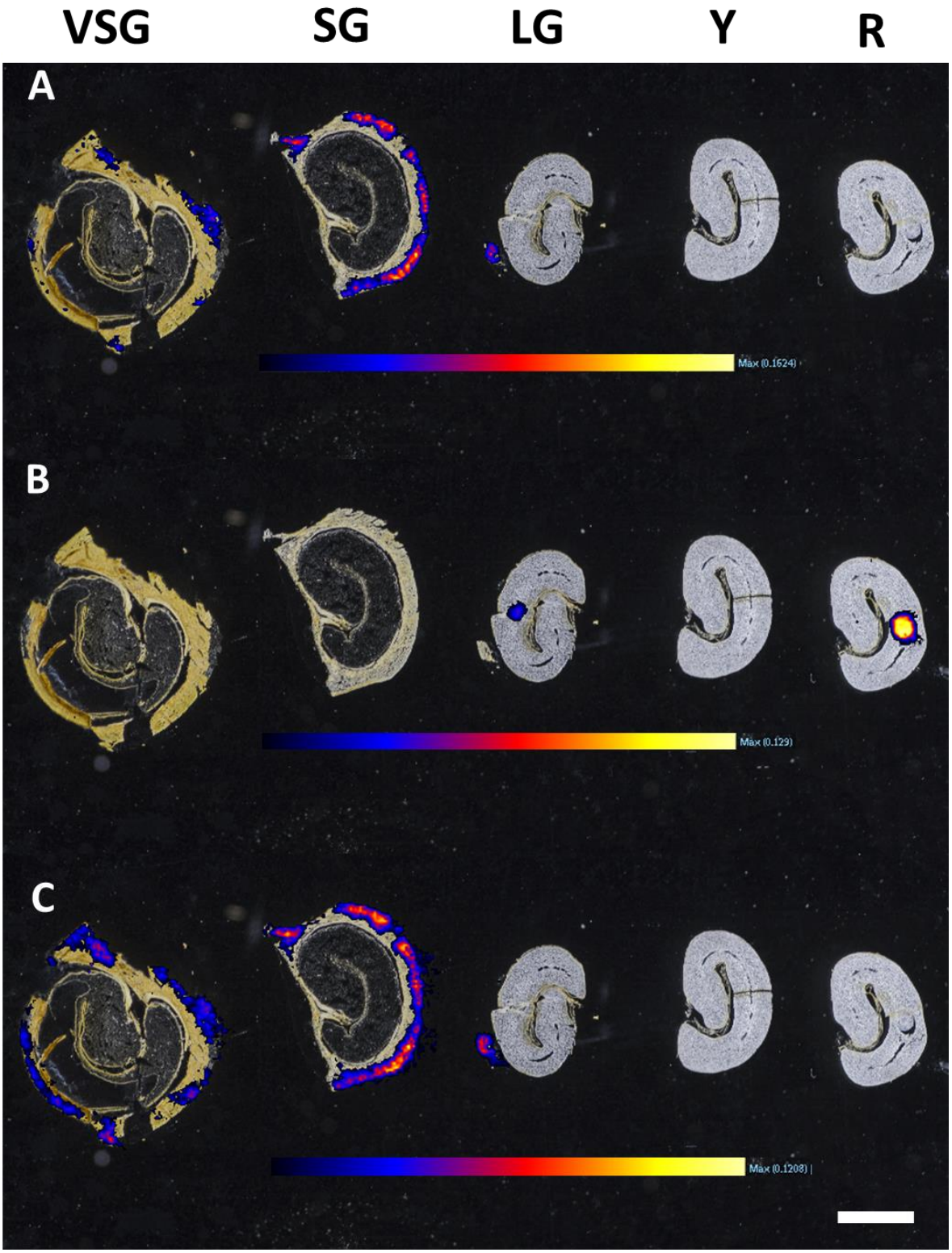


Figure 4.8. Epicatechin, catechin and procyanidin mapped in developing coffee beans during fruit development. A. epicatechin (m/z 289.0840). B. catcehin (m/z 499.1296). C. procyanidin. Bean sections are arranged per the stages of development (left to right) VSG, very small green; SG, small green; LG, large green; Y, yellow; R, red. Data are mapped as an overlay onto an image of the coffee bean sections. Both epicatechin and procyanidin localised to the pericarp in the early stages, no pericarp was present in the late stages of fruit development. Catechin was present in very low levels in the embryo in LG and R stages and was not found anywhere else. Bar = 3 mm

4.4 Discussion

The main chlorogenic acid groups present in coffee during fruit ripening showed different localisation patterns. During the early stages of development, they all co-localised to the inner perisperm and outer pericarp. They were not present at high levels in the immature endosperm, where only traces were observed. Sometime during the maturation of the endosperm, between stages SG and LG, CQA and diCQAs accumulated to high levels in the endosperm and low levels of FQA started to localise to the outer embryo. During later developmental stages, FQA localised strongly to the outer endosperm just like diCQAs but diCQAs were also concentrated in the embryo, this was the only type of chlorogenic acid which was strongly localised to embryos. CQA was evenly distributed throughout the endosperm, in all later stages of development and showed decreasing localisation to the embryo. The main differences between the accumulation of CQA and diCQA could be related to the two different pathways for diCQA synthesis characterised in the previous chapter.

Through *in vitro* assays, I showed that HQT and HCT, from coffee, could synthesise diCQAs in the presence of high levels of CQA at a low pH via their CCT activity. This activity could only occur when HCT and HQT localised to the vacuole, and 5-CQA acted as the acyl donor as well as the acyl acceptor. Another route for diCQA synthesis was a property of the HCT enzyme only. Here, caffeoyl CoA acted as the acyl donor, and 5-CQA as the acyl acceptor. This activity occurred at neutral pH probably when HCT was localised to the cytoplasm, where it could have this activity. We named this activity caffeoyl CoA chlorogenate transferase activity (CaAGT).

I showed through *in vitro* and *in vivo* assays that HQT alone synthesises 5-CQA. This 5-CQA may spontaneously isomerise to 3-CQA or 4-CQA. The dependency of spontaneous isomerisation is on pH suggesting that this occurred principally in the cytoplasm.

The highest levels of CQA were detected in the endosperm of the LG stage. This was where the highest levels of diCQA were also observed. These data suggested high activity of HQT at this stage resulted in high levels of CQA. Both can move to the vacuole of endosperm cells, and synthesis of diCQAs at this stage likely occurs through the CCT activity of HQT. By the Y stage, diCQA started to show higher levels at the outer edge of the folds of mature endosperm. Here, the endosperm had hardened and there was lower cell density and the cells are larger. The reason why these larger cell types accumulated more diCQA than the softer internal endosperm was unclear. HQT may localise to the vacuoles of these cells in higher concentrations than in other cells. FQA showed the same pattern of accumulation in later coffee fruit development. HQT has been shown to be responsible

for the synthesis of FQA [101] and presumably, the switch from CQAs to FQA is dependent on the supply and availability of feruloyl CoA relative to caffeoyl CoA. Feruloyl CoA may become more available later in coffee bean development. Caffeoyl CoA undergoes methylation by the enzyme caffeoyl CoA methyltransferase (CCoAMT) to form feruloyl CoA. So HQT probably uses quinic acid for the transesterification of the feruloyl CoA formed from caffeoyl CoA to form feuroylquinic acid (FQA) later in endosperm development.

DiCQAs were detected at high levels in embryos at late stages of fruit development when lower concentrations of CQAs were observed. The presence of high levels of diCQAs could be because of the CaAGT activity of HCT at neutral pH. A route for the synthesis of diCQAs in coffee embryos could be through the CaAGT activity of HCT using 5-CQA and caffeoyl CoA as substrates to form diCQA in the cytoplasm at a neutral pH. DiCQAs are unlikely to be synthesised by CCT activity of HQT in embryos because of the declining levels of CQA. The higher antioxidant activity of diCQA compared to CQA, could be a reason why diCQA accumulates at high levels in embryos, owing to the need for protection of the most delicate area of the seed. The likelihood of diCQAs being mobilised to form lignin is unclear, since no routes to diCQA synthesis have been reported as reversible. Consequently, a protective role is more probable.

The differential localisation of CQA and its precursor, *p*-coumaroyl quinate in the embryo is likely to be due to changes in the relative expression of C3'H. C3'H converts *p*-coumaroyl quinate to 5-CQA and *p*-coumaroyl shikimate into CSA respectively. Normally, a reduced expression of C3'H would result in accumulation of both *p*-coumaroyl quinate and *p*-coumaroyl shikimate, but *p*-coumaroyl shikimate was not detected in any sample. HCT, the enzyme that primarily forms *p*-coumaroyl shikimate is expressed in coffee during fruit development at all stages [24], and there was shikimic acid available in every tissue sampled. The reason why *p*-coumaroyl shikimate did not accumulate was probably that coffee has two different genes *CYP98A35* and *CYP98A36* encoding *C3'H*. Both are able to hydroxylate *p*-coumaroyl shikimate to form caffeoyl shikimic acid (CSA) whereas only one, *CYP98A35,* will hydroxylate *p*-coumaroyl quinate, forming CQA. Both reactions are slow and are reported to be irreversible [119]. Therefore, it is likely that this accumulation of *p*-coumaroyl quinate is because of the reduced expression of *CYP98A35.* The *CYP98A36* and *HQT* were found to be expressed in all stages of coffee during fruit development.

Low levels of caffeoyl shikimate were observed in the inner perisperm and the outer layer of the mature endosperm. During fruit development, the almost liquid-like inner perisperm and its cells are pushed away and dry to form the silver skin (testa). The pattern of accumulation observed for caffeoyl shikimate was consistent with areas corresponding to the silver skin in later stages of

development. Therefore, the caffeoyl shikimate observed in the silver skin, probably originated from the inner perisperm during early fruit development. The inner perisperm in coffee contributes to the formation of the lignified coffee bean husk [120]. The presence of caffeoyl shikimate in the inner perisperm is an indication that it is an intermediate in lignin biosynthesis. Incidentally, CQA was found also to localise in the perisperm during fruit development, where it may also supply caffeoyl CoA for lignin production. Notably, I did not observe any chlorogenic acid group localise together with caffeine. The distribution of caffeine was even throughout all developmental stages. Caffeine and 5-CQA associate with each other when in solution [121].

The spatial localisation of the different chlorogenic acid groups have shown that compounds relating to different groups have different tissue distributions and may consequently have different roles. A network of enzymes, which have different specificities for different substrates, controls the synthesis routes of different chlorogenic acid groups. Their cellular compartmentalisation and relative abundance can change the metabolic profile of chlorogenic acid compounds. As yet, it is unknown if active transport of different chlorogenic acids could contribute to the localisation patterns observed, but it is more likely that these are caused by the differential activities of HQT and HCT in developing beans. This study suggests that the different chlorogenic acids are likely to have different metabolic roles within the developing coffee seed.

Chapter 5: Identification of MYB transcription factors from coffee that potentially regulate the phenylpropanoid pathway and caffeoyl quinic acid (CQA) formation

5.1 Introduction

The phenylpropanoids are involved in a diverse range of plant responses towards biotic and abiotic stimuli. They are also essential to higher plants in providing monolignols for waterproofing and strengthening xylem to allow plants to grow tall and transport water (through the production of monolignols for lignin biosynthesis). Phenylpropanoid metabolism is required for the synthesis of the anthocyanin pigments, flavonol and flavone co-pigments. The diversity of phenylpropanoids is likely to be the result of the evolution of compounds which first appeared in order to protect the pioneering land plants from the damaging effects of UV light [68].

Transcriptional regulation of the phenylpropanoid pathway is of primary importance. It was first demonstrated by research on suspension cultures of parsley [122] and this was confirmed by the identification of many transcription factors (TFs) controlling different branches of phenylpropanoid metabolism [91, 123-127]. The majority of these TFs belong to the R2R3 MYB family of transcription factors.

The R2R3 MYB transcription factors comprise a large family of proteins found throughout the plant kingdom. The R2R3 region of MYB transcription factors comprises the DNA binding domain (DBD). Each Repeat (R) region consists of an imperfect repeat of 52 amino acids which form three helical regions arranged in a helix-helix-turn-helix (HHTH) fold. The plant-specific R2R3MYB family contains just two repeats compared to three found in the prototypic MYB protein, c-Myb [128]. These two repeats are most similar to R2 and to R3 in c-MYB and hence the plant proteins belong to the R2R3MYB family [129]. Three regularly spaced tryptophan residues form a tryptophan cluster in the tertiary structure of the HHTH which are characteristic of the R MYB repeat [130]. Situated in the third helix of each R region lies the amino acids that interact directly with DNA. The DBD situated in the R3 region is considered to form stronger interactions with DNA than that in R2 and is therefore considered more important for DNA recognition [80]. The R2R3 MYB transcription factors are divided into subgroups determined by motifs in their C-terminal domains which usually also contain the activation domain or occasionally a repression domain [82, 123]. Understanding to which subgroup a newly identified R2R3MYB transcription factor belongs, can facilitate identification of its putative function [74, 76, 126, 131]. R2R3MYB proteins have a diverse range of functions; they have been shown to be involved in regulating primary and secondary metabolism, cell fate and identity, developmental processes and responses to biotic and abiotic stimuli [86].

The function of R2R3MYBs is usually conserved within their respective subgroups across angiosperms. Some R2R3MYBs may influence pathways not directly related to those which

characterise their subgroup. For example, the StMTF1 R2R3MYB, although it phylogenetically belongs to subgroup 9 which are typically regulators of anthocyanin biosynthesis, when overexpressed in potato tubers, also affects flavonol and CGA levels although these effects may be indirect or related to requirements for partner TFs. This was not observed when StMTF1 was overexpressing in tissues above the ground, where it resulted anthocyanin accumulation [93].

Subgroup 7 MYBs have been found to regulate other branches of phenylpropanoid metabolism in different organisms, for example MYB12 from *Arabidopsis thaliana* (AtMYB12) causes flavonol accumulation when overexpressed in Arabidopsis, whereas when it is overexpressed in tomato fruit, it causes the accumulation of both flavonols and chlorogenic acids [90-92]. Subgroup 4 R2R3MYBs can also have different effects depending on their expression in different cell types. Members of subgroup 4 are all transcriptional repressors containing EAR repression motifs in their C-terminal domains, but whereas AtMYB32 controls pollen cell wall composition, eucalyptus EgMYB1 represses lignin production [83, 132].

R2R3MYBs have evolved to control transcription in different ways, while some require the formation of a complex with bHLH and WDR repeat proteins in order to activate or repress transcription, others can bind to DNA independently. MYBs which bind to bHLH cofactors have a conserved arrangement of amino acids or signature within the R3 repeat. This signature motif, $(D/E)LX_2(R-K)X_3LX_6LX_3R$, interacts with the N-terminal MYB-interacting region (MIR) of bHLH proteins belonging to subgroup IIIf [133]. When the WD repeat protein is bound, the complex is called the MBW complex. Some subgroup 4 R2R3MYB proteins such as EgMYB1 have been shown to interact with bHLH proteins whereas AtMYB6 and full length AtMYB4, which also belong to subgroup 4 MYBs do not bind bHLH proteins from subgroup IIIf [79]. The presence of the conserved signature required for bHLH interaction allows interaction specificities of MYB factors to be predicted from the sequence of R3 alone [79].

Species of the genus *Coffea* produce coffee beans which are important exports for many countries in the world. The chlorogenic acids, produced by the phenylpropanoid pathway, are present in relatively high levels in coffee plants and contribute to protection of the plant against abiotic and biotic stresses. Chlorogenic acids are also present in the coffee seed which are roasted to make the beverage. The high chlorogenic acid levels in *Coffea canephora* (*Robusta* coffee) contribute to some undesirable bitter tastes in the coffee made from its beans and are a contributing factor to the lower value of *Robusta* coffee beans compared to the other commercial species, *Coffea arabica* (*Arabica* coffee*)*. Although coffee contains very high levels of chlorogenic acids, the plant does not accumulate high levels of flavonols. I have not been able to detect any flavonols during

development of *Robusta* fruit. *Arabica* coffee fruit accumulates very low levels of rutin in a few varieties [18]. Transcription factors which control chlorogenic acid biosynthesis without affecting flavonol production have not been characterised. It is not understood why coffee accumulates just chlorogenic acids in this way, or what the evolutionary pressures were for it to do so. Understanding the control of the phenylpropanoid pathway in relation to chlorogenic acid biosynthesis in coffee is key for the development of marker assisted breading to select for improved chlorogenic acid levels in commercial coffee varieties. Understanding the control of chlorogenic acid biosynthesis in coffee will also help contribute to the future sustainability of these important crop species.

5.1.1 Aims

My aim was to understand how the phenylpropanoid pathway was controlled in coffee and to understand why coffee accumulates high levels of chlorogenic acids (CGAs) but maintains low levels of flavonols in its beans. With help from Nestlé my commercial sponsor, who provided me with EST sequences of R2R3MYBs from *Arabica* and *Robusta,* I cloned five different full length genes encoding R2R3MYB proteins likely to be regulating phenylpropanoid metabolism in coffee and for some, compared their relative expression to accumulation of CGAs in *Robusta*. I decided to analyse the function of these R2R3MYB transcription factors by overexpressing them in tobacco and measuring their effects on the phenylpropanoid pathway and CGA and flavonol contents. I also assessed how the overexpression of these R2R3MYB genes impacted the ability of tobacco seedlings to survive prolonged UV-B light exposure. To understand what the relative activities of these R2R3MYBs might be in coffee, I performed yeast 1 hybrid assays to determine those promoter regions of genes encoding enzymes of phenylpropanoid metabolism with which they interacted. These analyses suggested why flavonols do not accumulate to high levels in coffee beans and suggested why CGAs might be present at high levels in *Robusta* coffee.

5.2 Materials and methods

Nestlé identified eight R2R3MYB transcription factors from *Robusta* and *Arabica* which putatively regulated the phenylpropanoid pathway. I chose to characterise further the three best candidates at the start of my project, which were available only as partial EST sequences. These were, from *Arabica,* CaMYB12, and from *Robusta*, CcMYB1 and CcMYB2.

5.2.1 3' Rapid amplification of cDNA ends (3'RACE)

To acquire full length R2R3MYB cDNA sequences from the reported EST sequences I used 3'RACE. RNA from nearly open flowers, fully open flowers, and mature coffee grains was used to obtain CcMYB1, CcMYB2, and CaMYB12 cDNAs respectively. A 5'/3' RACE Kit, 2nd Generation (Roche) was used following the manufacturers guidelines. Total (2 μg) RNA was used for reverse transcription with the oligo d(T) anchor primer (5'-CACCACGCGTATCGATGTCTTTTTTTTTTTTTTTV-3') with cDNA synthesis buffer, dNTPs (50 μM), and Transcriptor Reverse Transcriptase (20 U) in a total volume of 20 μl to prepare first stand cDNA. Reactions were incubated for 60 minutes at 55 °C and the enzyme was inactivated by raising the temperature to 85 °C for 5 minutes.

Phusion polymerase (NEB) was used for direct amplification of cDNA using the first strand cDNA as a template. A total reaction volume of 50 μl contained 1X HF buffer (25 mM TAPS-HCl pH 9.3 at 25°C, 50 mM KCl, 2 mM MgCl2, 1 mM β-mercaptoethanol), 200 μM dNTPs 0.5 μM Forward primer (gene specific), 0.5 μM of Reverse primer (5'-CACCACGCGTATCGATGTC-3'), 1 μl of template, and 1 unit of Phusion DNA polymerase. A list of gene specific primers used for 3'RACE can be found in Appendix 1.

Amplifications of 3'RACE products were performed using a standard PCR program that consisted of 40 cycles of; denaturation 30 sec at 90 °C, annealing 30 sec at 50 - 60 °C (optimised for each primer pair), and extension 30 sec at 72 °C. PCR products were visualised on a 1 % agarose gel in 1 xTBE Buffer stained with ethidium bromide.

5.1.2 Stable transformation of *Nicotina tabaccum*

All the binary vectors were made using a Gateway™ compatible pBIN19 plasmid which drives expression of an insert between the att recombination sites by a double CaMV 35S promoter. This plasmid was created in the laboratory and has the name pJAM1502 (Figure 5.1). NPTII (neomycin phosphotransferase) exists between the left and right borders which confers kanamycin resistance in plants.

To insert the R2R3MYB gene of interest into pJAM1502, LR Gateway™ recombination was performed via pDONOR 207 with the MYB insert and the pJAM1502 plasmid. The reaction mixture was transformed into DH5 α *E.coli* which are sensitive to the *ccd* region within any un-recombined pJAM1052. The binary plasmids were extracted from overnight cultures of *E.coli* using a QIAGEN miniprep kit (method described in general materials and methods, 1.2.7). All of the binary plasmids were checked for the insertion of the coffee MYBs by a PCR reaction using a forward primer that binds to the 35S promoter and a reverse primer in the CaMV terminator region downstream (3') of the insert. The PCR product was visualised on a 1% agarose gel stained with ethidium bromide (0.5 μg/ml) and purified using a QIAGEN PCR clean up kit (method described in general materials and methods, 1.2.3). Each PCR product was sequenced using the Eurofins Mix to Seq kit service. The binary vectors were transformed by electroporation into electro-competent *A.tumefaciens* LBA4404. A colony was selected from the agar plates and grown overnight in liquid culture with selective antibiotics (rifampicin and kanamycin) at 28 °C shaking at 220 rpm in LB media. The next day the overnight cultures were resuspended in liquid MS media (with 3% sucrose). Young leaves from wild type *Nicotiana tabaccum* var. Samsun were sterilised with 10% bleach for 10 minutes, washed, and cut into 1-2 cm squares and placed in *Agrobacterium* solution for 5 minutes. Leaf discs were placed upside down on selection plates, 0.8 % agar MS (3% sucrose) with 100 μg/ml kanamycin, 0.1 μg/ml NAA (1-naphthaleneacetic acid), and 1 μg/ml BAP (6-benzylaminopurine) and subcultured every 2 weeks. Leaf discs were cultured in a growth room at 25 °C with a 16:8 hours light to dark light-cycle. Shoots from the callus appeared after 30 – 60 days and were cut and placed in rooting medium, 0.8 % MS agar, 3 % sucrose, and 100 μg/ml kanamycin. A list of recipes used is provided in Appendix 4. Once rooted, the shoots were genotyped by PCR, placed in soil, and transferred to the glasshouse.

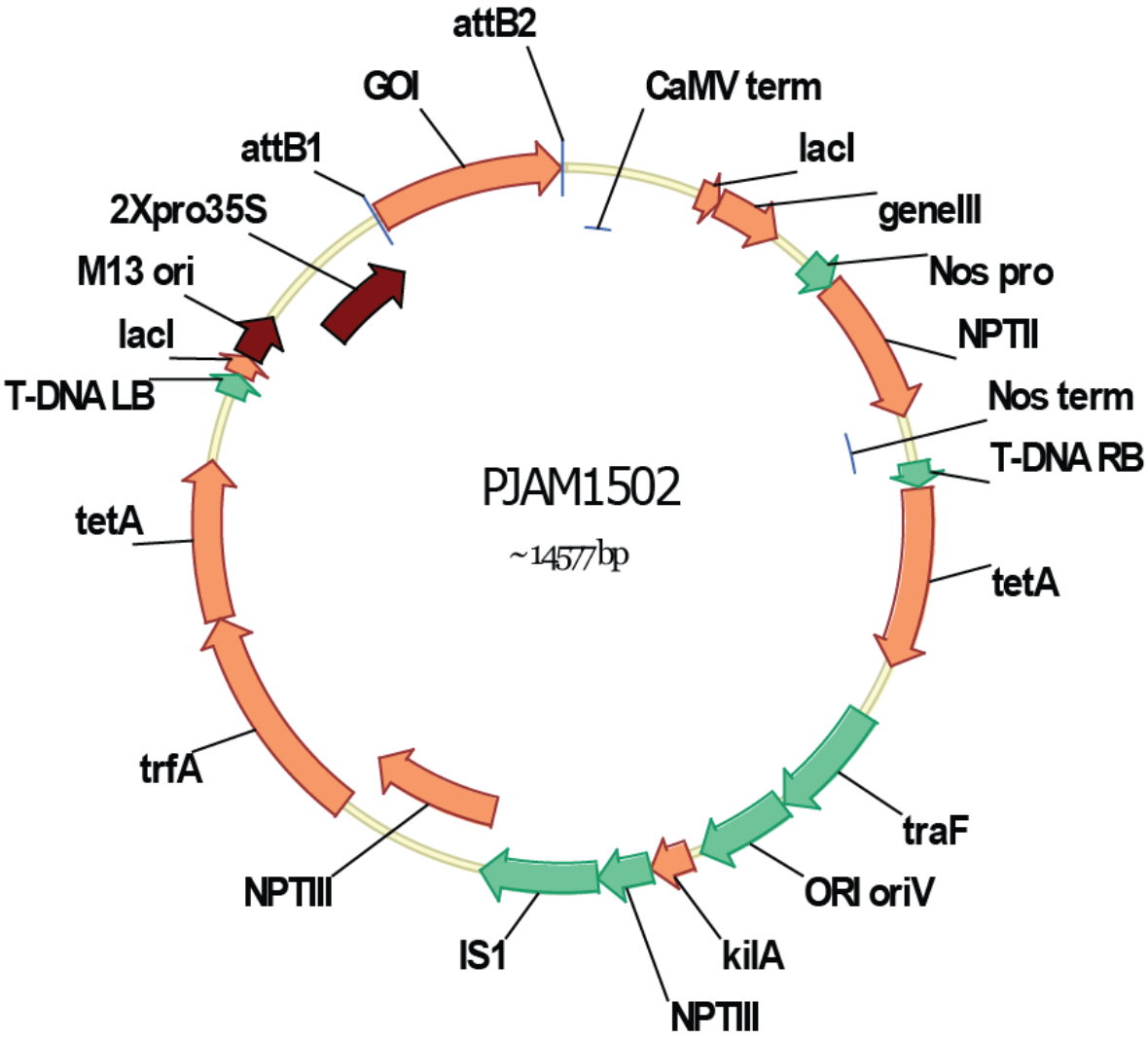


Figure 5.1. Gateway™ compatible pJAM 1502 binary vector used for the stable transformation of Coffee MYB transcription factors into tobacco. Expression in plants is driven by the **2 x 35S** CaMV promoter; **GOI**, gene of interest to be overexpressed; **CaMV term**, 35S promoter terminator sequence; **attB1** and **attB2**, gateway recombination sites; **Nos pro, NPTII** and **Nos term**, the promoter the gene and the terminator of the NPTII gene for kanamycin resistance in plants. T-DNA LB and T-DNA RB, left and right boarder sites of the insertion sequence. Image constructed using vector NTI software (Invitrogen).

5.1.3 High Performance Liquid Chromatography (HPLC)

High performance liquid chromatography was performed using a Dionex Ultimate 3000 (Thermo Scientific) with Luna 3 μm C18 100 mm x 2.0 mm (Phenomenex) column. PDA (Photodiode Array) data were collected at 330nm and the injection volume of each sample was 5 μl. The final concentrations of methanol extracts from leaf tissue were made up to a range of 1 – 5 g/l dry weight in 80 % methanol. The flow rate was 300 μl per minute and the column oven temperature was kept at 30 ^{o}C. The method consisted of a binary gradient with two solutions: Solution A consisted of 0.1 % formic acid, solution B was 100% acetonitrile (ACN) (Table 5.1). For quantification of compounds, 6 different concentrations of each standard were used to create a standard curve. A linear trend line produced from the standards was used to work out the unknown concentrations. Unknown concentrations fell within the range of the concentrations of standards. Data were extracted from Chromeleon™ software (ThermoScientific) using area of peaks created by monitoring at 330 nm. Three different isomers of caffeoyl quininc acid, 3-CQA, 4-CQA, and 5-CQA in addition to 3,5diCQA were used to work out the total chlorogenic acid content.

Table 5.1 Running method for HPLC analysis of methanol extracts from tobacco

Time (mins)	Solution A (%)	Solution B (%)
0	97	3
5	97	3
15	20	80
17	20	80
18	97	3
25	97	3

5.1.4 Quantitative Reverse Transcription Polymerase Chain Reaction (q-RT PCR)

I acquired samples of cDNA from various *Robusta* tissues from James McCarthy and Victoria Berry of Nestlé R&D centre Tours, France. I also synthesised cDNA from RNA extracted by myself from various coffee tissue samples.

For measuring CcMYB12 and CcMYB12b expression by qRT-PCR I used SYBR green as the fluorophore for detection using a CFX96 Touch™ Real-Time PCR Detection System (Bio-rad). SYBR green is a dye that emits green light at 520 nm when it is incorporated into double stranded DNA. RPL39 was used as a reference (housekeeping) gene. RPL39 is a large ribosomal subunit protein and

has been extensively used when comparing gene expression profiles in different organs of coffee plants [24, 25, 134-136]. (Primers used in qRT-PCR are found in Appendix 1). After an initial denaturation of 2 minutes at 95 °C, the reaction mixtures were subjected to 45 cycles of; denaturation, 95 °C, 20 sec; annealing, 60 °C, 20 sec; extension, 72 °C, 30 sec. Fluorescence was captured at the end of each extension step. A melt curve analysis was performed at the end of each run from 50 °C to 95 °C. Samples with more than one peak were dismissed because this indicated more than one PCR product. Threshold Ct values were normalised across all genes for each tissue before the data were exported.

For measuring the relative expression of CcMYB1, and CcMYB2, Taqman probes were used. FAM (6-carboxyfluorescein) was the fluorophore at the 5' end and TAMRA (6-carboxy-tetramethyl-rhodamine) was the quencher at the 3' end. The fluorescent tagged oligo provided an internal primer that gave off fluorescence as the 5' end was incorporated and was then quenched as the 3' end was combined. RPL39 was used as the reference (housekeeping) gene and the internal oligo was labelled with VIC (fluorophore) at the 5' end and TAMRA (quencher) at the 3' end. All the reactions contained 1x TaqMAN buffer (Applied Biosystems) which consisted of 5 mM $MgCl_2$, 800 mM dNTPs and 0.625 units of AmpliTaq Gold polymerase. The reactions contained 800 nM of forward and reverse gene-specific primers and 200 nM of the Taqman probe and a dilution of cDNA that corresponded to 0.1 μg of original RNA. Primers and probes were designed using PRIMER EXPRESS software (Applied Biosystems) by Maud Lepelley of Nestlé R&D Centre, Tours, France, (see Appendix 1). The reaction mixture was incubated for 2 minutes at 50 °C then 10 minutes at 95 °C, followed by 40 amplification cycles of 15 seconds at 95 °C, 60 seconds at 60 °C. The samples were quantified using the GeneAmp 7500 Sequence Detection System (Applied Biosystems). All samples were subjected to at least 3 technical replicates and two biological replicates.

I used SYBR green for the quantification of transcript levels in RNA from transgenic tobacco. Five biological replicates from 3 independent lines were analysed and there were at least 3 technical replicates for each analysis point. All primers used for qRT-PCR experiments were tested for their efficiency; only sets of primers which had an efficiency of 95 – 105 % were used.

5.1.5 Preparation of salmon sperm DNA

Salmon sperm DNA was dissolved in TE buffer to a concentration of 10 mg/ml and incubated overnight at 4 °C. The next day the DNA was sonicated for 2 x 30 seconds to reduce the DNA size to about 7 kb. DNA was then purified using phenol:chloroform:isoamyl alcohol (25:24:1) by adding at least 2 x the volume and vortexing. The samples were then centrifuged at 10000 x g for 5 minutes

and the upper layer (DNA fraction) was pipetted into a new tube. DNA was precipitated by adding 2.5 volumes of 100 % ethanol and 0.1 volume of 3M NaAc pH 5.5 and incubating overnight at 80 °C. The precipitate was pelleted by centrifuging at 17900 x g, the supernatant was removed, and the pellet was air dried in a Speed-Vac. The DNA was re-suspended in TE buffer (10mM tris HCl pH 8, 1 mM EDTA) to a concentration of 10 mg/ml.

5.1.6 Yeast transformation

To prepare competent yeast a fresh colony of strain Y187 that was 2-3mm in diameter was selected from a YPD plate and grown in liquid YPD (20 g/l bacto peptone, 10 g/l yeast extract, 20 g/l glucose) overnight at 28 °C with shaking at 220rpm. The next day the OD_{600} was checked to be more than 1.5. The overnight culture (200 ml) was used to inoculate 2 l of YPD and the culture was grown to an OD_{600} of 0.2 – 0.3 and then incubated at 28 °C, 220 rpm until the OD_{600} reached 0.4 – 0.6. Yeast were centrifuged at 1000 x g for 5 minutes and re-suspended in 50ml TE buffer. The yeast was then centrifuged again at 1000 x g for 5 minutes and re-suspended in 10 ml TE/LiAc (10mM Tris HCl pH 8, 1 mM EDTA, and 100 mM lithium acetate) buffer.

I added 100 ng of each plasmid to be transformed to a 1.5 ml tube with 0.1 mg of prepared salmon sperm carrier DNA. Competent yeast (100 µl) were added and then mixed. PEG/LiAc solution (600 µl) (40 % polyethylene glycol 6000, 10mM Tris HCl pH 8, 1 mM EDTA, 100 mM lithium acetate) buffer, was then added and then tubes were vortexed vigorously and incubated at 28 °C for 30 minutes shaking at 220 rpm. DMSO (70 µl) was then added and the sample was gently mixed and placed in a water bath for 15 minutes at 42 °C. The cells were placed on ice for 60 seconds to recover and centrifuged at 17900 x g for 5 seconds. The supernatant was removed and yeast were re-suspended in 100 µl TE buffer (10mM tris HCl pH 8, 1 mM EDTA). The yeast was pipetted onto SD – LT agar and spread by adding three 5 mm glass beads and rolling the beads across the surface until dry. A complete list of recipes used for yeast culture is given in Appendix 4.

5.1.7 Yeast 1 hybrid screen

A yeast 1 hybrid (Y1H) screen was used to identify MYB transcription factors binding to regulatory regions of genes encoding enzymes involved in the phenylpropanoid pathway. The promoter regions were amplified from *Robusta* coffee and were cloned by Gateway™ recombination (LR reaction) into the pHIS LEU2 vector. This vector (Figure 5.2), has a *HIS3* gene that encodes imidazole glycerol-phosphate dehydratase which catalyses the sixth step in histidine biosynthesis, promoting growth of yeast in histidine dropout medium, if transcribed. *HIS3* is transcribed if there is an

interaction between the test transcription factor and the inserted promoter regions. The transcription factor being tested is fused to the activation domain of *GAL4* which is a yeast-specific transcriptional activator which will turn on transcription of the *HIS3* gene. The R2R3MYB transcription factors cloned from Coffee were recombined by Gateway^{TM} LR reaction into pDEST 22 (Figure 5.2). This vector will promote transcription of the MYB gene with the *GAL4* fused activation domain in yeast. Fifteen promoter regions were analysed from *Coffea canephora* together with five different MYB transcription factors, a total 75 Y1H assays. A complete assay screen was repeated 3 times using different colonies each time to control for the presence of false positives.

Yeast transformed with plasmids containing the promoter:HIS3 and R2R3MYB:GAL4 AD cassettes were grown on dropout media (leucine, tryptophan and histidine [SD-LTH]) supplemented with increasing concentrations of 3-Amino-1,2,4-triazole (3-AT). An interaction between the MYB transcription factor and promoter region was scored when colonies appeared at a higher concentration of 3-AT, compared to the GUS control with the Coffee promoter.

5.1.8 UV exposure

Transgenic tobacco seeds were sterilised in 10 % bleach (Domestos) for 10 minutes and then washed 3 times in ddH_2O. Seeds were plated on 0.8 % agar MS with 3 % sucrose supplemented with 100 μg/μl kanamycin and the plates were kept in a controlled environment room at 23^{o}C with 16 hour light cycle (recipes for medium used are listed in Appendix 4). After 4 weeks the seedlings were exposed to UV-B light (3.2 mW/cm^2) for various lengths of time 200 mm below the inverted short wavelength transilluminator. The plates were resealed and the seedlings were pictured 10 days after the exposure to UV light.

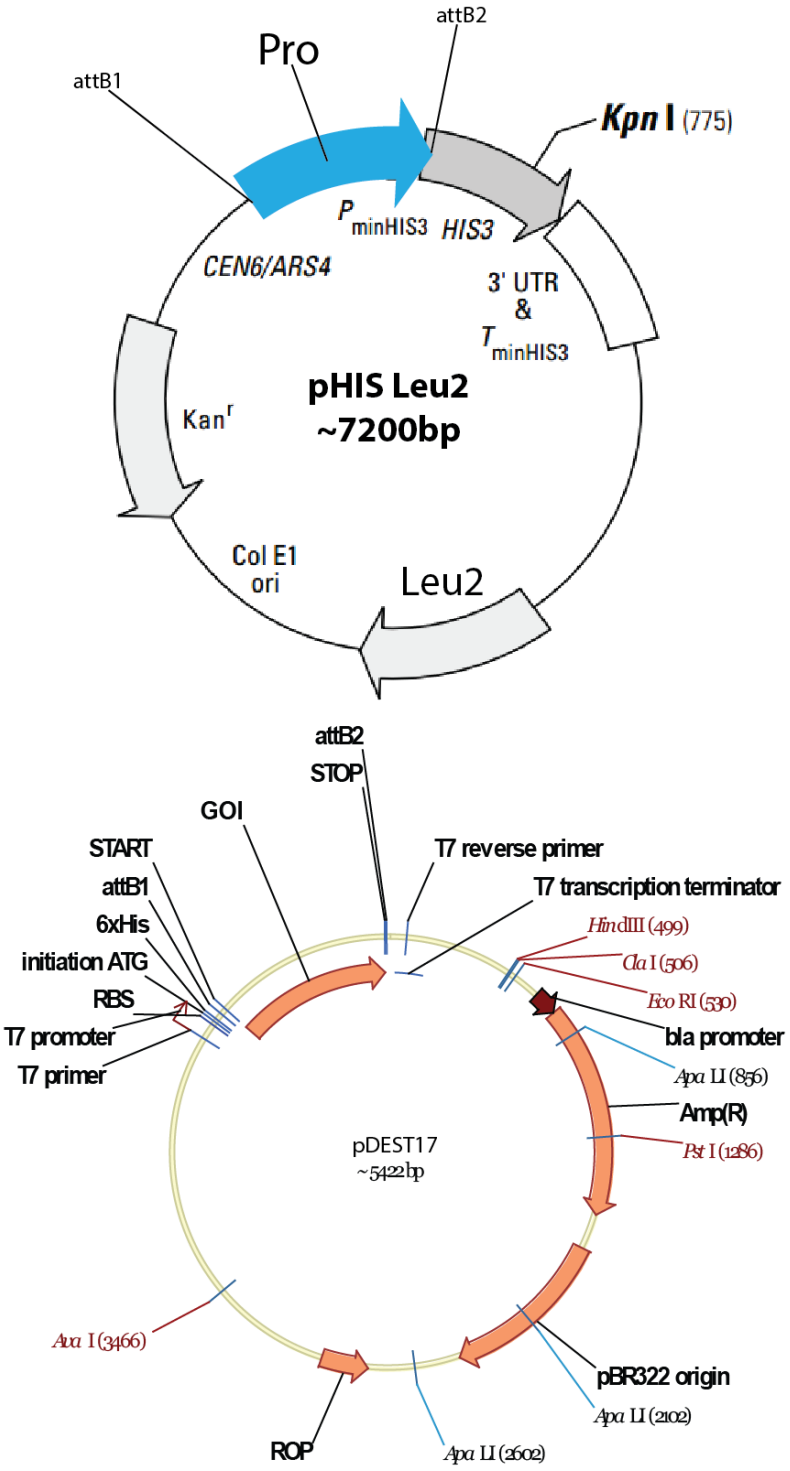


Figure 5.2. Vectors pHIS LEU2 and pDEST22 for yeast 1 hybrid assay. pHIS LEU2 contains the promoter (**Pro**) inserted by Gateway™ between the **attB1** and **attB2** sites. The **Leu2** gene produces leucine for selection in dropout media. The kanamycin gene (**Kanʳ**) is present for *E.coli* selection (image adapted from Contech Y1H instruction manual [online]). **pDEST22** expressed the gene to be tested, **GOI** (gene of interest) between Gateway™ attB1 and attB2 sites. The ampicillin gene (**Amp(R)**) for resistance in *E.coli;* and TRP1 gene for selection in yeast tyrosine dropout media. Expression is driven by the T7 promoter and terminated by the T7 terminator. Image from Vector NTI software (Invitrogen).

5.3 Results

To characterise the function of the R2R3MYB transcription factors I acquired the full length version of each of the selected EST sequences available and identified and cloned others. The R2R3MYB TFs were stably overexpressed in tobacco which exhibited various phenotypes. A predicted repressor of the pathway resulted in tobacco plants with stunted growth with white lesions appearing on its leaves and its flowers exhibited less anthocyanin pigmentation. Other R2R3MYBs from coffee caused the flowers to become white in colour with no other change to the appearance. Subsequent analysis of the flavonol and CGA contents of the different transgenic lines revealed that the repressor caused lower accumulation of the flavonol and CGA compounds and all others caused significantly higher levels. However, there was one R2R3MYB from *Robusta* that caused the higher accumulation of just the chlorogenic acids. Gene expression analysis of the transgenic tobacco revealed altered transcript levels of phenylpropanoid pathway genes. A UV-B sensitivity test revealed that seedlings of coffee R2R3MYB transformed tobacco plants showed either an increased or a decreased ability to survive prolonged periods of UV-B light stress.

I tested the ability of the selected R2R3MYB TFs from Coffee to bind to different promoter regions from structural phenylpropanoid genes from the *Robusta* genome. This resulted in the discovery that several of the selected R2R3MYB TFs showed a positive interaction with a promoter region from coffee that controls a non-functional version of the chalcone synthase gene (CHS), an important gene for the synthesis of all flavonoids. While other copies of CHS do exist in coffee, and flavonoids accumulate in low levels in some tissues, the promotion of expression of a truncated version of CHS by the selected R2R3MYB TFs is likely to be significant in explaining why such high levels of CGA are present throughout the plant. The promotion of transcription of a non-functional version of CHS by these important R2R3MYB TFs could be one of the main reasons why *Arabica* and *Robusta* do not accumulate high levels of flavonoids. They instead synthesise high levels of CGAs which are formed from a branch of phenylpropanoid metabolism that comes off prior to the flavonoid biosynthetic branch.

5.3.1 The cloning of full-length Coffee R2R3MYB genes

5.3.1.1 CcMYB1

Lepelley et al,. (2011) reported an R2R3MYB gene from coffee which they called CcMYB1. This EST sequence is listed in a database on the SOL genomics network, SGN-U613524 [137, 138]. They reported that the predicted CcMYB1 protein is 97% and 93% similar to AmMYB330 and AmMYB308 respectively across the R2R3 region. AmMYB330 and AmMYB308 are known repressors of the phenylpropanoid pathway [84]. To characterise CcMYB1 fully, the full length cDNA sequence was acquired by 3'RACE. RNA from open flowers from *Robusta* FRT32 was used to prepare cDNA and to amplify the full length cDNA sequence as quantitative reverse transcription PCR (qRT PCR) indicated realativly high expression in this tissue (Appendix 3 supplementary data)

The first strand cDNA used for the amplification of CcMYB1 was synthesised from RNA extracted from nearly open flowers of FRT 32 *Coffea canephora* (Nestlé var. FRT32). A 1500 base pair band (Figure 5.3), was observed following the 3'RACE which was determined to be 1482 bp long through subsequent cloning and sequencing. This sequence contained an open reading frame encoding a peptide of 315 amino acids in length with a 3'UTR of 168 bp. Alignment with SGN-U613524 showed that the acquired cDNA was 100% identical to the EST showing that the amplified gene was the same gene as that which had been described in the report [138]. When the *Coffea canephora* genome was published in 2014, I analysed my sequence which indicated that the 1232 bp fragment corresponded to a full length gene sequence. The complete cDNA sequence of CcMYB1 cloned by 3'RACE is shown in Appendix 2.

5.3.1.2 CcMYB2

CcMYB2 was the name given to another MYB transcription factor that putatively regulates the phenylpropanoid pathway, by Lepelley et al 2011. An EST sequence is listed on the SOL genomics database as SGN-U615272 [137]. Analysis of the R2R3 region of the EST sequence indicated that it shared 84% identity over its predicted amino acid sequence to StMf1 which is known to increase levels of chlorogenic acids, flavonols and anthocyanins in potato tubers [93]. To characterise this gene further, the full length cDNA sequence was required for functional analysis. 3'RACE was utilised because the EST sequence available encoded a full N-terminal R2 region with a predicted 5'UTR.

The 3'RACE amplification of CcMYB2 was performed using cDNA synthesised from RNA extracted from flowers of Robusta FRT 32 because relatively high transcript levels were detected here (Appendix 3 supplementary data). Oligos used for 3'RACE amplification are shown in Appendix 1.

The resulting fragment of DNA (Figure 5.3), suggested the amplification of a single gene product of about 1200 bp long, which was 1232 bp long as determined by sequencing. Analysis of the 3'RACE product indicated a full length sequence of an R2R3MYB TF that coded for a peptide of 272 amino acids.

5.3.1.3 CaMYB12

The first exon of a genomic DNA sequence from the assembly of the *Coffea arabica* genome encoded a peptide with homology to the R2R3 region of AtMYB12 was reported to me by Nestlé. AtMYB12 is an R2R3MYB transcription factor that, when overexpressed in tomato fruit, causes a 22-and 65-fold higher accumulation levels of caffeoyl quinic acids and flavonols respectively when compared to WT [90]. Primers to amplify the complete gene by 3'RACE are listed in Appendix 1. The cDNA used for the reaction was synthesised from the RNA of young leaves of *Coffea arabica* by 3'RACE PCR (Nestlé variety CccA35) as young leaves are known to accumulate relatively high levels of caffeoylquinic acid [114]. The amplification resulted in a DNA band about 1500bp (Figure 5.3), sequencing indicated that it encodes a homologue of AtMYB12. The longest open reading frame of amino acid sequence showed that the cDNA could be translated into a peptide of 391 amino acids. There is currently no published genome sequence of *Coffea arabica* so therefore I was unable to align the cDNA sequence with a complete gene sequence but, the presence of a 3'UTR suggested that the full length cDNA sequence had been amplified.

5.3.1.4 CcMYB12

I discovered genes encoding other R2R3MYB transcription factors that putatively regulate the phenylpropanoid pathway by performing a blast search using the cDNA sequence of CaMYB12 as bait in the published *Coffea canephora* genome sequence [53].

One gene with similarity to CaMYB12 is named Cc03_g01140 in the coffee genome hub and it is present on the + strand of chromosome 3. It is annotated to contain 3 introns and 4 exons and there does not seem to be any alternative splice variants from the RNA sequence data available on the coffee genome hub. Primers were designed (Appendix 1) and used to amplify this gene from cDNA synthesised from RNA extracted from the flowers of FRT32; the resulting fragment was cloned into the pGEM-T easy vector and sequenced. An analysis between the predicted peptide sequences (cloned vs Cc03_g01140) showed that the cloned cDNA sequence encoded a peptide which was 54 amino acids longer at the C terminus (Appendix 3 supplementary figures). Further investigations indicated that the cloned cDNA sequence included the 3rd predicted intron of Cc03_g01140 so therefore was a splice variant to the published sequence (Appendix 3 supplementary figures). The cloned cDNA sequence was called CcMYB12.

5.3.1.5 CcMYB12b

VvMYBF1 is another R2R3 MYB transcription factor from *Vitis vinifera* (grape) that belongs to subgroup 7 which when overexpressed in Arabidopsis mutant MYB12 resulted in complementation of the flavonol deficient phenotype [139]. By performing a blast search using VvMYBF1 as the query, against the coffee (*Coffea canephora)* genome database I was able to discover a gene encoding another MYB transcription factor closely related to AtMYB12 that is expressed in coffee. Initial searches using the AtMYB12 as the query did not result in finding this fragment as the genes share similarity only in the R2R3 region. Expression of VvMYBF1 in the Arabidopsis mutant *myb12* results in complementation of the flavonol-deficient phenotype showing VvMYBF1 is a transcriptional regulator of flavonol synthesis [139]. The RNA seq data on the coffee genome hub showed that the gene was expressed in various tissues. A blast search of the predicted amino acid sequence indicated that this encoded protein was similar to AtMYB12 (78% identical). For this reason the gene was named *CcMYB12b*. Cloning of *CcMYB12b* was from cDNA synthesised from the RNA of leaves of *Coffea canephora* FRT 32 because RNA seq data indicated the most reads of this gene were from that tissue. Alignment of the sequences of the predicted cDNA and cloned cDNA indicated that they were 100% identical. The full cDNA sequence is shown in Appendix 2.

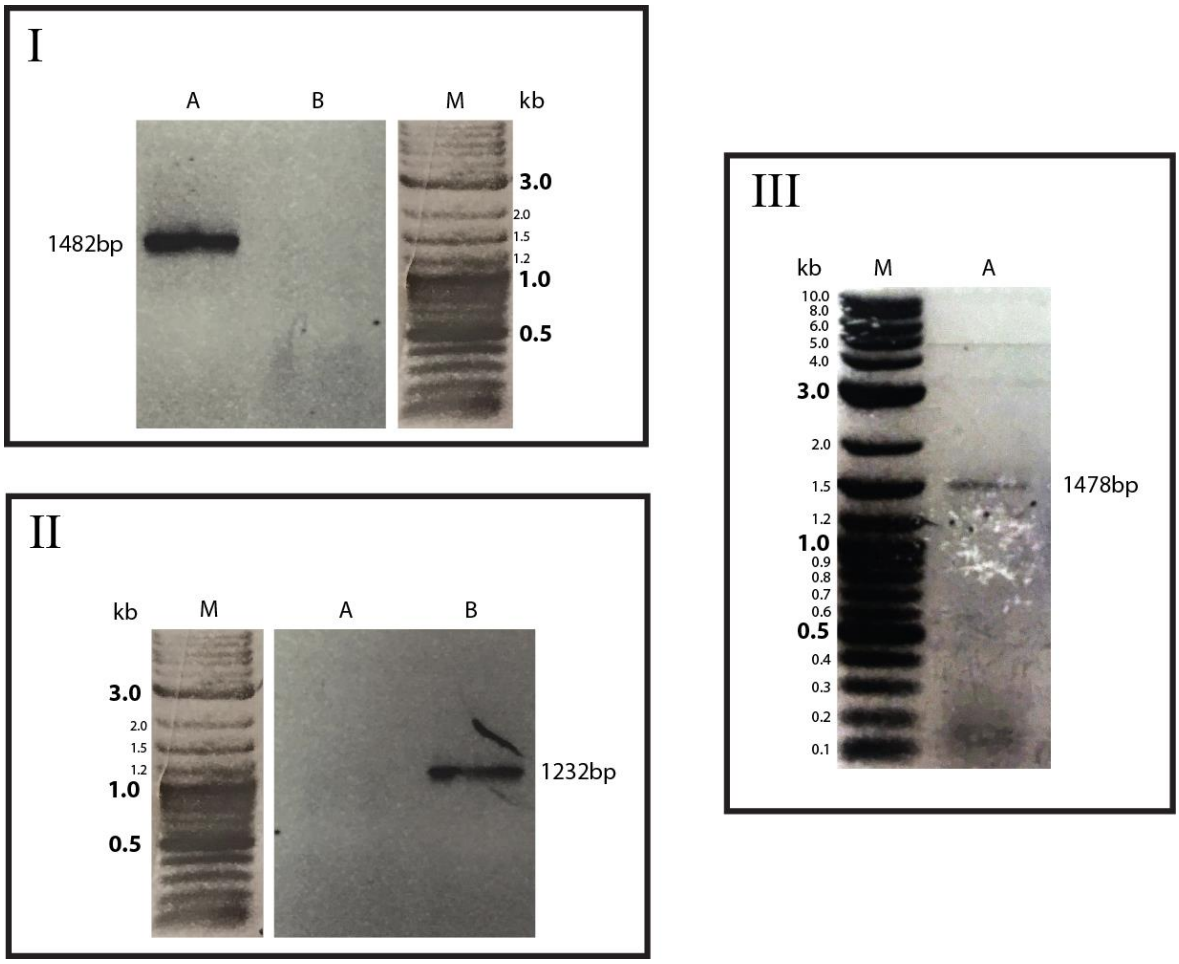


Figure 5.3. Agarose gels of the product from 3'RACE amplification of CcMYB1 (I), CcMYB2 (II) and CaMYB12 (III). CcMYB1 showed the presence of a single band of 1482 bp from amplification the cDNA template (A) and no band with no cDNA template (B). CcMYB2 showed the presence of a band at 1232 bp from reaction with the 3'RACE cDNA template (B), and no band when no cDNA was used (A). CaMYB12 has a band of 1478 bp from the PCR reaction (A). M = 2Log marker with annotated fragment sizes (NEB).

5.3.2 Functional gene expression analysis of cloned R2R3MYBs from Coffee

Before *in vivo* functional analysis of the cloned gene from coffee, I wanted to understand their relationship to other characterised R2R3MYB TFs by identifying the subgroups to which they belonged. The longest amino acid sequence of each gene was entered into the IT3F program to analyse phylogenetically to which MYB transcription factors they were most closely related. IT3F is a web based tool that allows the user to make comparisons between structurally related genes/proteins to assist in understanding their functions [131]. There is substantial data for the R2R3MYB TF subfamily listed on the website which allows for accurate identification of potential orthologues.

The *in silico* analysis revealed that CcMYB1 belongs to subgroup 4 of the R2R3 MYB family (Figure 5.4). It aligned most closely to MYB transcription factors that are reported to supress the phenylpropanoid pathway. AtMYB4 downregulates the expression of flavonols by supressing the promoter regions of C4H [82]. CaMYB1 was a good candidate for further functional analysis as a regulator of chlorogenic acid production.

The alignment of CcMYB2 showed that it belonged to subgroup 6 of R2R3MYBs (Figure 5.4). These genes function as regulators of anthocyanin biosynthesis. Subgroup 6 R2R3MYBs activate biosynthetic genes specific for anthocyanin biosynthesis. AtMYB75 (PAP1) upregulates the transcription of genes like dihydroflavanol 4-reductase (DFR) and anthocyanidin synthase (ANS). These proteins are involved in the production of anthocyanins and proanthocyanidins and change in their expression indirectly affects other branches of the phenylpropanoid pathway [86]. Overexpression of AtMYB113, AtMYB114, AtMYB75, and AtMYB90 subgroup 6 genes in Arabidopsis caused more anthocyanins to accumulate compared to WT [140]. The overexpression of the StMtf1 subgroup 6 R2R3MYB caused the accumulation of anthocyanins in tissues above ground. In the below ground tubers the accumulation of anthocyanins was in a polka-dot pattern. Overexpression of StMtf1 also caused the higher accumulation of flavonols and modest increases in CQA. This modest increase in CQA accumulation mainly occurred because of the change in flux of the phenylpropanoid pathway in tissues which were unable to accumulate anthocyanins from the overexpressed genes. The study also used a mutated version of the native protein designed to increase its activity [93]. Subgroup 6 R2R3MYB transcription factors require interaction with bHLH proteins (from subgroup IIIf) and WDR proteins for activity.

The phylogenetic analysis showed that CcMYB12, CaMYB12, and CcMYB12b belong to subgroup 7 (Figure 5.4). Genes within this subgroup are known to encode transcriptional activators of the phenylpropanoid pathway. These TFs activate structural genes which cause the increased accumulation of flavonols and CGAs. AtMYB12, when overexpressed in Arabidopsis, does not cause the accumulation of CGA because Arabidopsis lacks a gene encoding HQT which is required for CGA accumulation, whereas, when it is overexpressed in tomato fruit, it causes the accumulation of both flavonols and chlorogenic acids [57, 90, 91]. A deeper analysis of the activity of AtMYB12 in tomato fruit showed that it binds directly to 7 promoter regions of genes related to phenylpropanoid biosynthesis. Two of these target genes, *Eno* and *DAHPS* encode enzymes active in intermediary metabolism. Three of these targets are different copies of *PAL* at the start of general phenylpropanoid metabolism, and the last two are *CHS* and *F3H* genes which are involved in flavonol biosynthesis [92]. The overexpression of AtMYB12 in tomato fruit also caused greater transcription of genes directly related to CGA biosynthesis like *HQT* and *C3'H* [90]. Subgroup 7 R2R3MYBs regulate flavonol accumulation, some also regulate CGA synthesis, but none seam to regulate chlorogenic acid biosynthesis alone [91, 139, 141-145]. The objective of functionally characterising CaMYB12, CcMYB12 and CcMYB12b further was of particular interest because there was no reported evidence that both *Arabica* and *Robusta* coffee accumulates significant amounts of flavonols and other subgroup 7 MYB transcription factors increase flavonol accumulation.

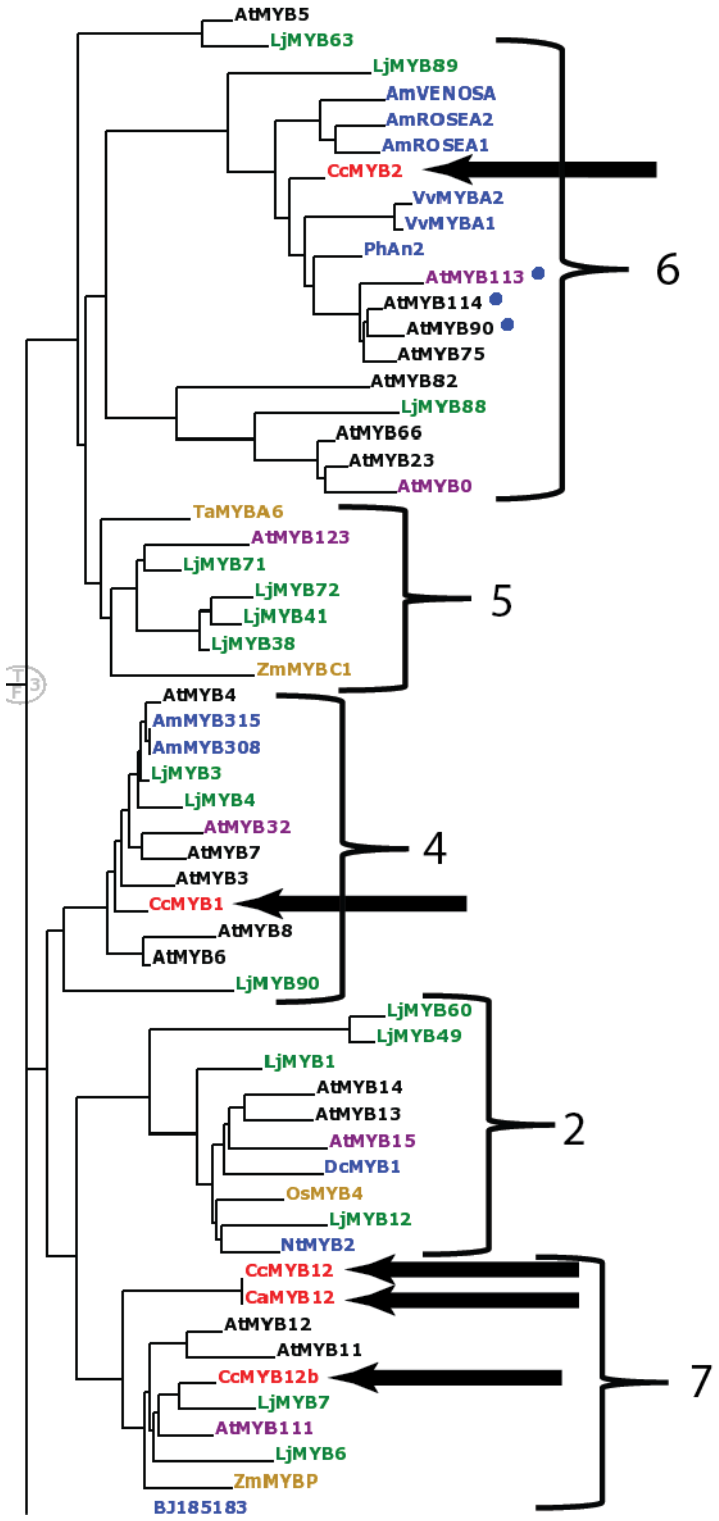


Figure 5.4 Phylogenetic analysis of cloned R2R3MYB genes. Online program IT3F which is a web based tool for functional analysis of transcription factors. The cloned R2R3MYBs from coffee are in red (arrowed), numbers represent different subgroups. **CcMYB2** aligns to subgroup 6. **CcMYB1** aligns to subgroup 4. **CcMYB12, CaMYB12,** and **CcMYB12b** align to subgroup 7 [131].

The relative expression of CcMYB1, CcMYB2, CcMYB12 and CcMYB12b was measured in different tissues of *Robusta* (Nestlé var. FRT32). I then compared the data to the accumulation patterns of chlorogenic acids as described in the previous chapter and assessed levels in different varieties. Generally, I found CGA accumulation patterns to be similar in different varieties of *Robusta*. Coffee grains showed maximal accumulation up until the LG stage, levels which then remained stable throughout the rest of seed development. There was maximal accumulation of CGA in young leaves which then declined throughout leaf maturation.

CcMYB1 showed relatively low expression levels during the first two stages of developing coffee grain which then increased to higher levels in the later stages (Figure 5.5). This pattern showed that CcMYB1 was possibly acting as a repressor of the phenylpropanoid pathway after the grain had accumulated maximal levels of CGA. Expression of CcMYB2 showed a clear association with tissues which accumulate anthocyanins especially the outer pericarp which progressively turns red towards reaching maturity and which I showed accumulated proanthocyanins and flavan3-ols in the MSI imaging described in chapter 4. The expression of CcMYB2 showed that it was expressed in increasingly higher levels in the pericarp of fruit during development (Figure 5.5). Its expression in the later stages of developing grain was not a sign that it promotes chlorogenic acid biosynthesis because coffee beans do not accumulate higher levels of CGA during the later stages of fruit development. A more likely reason for why it was detected in these stages was contamination from the wet pericarp during its removal to isolate the grain. I decided to not carry out further *in vivo* functional analysis of CcMYB2 because *in silico* analysis showed that it was very unlikely to be a direct regulator of chlorogenic acid biosynthesis.

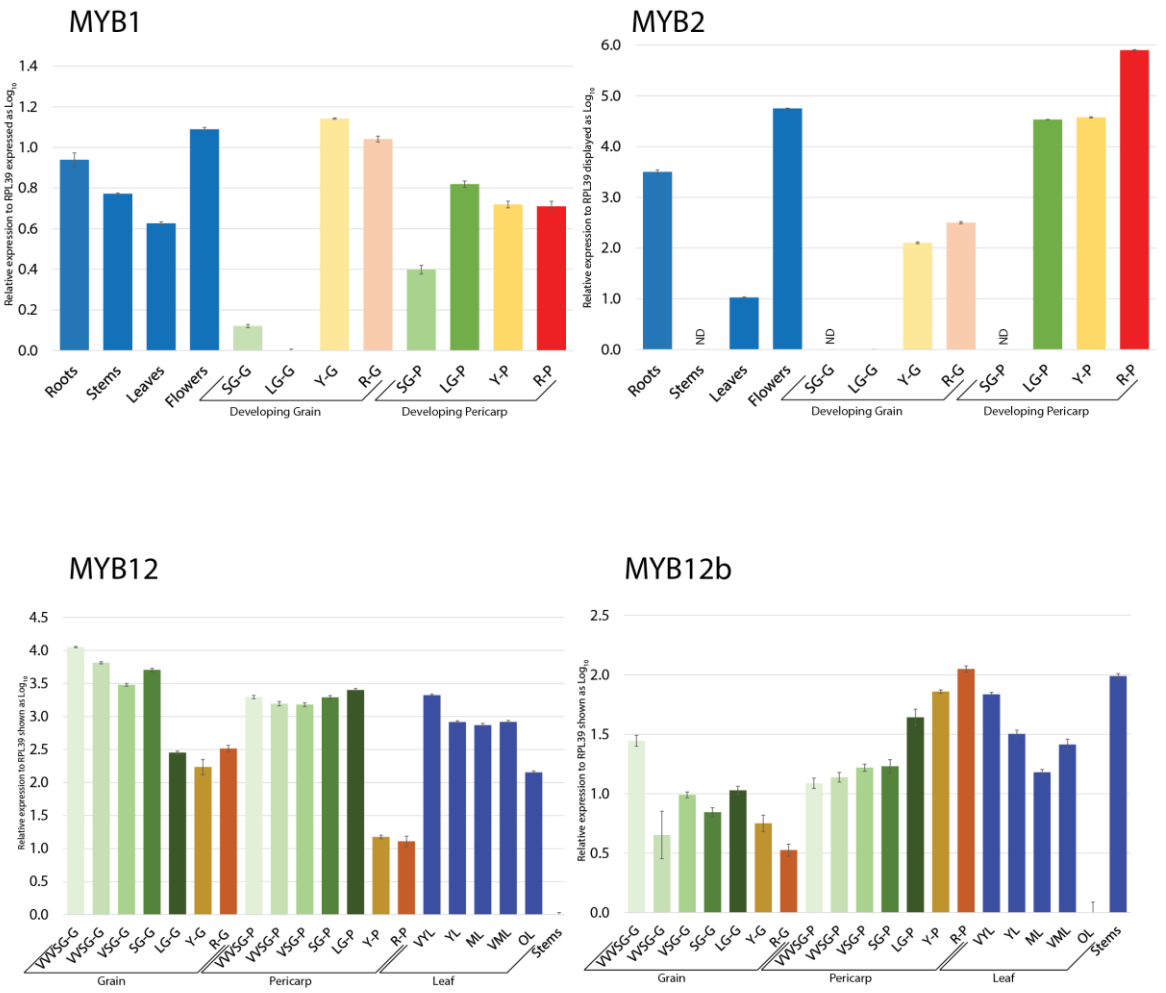


Figure 5.5. Relative transcription of MYB1, MYB2, MYB12 and MYB12b from *Robusta* (Cc) in developing tissues. The cDNA Robusta var. FRT32 was provided by Victoria Berry of Nestlé R&D centre, Tours, France. The data is displayed as Log_{10} values. Top panels - **SG-G, small green grain; LG-G, large green grain; Y-G, yellow grain; R-G, red grain; SG-P, small green pericarp; LG-P, large green pericarp; Y-P, yellow pericarp; R-P, red pericarp. Bottom panels** - Grain development - **VVVSG-G**, very very very small green; **VVSG-G**, very very small green; **VSG-G**, very small green; **SG-G**, small green; **LG-G**, large green; **YG-G**, yellow green; **R-G**, red; Pericarp development - **VVVSG-P**, very very very small green; **VVSG-P**, very very small green; **VSG-P**, very small green; **SG-P**, small green; **LG-P**, large green; **Y-P**, yellow; **R-P**, red; Leaf development - **VYL**, very young; **YL**, young; **ML**, mature; **VML**, very mature; **OL**, old. Error bars = standard error. ND = not determined.

CcMYB12 and CcMYB12b showed clearly different expression patterns (Figure 5.5). Of the two, CcMYB12 showed a closer association to CGA accumulation than CcMYB12b (chapter 4). In the developing grain, CcMYB12 was expressed at relatively high levels up until the LG stage, where after transcription levels dropped during the rest of fruit development. This was the same as the CGA accumulation pattern I observed (described in chapter 4). CcMYB12b expression showed an association to CGA accumulation in leaf development. CGA levels were highest in young leaves of coffee, presumably to protect the vulnerable leaves from UV light. CGA levels in leaves gradually dropped through leaf maturation. CcMYB12b showed decreasing transcript levels during leaf development (Figure 5.5). The gradual increase in transcript levels of CcMYB12b in the pericarp could also be associated with regulation of the phenylpropanoid pathway. Just like AtMYB12, which increases the transcription of upstream genes, CcMYB12b could have a similar effect in supplying precursors for hydroxycinnamate ester biosynthesis. CcMYB12 and CcMYB12b could both be regulators of chlorogenic acid biosynthesis, determining together the accumulation of CGA in different tissues.

An alignment of the coffee R2R3MYBs which belong to subgroup 7 showed that they all contain the typical imperfect repeats that make up the R2 and R3 regions (Figure 5.6). Just like other subgroup 7 genes, CcMYB12, CaMYB12 or CcMYB12b contain no amino acid residues associated with interaction with bHLH IIIf cofactors, the motif, [DE]Lx2[RK]x3Lx6Lx3R, was not present although it has been shown to be present in several other R2R3MYB subgroups (Figure 5.6)[79, 146, 147]. CcMYB12 and CcMYB12b and CaMYB12 contained the characteristic motif of subgroup 7 proteins ([K/R][R/x][R/K]xGRT[S/x][R/G]xx[M/x]K) following the R3 MYB repeat [139]. CcMYB12b contained another motif similar to this, which was not present in CaMYB12 and CcMYB12, at the C terminus (WLLS of the motif ([W/x][L/x]LS), previously highlighted, although not thought to be essential [139, 144] (Figure 5.6). The DNA binding domain (DBD) of these proteins showed almost complete identity. Within the R3 DBD CcMYB12 and CaMYB12 share a valine (103-V) whereas CcMYB12b and AtMYB12 have an isoleucine (103-I) reside in the same place. This is an conserved residue change which is not considered significant and is not associated with important residues determining DNA recognition [148]. CcMYB12 and CaMYB12 share key residues which are different to those found in AtMYB12 and CcMYB12b. CcMYB12 and CaMYB12 contain three extra amino acids with more positively charged side chains (arginine and lysine) which could cause differences in their interactions with DNA due these causing changes with the three α helix structures of the HHTH (Figure 5.6) [80, 130].

Subgroup 4 MYB transcription factors are identified by their conserved domains in the C terminus, C1 LIsrGIDPxT/SHRxI/L and the C2 motifs pdLNLD/ELxiG/S [7]. CcMYB1, which belongs to subgroup 4, had a R2R3 domain that was nearly identical to R2R3 domain of AtMYB4 (Figure 5.6). In addition, it contained the key residues within the motif characteristic of transcription factors which bind to bHLH proteins, although AtMYB4 also shares these residues and is reported to not interact with bHLH proteins. Consequently we cannot predict if CcMYB1 interacts with a bHLH protein (Figure 5.6) [79, 82, 146]. CcMYB1 also contains the C1 and C2 EAR motifs associated with transcriptional factors which repress transcription. Although the C2 motif for CcMYB1 has an aspartate (D) instead of an asparagine (N) in a residue considered important for repression, other subgroup 4 repressors, like conifer tree PtMYB21, also have this conservative substitution (Figure 5.6) [85, 87, 149].

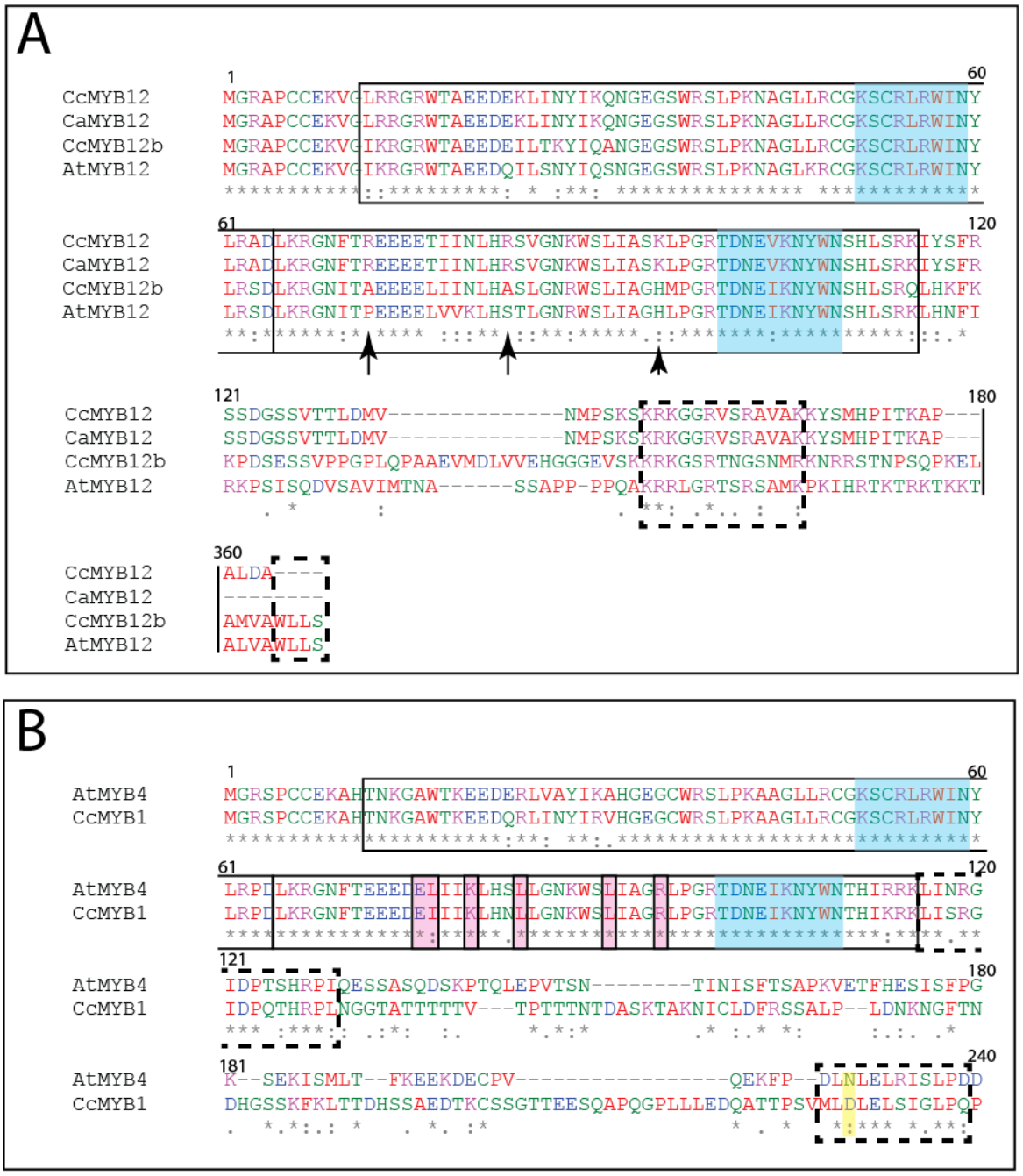


Figure 5.6. Peptide sequence alignments of R2R3MYBs for *in vivo* functional analysis against similar characterised R2R3MYBs from Arabidopsis. A. Longest peptide sequence alignment of the subgroup 7 coffee MYBs with AtMYB12. The R2 and R3 repeat is indicated by large sold line boxes. The DBDs (blue) are conserved in all genes. The two activation domains (dash line boxes) are conserved for the key residues in the first whereas CcMYB12 and CaMYB12 do not have the second activation domain present in some other subgroup 7 R2R3MYBs. Within the R3 regions the residues considered critical for bHLH interaction not present **B**. The longest peptide sequence from CcMYB1 was aligned to AtMYB4, these both belong to subgroup 4 of R2R3MYBs. The R2R3 region was almost completely conserved and the DBD was the same. The key residues which have been reported to be important for interaction with bHLH proteins for R2R3MYB are present in CcMYB1 and AtMYB4 in the R3 repeat (in pink). Although AtMYB4 has these key residues it is not though to interact with bHLH proteins. The C1 and C2 domains (dashed boxes respectively) show the EAR motifs for repression. The highlighted residue in C2 is where the D residue is present instead of the N residue which is considered important for EAR motif repression.

5.3.3 Stable overexpression of Coffee MYB genes in *Nicotina tabaccum* for functional characterisation

To analyse the function of the MYB transcription factors identified in coffee I decided to use *Nicotina tabaccum* as a surrogate plant host. This species was used because Arabidopsis was deemed unsuitable because it doesn't naturally accumulate chlorogenic acids and lacks a gene encoding HQT, which is the key enzyme required for CGA biosynthesis (See chapter 3 [57]). Direct transformation of coffee is a slow process and unsuited to a PhD timeline. *Nicotina tabaccum* is a good model species to use for several reasons. It is relatively easy to transform, it has a natural ability to accumulate significant levels of chlorogenic acid, relatively good genome data are available, and much work has already been undertaken to characterise the phenylpropanoid pathway in Solanaceous species [57, 90, 102, 111, 150-152].

I stably transformed tobacco plants with constructs for expression of each Coffee MYB transcription factor under the control of the double 35S promoter from CaMV. Plants were grown on kanamycin and then transferred to soil where they were genotyped. For the stable transformation of CcMYB1, CcMYB12 and CaMYB12, 10 independent transformants of each were obtained and analysed, each T0 plant having been regenerated from an individual callus. For the stable transformation of CcMYB12b, 7 independent plants were obtained, each from an individual callus.

Plants from the T1 generation of CcMYB1, CaMYB12, and CcMYB12 were used for all analysis. The seeds from T0 were germinated on 0.8 % agar MS with 100 μg/μl kanamycin. Seedlings that were not sensitive to kanamycin were potted on in soil and placed in the glasshouse. If the seedlings are sensitive to kanamycin, they never develop true leaves and they start to turn yellow shortly after germination and eventually die.

Relative quantification of transcript levels from the overexpression lines showed all the genes were overexpressed in the independent transformed lines (Figure 5.7). Transcripts measured in WT plants did not show the presence of any significant levels of Coffee MYB gene transcripts.

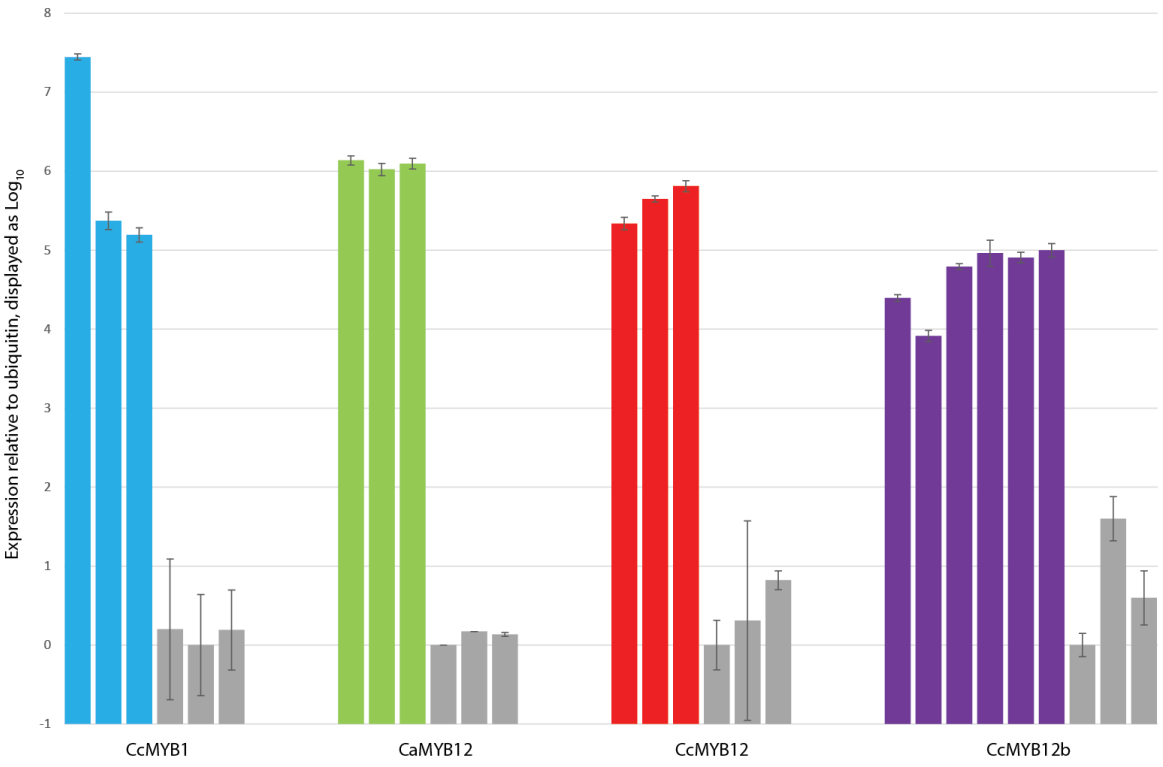


Figure 5.7. Relative transcription of CcMYB1, CaMYB12, CcMYB12, and CcMYB12b overexpression in tobacco leaves. Data represents Log_{10} ΔΔCt values relative to RPL39. Coloured bars represent overexpressed lines, grey bars represent WT relative quantification from primers for respective gene overexpression. Data for CcMYB1, CaMYB12, and CcMYB12 represent 5 biological replicates from T1. Data for CcMYB12b represent 6 independent lines from T0. Error bars = standard error.

5.3.3.1 Phenotypic changes in tobacco transformed with 35S:CcMYB1

The stable transformations resulted in some changes to the appearance of CcMYB1 flowers which appeared to accumulate lower levels of anthocyanins which were present only in the areas of the main veins of the petals that normally accumulate the most anthocyanin (Figure 5.9, B).

The leaves of the plants overexpressing CcMYB1 showed a similar phenotype to that observed in the overexpression of AtMYB4, AmMYB308 and AmMYB330 in tobacco [82, 84] (Figure 5.8 left). White lesions appeared on the older leaves of mature CcMYB1 overexpression lines but did not appear on the wild type. The white lesions were distinct from the brown ones seen on mature wild type tobacco leaves; they appeared earlier and they were distinctly white in comparison. The phenotype in AtMYB4-expressing tobacco has been reported to be the result of the negative regulation of hydroxycinnamic acid metabolism primarily by reducing the expression of C4H, it is very likely that CcMYB1 overexpression is acting in the same way [82]. The young leaves of these plants also looked paler than the controls while the veins remained dark green (data not shown). Another effect of CcMYB1 overexpression was the reduced overall size of the mature leaves. Young leaves expanded at the same rate as the controls, but slowed down as they reached maturity. The plants expressing CcMYB1 were also not as tall as other plants, they did not grow as quickly as the controls or other MYB overexpression lines (Figure 5.8 right). Plants also showed a reduced elongation of the stem internodes during development (Figure 5.8 right).

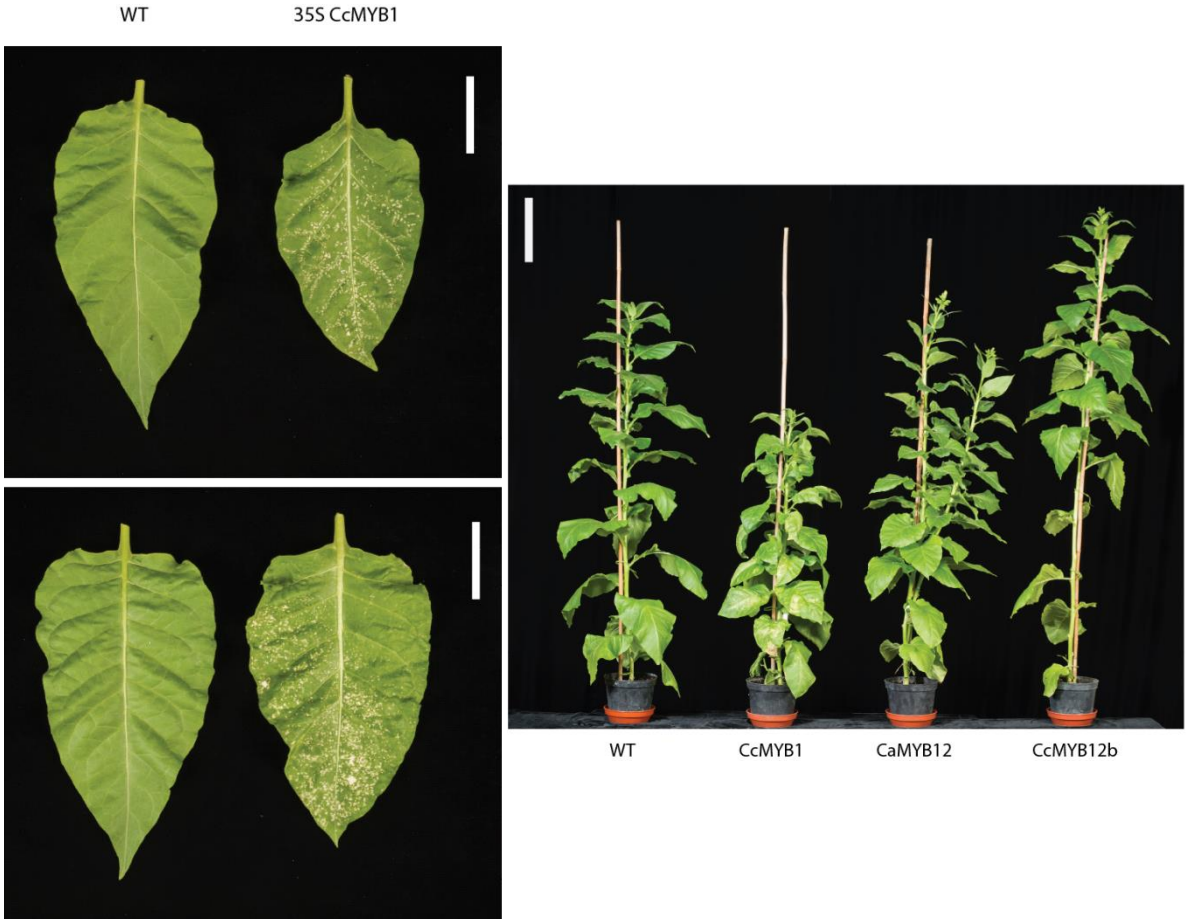


Figure 5.8. Phenotype of CcMYB1 overexpression in tobacco. Left, leaves, white lesions were observed for 35S:CcMYB1. Bar = 50 mm. **Right**, reduced height of 35S:CcMYB1 tobacco was observed which showed a reduced elongation of the stem internodes. The tobacco plants expressing other coffee MYB transcription factors remained unaffected. Bar = 200 mm.

5.3.3.2 Phenotypic changes in tobacco transformed with 35S:CaMYB12

Tobacco plants overexpressing CaMYB12 showed a strong phenotype in the appearance of the flowers compared to WT (Figure 5.9, A vs E). The flowers of CaMYB12 overexpression lines were paler in colour because they did not accumulate anthocyanins. Apart from the phenotype seen in the flowers, CaMYB12 overexpression in tobacco did not show any other phenotype (Figure 5.8 right). The phenotype observed was very similar to tobacco overexpressing AtMYB12 [90].

5.3.3.3 Phenotypic changes in tobacco transformed with 35S:CcMYB12

Tobacco plants that were overexpressing CcMYB12 showed a similar phenotype to tobacco overexpressing CaMYB12 and AtMYB12 [90]. Flowers of CcMYB12 were very pale in colour, more so than CaMYB12 overexpression lines (Figure 5.9 D vs E respectively). Visually, there was no anthocyanin accumulation observed from the flowers of tobacco plants overexpressing CcMYB12. The change in flower colour was the only phenotype observed for overexpression of CcMYB12.

5.3.3.4 Phenotypic changes in tobacco transformed with 35S:CcMYB12b

Tobacco plants transformed with 35S:CcMYB12b showed a slight phenotype in the flowers that they produced compared to WT controls. All of the flowers looked to be paler than the wild type (Figure 5.9, C vs A respectively). Some anthocyanin accumulation was observed in the flowers for CcMYB12b overexpression lines but clearly the levels were less compared to WT. These flowers were not as pale as tobacco plants expressing CaMYB12, CcMYB12 or AtMYB12 [90]. Apart from the change in the colour of the flowers, the plants did not show any other significant phenotype (Figure 5.8 right).

5.3.3.5 Crosses of transgenic tobacco plants expressing the candidate R2R3MYB TFs with tobacco plants expressing bHLH (basic helix loop helix) proteins from Antirrhinum majus.

To test the hypothesis that some of the MYB transcription factors isolated from coffee might interact with a bHLH (basic helix loop helix) transcription factor, tobacco plants overexpressing CcMYB1 and CaMYB12 were crossed with tobacco plants expressing two bHLH proteins from *Antirrhinum majus*; Delila (35S:DEL) and Mutablis (35S:MUT) [153]. I performed four crosses, 35S:CcMYB1 X 35S:DEL, 35S:CcMYB1 X 35S:MUT, 35S:CaMYB12 X 35S:DEL, and 35S:CaMYB12 X 35S:MUT. None of the crosses resulted in changes to the phenotypes conferred by dominant overexpression of each MYB gene, suggesting no functional interaction of the R2R3MYB proteins with either of the bHLH proteins.

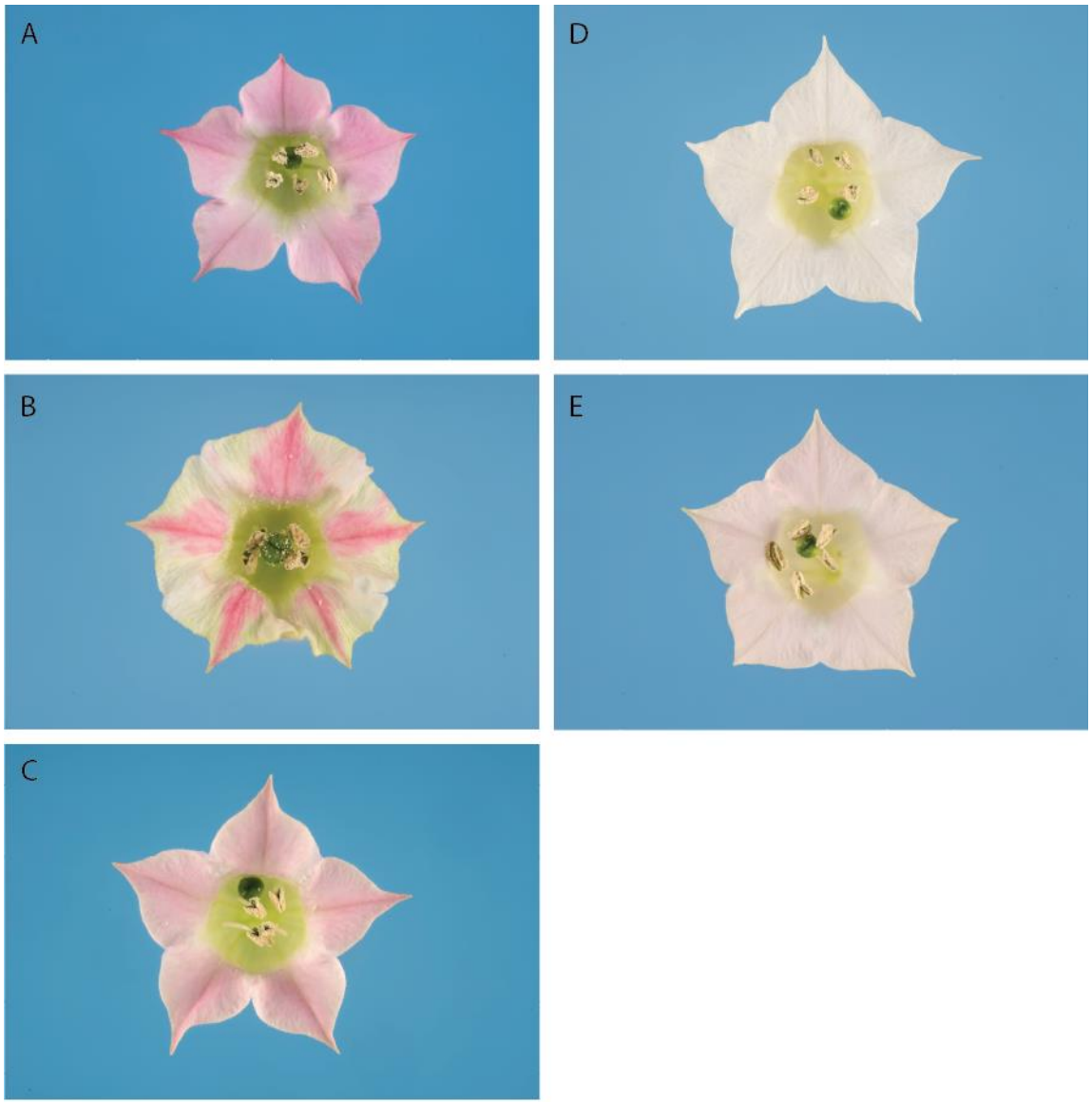


Figure 5.9. Phenotype of tobacco flowers overexpressing different MYB transcription factors from coffee. A, WT; **B**, CcMYB1; **C**, CcMYB12b; **D**, CcMYB12; **E**, CaMYB12

5.3.3.6 Transgenic tobacco plants exhibited changes in polyphenol accumulation

Tobacco plants overexpressing R2R3MYB transcription factors from coffee were measured for differences in the accumulation of the major flavonols and chlorogenic acids present in leaves. Kaempferol, rutin, and chlorogenic acid were measured by HPLC. The three isomers of caffeoyl quinic acid, (3-caffeoylquinic acid, 4-caffeoylquinic acid, and 5-caffeoylquinic acid (CGA) and 3,4-dicaffeoylquinic acid were quantified and shown together as total chlorogenic acids (CGA).

5.3.3.6.1 Accumulation of polyphenols in CcMYB1 overexpression lines

Leaves of tobacco plants overexpressing CcMYB1 showed a significant repression of all measured compounds compared to WT (Figure 5.10). The levels of rutin dropped on average by 6 fold, kaempferol and chlorogenic acid levels showed an average 4 fold decrease compared to WT. These data suggested that CcMYB1 was acting as a repressor of genes responsible for the biosynthesis of these compounds. It was not clear which genes CcMYB1 was repressing directly in order to cause this effect, although AtMYB4, a functional homologue, represses the transcription of C4H through specific promoter binding [82], and it was likely that CcMYB1 was repressing C4H in a similar way.

5.3.3.6.2 Accumulation of polyphenols in CaMYB12 and CcMYB12 overexpression lines

Tobacco leaves expressing CaMYB12 and CcMYB12 showed significantly enhanced levels of rutin, kaempferol and chlorogenic acids compared to WT (Figure 5.10). These enhanced levels were an indication that both CaMYB12 and CcMYB12 overexpression upregulates transcription of genes encoding enzymes of the phenylpropanoid pathway that control flavonol and chlorogenic acid biosynthesis. There was no significant difference between levels of rutin, kaempferol and chlorogenic acids present in plants overexpressing CaMYB12 or overexpressing CcMYB12, and the differences observed between these overexpression lines were likely to be due to differences in transcription of the overexpressed coffee MYB genes. The enrichment of these compounds was similar to overexpression of AtMYB12 in tobacco and tomato fruit [90] (Figure 5.10). AtMYB12 directly binds to promoters of enzymes in the early stages of the phenylpropanoid pathway in Solanaceous species [151], there is a strong possibility that CaMYB12 and CcMYB12 both upregulate the pathway through interaction with the same target promoters.

5.3.3.6.3 Accumulation of polyphenols in CcMYB12b overexpression lines

Overexpression of CcMYB12b had a different effect on the levels of flavonols and hydroxycinnamic acid esters compared to CaMYB12 or CcMYB12 overexpression lines. Overexpression of CcMYB12b resulted in a significant reduction in the levels of rutin and kaempferol while significantly increasing levels of chlorogenic acid levels in leaves. However, there was no significant difference between the chlorogenic acid content of CcMYB12b leaves and CaMYB12 and CcMYB12 leaves. It was not clear whether the reduction in rutin and kaempferol synthesis was through CcMYB12b acting as a repressor of flavonol biosynthetic genes and increasing flux to the production of chlorogenic acids. Equally plausible, CcMYB12b might act as an activator only of genes related to chlorogenic acid production, pulling flux away from flavonol production. The fact that CcMYB12b overexpression enhanced only chlorogenic acid production in tobacco leaves was an important observation because coffee beans do not accumulate flavonols so CcMYB12b might be a key transcription factor for controlling CGA biosynthesis in coffee.

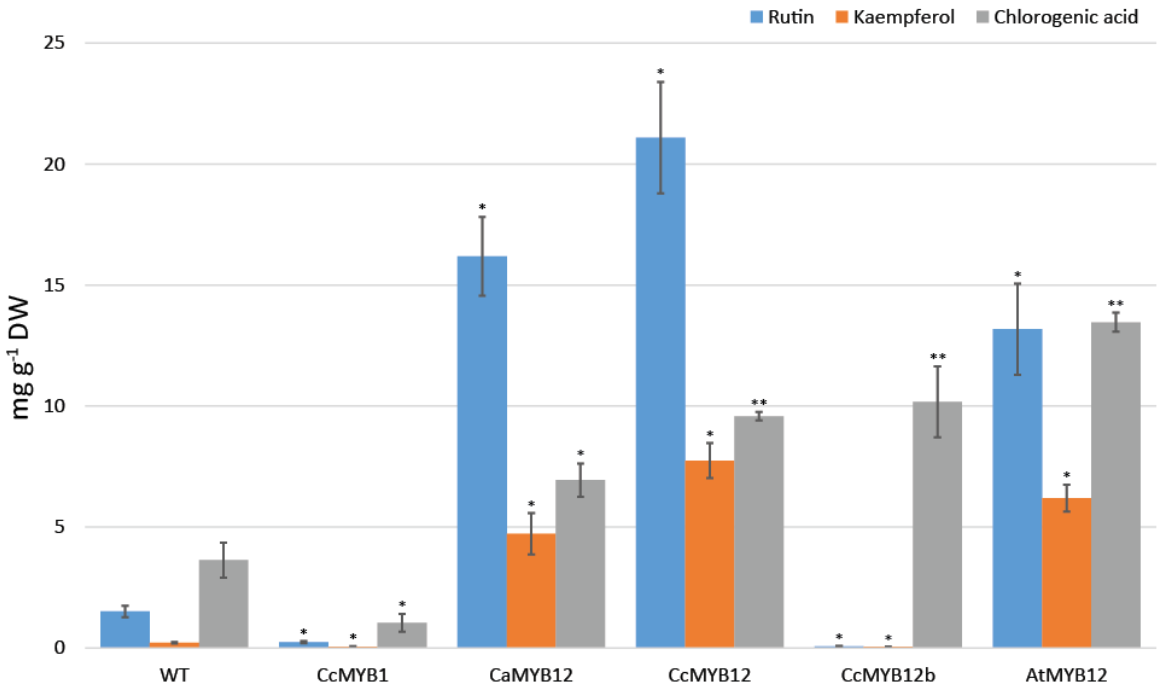


Figure 5.10. Quantification of major flavonols and chlorogenic acid in transgenic tobacco plants. Chlorogenic acid data represents the total of, 3-caffeoyl quinic acid (3-CQA), 4-caffeoyl quinic acid (4-CQA), 5-caffeoylquinic acid (5-CQA), and 3,4-dicaffeoylquinic acid. Data represent the mean values (±SE) of 5 biological replicates from 3 independent lines from T1, except CcMYB12b, which represent 6 independent lines from T0. *p<0.05, **p<0.01 compared to WT.

5.3.3.7 Overexpression of the selected MYB transcription factors from Coffee caused changes in the expression of genes in the phenylpropanoid pathway

The overexpression of the coffee MYB transcription factors resulted in changes in the expression of genes encoding enzymes in the phenylpropanoid pathway (Figure 5.11). The R2R3MYBs from coffee overexpressed in tobacco showed changes in transcript levels of major structural genes encoding enzymes related to rutin, kaempferol and chlorogenic acid biosynthesis.

5.3.3.7.1 Gene expression changes resulting from CcMYB1 overexpression in tobacco

Tobacco leaves overexpressing CcMYB1 showed reduced transcript levels of 4-coumarate:CoA ligase (4CL) (Figure 5.11 C, compared to WT). 4CL is an enzyme in the phenylpropanoid pathway that catalyses the conversion of p-coumaric acid to p-coumaroyl-CoA, a central molecule in the diversification of the phenylpropanoid pathway. The suppression of 4CL in tobacco and in a coniferous gymnosperm resulted in dwarfed plants with reduced lignin content [154, 155]. This phenotype is similar to the phenotype observed for CcMYB1 overexpression in tobacco. The reduction in 4CL transcript levels was also observed for AtMYB4 overexpression in Arabidopsis. Cinnamate 4-hydroxylase (C4H) is the principle target of AtMYB4 in Arabidopsis but the overexpression of CcMYB1 in tobacco did not show any effect on C4H transcript levels. CcMYB1 might interact with different promoters than AtMYB4 although their overexpression resulted in similar phenotypes [82]. AtMYB4 overexpression in Arabidopsis also lowered transcript levels of CHS although this was not observed for CcMYB1 overexpression in tobacco. Transcript levels of dihydroflavonol 4-reductase (DFR) were reduced in tobacco plants overexpressing CcMYB1.

CcMYB1 overexpression in tobacco showed no changes in transcript levels of HQT or HCT, these are both enzymes that are related to the biosynthesis of chlorogenic acids and monolignols, respectively. Their repression might be expected when CcMYB1 overexpression in tobacco caused lower accumulation of chlorogenic acids and reduced height (a sign of monolignol repression) [84]. However, because CGA biosynthesis is dependent on using coumaroyl CoA as an intermediate, the suppression of 4CL transcript levels by overexpression of CaMYB1 might be the cause of the lower accumulation of CGA in these lines. My results confirmed the interpretation of the phylogenetic alignments that CcMYB1 is a transcriptional repressor of phenylpropanoid metabolism.

5.3.3.7.2 Gene expression changes due to CaMYB12 and CcMYB12 overexpression in tobacco

Transcript levels of phenylalanine ammonia lyase (PAL), cinnamate 4 hydroxylase (C4H), and 4-coumarate:CoA ligase (4CL) (Figure 5.11 A, B, and C respectively), were all upregulated when CaMYB12 and CcMYB12 were overexpressed. These data were consistent with observations for AtMYB12 overexpression in tobacco [90]. PAL, C4H and 4CL are the enzymes in the general phenylpropanoid pathway and their activities lead to the production of 4-coumaroyl CoA which is commonly considered as the central molecule in the diversification of the phenylpropanoid pathway. In this part of the pathway AtMYB12 has been observed to bind to the promoter region of PAL, and the upregulation of C4H and 4CL transcripts by CcMYB12 and CaMYB12 overexpression might be indirect [92].

Chalcone synthase (CHS), chalcone isomerase (CHI), flavanone 3-hydroxylase (F3H), and flavonoid 3'-hydroxylase (F3'H) (Figure 5.11 D, E, F, and G respectively), are 4 enzymes which in turn create dihydrokaempferol, and through F3'H activity, dihydroquercetin. CHS and CHI transcript levels were elevated in leaves of CaMYB12 and CcMYB12 overexpression lines when compared to wild type, similar to the levels observed for AtMYB12 overexpression in lines of tobacco [90].

Coumaroylshikimate 3' hydroxylase quinate (C3'H), hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase (HCT), and hydroxycinnamoyl-CoA quinate hydroxycinnamoyltransferase (HQT) are enzymes that lead to the production of chlorogenic acids and monolignols (Figure 5.11 K, L and M). HCT levels were elevated significantly in CaMYB12 and CcMYB12 overexpression lines, just like levels observed in AtMYB12 overexpression lines. C3H levels were not changed significantly, whereas in AtMYB12 overexpression in tomato fruit transcript levels did increase [90].

CaMYB12 increased F3'5'H levels whereas CcMYB12 had no effect, but rather reduced levels of F3'5'H similar to the effects observed for AtMYB12 overexpression in tobacco. The levels of ANS transcripts differed between CaMYB12 and CcMYB12 overexpression lines, with CcMYB12 showing reduced production of ANS compared to the levels in WT, CcMYB1, CaMYB12 and AtMYB12 overexpression lines which all showed similar levels of expression.

There was no significant change in expression of F3H caused by overexpression of CaMYB12, CcMYB12 or AtMYB12 (Figure 5.11 F), although it might be expected that AtMYB12 overexpression caused an increase in F3H levels because AtMYB12 overexpression in tomato fruit causes an upregulation of F3H transcript levels [90]. The results for F3H expression were rather variable so any differences observed could be due to a small change in environmental conditions.

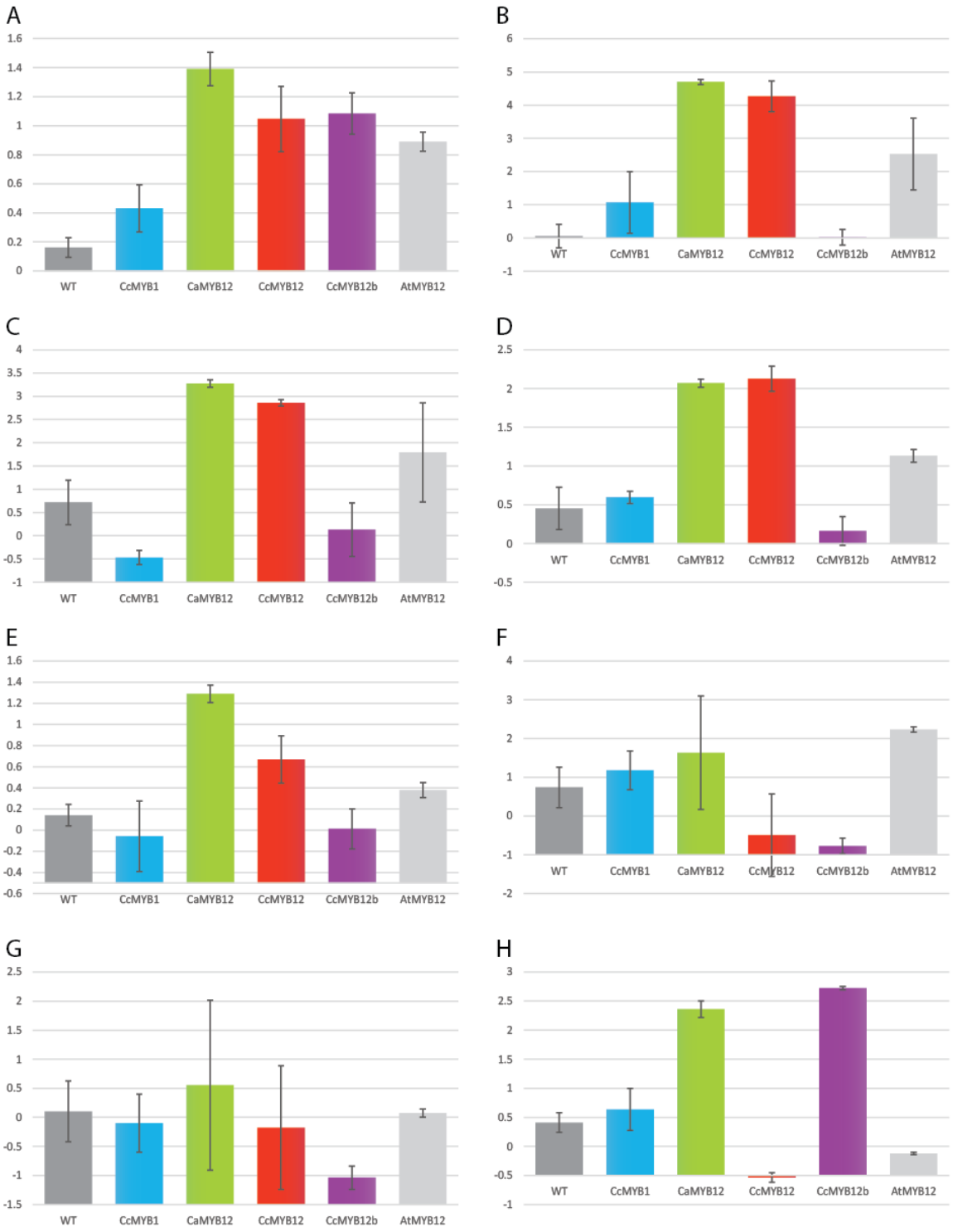
5.3.3.7.3 Gene expression changes due to CcMYB12b overexpression in tobacco

CcMYB12b overexpression induced similar changes in PAL transcripts as CaMYB12 and CcMYB12 overexpression, but C4H and 4CL transcript levels in CcMYB12b plants were similar, if not lower, than transcript levels found in WT plants.

The overexpression of CcMYB12b did not significantly change the overall transcript levels of CHS and CHI. Levels of F3'H were downregulated compared to WT expression, the lower levels could be another reason why CcMYB12b showed very low levels of quercetin accumulation.

Flavonoid 3' 5' hydroxylase (F3'5'H), dihydroflavonol 4-reductase (DFR), and anthocyanidin synthase (ANS) (Figure 5.11 H, I, and J respectively), are enzymes that lead to the production of anthocyanins. CcMYB12b overexpression showed elevated levels of F3'5'H and ANS transcripts which was not an effect seen for overexpression of AtMYB12.

Expression of genes involved specifically in monolignol and chlorogenic acid biosynthesis was measured, including p-coumaroyl quinate/shikimate 3'-hydroxylase (C3'H), hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase (HCT) and hydroxycinnamoyl-CoA quinate hydroxycinnamoyltransferase (HQT). Overexpression of CcMYB12b exhibited changes to transcript levels of C3'H and HCT unlike CaMYB12, CcMYB12 and AtMYB12. C3'H transcript levels were very low compared to all other lines (controls and other overexpression lines). This was an unexpected observation considering the elevated levels of chlorogenic acid and the integral role C3'H has in its biosynthesis. This observation might suggest that chlorogenic acid biosynthesis could occur independently of C3'H, or that C3'H is regulated independently of HQT which responds to CcMYB12b. HCT transcript levels were similar to those observed in WT. HCT is an enzyme that has the ability to synthesise chlorogenic acid from quinate and caffeoyl CoA but has a far greater affinity for shikimate in the reaction, creating caffeoyl shikimate. HQT is the major protein responsible for chlorogenic acid (CGA) biosynthesis and its substantial upregulation in CcMYB12b overexpression lines suggested it was responsible for the elevated levels of CGA in these lines. The high levels of HQT transcripts observed for CcMYB12b overexpression lines suggested that unlike the AtMYB12 functional homologues, CaMYB12 and CcMYB12, CcMYB12b may selectively activate HQT to promote the synthesis of chlorogenic acid in *Robusta* coffee beans.


A
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
WT
CcMYB1
CaMYB12
CcMYB12
CcMYB12b
AtMYB12
B
6
5
4
3
2
1
0
-1
WT
CcMYB1
CaMYB12
CcMYB12
CcMYB12b
AtMYB12
C
4
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
WT
CcMYB1
CaMYB12
CcMYB12
CcMYB12b
AtMYB12
D
2.5
2
1.5
1
0.5
0
-0.5
WT
CcMYB1
CaMYB12
CcMYB12
CcMYB12b
AtMYB12
E
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
WT
CcMYB1
CaMYB12
CcMYB12
CcMYB12b
AtMYB12
F
4
3
2
1
0
-1
-2
WT
CcMYB1
CaMYB12
CcMYB12
CcMYB12b
AtMYB12
G
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
WT
CcMYB1
CaMYB12
CcMYB12
CcMYB12b
AtMYB12
H
3
2.5
2
1.5
1
0.5
0
-0.5
-1
WT
CcMYB1
CaMYB12
CcMYB12
CcMYB12b
AtMYB12

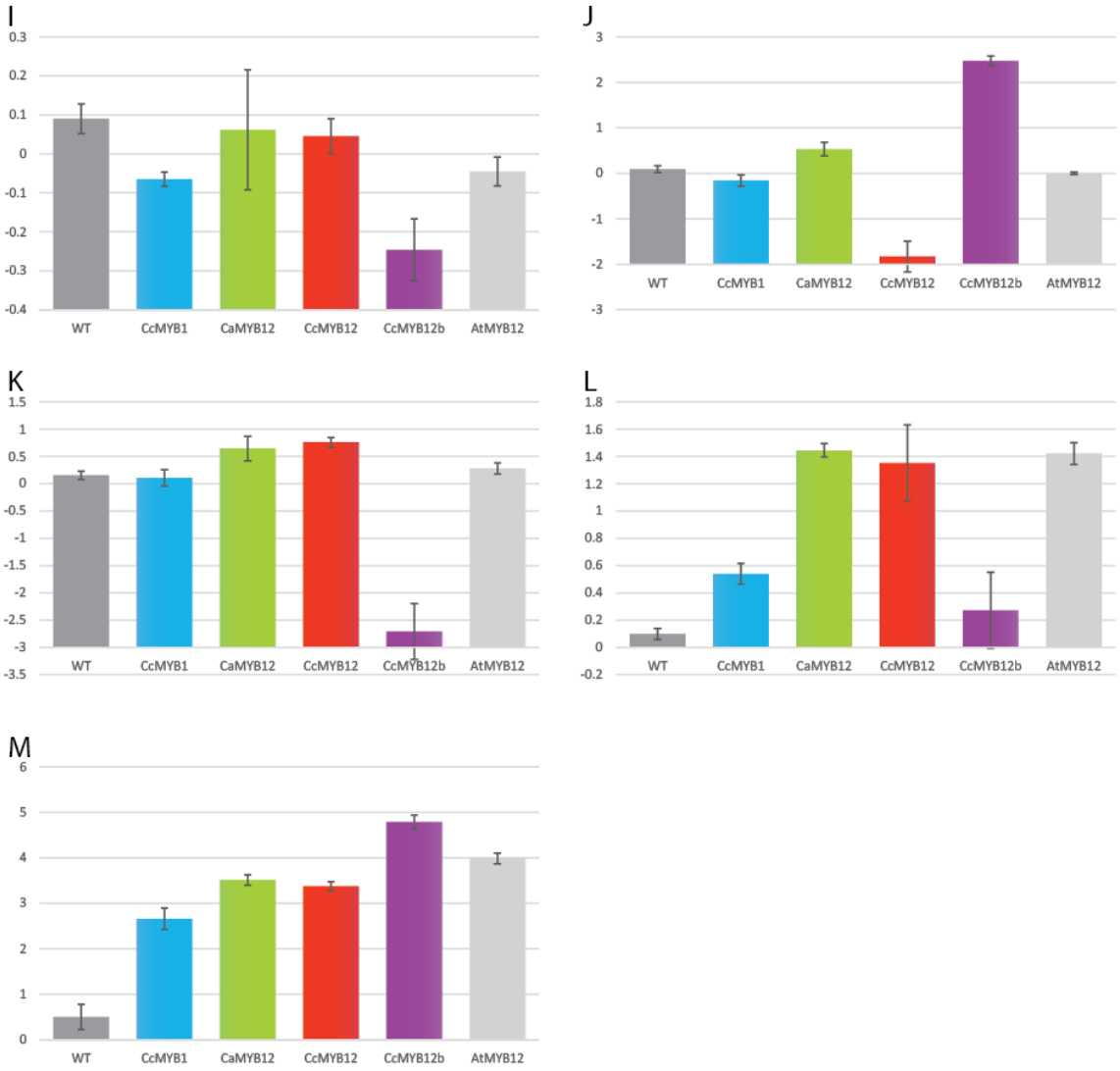


Figure 5.11. Relative transcript levels of major phenylpropanoid genes in transgenic tobacco. cDNA synthesised from leaf RNA using 5 biological replicates in 3 independent lines. Data shown as ΔΔCt values relative to ubiquitin represented as $\log_{10}$. **A**, phenylalanine ammonia lyase (PAL); **B**, cinnamate 4 hydroxylase (C4H); **C**, 4-coumarate:CoA ligase (4CL); **D**, chalcone synthase (CHS); **E**, chalcone isomerase (CHI); **F**, flavanone 3-hydroxylase (F3H); **G**, flavonoid 3'-hydroxylase (F3'H); **H**, flavonoid 3' 5' hydroxylase (F3'5'H); **I**, dihydroflavonol 4-reductase (DFR); **J**, anthocyanidin synthase (ANS); **K**, p-coumaroyl quinate/shikimate 3'-hydroxylase (C3'H); **L**, hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase (HCT); **M**, hydroxycinnamoyl-CoA quinate hydroxycinnamoyltransferase (HQT). The data represented as a heat map is in supplementary figures, Appendix 3.

5.3.4 Candidates that control HQT expression in tomato through analysis of RNA sequence data from introgression lines

I used an introgression line population of *S.lycopersicum* var. M82 with insertions of the *S.pennellii* genome RNAseq data from leaves to identify *cis* and *trans* acting QTL which regulate HQT, HCT and C3'H (Figure 5.12) [156]. The region of the *S.pennellii* genome which contained HQT, which is present in ILs 7-4, 7-4-1, 7-5 and 7-5-5, showed that the *S.pennellii* HQT gene was transcribed less than the equivalent HQT in M82 (*cis*-eQTL). SlMYB12 is a candidate for *trans* regulation of HQT because where it is located, it caused the upregulation of HQT (trans-eQTL in IL1-1-2). IL1-1-2 also shows higher SlMYB12 transcript levels in RNA-seq data from leaves than M82 leaves. Both HCT and C3'H were not upregulated by the increased in expression of SlMYB12 in IL1-1-2 (Figure 5.12). Therefore SlMYB12 is a likely candidate for direct regulation of HQT transcription in tomato. HCT and C3'H were also upregulated by *cis* elements in *S.pennellii* (*cis*-eQTL). I also found several candidates present in different areas of the genome for the upregulation of HQT transcription, these are listed in Table 5.2. To limit search areas, I restricted identification of potential regulators to regions which showed a 20% increase in transcription of HQT. ILs which showed increased levels of HQT transcripts (*trans*-eQTLs) can be mined further to identify possible regulators of its transcription.

Table 5.2 eQTL candidates for the upregulation of HQT transcription

IL line gene position	Solyc gene name	Annotation to TF family
1.1.2	Solyc01g079360	WRKY domain
	Solyc01g079560	B3 domain
	Solyc01g005790	MYB domain
	Solyc00g050430	bHLH family
	Solyc01g079620	SlMYB12 - known
3.4	Solyc03g112390	MYB domain
	Solyc12g049350	MYB12 - like
8.2	Solyc08g076930	MYB domain
	Solyc08g078170	Ethylene-responsive transcription factor
	Solyc08g076700	MYB domain
	Solyc08g080580	MYB domain
	Solyc08g076930	MYB domain

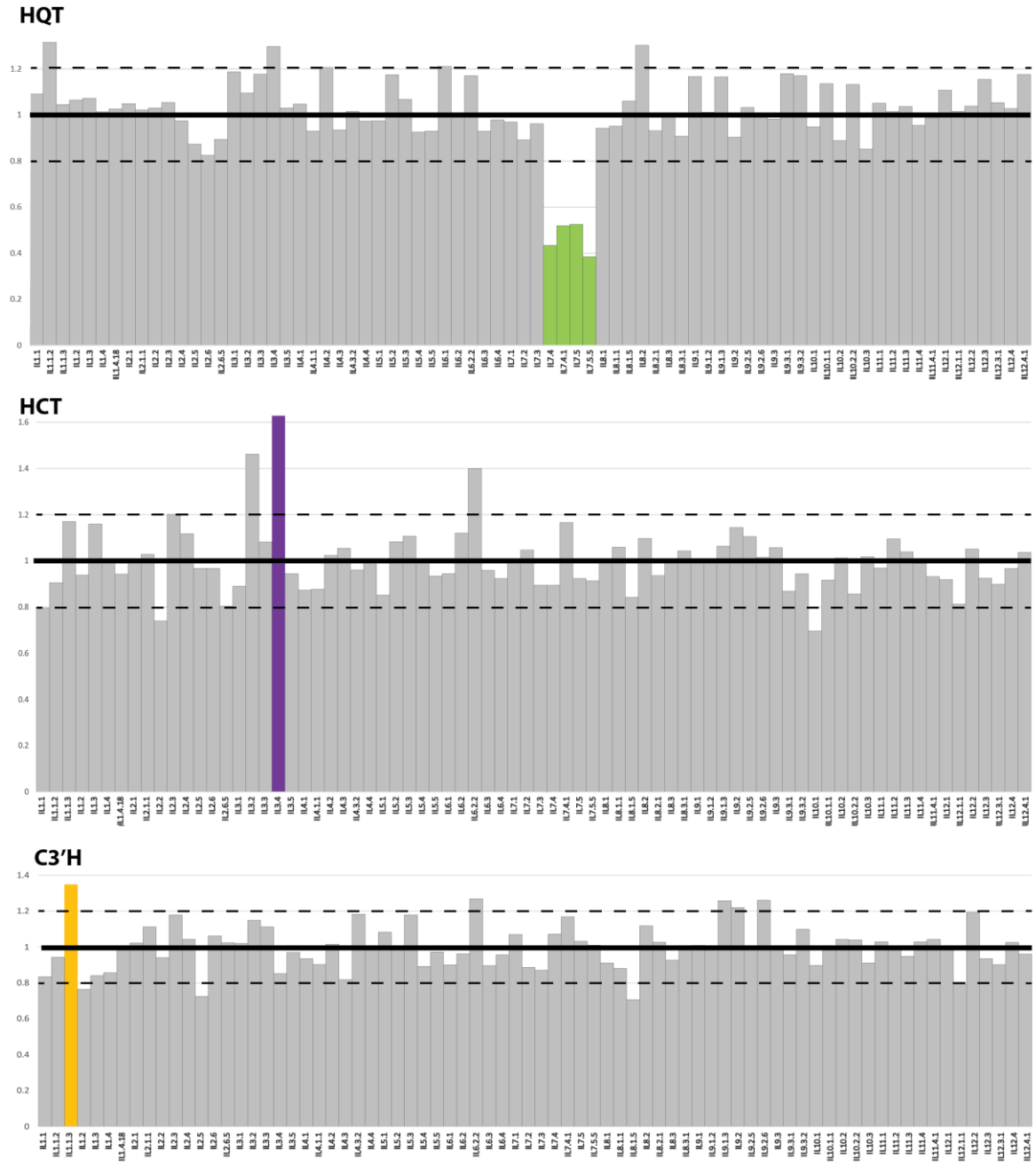


Figure 5.12. *Solanum lycopersicum* introgression lines (IL) of *Solanum pennellii* show *cis* and *trans* regulation of HQT, HCT and C3'H. X axis is each line of introgression of the *S.pennellii* genome fragments replacing *S.lycopersicum* in the same position. Each number is an indication of where the insert of the *S.pennellii* genome fragment is within the *S.pennellii* genome. Y axis is the normalised RPKM values for each respective gene. The solid line shows the normalised average and dashed lines indicates a 20 % difference to the average. The coloured bar indicates the position(s) of the respective genes within the genome. HQT showed *cis* regulation where the HQT gene from *S.pennellii* was not transcribed as much as the HQT from *S.lycopersicum.* Both HCT and C3'H showed *cis* upregulation, the copies of these genes present in *S.pennellii* are transcribed more than HCT and C3'H from *S.lycopersicum*. The insertion of *S.pennellii* genome in position 1.1.2 caused the increased transcription of HQT. SlMYB12 was regulated by its *cis* element in this IL line its increased transcription likely targeted HQT and increased its expression. Other candidates for *trans* regulation could be found within areas which show upregulation or repression. Likely areas are within regions exceeding a 20% difference in transcription.

5.3.5 The overexpression of Coffee MYB transcription factors changes the ability of seedlings to survive high levels of UV light

Tobacco seedlings stably overexpressing CcMYB1, CaMYB12 and CcMYB12 were exposed to UV light for different lengths of time to assess the affect the MYB genes had on the seedlings ability to survive prolonged exposure to UV light. Seedlings overexpressing CcMYB1 showed increased sensitivity to UV light in comparison to the WT (Figure 5.13). This was similar to observations made when AtMYB4 was overexpressed in Arabidopsis and exposed to UV light [82]. Tobacco seedlings overexpressing CcMYB1 showed more yellowing, shrivelling of leaves and death from both selected lines when compared to the WT seedlings exposed for the same length of time. This effect was likely due to the reduced levels of flavonols and chlorogenic acids.

Tobacco seedlings overexpressing CaMYB12 and CcMYB12 showed an increased ability to survive when compared to the WT from exposure to UV light for 40 minutes. Both selected lines from both transformations stayed greener and less shrivelled when compared to the WT (Figure 5.14). The ability of these stable transformations to cope with a prolonged period of UV light was probably due to the higher accumulation of chlorogenic acids and flavonols. There was no visual difference between CaMYB12 seedlings and CcMYB12 seedlings following their exposure to UV-B light.

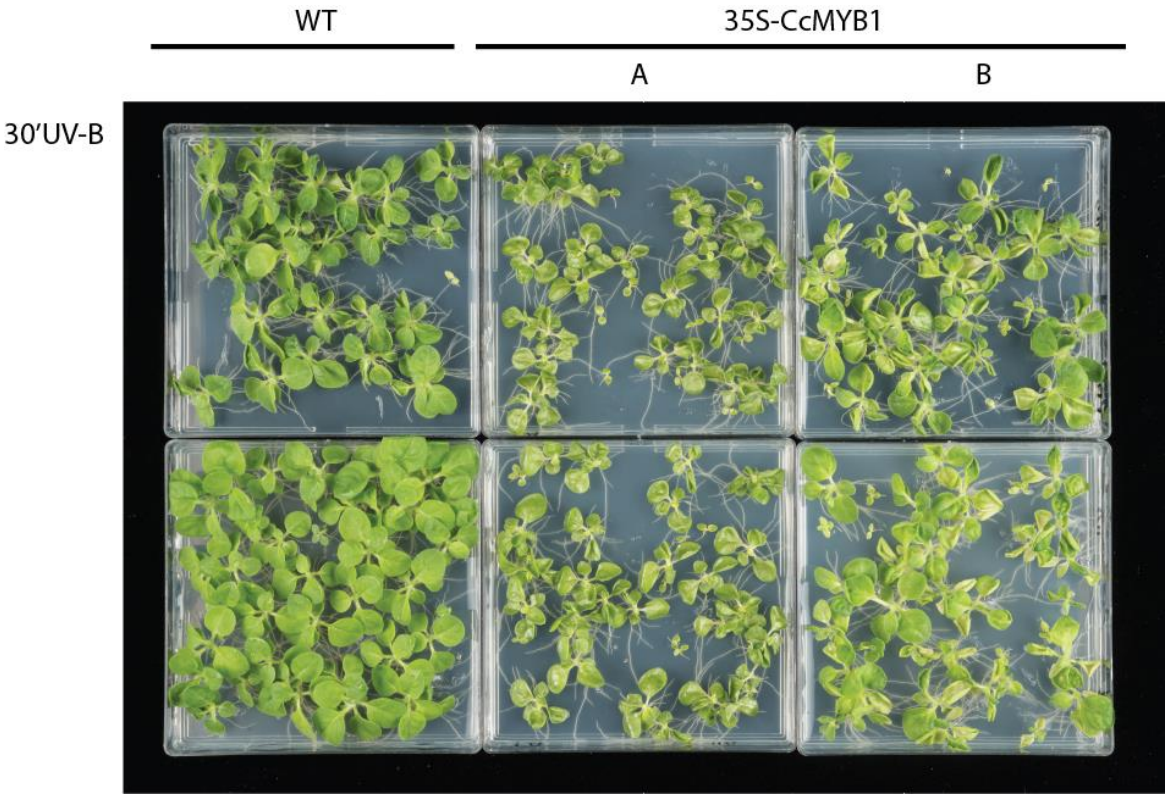


Figure 5.13. Tobacco 35S:CcMYB1 seedlings are more sensitive to UV-B light. Seedlings expressing CcMYB1 showed an enhanced sensitivity compared to WT when exposed to 30 min UV-B (3.2 mW/cm2). Two replicates of two independent transgenic lines were used for the experiment

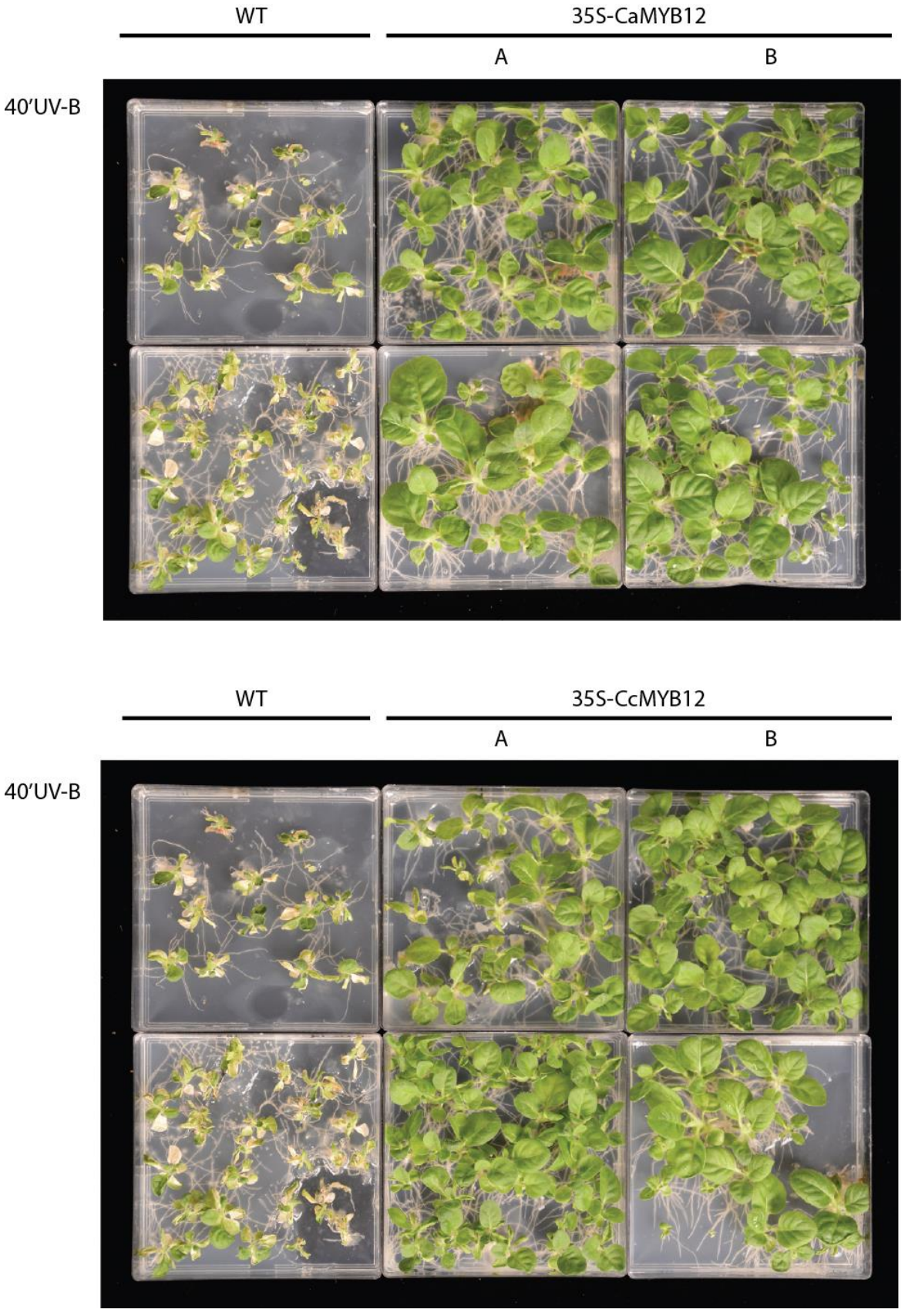


Figure 5.14 Tobacco 35:CaMYB12 and CcMYB12 showed enhanced protection from UV-B. Tobacco seedlings expressing CaMYB12 and CcMYB12 maintained a healthier growth compared to WT when exposed to 40 min UV-B (3.2 mW/cm2). Two replicates of two independent transgenic lines were used for each experiment.

5.3.6 Coffee MYB transcription factors bind to the promoter regions of coffee genes related to the phenylpropanoid pathway

To show that the Coffee MYB transcription factors are likely to be functionally active in coffee and play a part in regulating the phenylpropanoid pathway I performed yeast 1 hybrid (Y1H) assays on a range of promoters isolated from *Coffea canephora* FRT32. A Y1H assay is an indication that there is interaction between the transcription factor overexpressed and the promoter region. The limitations of a Y1H assay are that it only shows if there is an interaction between the transcription factor and the promoter, it does not show if the interaction is capable of upregulation or repression of polymerase II activity. Where no interaction is observed in a Y1H assay it does not mean that there is not interaction between transcription factor and promoter *in vivo*.

Both CcMYB2 and CcMYB1 did not show any interaction with any of the promoters tested. CcMYB2 belongs to subgroup 6 R2R3MYBs which require the interaction with bHLH in order to function [79] and CcMYB1 has the residues which indicate that it may require interaction with a bHLH. The lack of an appropriate bHLH protein in the yeast 1 hybrid assays is a possible reason why no interaction was observed. This is not proof that CcMYB1 (subgroup 4), requires the formation of MBW complex to repress transcription, further experiments will need to be performed to confirm this. It has been suggested but not proven that some subgroup 4 R2R3MYBs require interaction with bHLH proteins [79].

5.3.6.1 *CaMYB12 and CcMYB12b interacted with the cinnamate 4 hydroxylase (C4H) promoter from* Robusta *coffee.*

CaMYB12 and CcMYB12b showed a positive interaction with the promoter regions of cinnamate 4 hydroxylase (C4H). No growth of yeast was observed in concentrations of 3-AT of 5 mM and or in the control GUS expression with the C4H promoter when compared to CaMYB12 and CcMYB12b expression (Figure 5.15, panel A). This promoter was found next to a copy of C4H that was not assigned to a particular chromosome (identified as chromosome 0 on searches), out of the 4 putative C4H genes found in the *Coffea canephora* this was the only promoter region that showed positive interactions. *In silico* analysis of the promoter region showed that there were two possible motifs to which AtMYB12 might bind (Figure 5.16). The motif closest to the TATA box was the most similar motif to the AtMYB12 consensus motif and was located about 30 base pairs upstream of the predicted transcription start site, similar to positions observed from AtMYB12 binding in tomato [151]. Coffee MYB transcription factors CaMYB12 and CcMYB12b showed very close homology in their R2R3 regions to AtMYB12 so their binding motif was most likely similar or identical to that of AtMYB12.

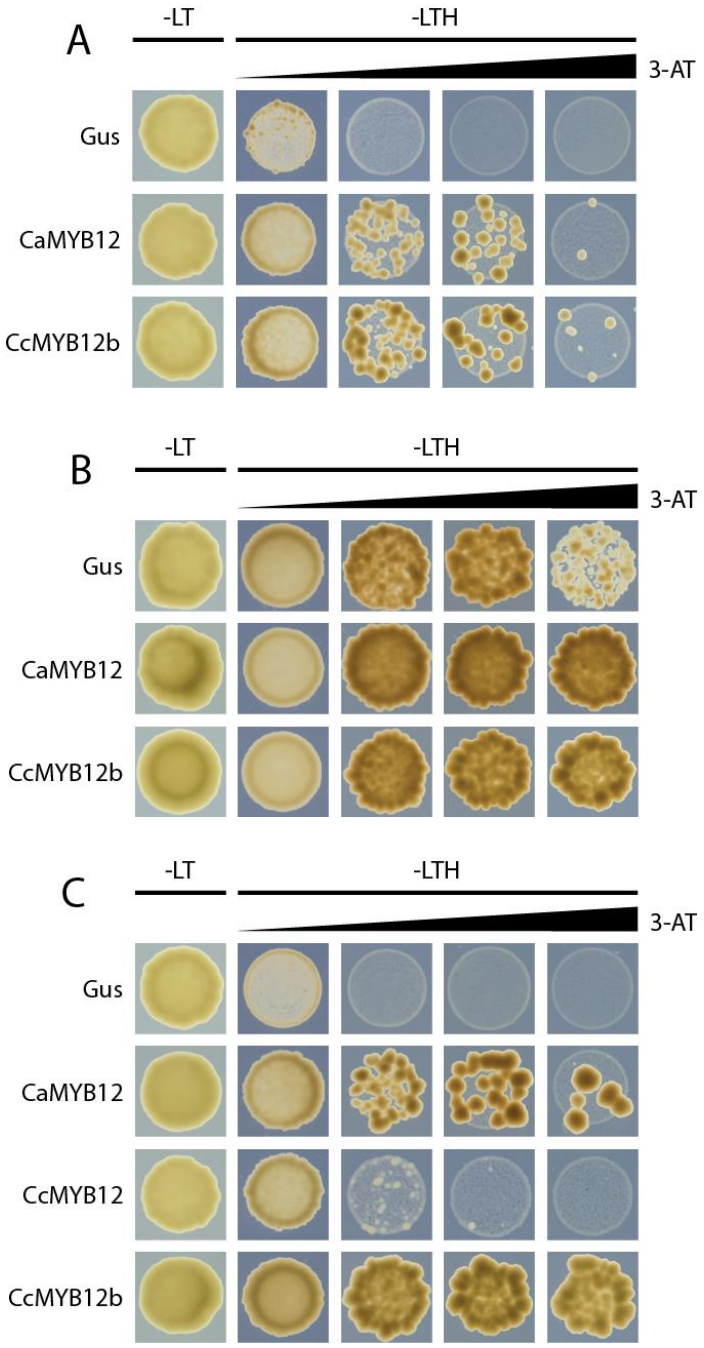


Figure 5.15. Several coffee R2R3MYBs showed interactions with promoter regions with structural genes of the phenylpropanoid pathway from Robusta. Yeast 1 hybrid assay, transcription factor showed interaction with promoter if yeast colonies appeared in increasing concentrations of 3-AT compared to Gus control. 3-AT at 0 mM, 5 mM, 10 mM, and 20 mM. **A**, cinnamate 4 hydroxylase (C4H), CaMYB12 and CcMYB12b showed an interaction; **B**, chalcone synthase (CHS), CaMYB12 and CcMYB12b showed a positive interaction; **C**, chalcone isomerase (CHI), CaMYB12, CcMYB12 and CcMYB12b interacted. Promoters tested represented 1 Kb upstream of ATG translation start site.

5.3.6.2 *CaMYB12, CcMYB12 and CcMYB12b interact with the chalcone isomerase (CHI) promoter from Robusta*

The promoter of chalcone isomerase from *Coffea canephora* showed an interaction with CaMYB12, CcMYB12 and CcMYB12b transcription factors. Yeast carrying the chalcone isomerase (CHI) promoter and overexpressing GUS protein showed complete inhibition of growth on plates supplemented with 5 mM 3-AT (Figure 5.15, panel C). Yeast overexpressing CaMYB12 with the CHI promoter showed growth on concentrations of 3-AT reaching 20 mM. CcMYB12 overexpression in yeast with the CHI promoter showed small amounts of growth in concentrations of 3-AT of 5 mM and less in 10 mM. Yeast overexpressing CcMYB12b with the CHI promoter showed good growth on all 3-AT concentrations up to 20 mM. The copy of CHI is present on Chromosome 2 of the *Robusta* genome. I found 2 copies of CHI to be present in the *Coffea canephora* genome sequence and this was the only promoter that showed an interaction with the tested MYBs. *In silico* analysis of the promoter region showed a motif that was the same as the predicted binding motif of AtMYB12 present 260 bp upstream of the transcription start site (Figure 5.16, top). In tomato, the furthest AtMYB12 binding motif was 144 bp upstream from the start of transcription of F3H (Figure 5.16, top [92]). Although the predicted motif of AtMYB12 in the CHI promoter is a further 116 bp upstream of the furthest AtMYB12 binding motif in tomato, transcription factors have been observed to bind to promoters this far away from the transcriptional start site.

5.3.6.3 *CaMYB12 and CcMYB12b interact with the chalcone synthase (CHS) promoter from Robusta*

There was an interaction between CaMYB12, CcMYB12b and the promoter region of chalcone synthase (CHS) from *Coffea canephora.* Yeast overexpressing GUS with the CHS promoter showed less growth on a concentration of 20 mM 3-AT compared to yeast overexpressing CaMYB12 and CcMYB12b (Figure 5.15, panel B). This copy of CHS lies on chromosome 3 of the *Coffea canephora* genome. A total of 4 copies of CHS were identified during searches. This was the only CHS promoter region that showed an interaction with the tested MYBs. RNA sequence reads available in the coffee genome database indicated that this gene has a coding region that translates to form only a 64 amino acid peptide. These data are from many reads of RNA sequence data from tissues that include pistil, root, stamen, young leaves, old leaves, and, stem and flower [53]. This is far smaller than the full length CHS protein found in Arabidopsis (395 aa) so therefore it is likely that this peptide has little or no function despite being expressed in many tissues.

Name		Strand	*P* value	Sites
Solyc03g114500	*ENO*	+	5.35e−05	$^{-77}$TGTCCATTA CACCTACC AACTCATTG^{-52}
Solyc04g074480	*DAHPS*	+	4.13e−05	$^{-108}$AATACCCTT TACCTAAC TATTGTCCA^{-83}
Solyc09g007920	*PAL5A*	+	1.09e−05	$^{-57}$AATTTTCTT TACCTACC ATCCTTTGT^{-32}
Solyc09g007900	*PAL5C*	+	1.93e−04	$^{-67}$GTATTTTCC TAGCAACC CCCTCTCAC^{-42}
Solyc09g007890	*PAL5D*	+	4.13e−05	$^{-68}$TTTTTTTTT TACCTAAC CATCATTTG^{-43}
Solyc09g091510	*CHS-1*	+	4.13e−05	$^{-66}$CGTGATCTC TAGCTACC ATTCTTTTT^{-41}
Solyc02g083860	*F3H*	+	1.09e−05	$^{-159}$GTATGGTGG TACCTACC CTACAATTG^{-134}

C4H

CAAAAGACCAACATATGATATGAGCATCCAATACATATCTGATTGTTGATCTCAGTTT
CACAAGTACTTAAATTTATTTTCTCCACCAAAGCCAAATTAAGGTAAAAACCGGCAAG
AAATTATCTAGTTCCTCGTGTGATTTCCACACAAATTGTAGAAAGGCAAGTTTGAATT
GTTTCAAGACTTTGTACAAAGATGGATGGCCTTGGAGAGGCACTTCAACTACCAAAC
CTGCAAGTGCAACCACCAAATTTGACTTGAGAAAATGGTATTTCATTCTCCTAACCAG
ATTCATAAGCAAGCCCGTGGATTTTCCTAAAACCAGCGGATACCTACCTTCCAAACCC
ACTTACATTAGTCATATATAAACTCAATGTCCTTACCTTTCCTTCCTCGTACTGATCAGT
TTCCTATCCCTCTTACAACATG

CHS

AACCATTAATGAAATACTAGTTAAAAAAATAGTAGTAATACTAGTGAACCAAAATTTT
TTTCTTTTTTGTCCTTCCTCTTCTCTTCACACCCTTTTTTTCCTCAAAATTCGGACCGTAC
ACCGGACAGATAGCATGAAAAACTAGTAAATCAATTTTTTCTCAAAAAACCACCATAT
CGAAATAGGTTTATTATAGCTTTATTACTAGCTTGTGTGTTGATTGATGACACTAAGG
TCCAAACATGGCAACAAAAATTGAGAGTGTACGCAGGGAAATTGGAACTCGGTGTT
AGGTACCAAATTTCGGGGTCATGTAGATGGTTTGAGGGGCAAGTGCAAACCACACG
AAACTATCACGTGCGGCCGGCAACCTTACGTTCGCTAGGCATGTGATTCTAAGCTACC
ATTATAGTAATTACCCCATCTTCGAGTACTATATATACCAAACCATTTTGCAGACTTGC
ACAGCCACAACTTCTTCCGGAGCTACAAAGCTAGCAGCTACCTGCAAAACTCTACGTA
CCTTAATTTTTCTCTTTTTCCGAGCAAAAATG

CHI

ACATCAACATCAAAAAATCAAATCCCCCAAAAGACTTTGAACGACCTGGTTCTGGTTG
GCGAACTCGTGGCCACGTTTTAACATCTACCAAACTAAACCCCCTGCAACAGCTGAAT
CTACTCGTCACTCTCACCTTCATTAATTTCCCCATCAAAATTTCCAAGCTCCTACAGCG
AGTTTTCGTGTGTTTGCATAACTTCGCCGGCAGTTTAGCGTCGGTGTCATGGTACTAG
GCCATCCTACTCTCTGGGGAAGTACTCTACTGCATTTTTTTTTGGTCCATACTCCCCAG
CTATATATATATATATAAAGCCATAACCTGACCCTGCTTTTTTCCAATAACAAGAAG
AACATTTTATTGATAACAACAGATATCGTCCATTCCTCCTACTTTCTCACTAACCAGTC
CACCCATCTCCAATTCTCCCTCCACCACCCTTCGGTCAATTTTGCTTCTCTACCCCCCAA
AAAAAAAAAAAAGTTTCAATCTCCAAGGCTCCAAATTCTGCCCATCCCAACTAACCCC
TTCGTTTTAAACGACTCTATATATAGCCTCATACCACCACTACCACTGCTACAAACTGC
AAAACTCCCAGACTGCCAACTCTTCAAAGGTGCCGCGAAATCTCTGAGTGCTAGTTAC
AATATG

Figure 5.16. Binding motif of AtMYB12 and promoter regions of coffee showing interaction with R2R3MYBs. Top, reproduced from Zhang et al,. (2015) binding motif of AtMYB12 in tomato and respective promoters it binds [92] **Bottom**, promoters of Robusta phenylpropanoid genes from Y1H interactions. **Blue** predicted CAAT sequence, **Red** TATA binding sequence. Grey the 5'UTR and green is methionine start of translation. Purple is the best possible AtMYB12 binding motif present.

The predicted truncated version of CHS encodes a protein which aligns to the first 64 amino acids of characterised CHS proteins like those from *Arabidopsis* and *S.lycopersicum.* There is a TAA stop codon after the 64th amino acid. Two other copies of CHS within the genome are also transcribed but in lower levels (Figure 5.17). These other versions were also truncated, encoding peptides which are 328 amino acids. Alignment of these peptides indicated that they are missing about 64 amino acids of the N terminus when compared to characterised full length versions. These other versions of CHS transcripts started from where the short truncated version finished (Figure 5.17). The CHS protein is a large protein which has several interaction domains [157]. It is possible that these longer but truncated versions of CHS encode proteins which have enzyme activity, as most of the interaction domains of CHS proteins are located further upstream and truncated versions of CHS have been reported to be active *in vivo* [158]. There was one CHS gene present within the *Robusta* genome which encoded a protein which was likely to be full length, at 392 amino acids, although this was expressed only at trace levels, in the roots.

I found the Reads Per Kilobase of transcript per Million mapped reads (RPKM) for the 4 copies of the CHS transcripts in the published *Robusta* genome [53]. Two of the CHS transcripts were present for which the RPKM values could be compared. The truncated version of CHS showed very high levels of transcripts in all tissues compared to the full length version of CHS (Figure 5.17). CHS is an essential enzyme for the synthesis of anthocyanins and flavonols. Since coffee accumulates little anthocyanins and there is currently no evidence of significant flavonol accumulation, this observation could be significant.

The promoter region of CHS that showed interaction with CaMYB12 and CcMYB12b has a motif that is very similar to the AtMYB12 motif in tomato (Figure 5.16, top [92]). All but the first T/C is present, 81 bp before the start of transcription, between the CAAT and the TATA boxes.

5.3.6.4 Other promoters that showed no interaction in the yeast 1 hybrid assay

Other promoter regions were also tested in the yeast 1 hybrid assay; including 4 promoters from 4 copies of PAL, a promoter from 4CL, F3H, FLS, HCT, and HQT. None of these promoters showed interactions with any of the R2R3MYB transcription factors. Despite this, a search of the promoter regions indicated that PAL has a CACCTACC 295 bp motif upstream of the ATG transcription start site, this is the same as the predicted binding motif of AtMYB12 in the promoter region ENO (plastidial enolase) in tomato [92].

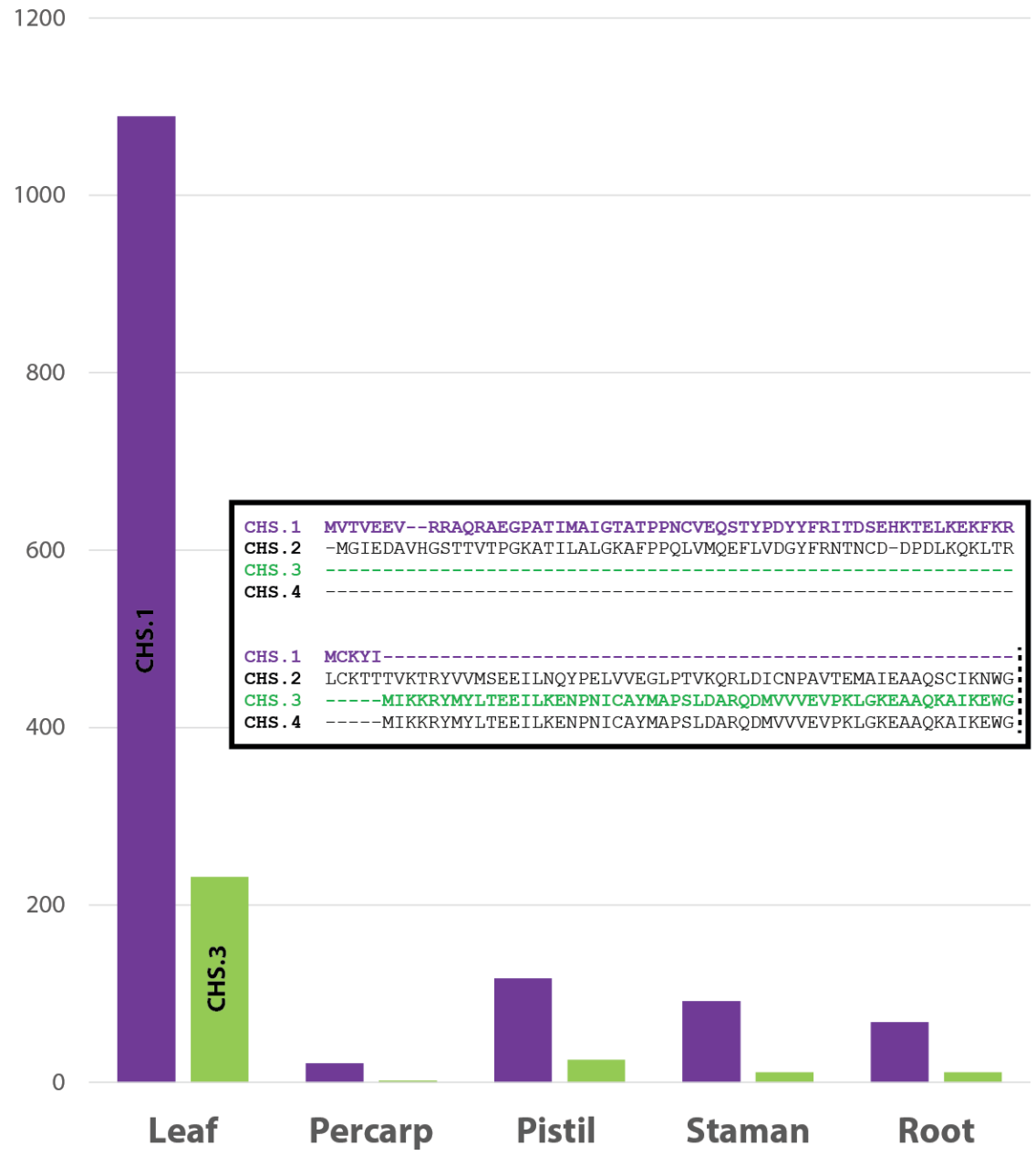


Figure 5.17 Relative RPKM values of two transcribed copies of CHS present in *Robusta* expressed in different tissues and an alignment of all copies present. Purple represents the RPKM values obtained from the truncated version of CHS to which CcMYB12b and CaMYB12 bind the promoter of. Green represents a larger truncated version of CHS present in the genome to which no tested R2R3MYBs bound to the promoter. The truncated version of CHS showed higher RPKM values in all tissues measured. **Black box** shows an alignment of all 4 copies of CHS present within the *Robusta* genome. Only one full length version exists, CHS.2 which did not show high expression in any tissue. CHS.1 is truncated after 64 amino acids, CHS.3 and CHS.4 peptide sequences starts where CHS.1 finishes.

5.4 Discussion

I identified 5 R2R3MYB transcription factors from *Coffea canephora* and *Coffea arabica* encoding candidate regulators of phenylpropanoid metabolism in coffee. One, CcMYB1 belonged to R2R3MYB subgroup 4, a family characterised by transcriptional repressors of phenylpropanoid metabolism. A second, CcMYB2 resembled regulators of anthocyanin biosynthesis (subgroup 6). The other 3, CaMYB12, CcMYB12 and CcMYB12b, belonged to subgroup 7 resembling AtMYB12 and were considered likely to regulate flavonol and hydroxycinnamic acid ester formation, based on data from another species.

Overexpression of CcMYB1 resulted in tobacco plants with a clear phenotype, most noticeable were the white lesions that appeared on the leaves but also the reduction in overall growth of plants and flowers showing lower, patterned anthocyanin accumulation. The lower anthocyanin accumulation, suggested that the reduced anthocyanin production was a result of lower general phenylpropanoid metabolism. These changes were not observed clearly in primary transformants (T0) but became apparent in a significant number of the next generation which were germinated on kanamycin plates to ensure the presence of the neomycin phosphotransferase II gene from the T-DNA insertion. This observed phenotype appeared only in lines with the highest transcript levels of CcMYB1 showing that protein concentration is likely important for the exhibition of the phenotype. This could be because the C2 EAR domain in CcMYB1 contains an aspartate residue instead of an asparagine which is considered essential for transcriptional repression through an EAR motif [87]. The overall stunted growth of CcMYB1 overexpression lines may reflect a lower level of monolignol biosynthesis limiting lignin production, although due to time constraints I did not have time to measure lignin levels quantitatively or by staining with phloroglucinol.

Analysis of tobacco lines overexpressing the CcMYB1 gene demonstrated that CcMYB1 acts as a negative regulator of hydroxycinnamic acid and flavonol metabolism. The accumulation of chlorogenic acid, kaempferol, and rutin were lower than levels seen in WT. This negative regulation is likely to occur through the control of transcription of 4-coumarate:CoA ligase (4CL), principally because transcript levels of 4CL in CcMYB1 overexpression were less than transcript levels observed in WT and PAL and C4H transcript levels remained unaffected. Overexpression of AtMYB4 in Arabidopsis resulted in a primary inhibition of C4H transcript levels, but at higher levels of AtMYB4, 4CL transcript levels were also reduced [82]. Changes in transcription of these genes affects all the branches of the phenylpropanoid pathway downstream of p-coumaroyl CoA which is a central compound in the diversification of phenylpropanoids.

Although no activity was detected in the Yeast 1 hybrid assay for CcMYB1 against any promoter regions of phenylpropanoid pathway genes, it is possible that CcMYB1 does bind to the 4CL and/or C4H promoter regions in coffee or CcMYB1 represses a repressor of 4CL. Transactivation assays using the respective promoters driving expression of a reporter gene could be performed in CcMYB1 overexpressing tobacco plants to investigate the direct interaction with the phenylpropanoid pathway gene promoters. The activity of CcMYB1 on the repression of transcription of genes could be through competitive repression through displacing activators binding to the promoter or as a direct repressor through the EAR repression domain present in the C-terminal domain of the CcMYB1 protein. Also, I observed that the key residues for bHLH binding were present in CcMYB1 which might be required as a cofactor for binding which was not present in the yeast 1 hybrid assay. It has been suggested, that AtMYB4 also binds to bHLH proteins [79].

To test the theory that CcMYB1 requires a bHLH in order to operate as a repressor, tobacco plants stably expressing CcMYB1 were crossed with several other tobacco plants expressing bHLH from different plant species (Delila and Mutabilis from *Antirrhinum majus*; [81]). The resulting plants maintained the same phenotypes as CcMYB1 alone, meaning that CcMYB1 is unlikely to bind to bHLH proteins to increase repression of its target genes. Other evidence that CcMYB1 represses the 4CL promoter in tobacco directly is the fact that AmMYB308 (a functional homologue of CcMYB1) requires its C-terminal domain in order to exhibit the stunted growth phenotype in tobacco. This was tested by the overexpression of just the N-terminal DNA binding domain in tobacco which did not result in the stunted phenotype that overexpression of the full length protein caused [84]. The C-terminal domains of subgroup 4 R2R3 contain a conserved EAR motif that is essential to their active repression of their targets. The R2R3MYB can bind to its *cis* element and the EAR motif will interact with co-repressors which probably occurs through chromatin modification of regulatory regions by histone deacetylation [88].

The overexpression of CcMYB1 in tobacco caused seedlings to become more sensitive to UV-B light, primarily because of reduced chlorogenic acid and flavonol levels. This suggested that the primary function of CcMYB1 is to act as a direct repressor of general phenylpropanoid metabolism.

Overexpression of CaMYB12 and CcMYB12 in tobacco caused a dramatic increase in flavonols and chlorogenic acid accumulation very similar to the effect of AtMYB12 overexpression in Solanaceous species [90]. This increased accumulation of chlorogenic acids and flavonols caused seedlings to be more tolerant to UV-B light. There were no significant differences in the flavonols, chlorogenic acids or the phenotypes observed or the resultant changes in expression of phenylpropanoid pathway genes between CaMYB12 and CcMYB12 or between the two MYB12 genes from coffee and

AtMYB12 overexpression in tobacco. The only differences in the amino acid sequences of the proteins encoded by these genes lay in the C-terminal domains, suggesting any functional differences would be due to differences in their activation domains. It was not possible to determine whether the C-terminus of CaMYB12 or that of CcMYB12 was better at promoting transcription through overexpression of these genes in tobacco although, because the C-terminus of CaMYB12 is shorter than that of CcMYB12 it was hypothesised that the two genes might show a difference in their activation domain and exhibit different activation characteristics. However, from the metabolic data, phenotypic data and target gene expression analyses no differences in activity between CaMYB12 and CcMYB12 could be detected, nor indeed differences between these transcription factors and AtMYB12.

The DNA binding domains of CaMYB12, CcMYB12 and AtMYB12 are almost identical and it is very likely that CaMYB12 and CcMYB12 target exactly the same promoters as AtMYB12 in Solanaceous species. My yeast 1 hybrid assay of promoter regions of *Robusta* revealed that CaMYB12 and can bind to the promoter of CHS which is also a target of AtMYB12. It was also observed that CaMYB12 can bind to the promoter of C4H in *Robusta* and both CaMYB12 and CcMYB12 can bind to the promoters of CHI. To my knowledge, this is the first time that MYB12 homologues have been reported to bind to these promoters which was not observed in ChIP seq analysis of AtMYB12 overexpression in tomato fruit [92]. Although I did not see an interaction between CcMYB12 and the promoter region of CHS, whereas CaMYB12 did interact with this promoter, all the functional analysis indicated that these genes are identical so it is likely that CcMYB12 does interact with the CHS promoter. The overexpression of CcMYB12 in tobacco also caused the upregulation of transcription of CHS to the same levels observed from CaMYB12 overexpression. There were no significant differences between the levels of flavonols accumulated between CcMYB12 and CaMYB12. Failure to observe an interaction between a transcription factor and a promoter in a yeast 1 hybrid assay is well known not to reflect *in planta* activity and that an interaction *in vivo* may be possible.

The chalcone synthase (CHS) promoter from *Robusta* that was bound by CaMYB12 and CcMYB12b in my yeast one hybrid assays, drives the expression of a putative protein that is only 64 amino acids long, this is shorter than other reported functional versions of CHS and it is likely that this version of CHS does not have functional activity when expressed. I also found two other copies of the CHS gene in the *Robusta* genome sequence. None of the promoter regions of these versions of CHS were found to interact with any of the tested R2R3MYB transcription factors in the Yeast 1 hybrid assay. I used the RPKM values for the genes to understand in what tissues different versions of CHS are

expressed. Only two of the CHS genes were present in the RNA-seq data. The RPKM values showed that the truncated version of CHS was present at much greater levels than the full length version. The highest levels of expression were in the leaves. The leaves of coffee accumulate high levels of chlorogenic acid and CcMYB12b, which binds to the promoter of this version of CHS is also transcribed at high levels in the leaves. CcMYB12b could be controlling chlorogenic acid biosynthesis in the leaves by upregulating general phenylpropanoid pathway genes. CcMYB12b would also promote the transcription of the truncated CHS protein, so blocking the diversification of p-coumaroyl CoA to form flavonols through the activity of CHS. A functional copy of CHS is also expressed in leaves, but at much lower levels. The anthocyanins, cyanidin-3-rutinoside and delphinidin glycoside accumulate at low levels during early leaf development [159, 160]. This functional copy of CHS could be facilitating the production of anthocyanins at low levels in young leaves, but perhaps is not controlled by CcMYB12b, such that anthocyanin accumulation can happen independently from chlorogenic acid biosynthesis in coffee. Unfortunately, there were no data available for the RPKM values of CHS in the endosperm of coffee beans. If the truncated version of CHS is expressed in high levels in the endosperm, together with the fact that I could not detect flavonols in the coffee fruit during development, this would suggest that CcMYB12b is responsible for the high levels of chlorogenic acid in beans and that although CcMYB12 and CcMYB12b activate transcription of CHS this results in activation of expression of a short CHS peptide that would have no CHS activity, blocking any competing production of flavonols or other flavonoids.

The observation that CaMYB12 and CcMYB12b drive transcription of a putative non-functional version of CHS could be a reason why coffee accumulates such high levels of chlorogenic acid while maintaining little or no flavonol accumulation. Whilst other MYB12 homologues bind to - and upregulate a functional version of CHS including other upstream phenylpropanoid pathway genes, which upregulate both flavonoid and chlorogenic acid synthesis pathways, the coffee MYB12 homologues do not activate the branch leading to the biosynthesis of flavonoids. Therefore, they likely upregulate only the pathway leading to chlorogenic acid biosynthesis.

CaMYB12, CcMYB12 and CcMYB12b all bound to the promoter of CHI from *Robusta* in the yeast one hybrid assay which is a gene that is involved downstream of CHS in the biosynthesis of anthocyanins and flavonols. The RNA sequencing data available on the coffee genome hub showed that this gene is expressed in many tissues and the RNA sequence reads make up a gene that encodes a peptide that is 260 amino acid long, CHI from Arabidopsis is 246 amino acids long so it is very likely that this version of CHI in *Robusta* is functional.

Coffee as a plant genus, accumulates very low levels of flavonols, immature pericarp of *Coffea arabica* only accumulates 0.250 μg/mL and 1.154 μg/mL of kaempferol and rutin respectively. In the same tissue, chlorogenic acid levels are 298.82 μg/mL. In young leaves of *Coffea arabica* kaempferol and rutin account for 1.544 μg/mL and 1.599 μg/mL and chlorogenic acid levels are 605.65 μg/mL [161]. Interestingly, an analysis of polyphenol content did not detect any flavonols in whole coffee fruit of *Robusta* obtained from plants grown in Mexico, India and China whereas the flavonol rutin was detected in *Arabica* coffee fruit from the same countries [18]. This, coupled with the fact that CcMYB12 and CcMYB12b are expressed in the coffee fruit (pericarp and grain), indicate that the expression of the predicted non-functional version of CHS likely induced by CcMYB12 and CcMYB12b does not result in significant flavonol accumulation in beans. This was confirmed by the MS imaging described in chapter 4. Interesting future research objectives would be to investigate whether the coffee flavonol deficiency is due to a lack of a functional copy of CHS being induced by CaMYB12, CcMYB12 and CcMYB12b. The truncated version of CHS could possibly act as an inhibitor of functional CHS activity [157]. The full CHS protein exists as a homodimer which interacts through two α helices situated within the first 50 amino acid residues of the N-terminus [158]. If the truncated version of CHS which is 64 amino acids long folds to form the two α helices, it could interact with a full length version of CHS and prevent it from forming its homodimer [158].

The differences in chlorogenic acid levels in *Robusta* and *Arabica* cannot be deduced from any differences observed from the functional analysis of CcMYB12 and CaMYB12 in these studies. For future investigation into why *Robusta* accumulates more chlorogenic acid than *Arabica,* CaMYB12 gene expression analysis would need to be performed and compared to CcMYB12 expression in comparable tissues. I would also need to determine whether the *C.arabica* genome contains a gene encoding CaMYB12b. Many R2R3MYB transcription factors control the phenylpropanoid pathway and it is unlikely that the difference in chlorogenic acid accumulation between the species is due the differences in just these two transcription factors.

The gene CcMYB12b, when overexpressed in tobacco upregulated the accumulation of only chlorogenic acid biosynthesis. To my knowledge, this is the first example of a R2R3MYB transcription factor causing an increase in accumulation of only chlorogenic acids in any plant species. Considering the DNA binding domains of CcMYB12 and CcMYB12b are very similar, this effect was unexpected. CcMYB12b also caused the levels of flavonol accumulation to drop compared to WT when overexpressed in tobacco. The drop in flavonol accumulation observed in CcMYB12b compared to WT could be for several reasons: because of the change in flux characteristics of the pathway genes, or CcMYB12b could be competitively binding to the promoter

regions of flavonol specific pathway, or CcMYB12b could be activating the transcription of a specific inhibitor of the flavonol biosynthetic genes. This observation is similar to the effects observed from StMtf1 overexpression in potato tubers where they observed that a mutated StMf1, which was predicted to cause anthocyanin accumulation when expressed in all tissues, caused the accumulation of CGA and flavonols in tubers [93]. They showed that the HQT gene was upregulated which was not initially thought to be a target of StMtf1. Microarray experiments showed the upregulation of important precursors of phenylpropanoid biosynthesis such as proline dehydrogenase (PDH) and S-Adenosyl-l-methionine synthetase (SAMS) genes. This triggered the modifications to the expression levels of genes involved in CGA, lignin and flavonoid biosynthesis.

Because *Arabica* shares half its genome with *Robusta* (tetraploid made up from *C.eugenioides* and *C.canephora*) it is likely that a version of the *CcMYB12b* gene exists in *Arabica,* or if no functional version is present, this could be a reason why *Arabica* accumulates less chlorogenic acid. Since *Arabica* accumulates more flavonols than *Robusta* it is possible that CaMYB12 activates a functional version of CHS, this could be another reason why there is less chlorogenic acid present. Analysis of CcMYB12b through the yeast 1 hybrid assay revealed that this R2R3MYB most probably acts as an activator of early biosynthetic genes of the phenylpropanoid pathway primarily through direct binding as illustrated by its binding to the promoter of CHI in *Robusta*.

MYB transcription factors from coffee reported here have a dramatic effect on the phenylpropanoid pathway in tobacco. Although I have proven that some of these MYB transcription factors bind directly to the promoter regions of some phenylpropanoid pathway genes in coffee, it does not mean that they have the same effects in coffee as in tobacco. For example, when AtMYB12 is overexpressed in Arabidopsis, it has not been reported to bind to one of the 20 PAL promoters whereas in tomato, AtMYB12 binds to several PAL promoters which results in a higher fold change in flavonol content than in Arabidopsis [90, 91]. Future functional characterisation must be performed in coffee to fully understand the role of these R2R3MYB transcription factors in the control of chlorogenic acid biosynthesis in coffee. The route to completing flavonol biosynthesis might be restricted in *Robusta* because both CcMYB12 and CcMYB12b activate a defective CHS gene.

Both CcMYB12 and CcMYB12b are probably involved in regulating CGA production in coffee beans because 5-CQA synthesis requires the activity of C3'H as well as HQT and C3'H was activated only by CcMYB12. The route to completing flavonol biosynthesis might be cut off because CcMYB12b protein activates a defective CHS gene. At this point it is not clear if *Arabica* also has this truncated

version of CHS. If it does not, it could be a clear reason why *Arabica* accumulates less chlorogenic acids than *Robusta* and modest levels of flavonols are observed in several tissues of *Arabica*.

In tomato, the observation that SlMYB12 is a regulator of HQT by eQTL mapping showed that MYB12 homologues have responsibility for regulating CQA synthesis. This supports a report by Luo et al., (2008) where the upregulation of AtMYB12 in tomato fruit resulted in the increased transcription of HQT and subsequently 5-CQA [90]. Although it was discovered that AtMYB12 targets 7 promoters in tomato, through ChIP seq analysis, MYB12 homologues could be targeting more which were not identified, due to limitations in the technique [92]. Through using the eQTL mapping data I was also able to identify other transcription factor candidates which can be cloned and functionally tested in the role of regulating HQT which is the key enzyme for chlorogenic acid biosynthesis. These candidate regulators could be cloned and functionally analysed from *Robusta* also to investigate further the regulation of chlorogenic acid biosynthesis in coffee.

Chapter 6: General discussion

6.1 Introduction

Coffee is an important agricultural crop and many of the poorest people around the world rely on it for an income. The sustainability of future coffee production is an important avenue to research as the genetic diversity of both *Arabica* and *Robusta* puts global production at risk from pandemic diseases. Understanding how chlorogenic acids are produced within coffee will help provide useful information which can be used to develop new varieties of coffee to improve and safeguard future production.

6.2 Roles of HCT and HQT in chlorogenic acid biosynthesis

It is well established that HQT is directly involved in the synthesis of CGA which uses quinate as a preferred acyl acceptor [57, 62] and that HCT is involved in the synthesis of precursors for lignin biosynthesis which exhibits a preference for shikimate [63]. *In vitro* assays have shown that HCT is able to directly synthesise 5-CQA via acyltransferase activity from caffeoyl CoA to quinate [101]. We showed through the generation of a homozygous knockout of the HQT gene and analysis of CGA content that only under exceptional circumstances HCT can directly synthesise CGA *in planta*. Only trace levels of CGA were detected in tissues with maximal HCT transcription. Therefore, tomato HCT does not have a significant role in directly synthesising CGA in tomato. Through complementation assays I could reintroduce 5-CQA biosynthesis in tomato plants with a homozygous knockout of the HQT gene by overexpressing HQT from either coffee or tomato.

Previously, we reported the route to diCQA synthesis via chlorogenate:chlorogenate transferase (CCT) activity occurred in the presence of the HQT enzyme in tomato [102]. I found that CCT activity is a feature of both HQT and HCT but likely only of significance in the vacuole, where there are high concentrations of 5-CQA, an absence of acyl donors and an acid pH. Kinetic analysis showed that HQT has a predominant role compared to HCT in CCT activity.

DiCQA synthesis was previously reported to occur by another, not fully characterised, route. Lallemand et al., (2012) described an activity mediated by the coffee HCT enzyme. They detected diCQAs when they incubated recombinant HCT with coenzyme A and 5-CQA overnight [101]. When they used a version of HCT which was mutated at a residue responsible for coenzyme A binding, they showed an increase in diCQA production [101]. I was motivated to explore this activity further to understand the substrates and conditions required to synthesise diCQAs via this pathway, I also wanted to understand if this activity was unique for coffee HCT. We found that caffeoyl CoA and 5-CQA are the substrates for this synthesis route which we called caffeoyl CoA chlorogenate transferase activity (CaAGT). I found that this activity is a property of only HCT (both coffee and tomato) and occurs at a neutral pH. Presumably therefore this activity occurs when HCT is localised the cytoplasm in the presence 5-CQA and caffeoyl CoA as the acyl donor. Although enzyme kinetic analysis was not performed with the CaAGT activity it is possible that this pathway can occur in the presence of lower concentrations of 5-CQA compared to CCT activity. Any future investigations should include an enzyme kinetic analysis of CaAGT activities to assess its biological relevance.

MSI is a relatively new technique, previously MSI of CQA in coffee have shown the effects of post-harvest processing and the coffee berry borer [115]. We used the technique to show how changing

metabolite spatial localisation patterns in developing coffee fruit can help explain how diCQAs are synthesised by separate pathways. The pattern of diCQA localisation in coffee during fruit development showed that diCQAs were principally localised to areas of the endosperm where there were also high concentrations of 5-CQA. This suggested that the principle route of synthesis there was through the BAHD activity of HQT and then the CCT activity of HQT. Even though CCT activity is likely to be slow and inefficient, the development of coffee fruit is a process which takes almost 9 – 11 months for Robusta offering plenty of time for CCT activity of HQT to occur during endosperm development. DiCQAs were present at higher concentrations then 5-CQA in the embryo during the late stages of coffee fruit development. It was therefore unlikely that diCQAs were synthesised via the CCT activity of either HQT or HCT in embryos because this activity relies on high concentrations of CQA. A more likely route to diCQA synthesis in embryos was through the CaAGT activity of HCT.

6.3 How coffee MYB transcription factors regulate chlorogenic acid biosynthesis

The first plant MYB transcription factor was described by Paz-Ares et al., (1987) who reported a gene which encodes a protein with 40% sequence homology to animal *myb* proto-oncogenes [123]. This C1 (*COLORED1*) locus from *Zea mays* was predicted to have a role in regulating anthocyanin biosynthesis. Since, many plant MYB transcription factors have been characterised to regulate the phenylpropanoid pathway, and because of their highly conserved R regions which are grouped - have similar functions, allows for confident prediction of unknown genes [86, 131]. Five R2R3MYB transcription factors were cloned from coffee, four from Robusta and one from *Arabica*, which were predicted to regulate the phenylpropanoid pathway. Four of these were overexpressed in tobacco plants and their effects were analysed. Like other members of subgroup 4 R2R3MYBs CcMYB1, repressed the phenylpropanoid pathway and its overexpression resulted in a prominent phenotype associated with repression of general phenylpropanoid metabolism, possibly through the transcriptional repression of 4CL. CaMYB12 and CcMYB12 belong to subgroup 7 of R2R3MYBs, they upregulated the phenylpropanoid pathway which resulted in the higher accumulation of kaempferol glycosides, rutin, and the chlorogenic acids. This activity is common with another characterised subgroup 7 MYBs [90-92]. CcMYB12b, another subgroup 7 MYB, caused a different effect to the other MYB12-like genes by upregulating chlorogenic acid and downregulating flavonol levels in tobacco. I am unaware of any other transcriptional regulators factor which has this effect, albeit in a surrogate host. The gene expression analysis of the overexpression of CcMYB12b indicated that CcMYB12b might be directly upregulating the HQT gene to cause the observed higher levels of CGA accumulation. To understand whether MYB12 homologues also regulate HQT expression in tomato, I showed using eQTL mapping in introgression lines, that the higher expression of SlMYB12 is linked to higher levels of HQT transcripts. HQT itself was expressed at lower levels in leaves of the IL line carrying the HQT gene from *S. pennellii*.

Unusually, coffee accumulates high levels of CGA and maintains low levels of flavonoids [18, 22, 162]. The CHS gene encodes an important enzyme which is responsible for the first committed step in the flavonoid biosynthetic pathway [163]. CHS acts upon coumaroyl CoA which is a major branch point for diversification within the phenylpropanoid pathway [54]. Through yeast 1 hybrid assays, I could show that CcMYB12b and CaMYB12 can interact with a promoter which controls the expression of a truncated version of the CHS protein. Other copies of CHS exist in *Robusta* but their promoter regions do not contain the MYB12 binding motif and did not show an interaction with CaMYB12 and CcMYB12b. The truncated version of CHS was expressed in high levels compared to

the other three versions present within the current published *Robusta* genome [53]. This is a likely reason to why *Robusta* does not accumulate high levels of flavonoids. *Arabica* accumulates more flavonols then *Robusta* but still does not accumulate significant levels compared to its CGA content [18]. As no genome sequence was available currently for *Arabica* it is unknown if *Arabica* also has the gene which produces this truncated version of CHS. It could be hypothesised that *Arabica* does have the truncated version of CHS present and it is upregulated by CaMYB12 but accumulates more flavonols than *Robusta* because half of its genome is derived from the wild *C.eugenioides* [6]. This genome may contain other copies of CHS which are regulated by other mechanisms.

6.4 Future studies and applications of findings

AtMYB12 interacts with several promoter regions of genes encoding enzymes associated with the general phenylpropanoid pathway, flavonoid biosynthesis and primary metabolism [92]. CcMYB12 and CaMYB12 show functional similarity to AtMYB12 and CcMYB12b does not (despite having an identical DBD to AtMYB12). The binding motif for CcMYB12b should be elucidated to explore this phenomenon further. CcMYB12b could be used as a tool to manipulate the phenylpropanoid pathway in other hosts for specific increased accumulation of chlorogenic acid.

CcMYB12b overexpression caused the upregulation of chlorogenic acids but not flavonols in tobacco. Consequently, CcMYB12b could serve as a tool for the analysis of differentiating protective roles of flavonols and chlorogenic acids in plants. Tobacco seedlings which are overexpressing CcMYB12b should be exposed to UV-B light. This will indicate whether CGA plays a bigger role in the protection of plants to UV light compared to flavonols. These plants could also serve as a tool for plant-pathogen interaction studies to analyse which type of polyphenols serve what roles in plant protection.

The understanding of chlorogenic acid biosynthesis in *Robusta* could lead to strategies to reduce the high levels of CGAs in coffee. Currently, *Robusta* coffee is worth around half the price of *Arabica* and accounts for only 30% of world's production. *Robusta* is more sustainable than *Arabica* for many reasons including, better drought tolerance, disease resistance, and ease of cultivation. One of the main reasons why *Robusta* is less desirable and hence not worth as much as *Arabica* is because of its high chlorogenic acid content. The high chlorogenic acid content contributes to the undesirable bitter flavours present in the final beverage. Post-harvest strategies have been implemented in an effort to reduce the bitter flavours which are costly and untargeted [162]. Target assisted breading, or a genome editing strategy on key genes identified in this thesis could result in *Robusta* coffee which accumulated less chlorogenic acids. This may increase demand and help ensure future sustainability of coffee agriculture.

References

1. Rotaris, L. and R. Danielis, *Willingness to pay for fair trade coffee: A conjoint analysis experiment with Italian consumers.* Journal of Agricultural and Food Industrial Organization, 2011. **9**(1).
2. Levy, D., J. Reinecke, and S. Manning, *The Political Dynamics of Sustainable Coffee: Contested Value Regimes and the Transformation of Sustainability.* Journal of Management Studies, 2016. **53**(3): p. 364-401.
3. Ubilava, D., *El Niño, La Niña, and world coffee price dynamics.* Agricultural Economics, 2012. **43**(1): p. 17-26.
4. Davis, A.P., et al., *An annotated taxonomic conspectus of the genus Coffea (Rubiaceae).* Botanical Journal of the Linnean Society, 2006. **152**(4): p. 465-512.
5. Vossen, H., B. Bertrand, and A. Charrier, *Next generation variety development for sustainable production of arabica coffee (Coffea arabica L.): a review.* Euphytica, 2015. **204**(2): p. 243-256.
6. Geleta, M., et al., *Genetic Diversity of Arabica Coffee (Coffea arabica L.) in Nicaragua as Estimated by Simple Sequence Repeat Markers.* The Scientific World Journal, 2012. **2012**: p. 939820.
7. Cenci, A., M.-C. Combes, and P. Lashermes, *Genome evolution in diploid and tetraploid Coffea species as revealed by comparative analysis of orthologous genome segments.* Plant Molecular Biology, 2011. **78**(1): p. 135-145.
8. Lashermes, P., et al., *Genome rearrangements derived from homoeologous recombination following allopolyploidy speciation in coffee.* The Plant Journal: For Cell And Molecular Biology, 2014. **78**(4): p. 674-685.
9. Yu, Q., et al., *Micro-collinearity and genome evolution in the vicinity of an ethylene receptor gene of cultivated diploid and allotetraploid coffee species (Coffea).* The Plant Journal: For Cell And Molecular Biology, 2011. **67**(2): p. 305-317.
10. Bertrand, B. and F. Anthony, *Genetics of Resistance to Root-Knot Nematodes (Meloidogyne spp.) and Breeding*, in *Plant-Parasitic Nematodes of Coffee*, R.M. Souza, Editor. 2008, Springer Netherlands: Dordrecht. p. 165-190.
11. Wolf, A., G.A. Bray, and B.M. Popkin, *A short history of beverages and how our body treats them.* Obesity Reviews, 2008. **9**(2): p. 151-164.
12. *The Merchant Houses of Mocha: Trade and Architecture in an Indian Ocean Port*. 2009, Seattle: University of Washington Press.
13. Ukers, W.H., *All about coffee*. 1922: Tea and Coffee Trade Journal Company.
14. Fredholm, B.B., *Notes on the History of Caffeine Use*, in *Methylxanthines*. 2011, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 1-9.
15. Jaiswal, R., et al., *Profile and Characterization of the Chlorogenic Acids in Green Robusta Coffee Beans by LC-MSn: Identification of Seven New Classes of Compounds.* Journal of Agricultural and Food Chemistry, 2010. **58**(15): p. 8722-8737.

16. Raichle, B.W., R.S. Sinclair, and J.C. Ferrell, *Design and construction of a direct hydro powered coffee depulper.* Energy for Sustainable Development, 2012. **16**(4): p. 401-405.
17. Baggenstoss, J., et al., *Coffee Roasting and Aroma Formation: Application of Different Time–Temperature Conditions.* Journal of Agricultural and Food Chemistry, 2008. **56**(14): p. 5836-5846.
18. Mullen, W., et al., *Polyphenolic and Hydroxycinnamate Contents of Whole Coffee Fruits from China, India, and Mexico.* Journal of Agricultural and Food Chemistry, 2013. **61**(22): p. 5298-5309.
19. Ky, C.L., et al., *Caffeine, trigonelline, chlorogenic acids and sucrose diversity in wild Coffea arabica L. and C. canephora P. accessions.* Food Chemistry, 2001. **75**(2): p. 223-230.
20. Clifford, M.N., et al., *Characterization by LC-MSn of Four New Classes of Chlorogenic Acids in Green Coffee Beans: Dimethoxycinnamoylquinic Acids, Diferuloylquinic Acids, Caffeoyl-dimethoxycinnamoylquinic Acids, and Feruloyl-dimethoxycinnamoylquinic Acids.* Journal of Agricultural and Food Chemistry, 2006. **54**(6): p. 1957-1969.
21. Campa, C., et al., *Qualitative relationship between caffeine and chlorogenic acid contents among wild Coffea species.* Food Chemistry, 2005. **93**(1): p. 135-139.
22. Perrone, D., et al., *Comprehensive analysis of major and minor chlorogenic acids and lactones in economically relevant Brazilian coffee cultivars.* Food Chemistry, 2008. **106**(2): p. 859-867.
23. Kremr, D., et al., *Unremitting problems with chlorogenic acid Nomenclature: A review.* Química Nova, 2016. **39**: p. 530-533.
24. Lepelley, M., et al., *Chlorogenic acid synthesis in coffee: An analysis of CGA content and real-time RT-PCR expression of HCT, HQT, C3H1, and CCoAOMT1 genes during grain development in C. canephora.* Plant Science, 2007. **172**(5): p. 978-996.
25. Mills, C.E., et al., *The effect of processing on chlorogenic acid content of commercially available coffee.* Food Chemistry, 2013. **141**(4): p. 3335-3340.
26. Farah, A., et al., *Effect of Roasting on the Formation of Chlorogenic Acid Lactones in Coffee.* Journal of Agricultural and Food Chemistry, 2005. **53**(5): p. 1505-1513.
27. Ludwig, I.A., et al., *Extraction of coffee antioxidants: Impact of brewing time and method.* Food Research International, 2012. **48**(1): p. 57-64.
28. Caprioli, G., et al., *The influence of different types of preparation (espresso and brew) on coffee aroma and main bioactive constituents.* International Journal of Food Sciences and Nutrition, 2015. **66**(5): p. 505-513.
29. Gloess, A.N., et al., *Comparison of nine common coffee extraction methods: Instrumental and sensory analysis.* European Food Research and Technology, 2013. **236**(4): p. 607-627.

30. Farah, A., et al., *Chlorogenic acids and lactones in regular and water-decaffeinated arabica coffees.* Journal of Agricultural and Food Chemistry, 2006. **54**(2): p. 374-381.
31. Jeszka-Skowron, M., et al., *Chlorogenic acids, caffeine content and antioxidant properties of green coffee extracts: influence of green coffee bean preparation.* European Food Research and Technology, 2016: p. 1-7.
32. Frank, O., G. Zehentbauer, and T. Hofmann, *Bioresponse-guided decomposition of roast coffee beverage and identification of key bitter taste compounds.* European Food Research and Technology, 2006. **222**(5-6): p. 492-508.
33. Farah, A., et al., *Correlation between cup quality and chemical attributes of Brazilian coffee.* Food Chemistry, 2006. **98**(2): p. 373-380.
34. Ranheim, T. and B. Halvorsen, *Coffee consumption and human health – beneficial or detrimental? – Mechanisms for effects of coffee consumption on different risk factors for cardiovascular disease and type 2 diabetes mellitus.* Molecular Nutrition & Food Research, 2005. **49**(3): p. 274-284.
35. Cano-Marquina, A., J.J. Tarín, and A. Cano, *The impact of coffee on health.* Maturitas, 2013. **75**(1): p. 7-21.
36. Higdon, J.V. and B. Frei, *Coffee and Health: A Review of Recent Human Research.* Critical Reviews in Food Science and Nutrition, 2006. **46**(2): p. 101-123.
37. van Dam, R.M. and E.J.M. Feskens, *Coffee consumption and risk of type 2 diabetes mellitus.* The Lancet, 2002. **360**(9344): p. 1477-1478.
38. Yagasaki, K., et al., *Inhibitory effects of chlorogenic acid and its related compounds on the invasion of hepatoma cells in culture.* Cytotechnology, 2000. **33**(1-3): p. 229-235.
39. Olthof, M.R., et al., *Consumption of high doses of chlorogenic acid, present in coffee, or of black tea increases plasma total homocysteine concentrations in humans.* The American Journal of Clinical Nutrition, 2001. **73**(3): p. 532-538.
40. Murase, T., et al., *Coffee polyphenols suppress diet-induced body fat accumulation by downregulating SREBP-1c and related molecules in C57BL/6J mice.* American Journal of Physiology - Endocrinology And Metabolism, 2011. **300**(1): p. E122-E133.
41. Bakuradze, T., et al., *Antioxidant effectiveness of coffee extracts and selected constituents in cell-free systems and human colon cell lines.* Molecular Nutrition & Food Research, 2010. **54**(12): p. 1734-1743.
42. Hartmann, T., *From waste products to ecochemicals: Fifty years research of plant secondary metabolism.* Phytochemistry, 2007. **68**(22–24): p. 2831-2846.
43. Yanhui, C., et al., *The MYB transcription factor superfamily of Arabidopsis: Expression analysis and phylogenetic comparison with the rice MYB family.* Plant Molecular Biology, 2006. **60**(1): p. 107-124.

44. Dixon, R.A. and N.L. Paiva, *Stress-Induced Phenylpropanoid Metabolism.* The Plant Cell Online, 1995. **7**(7): p. 1085-1097.
45. Douglas, C.J., *Phenylpropanoid metabolism and lignin biosynthesis: from weeds to trees.* Trends in plant science, 1996. **1**(6): p. 171-178.
46. Butelli, E., et al., *Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors.* Nat Biotech, 2008. **26**(11): p. 1301-1308.
47. Yao, L.H., et al., *Flavonoids in Food and Their Health Benefits.* Plant Foods for Human Nutrition, 2004. **59**(3): p. 113-122.
48. Farah, A., et al., *Chlorogenic Acids from Green Coffee Extract are Highly Bioavailable in Humans.* The Journal of Nutrition, 2008. **138**(12): p. 2309-2315.
49. Graf, B.A., P.E. Milbury, and J.B. Blumberg, *Flavonols, Flavones, Flavanones, and Human Health: Epidemiological Evidence.* Journal of Medicinal Food, 2005. **8**(3): p. 281-290.
50. Organisation, W.H., *Preventing Chronic Diseases: A Vital Investment*. 2005, WHO: Geneva, Switzerland.
51. Crozier, A., I.B. Jaganath, and M.N. Clifford, *Dietary phenolics: chemistry, bioavailability and effects on health.* Natural Product Reports, 2009. **26**(8): p. 1001-1043.
52. Herrmann, K.M. and L.M. Weaver, *THE SHIKIMATE PATHWAY.* Annual Review of Plant Physiology and Plant Molecular Biology, 1999. **50**(1): p. 473-503.
53. Denoeud, F., et al., *The coffee genome provides insight into the convergent evolution of caffeine biosynthesis.* Science, 2014. **345**(6201): p. 1181-1184.
54. Vogt, T., *Phenylpropanoid Biosynthesis.* Molecular Plant, 2010. **3**(1): p. 2-20.
55. Dao, T.T.H., H.J.M. Linthorst, and R. Verpoorte, *Chalcone synthase and its functions in plant resistance.* Phytochemistry Reviews, 2011. **10**(3): p. 397-412.
56. Zhang, Y., E. Butelli, and C. Martin, *Engineering anthocyanin biosynthesis in plants.* Current Opinion in Plant Biology, 2014. **19**: p. 81-90.
57. Niggeweg, R., A.J. Michael, and C. Martin, *Engineering plants with increased levels of the antioxidant chlorogenic acid.* Nat Biotech, 2004. **22**(6): p. 746-754.
58. Villegas, R.J. and M. Kojima, *Purification and characterization of hydroxycinnamoyl D-glucose. Quinate hydroxycinnamoyl transferase in the root of sweet potato, Ipomoea batatas Lam.* Journal of Biological Chemistry, 1986. **261**(19): p. 8729-33.
59. D'Auria, J.C., *Acyltransferases in plants: a good time to be BAHD.* Curr Opin Plant Biol, 2006. **9**(3): p. 331-40.
60. Bayer, A., X. Ma, and J. Stockigt, *Acetyltransfer in natural product biosynthesis--functional cloning and molecular analysis of vinorine synthase.* Bioorg Med Chem, 2004. **12**(10): p. 2787-95.

61. Legrand, G., et al., *Identification and Characterization of Five BAHD Acyltransferases Involved in Hydroxycinnamoyl Ester Metabolism in Chicory.* Frontiers in Plant Science, 2016. **7**(741).
62. Sonnante, G., et al., *Novel Hydroxycinnamoyl-Coenzyme A Quinate Transferase Genes from Artichoke Are Involved in the Synthesis of Chlorogenic Acid.* Plant Physiology, 2010. **153**(3): p. 1224-1238.
63. Hoffmann, L., et al., *Silencing of Hydroxycinnamoyl-Coenzyme A Shikimate/Quinate Hydroxycinnamoyltransferase Affects Phenylpropanoid Biosynthesis.* The Plant Cell Online, 2004. **16**(6): p. 1446-1465.
64. Chedgy, R.J., T.G. Kollner, and C.P. Constabel, *Functional characterization of two acyltransferases from Populus trichocarpa capable of synthesizing benzyl benzoate and salicyl benzoate, potential intermediates in salicinoid phenolic glycoside biosynthesis.* Phytochemistry, 2015. **113**: p. 149-59.
65. Mahesh, V., et al., *Functional characterization of two p-coumaroyl ester 3'-hydroxylase genes from coffee tree: evidence of a candidate for chlorogenic acid biosynthesis.* Plant Molecular Biology, 2007. **64**(1-2): p. 145-159.
66. Kim, Y.H., et al., *Gene engineering, purification, crystallization and preliminary X-ray diffraction of cytochrome P450 p-coumarate-3-hydroxylase (C3H), the Arabidopsis membrane protein.* Protein Expression and Purification, 2011. **79**(1): p. 149-155.
67. Waldhauser, S.S.M. and T.W. Baumann, *Compartmentation of caffeine and related purine alkaloids depends exclusively on the physical chemistry of their vacuolar complex formation with chlorogenic acids.* Phytochemistry, 1996. **42**(4): p. 985-996.
68. Weng, J.-K. and C. Chapple, *The origin and evolution of lignin biosynthesis.* New Phytologist, 2010. **187**(2): p. 273-285.
69. Davin, L.B. and N.G. Lewis, *An historical perspective on lignan biosynthesis: Monolignol, allylphenol and hydroxycinnamic acid coupling and downstream metabolism.* Phytochemistry Reviews, 2003. **2**(3): p. 257-288.
70. Boerjan, W., J. Ralph, and M. Baucher, *LIGNIN BIOSYNTHESIS.* Annual Review of Plant Biology, 2003. **54**(1): p. 519-546.
71. Smith, A.M., *Plant Biology*. 2010: Garland Science.
72. Alberts, B., et al., *Molecular Biology of the Cell*. 2007: Garland Science.
73. Wolberger, C., *MULTIPROTEIN-DNA COMPLEXES IN TRANSCRIPTIONAL REGULATION.* Annual Review of Biophysics and Biomolecular Structure, 1999. **28**(1): p. 29-56.
74. Stracke, R., M. Werber, and B. Weisshaar, *The R2R3-MYB gene family in Arabidopsis thaliana.* Current Opinion in Plant Biology, 2001. **4**(5): p. 447-456.
75. Mehrtens, F., et al., *The Arabidopsis transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis.* Plant Physiol, 2005. **138**(2): p. 1083-96.

76. Shelton, D., et al., *Transcription Factors of Lotus: Regulation of Isoflavonoid Biosynthesis Requires Coordinated Changes in Transcription Factor Activity.* Plant Physiology, 2012. **159**(2): p. 531-547.
77. Du, H., et al., *The R2R3-MYB Transcription Factor Gene Family in Maize.* PLoS ONE, 2012. **7**(6): p. e37463.
78. Liu, J., A. Osbourn, and P. Ma, *MYB transcription factors as regulators of phenylpropanoid metabolism in plants.* Molecular Plant, 2015. **8**(5): p. 689-708.
79. Zimmermann, I.M., et al., *Comprehensive identification of Arabidopsis thaliana MYB transcription factors interacting with R/B-like BHLH proteins.* The Plant Journal, 2004. **40**(1): p. 22-34.
80. Ogata, K., et al., *Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices.* Cell, 1994. **79**(4): p. 639-648.
81. Schwinn, K., et al., *A Small Family of MYB-Regulatory Genes Controls Floral Pigmentation Intensity and Patterning in the Genus Antirrhinum.* The Plant Cell, 2006. **18**(4): p. 831-851.
82. Jin, H., et al., *Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis.* The EMBO journal, 2000. **19**(22): p. 6150-6161.
83. Preston, J., et al., *AtMYB32 is required for normal pollen development in Arabidopsis thaliana.* The Plant Journal, 2004. **40**(6): p. 979-995.
84. Tamagnone, L., et al., *The AmMYB308 and AmMYB330 transcription factors from antirrhinum regulate phenylpropanoid and lignin biosynthesis in transgenic tobacco.* Plant Cell, 1998. **10**(2): p. 135-54.
85. Aharoni, A., et al., *The strawberry FaMYB1 transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco.* The Plant Journal, 2001. **28**(3): p. 319-332.
86. Dubos, C., et al., *MYB transcription factors in Arabidopsis.* Trends in plant science, 2010. **15**(10): p. 573-581.
87. Ohta, M., et al., *Repression Domains of Class II ERF Transcriptional Repressors Share an Essential Motif for Active Repression.* The Plant Cell, 2001. **13**(8): p. 1959-1968.
88. Kagale, S. and K. Rozwadowski, *EAR motif-mediated transcriptional repression in plants: an underlying mechanism for epigenetic regulation of gene expression.* Epigenetics, 2011. **6**(2): p. 141-6.
89. Albert, N.W., et al., *A Conserved Network of Transcriptional Activators and Repressors Regulates Anthocyanin Pigmentation in Eudicots.* The Plant Cell, 2014. **26**(3): p. 962-980.
90. Luo, J., et al., *AtMYB12 regulates caffeoyl quinic acid and flavonol synthesis in tomato: expression in fruit results in very high levels of both types of polyphenol.* The Plant Journal, 2008. **56**(2): p. 316-326.

91. Mehrtens, F., et al., *The Arabidopsis Transcription Factor MYB12 Is a Flavonol-Specific Regulator of Phenylpropanoid Biosynthesis.* Plant Physiology, 2005. **138**(2): p. 1083-1096.
92. Zhang, Y., et al., *Multi-level engineering facilitates the production of phenylpropanoid compounds in tomato.* Nat Commun, 2015. **6**.
93. Rommens, C.M., et al., *Engineered native pathways for high kaempferol and caffeoylquinate production in potato.* Plant Biotechnology Journal, 2008. **6**(9): p. 870-886.
94. Kneissl, J., et al., *A Rice Phytochrome A in Arabidopsis: The Role of the N-terminus under red and far-red light.* Molecular Plant, 2008. **1**(1): p. 84-102.
95. Martin, C., et al., *How Can Research on Plants Contribute to Promoting Human Health?* The Plant Cell Online, 2011. **23**(5): p. 1685-1699.
96. Martin, C., et al., *Plants, diet, and health.* Annu Rev Plant Biol, 2013. **64**: p. 19-46.
97. Cho, A.-S., et al., *Chlorogenic acid exhibits anti-obesity property and improves lipid metabolism in high-fat diet-induced-obese mice.* Food and Chemical Toxicology, 2010. **48**(3): p. 937-943.
98. Choi, J., et al., *In Vivo Antihepatotoxic Effects of Ligularia fischeri var. spiciformis and the Identification of the Active Component, 3,4-Dicaffeoylquinic Acid.* Journal of Medicinal Food, 2005. **8**(3): p. 348-352.
99. Gu, R., et al., *Simultaneous determination of 1,5-dicaffeoylquinic acid and its active metabolites in human plasma by liquid chromatography–tandem mass spectrometry for pharmacokinetic studies.* Journal of Chromatography B, 2007. **852**(1–2): p. 85-91.
100. Comino, C., et al., *The isolation and mapping of a novel hydroxycinnamoyltransferase in the globe artichoke chlorogenic acid pathway.* BMC Plant Biology, 2009. **9**.
101. Lallemand, L.A., et al., *A Structural Basis for the Biosynthesis of the Major Chlorogenic Acids Found in Coffee.* Plant Physiology, 2012. **160**(1): p. 249-260.
102. Moglia, A., et al., *Dual Catalytic Activity of Hydroxycinnamoyl-Coenzyme A Quinate Transferase from Tomato Allows It to Moonlight in the Synthesis of Both Mono- and Dicaffeoylquinic Acids.* Plant Physiology, 2014. **166**(4): p. 1777-1787.
103. Sander, M. and M. Petersen, *Distinct substrate specificities and unusual substrate flexibilities of two hydroxycinnamoyltransferases, rosmarinic acid synthase and hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl-transferase, from Coleus blumei Benth.* Planta, 2011. **233**(6): p. 1157-71.
104. Besseau, S., et al., *Flavonoid Accumulation in Arabidopsis Repressed in Lignin Synthesis Affects Auxin Transport and Plant Growth.* The Plant Cell, 2007. **19**(1): p. 148-162.
105. Menin, B., et al., *Identification and mapping of genes related to caffeoylquinic acid synthesis in Cynara cardunculus L.* Plant Science, 2010. **179**(4): p. 338-347.

106. Lallemand, L.A., et al., *Purification, crystallization and preliminary X-ray diffraction analysis of a hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase (HCT) from Coffea canephora involved in chlorogenic acid biosynthesis.* Acta Crystallographica Section F: Structural Biology and Crystallization Communications, 2012. **68**(7): p. 824-828.
107. Friedman, M. and H.S. Jürgens, *Effect of pH on the Stability of Plant Phenolic Compounds.* Journal of Agricultural and Food Chemistry, 2000. **48**(6): p. 2101-2110.
108. Winter, D., et al., *An "Electronic Fluorescent Pictograph" Browser for Exploring and Analyzing Large-Scale Biological Data Sets.* PLoS ONE, 2007. **2**(8): p. e718.
109. Joët, T., et al., *Metabolic pathways in tropical dicotyledonous albuminous seeds: Coffea arabica as a case study.* The New Phytologist, 2009. **182**(1): p. 146-162.
110. Jiang, D. and D.G. Peterson, *Role of hydroxycinnamic acids in food flavor: A brief overview.* Phytochemistry Reviews, 2010. **9**(1): p. 187-193.
111. Clé, C., et al., *Modulation of chlorogenic acid biosynthesis in Solanum lycopersicum; consequences for phenolic accumulation and UV-tolerance.* Phytochemistry, 2008. **69**(11): p. 2149-2156.
112. Leiss, K.A., et al., *Identification of Chlorogenic Acid as a Resistance Factor for Thrips in Chrysanthemum.* Plant Physiology, 2009. **150**(3): p. 1567-1575.
113. Mondolot, L., et al., *Evolution in caffeoylquinic acid content and histolocalization during Coffea canephora leaf development.* Annals of Botany, 2006. **98**(1): p. 33-40.
114. Conéjéro, G., et al., *Spectral analysis combined with advanced linear unmixing allows for histolocalization of phenolics in leaves of coffee trees.* Frontiers in Plant Science, 2014. **5**: p. 39.
115. Garrett, R., C.M. Rezende, and D.R. Ifa, *Revealing the spatial distribution of chlorogenic acids and sucrose across coffee bean endosperm by desorption electrospray ionization-mass spectrometry imaging.* LWT - Food Science and Technology, 2016. **65**: p. 711-717.
116. Bodzon-Kulakowska, A. and P. Suder, *Imaging mass spectrometry: Instrumentation, applications, and combination with other visualization techniques.* Mass Spectrometry Reviews, 2016. **35**(1): p. 147-169.
117. Koshiro, Y., et al., *Biosynthesis of chlorogenic acids in growing and ripening fruits of Coffea arabica and Coffea canephora plants.* Z Naturforsch C, 2007. **62**(9-10): p. 731-42.
118. Hölscher, D., et al., *Matrix-free UV-laser desorption/ionization (LDI) mass spectrometric imaging at the single-cell level: distribution of secondary metabolites of Arabidopsis thaliana and Hypericum species.* The Plant Journal, 2009. **60**(5): p. 907-918.
119. Mahesh, V., et al., *Isolation and genetic mapping of a Coffea canephora phenylalanine ammonia-lyase gene (CcPAL1) and its involvement in the*

accumulation of caffeoyl quinic acids. Plant Cell Reports, 2006. **25**(9): p. 986-992.
120. De Castro, R.D. and P. Marraccini, *Cytology, biochemistry and molecular changes during coffee fruit development.* Brazilian Journal of Plant Physiology, 2006. **18**: p. 175-199.
121. Sondheimer, E., F. Covitz, and M.J. Marquisee, *Association of naturally occurring compounds, the chlorogenic acid-caffeine complex.* Archives of Biochemistry and Biophysics, 1961. **93**(1): p. 63-71.
122. Chappell, J. and K. Hahlbrock, *Transcription of plant defence genes in response to UV light or fungal elicitor.* Nature, 1984. **311**(5981): p. 76-78.
123. Paz-Ares, J., et al., *The regulatory c1 locus of Zea mays encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators.* The EMBO Journal, 1987. **6**(12): p. 3553-3558.
124. Spitzer-Rimon, B., et al., *EOBII, a Gene Encoding a Flower-Specific Regulator of Phenylpropanoid Volatiles' Biosynthesis in Petunia.* The Plant Cell, 2010. **22**(6): p. 1961-1976.
125. Medina-Puche, L., et al., *An R2R3-MYB Transcription Factor Regulates Eugenol Production in Ripe Strawberry Fruit Receptacles.* Plant Physiology, 2015. **168**(2): p. 598-614.
126. Kranz, H.D., et al., *Towards functional characterisation of the members of the R2R3-MYB gene family from Arabidopsis thaliana.* The Plant Journal, 1998. **16**(2): p. 263-276.
127. Verdonk, J.C., et al., *ODORANT1 Regulates Fragrance Biosynthesis in Petunia Flowers.* The Plant Cell, 2005. **17**(5): p. 1612-1624.
128. Martin, C. and J. Paz-Ares, *MYB transcription factors in plants.* Trends Genet, 1997. **13**(2): p. 67-73.
129. Romero, et al., *More than 80R2R3-MYB regulatory genes in the genome of Arabidopsis thaliana.* The Plant Journal, 1998. **14**(3): p. 273-284.
130. Ogata, K., et al., *Solution structure of a DNA-binding unit of Myb: a helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core.* Proceedings of the National Academy of Sciences, 1992. **89**(14): p. 6428-6432.
131. Bailey, P.C., et al., *IT3F: A web-based tool for functional analysis of transcription factors in plants.* Phytochemistry, 2008. **69**(13): p. 2417-2425.
132. Legay, S., et al., *Molecular characterization of EgMYB1, a putative transcriptional repressor of the lignin biosynthetic pathway.* Plant Science, 2007. **173**(5): p. 542-549.
133. Xu, W., C. Dubos, and L. Lepiniec, *Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes.* Trends Plant Sci, 2015. **20**(3): p. 176-85.
134. Mubarak, A., et al., *Acute Effects of Chlorogenic Acid on Nitric Oxide Status, Endothelial Function, and Blood Pressure in Healthy Volunteers: A*

Randomized Trial. Journal of Agricultural and Food Chemistry, 2012. **60**(36): p. 9130-9136.
135. Czemmel, S., et al., *The Grapevine R2R3-MYB Transcription Factor VvMYBF1 Regulates Flavonol Synthesis in Developing Grape Berries.* Plant Physiology, 2009. **151**(3): p. 1513-1530.
136. Liu, Y. and M.G. Nair, *An Efficient and Economical MTT Assay for Determining the Antioxidant Activity of Plant Natural Product Extracts and Pure Compounds.* Journal of Natural Products, 2010. **73**(7): p. 1193-1195.
137. Jin, U.-H., et al., *A phenolic compound, 5-caffeoylquinic acid (chlorogenic acid), is a new type and strong matrix metalloproteinase-9 inhibitor: Isolation and identification from methanol extract of Euonymus alatus.* Life Sciences, 2005. **77**(22): p. 2760-2769.
138. M. Lepelley, C.P., T. Vinos Poyo, and J. McCarthy, *Identification of C. canephora R2R3-type MYB like gen candidates potentially involved in the regulation of the levels of phenylpropanoid/CGA compounds in coffee*, in *R&D Report*. 2011, R&D Centre, Tours, France R&D Centre, Tours, France p. 1-96.
139. Czemmel, S., et al., *The grapevine R2R3-MYB transcription factor VvMYBF1 regulates flavonol synthesis in developing grape berries.* Plant Physiol, 2009. **151**(3): p. 1513-30.
140. Comino, C., et al., *The isolation and mapping of a novel hydroxycinnamoyltransferase in the globe artichoke chlorogenic acid pathway.* BMC Plant Biology, 2009. **9**(1): p. 30.
141. Stracke, R., et al., *Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the Arabidopsis thaliana seedling.* Plant J, 2007. **50**(4): p. 660-77.
142. Yuan, Y., et al., *The Scutellaria baicalensis R2R3-MYB Transcription Factors Modulates Flavonoid Biosynthesis by Regulating GA Metabolism in Transgenic Tobacco Plants.* PLoS ONE, 2013. **8**(10): p. e77275.
143. Nakatsuka, T., et al., *Isolation and characterization of GtMYBP3 and GtMYBP4, orthologues of R2R3-MYB transcription factors that regulate early flavonoid biosynthesis, in gentian flowers.* Journal of Experimental Botany, 2012.
144. Huang, W., et al., *A R2R3-MYB Transcription Factor Regulates the Flavonol Biosynthetic Pathway in a Traditional Chinese Medicinal Plant, Epimedium sagittatum.* Frontiers in Plant Science, 2016. **7**(1089).
145. Yoshida, K., et al., *Functional differentiation of Lotus japonicus TT2s, R2R3-MYB transcription factors comprising a multigene family.* Plant Cell Physiol, 2008. **49**(2): p. 157-69.
146. Feller, A., et al., *Evolutionary and comparative analysis of MYB and bHLH plant transcription factors.* The Plant Journal, 2011. **66**(1): p. 94-116.
147. Kui, L.W., et al., *An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae.* Bmc Plant Biology, 2010. **10**: p. 17.

148. Kelemen, Z., et al., *Analysis of the DNA-Binding Activities of the Arabidopsis R2R3-MYB Transcription Factor Family by One-Hybrid Experiments in Yeast.* PLoS ONE, 2015. **10**(10): p. e0141044.
149. Bedon, F., et al., *Subgroup 4 R2R3-MYBs in conifer trees: gene family expansion and contribution to the isoprenoid- and flavonoid-oriented responses.* Journal of Experimental Botany, 2010. **61**(14): p. 3847-3864.
150. Docimo, T., et al., *Phenylpropanoids accumulation in eggplant fruit: Characterization of biosynthetic genes and regulation by a MYB transcription factor.* Frontiers in Plant Science, 2016. **6**(JAN2016).
151. Comino, C., et al., *Isolation and functional characterization of a cDNA coding a hydroxycinnamoyltransferase involved in phenylpropanoid biosynthesis in Cynara cardunculus L.* BMC Plant Biology, 2007. **7**(1): p. 1-14.
152. Walker, A.M., et al., *Elucidation of the Structure and Reaction Mechanism of Sorghum Hydroxycinnamoyltransferase and Its Structural Relationship to Other Coenzyme A-Dependent Transferases and Synthases.* Plant Physiology, 2013. **162**(2): p. 640-651.
153. Butelli, E., et al., *Retrotransposons Control Fruit-Specific, Cold-Dependent Accumulation of Anthocyanins in Blood Oranges.* The Plant Cell Online, 2012. **24**(3): p. 1242-1255.
154. Kajita, S., Y. Katayama, and S. Omori, *Alterations in the Biosynthesis of Lignin in Transgenic Plants with Chimeric Genes for 4-Coumarate: Coenzyme A Ligase.* Plant and Cell Physiology, 1996. **37**(7): p. 957-965.
155. Wagner, A., et al., *Suppression of 4-Coumarate-CoA Ligase in the Coniferous Gymnosperm Pinus radiata.* Plant Physiology, 2009. **149**(1): p. 370-383.
156. Chitwood, D.H., et al., *A Quantitative Genetic Basis for Leaf Morphology in a Set of Precisely Defined Tomato Introgression Lines.* The Plant Cell, 2013. **25**(7): p. 2465-2481.
157. Ferrer, J.-L., et al., *Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis.* Nat Struct Mol Biol, 1999. **6**(8): p. 775-784.
158. Austin, M.B. and J.P. Noel, *The chalcone synthase superfamily of type III polyketide synthases.* Natural Product Reports, 2003. **20**(1): p. 79-110.
159. Domingues Junior, A.P., et al., *Looking for the physiological role of anthocyanins in the leaves of Coffea arabica.* Photochem Photobiol, 2012. **88**(4): p. 928-37.
160. Murthy, P.S., et al., *Extraction, characterization and bioactivity of coffee anthocyanins.* European Journal of Biological Sciences, 2012. **4**(1): p. 13-19.
161. Patay, É.B., et al., *Histological and phytochemical studies of Coffea benghalensis B. Heyne ex Schult., compared with Coffea arabica L.* Farmacia, 2016. **64**(1): p. 125-130.
162. Farah, A. and C.M. Donangelo, *Phenolic compounds in coffee.* Brazilian Journal of Plant Physiology, 2006. **18**: p. 23-36.

163. Chen, S., et al., *Identification and Characterization of Chalcone Synthase Gene Family Members in Nicotiana tabacum.* Journal of Plant Growth Regulation, 2016: p. 1-11.

Appendix 1: Oligonucleotides

sgRNA for HQT KO plants

guide 1A CCTTATCTCTTTAGCTCTTTCTC
guide 1B GAAGTTAATTGTAATGGTGAAGG

Oligo Name	Purpose	Sequence
11ccMYB1 3’ RACE 1	3’RACE PCR	TATAGAGAGAAAAGACAAGGGCTACTG
12ccMYB1 3’ RACE 2	3’RACE PCR	AGTTTAAGCTGACAACAGATCACTCAT
13ccMYB2 3' RACE 1	3’RACE PCR	CTATTAGTATGCGTAACACTGACGGACTA
14ccMYB2 3' RACE 2	3’RACE PCR	GTTAATTACAGCTCTACTGATGCAGACTT
23 3' RACE AP	3’RACE anchor	GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTT
24 3' RACE SEQ	3’RACE 2nd step	GGCCACGCGTCGACTAGTAC
25 caMYB12 F1	3’RACE PCR	TTAATAATGGGAAGAGCACCTTGCTGTG
26 caMYB12 F2	3’RACE PCR	ATGGGAAGAGCACCTTGCTGTGAG
ROCHE PCR anchor	3’RACE 2nd step	GACCACGCGTATCGATGTCGAC
28 caMYB12 F3	3’RACE PCR	CTGAGGTGTGGGAAGAGTTGCAG
29 caMYB12 F4	3’RACE PCR	AAACAAAATGGGGAAGGGTCCTG
30 caMYB12 F5	3’RACE PCR	GCTGTGAGAAGGTAGGGTTGAGGA
31 caMYB12 R1	3’RACE PCR	GGATCGATGCAGGTTGATGATTG
CaMYB12attB1	Gateway™ recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGG GAAGAGCACCTTGCTGTG
CaMYB12attB2	Gateway™ recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTATTAAC TCTCCCATGGCCATAGT
CcMYB1attB1	Gateway™ recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGG GACGTTCACCTTGCTGTG
CcMYB1attB2	Gateway™ recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTACTAAC CCTGAAATCTATCTTCG
AtMYB12 CDS f	Partial CDS	GGTCCGCTATGAAACCAAAA

AtMYB12 CDS r	Partial CDS	TGTTATGCAACTCCCCATCA
CcMYB1 CDS-F	Partial CDS	CAGACTTGAAGAGAGGAAACTTCAC
CcMYB1 CDS-R	Partial CDS	GTAATAGGAGGAGCTTTTGTGATTG
CaMYB12 CDS-F	Partial CDS	CTGATCTCAAGAGAGGAAACTTCAC
CaMYB12 CDS-R	Partial CDS	ACCGCTACTACATTTTGTATCTTCG
CcMYB2attB1	Gateway^{tm} recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGG AGAAGTATGGTGAAGGAA
CcMYB2attB2	Gateway^{tm} recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTATTGCA TGGTTTGGTGATCGGAA
CcMYB12attB1	Gateway^{tm} recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGG GAAGAGCACCTTGCTGTG
CcMYB12attB2	Gateway^{tm} recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTATGCAT CCAACGCGTTGGGAGCT
CcMYB12f1	Full length CDS	GGAAATTAATAATGGGAAGAGCACCT
CcMYB12r1	Full length CDS	AACTACTGCAACAGGTCTTTGTTTTT
CcMYB12f2	Full length CDS	ATGGGAAGAGCACCTTGCTGTG
CcMYB12r2	Full length CDS	AACTACTGCAACAGGTCTTTGTTTTTG
pBIN19 fwd	Plasmid insert PCR	aaaaggaaggtggctcctacaaat
pBIN19 rev	Plasmid insert PCR	GAGAGACTGGTGATTTCAGCGAAT
pDEST17 fwd	Plasmid insert PCR	ccgcgaaattaatacgactcacta
PDEST17 rev	Plasmid insert PCR	ggatatagttcctcctttcagcaa
pDONOR207 fwd	Plasmid insert PCR	gttttatttgatgcctggcagttc
pDONOR207 rev	Plasmid insert PCR	gaatatggctcataacacccctttg
CcHCTFWD	Full length CDS	ATGAAAATCGAGGTGAAGGA

CcHCT REV	Full length CDS	AATGTCATACAAGAAACTCTGG
CcHQT fwd	Full length CDS	GTGGGAAGCGAGAAAATGAA
CcHQT REV	Full length CDS	AATTTCTCAGAAATCGTACAGG
CcHCT att R2 stop	Gateway^{tm} recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAAA TGTCATACAAGAA
CcHQT att R2 stop	Gateway^{tm} recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTACTAGT GATTCAGGAACTT
NtPAL fwd CDS	Partial CDS (qRTPCR)	TGCGAAAAGGACTTGCTTCG
NtPAL rev CDS	Partial CDS (qRTPCR)	TGTTGGCAATTGCAGGGTTC
NtCH4 fwd cds	Partial CDS (qRTPCR)	AGAAAGAAACTTGCAAACACGAC
NtCH4 rev CDS	Partial CDS (qRTPCR)	TCGCATCGTGAAGGTTCATG
Nt4CL fwd CDS	Partial CDS (qRTPCR)	ATGTGTTTTGTGTGTGTTGCC
Nt4CL rev CDS	Partial CDS (qRTPCR)	CAGCGTGAGGAAAGTTCTTAGC
NtCHS fwd CDS	Partial CDS (qRTPCR)	GGAAAGCCTCTGCAAAAGAAGG
NtCHS rev CDS	Partial CDS (qRTPCR)	GCAACACTGTGGAGAACAACAG
NtCHI fwd CDS	Partial CDS (qRTPCR)	ATCGTCACAGGTCCCTTTGAG
NtCHI rev CDS	Partial CDS (qRTPCR)	AATCGTCAATGCCCCAACAG
NtF3H fwd CDS	Partial CDS (qRTPCR)	AGCCATCTACAGGGTGAAGTG
NtF3H fwd CDS	Partial CDS (qRTPCR)	TGTGTCGTTTCAGTCCAAGG
NtF3'H fwd CDs	Partial CDS (qRTPCR)	TCCGCCAGGATGAAGTAAGAAC
NtF3'H FWD CDS	Partial CDS (qRTPCR)	TGGATCAACACCGCCATTTG
NtF3'5'H FWD CDS	Partial CDS (qRTPCR)	ATGTTGACATTCGCCATGGC
NtF3'5'H rev CDS	Partial CDS (qRTPCR)	AGTTGTACTGAGCCTTTCGC

NtDFR fwd CDS	Partial CDS (qRTPCR)	ATGGCAAGTGAAGCTCATGC
NtDFR rev CDS	Partial CDS (qRTPCR)	GTTGTTCTCAGGATCACGAACAG
NtANS fwd CDS	Partial CDS (qRTPCR)	TCGCGAGAAATGCCACAAAG
NtAND rev CDS	Partial CDS (qRTPCR)	TTCAAGCTGACCACAAGCAC
NtGT fwd CDS	Partial CDS (qRTPCR)	ATGGCCACTATTTGCTGAGC
Nt GT rev CDS	Partial CDS (qRTPCR)	AACCACCTTCTCCAAATGCC
NtRT fwd CDS	Partial CDS (qRTPCR)	AATGTTCAACGGACCAAGCC
NtRT revCDS	Partial CDS (qRTPCR)	TTCACGTTGCATGGCTAACC
NtC3H fwd CDS	Partial CDS (qRTPCR)	ATTGGCCTCCTCTGGGATATG
NtC3H rev CDS	Partial CDS (qRTPCR)	ACATGATTCGATCCGTTCCG
Nt HCT fwd CDS	Partial CDS (qRTPCR)	TTGTTCGTGGTGCACATACG
NtHCT rev CDS	Partial CDS (qRTPCR)	AACCCCAAAAGAAGCAGCAC
NtHQT fwd CDS	Partial CDS (qRTPCR)	CCCCTCATCATCTTTCAGGTTAC
NtHQT rev CDS	Partial CDS (qRTPCR)	AAGCCCAAGTTGGTCACTAGAG
CaMYB12 fwd CDS	Partial CDS (qRTPCR)	
CaMYB12 rev CDS	Partial CDS (qRTPCR)	AGCAGCTGCAGCAAGAAAAG
CcMYB1 fwd CDS	Partial CDS (qRTPCR)	CCAATACTGACGCCTCGAAAAC
CcMYB1 rev CDS	Partial CDS (qRTPCR)	ACCTTGTGTCAGCTGTTGTG
CcMYB2 fwd CDS	Partial CDS (qRTPCR)	gggcaacgtaactagtttgtgg
CcMYB2 rev CDS	Partial CDS (qRTPCR)	ttgcatggtttggtgatcgg
NtUBI fwd CDS	Partial CDS (qRTPCR)	ATTCTCGCCGCTTTTTGCTC
Nt Ubi rev CDs	Partial CDS (qRTPCR)	TCATCAGGGCCGAAAATGAC
Nt HQT CDS fwd 2	Partial CDS (qRTPCR)	GCCCAAAGCCTAGTACCACA

Nt HQT CDS rev 2	Partial CDS (qRTPCR)	GTACGTCGGTCCACGGATTA
CcMYB2 att f2	Gateway^{tm} recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGG AGAAGTATGGTGAA
CcMYB2 att r2 (stop)	Gateway^{tm} recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTACTATT GCATGGTTTGGTG
CcHQT att F2	Gateway^{tm} recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGA AGATAACCGTGAAG
CcHQT att R2	Gateway^{tm} recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTAGAAAT CGTACAGGAACTT
CcHQT att R2 Stop	Gateway^{tm} recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAGA AATCGTACAGGAA
GFP att f1	Gateway^{tm} recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGG TGAGCAAGGGCGAG
GFP att r1	Gateway^{tm} recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTATTACT TGTACAGCTCGTCCATGCC
NtHQT att F1	Gateway^{tm} recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGG GAAGTGAAAAAATG
NtHQT att r1	Gateway^{tm} recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAAA ATTCATACAAATA
CcMYB12b att F1	Gateway^{tm} recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGG GAAGAGCCCCTTGCTGTGAG
CcMYB12 b att R1	Gateway^{tm} recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAAG AGAGAAGCCAGGC
CcMYB12 b mid f1	Partial CDS	TGGAAGAGGGAGGTACGAGTAA
CcMYB12 b mid r1	Partial CDS	CTTTTGAAACGTCCATTTTTCC
CcMYB12b F1	Full length CDS	ATGGGAAGAGCCCCTTGCT

CcMYB12b R1	Full length CDS	TCAAGAGAGAAGCCAGGCAAC
CcPalUnk pro fwd 1	Promoter PCR	GCCAAATCATATTTGCTAGATGC
CcPalUnk pro rev 1	Promoter PCR	CTTTACCACCGGCTTCCTAAACT
CcPAL2pro fwd 1	Promoter PCR	CGCAAGAAATTTACCCTCCTACT
CcPAL2pro rev 1	Promoter PCR	TGACTTCATCAAGATGACTTCCTT
CcPAL3pro fwd 1	Promoter PCR	ATTTGTTTCGTTACGTTGGATTG
CcPAL3pro rev 1	Promoter PCR	ATATCCAGATTTTTGACGGAAGG
CcC4HCh1 pro fwd 1	Promoter PCR	GCACAACCCTTTACTTTACAACG
CcC4HCh1 pro rev 1	Promoter PCR	CAGATCTTCAAGTAACCCCTCAA
CcC4HCh0-1 pro fwd 1	Promoter PCR	TAGGATTAACCTGCTTTGGTTCA
CcC4HCh0-1 pro rev 1	Promoter PCR	GTGGTTCAAATCATCACCAACTT
CcC4HCh0-2pro fwd 1	Promoter PCR	TGAACTCTTGAATGTCAACCAGA
CcC4HCh0-2pro rev 1	Promoter PCR	TTGCTACAAGACTTCCATATCAGG
CcC4HCh0-3 pro fwd 1	Promoter PCR	GTTGATGCAAATCCAAATTCCTA
CcC4HCh0-3 pro rev 1	Promoter PCR	CCAAAACAAGTAATGGGATGGTA
Cc4CLpro fwd 1	Promoter PCR	TGGTGAGGTGGTAGATTAAATGG
Cc4CLpro rev 1	Promoter PCR	ATGATTCACCGATAACAGCATCT
CcCHSCh3 pro fwd 1	Promoter PCR	TTCGAAATAGGTTTCAAAGGTCA
CcCHSCh3 pro rev 1	Promoter PCR	TGCTCACTATCAGTAATGCGAAA

CcCHS Ch0-1 pro fwd 1	Promoter PCR	GCTAGCCCTTCTTGCTCTTATTC
CcCHS Ch0-1 pro rev 1	Promoter PCR	CTAGATGGGTGATCTTGGACTTG
CcCHS Ch0-2 pro fwd 1	Promoter PCR	CTGTATGGGTAGAGAAACGAACG
CcCHS Ch0-2 pro rev 1	Promoter PCR	CTAGATGGGTGATCTTGGACTTG
CcCHI Ch2 pro fwd 1	Promoter PCR	AGTTCTAATTGCTTCCCTTTTGG
CcCHI Ch2 pro rev 1	Promoter PCR	CATTCCAAGGACATTGAAAAGAG
CcCHI Ch3 pro fwd 1	Promoter PCR	AAAGAAATGGGCTTCAAAAAGTC
CcCHI Ch3 pro rev 1	Promoter PCR	AATTGCAGTCTCCTCCAAATACA
CcF3H pro fwd 1	Promoter PCR	GGGCTCAACCTTTCTATTTCACT
CcF3H pro rev 1	Promoter PCR	ATCAACCCCATGATCAATAACCT
CcHCT pro fwd 1	Promoter PCR	ATGAAATTGGATTTTGACCCTCT
CcHCT pro rev 1	Promoter PCR	TAGCACTTTGGCATCAAAGAAAT
CcHQT pro fwd 1	Promoter PCR	AATTTATTGAGCGTGAAAGGTGA
CCHQT pro rev 1	Promoter PCR	TTCTTGCTACCAGAAGATCCAAA
165 CcPalUnk pro ATT f1	Promoter PCR	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAAATC CATTGAATTTTCTTCCTTTG
166 CcPalUnk pro ATT r1	Promoter PCR	GGGGACCACTTTGTACAAGAAAGCTGGGTACTTGC AGCTTTCTCTGTTCAGGA
167 CcPAL2proATTf1	Gateway^{tm} promoter recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTATGTT AATGCAAGTTAGCT

168 CcPAL2proATTr1	Gateway^{tm} promoter recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTATTTTG CTTGCTTGCTGGC
169 CcPAL3proATTf1	Gateway^{tm} promoter recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAAGGA GATATTTTATTTTA
170 CcPAL3proATTr1	Gateway^{tm} promoter recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTAGCTCT CCACCACCCACTCACTGCTG
171 CcC4HCh1proATTf1	Gateway^{tm} promoter recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTATGAT GAGCCATGAGGATTA
172 CcC4HCh1proATTr1	Gateway^{tm} promoter recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTAATTAT TGTACATCATGAGCTG
173 CcC4HCh0-1proATTf1	Gateway^{tm} promoter recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTATGAG TTTTTTAAATATTTAT
174 CcC4HCh0-1proATTr1	Gateway^{tm} promoter recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTAGACGG TGTGGAGGGCGCGT
177 CcC4HCh0-3 pro ATT f1	Gateway^{tm} promoter recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTACATG TGGCAGCATACAAA
178 CcC4HCh0-3 pro ATT r1	Gateway^{tm} promoter recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTAGTTGT AAGAGGGATAGGA
179 Cc4CLpro ATT f1	Gateway^{tm} promoter recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTATGCA AAAGATGTGAAAGT
180 Cc4CLpro ATT r1	Gateway^{tm} promoter recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTACAAGC TGGAAGCTTTTTC
181 CcCHSCh3 pro ATT f1	Gateway^{tm} promoter recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAACAT AACTACAGAGTAAT
182 CcCHSCh3 pro ATT r1	Gateway^{tm} promoter recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTATTTTG CTCGGAAAAAGAG

183 CcCHS Ch0-1 pro ATT f1	Gateway[tm] promoter recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAAACT TCTACATTTTAAGG
184 CcCHS Ch0-1 pro ATT r1	Gateway[tm] promoter recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTAGGATT TGTCACCTGAAAT
185 CcCHS Ch0-2 pro ATTf1	Gateway[tm] promoter recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGAAT CCGATCAACTGATT
186 CcCHS Ch0-2 pro ATTr1	Gateway[tm] promoter recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTAGGATT TGTCACCTGAAAT
187 CcCHI Ch2 pro ATTf1	Gateway[tm] promoter recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGAGA AGCACATTATTGTT
188 CcCHI Ch2 pro ATTr1	Gateway[tm] promoter recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTAATTGT AACTAGCACTCAG
189 CcCHI Ch3 pro f1	Gateway[tm] promoter recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAAGG GCTCAGCCGAACTGC
190 CcCHI Ch3 pro r1	Gateway[tm] promoter recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTATAGTA TCATACAACCTCA
191 CcF3H pro ATT f1	Gateway[tm] promoter recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTATGCC TTCAATATATAGTA
192 CcF3H pro ATT r1	Gateway[tm] promoter recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTACTGTC TCACTGTTTCAGT
193 CcHCT pro ATT f1	Gateway[tm] promoter recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTATGGT CAGACTATTAGTAG
194 CcHCT pro ATT r1	Gateway[tm] promoter recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTAGGTTG AAAGTTGTATGAT
195 CcHQT pro ATT f1	Gateway[tm] promoter recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAAGTA CAGGAGTGGCCGGG

196 CcHQT pro ATT r1	Gateway^{tm} promoter recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTATTTCTC GCTTCCCACGCTGATCAA
197 CcPalUnk pro ATT f2	Gateway^{tm} promoter recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTACCAT TACACCCCTCCATG
198 CcPalUnk pro ATT r2	Gateway^{tm} promoter recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTACTTGC AGCTTTCTCTGTTCAGGA
199 CcPAL2 pro ATT f2	Gateway^{tm} promoter recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAACAT GGTGCTTTCGTGCA
200 CcPAL2 pro ATT r2	Gateway^{tm} promoter recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTATTTTG CTTGCTTGCTGGCTGGC
201 Cc4CL pro ATT f2	Gateway^{tm} promoter recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAAGAA AAGAGGAAAAAGCTTCCAGCT
202 Cc4CL pro ATT r2	Gateway^{tm} promoter recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTACAATT CCCTCCCCAGTGG
203 CcCHS Ch0-2 ATT f2	Gateway^{tm} promoter recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTACAAG AATCAAGTAATACT
204 CcCHS Ch0-2 ATT r2	Gateway^{tm} promoter recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTAGGATT TGTCACCTGAAAT
208 CcMYB12b mid f1	Partial CDS	acgtttcaaaaggaaggaatca
209 CcMYB12b mid r1	Partial CDS	tgttgttgtcagaatcccagag
210 CcMYB12 mid f1	Partial CDS	ggagagcggaaaatatcagttg
211 CcMYB12 mid r1	Partial CDS	ttttctccttccagcttttctg
214 CcHQT R Mid	Partial CDS	CCACGAGCTATGTCTGACCA
215 CcHCT R Mid	Partial CDS	TGCAGCACTAGGAGAGCGTA

216 GFP r mid	Partial CDS	AAGTCGTGCTGCTTCATGTG
217 SlHCT r mid	Partial CDS	CGGACCATGTGTTGATGAAG
218 SlHQT r mid	Partial CDS	CCAGCCATTGGATAAAATGAA
219 Cc4CLpro fwd 2	Promoter PCR	GAAGGGTCCTCTTAGTCAAGCAT
220 Cc4CLpro rev 2	Promoter PCR	CAGACAATGATTCACCGATAACA
221 CcFLS pro F1	Promoter PCR	ACCGCTCGAGATTGAATAGAAG
222 CcFLS pro R1	Promoter PCR	TGTTACTGCAGGTTGCTCATTT
223 CcFLS pro att f1	Gateway^{tm} promoter recombination	GGGGACAAGTTTGTACAAAAAAGCAGGCTTATGTT GTATTTGTGTATTAC
224 CcFLS pro att r1	Gateway^{tm} promoter recombination	GGGGACCACTTTGTACAAGAAAGCTGGGTATCCTT CCCTCACTCTCCA
245 CcRPL39 q F1	qRT-PCR	AAGAAGAAGCTGGCGAAGAA
246 CcRPL39 q R1	qRT-PCR	CATTGTAGCGGATGGTGTTG
247 CcRPL39 q F2	qRT-PCR	CGAGTACTACCAACCCATCCTC
248 CcRPL39 q R2	qRT-PCR	TCTTCGCCAGCTTCTTCTTT
249 CcRPL39 q F3	qRT-PCR	CTTACAGCAACCGAAGATGC
250 CcRPL39 q R3	qRT-PCR	TTGTAGCGGATGGTGTTGTC
251 CcMYB12b q F1	qRT-PCR	ggaggcggtgaagtatcaaa
252 CcMYB12b q R1	qRT-PCR	aattcttttggttgcgatgg
253 CcMYB12b q F2	qRT-PCR	cgggatgttgtgtcttgatg
254 CcMYB12b q R2	qRT-PCR	gggtccatattatcgcatcg
255 CcMYB12b q F3	qRT-PCR	atcagatcggctgcacttct
256 CcMYB12b q R3	qRT-PCR	agagccacgacagcaagtct

257	CcMYB12 q F1	qRT-PCR	ggacctgagcgagatttcag
258	CcMYB12 q R1	qRT-PCR	aacagtcgtacaggccatcc
259	CcMYB12 q F2	qRT-PCR	ggaaggagaaaagggagtcg
260	CcMYB12 q R2	qRT-PCR	caggtccatcaattcgctct
261	CcMYB12 q F3	qRT-PCR	cacatggggtacctgagtcc
262	CcMYB12 q R3	qRT-PCR	attttccgctctcctcaacc

Appendix 2: Cloned R2R3MYB sequences

Robusta CcMYB1

TGGGGGTACGACGGGCAGTGATTGTAATACGACTCACTATAGGGCGAATTGGGCCCGACGTCGCATGCTC
CCGGCCGCCATGGCGGCCGCGGGAATTCGATTGGGCTACTGCTTACTAGCACTTATCCATTTTTTCTCTTTG
ATGTTTTGAGCGCTGCCTTTCTTCTTCTTCTTCTTCTTTTTCTTCTAGGAAAGCAAGAAAGAAAGAAAGAAAG
AGAGAGAGAGAGGTTAATATTCCATCATTCATTCATGGGACGTTCACCTTGCTGTGAGAAAGCTCACACCA
ACAAAGGAGCATGGACTAAAGAGGAAGACCAGCGTCTCATCAACTATATTCGTGTTCATGGTGAAGGCTG
CTGGAGATCCCTCCCTAAAGCTGCAGGGTTGCTTAGATGTGGCAAGAGTTGTAGGCTAAGATGGATAAATT
ATCTTAGGCCTGATCTCAAGAGAGGAAACTTCACTGAAGAAGAAGATGAGATAATCATTAAGCTTCATAAT
CTTCTTGGAAACAAATGGTCTCTAATAGCAGGAAGATTGCCCGGACGAACAGACAATGAGATCAAGAATT
ACTGGAACACACACATTAAACGCAAGCTCATTAGCCGTGGCATTGACCCCCAAACGCACCGCCCTCTTAAT
GGTGGCACGGCCACCACGACAACTACAGTCACGCCCACGACCACCAATACTGACGCCTCGAAAACAGCCA
AAAACATTTGCCTAGACTTCAGAAGTTCAGCATTGCCATTGGATAACAAAAATGGTTTCACGAATGATCATG
GAAGCAGCAAGTTTAAGCTGACAACAGATCACTCATCAGCCGAAGATACAAAATGTAGTAGCGGTACAAC
AGAGGAGAGTCAGGCACCACAGGGACCACTTCTACTGGAAGATCAAGCCACCACCCCATCAGTGATGCTT
GATCTTGAGCTATCTATTGGACTTCCACAGCCCAAATCTAGTATTAGCTCCTTTTCAAGTTCTGCTGAATCTA
AGGTTTCTCAAGGTTTCTGGGCCGCGGCAGCACCACCACAACAGCTGACACAAGGTGCCGCCGCCGCCGC
CGCCGCCGCAACAGCGACGGGAAAGCCAGTTTGTTTGTGCTGGCAACTGGGGTATAAGAGTGGTCAGTTA
TGTAGAAATTGTGATATGGCATTACAGATATCATTGAAGATCGAAGATAGATTTCAGGGTTAGATAATACT
AATAAGTTTTTTTTTTGTTTTTTAAATTCCTGAAACAAATTTTCCTTTTGGGTATTGGTACTAAGTAGTTAGGA
GTTAGTACCTTTCTCCAACCGAATGTAATCAGTCCCAGTGGCGACAAATGTAATATTTGCCAAACCCAAATT
TACCAAGATTTAAAAAAAAAAAAAAA

Robusta CcMYB2

ATGGAGAAGTATGGTGAAGGAAAGTGGCATCAAGTTCCAGTTCGGGCAGGGTTAAATAGGTGCCGCAAA
AGTTGTAGACTACGGTGGCTCAACTATCTCAGGCCAGATATAAAGCGAGGAGACTTCGCTTCTGATGAAGT
CGATCTTCTCATCAGGCTTCAAAAGCTGTTGGGCAACAGATGGTCACTAATTGCCGGTAGACTTCCTGGAA
GAACTGCAAATGATGTGAAAAATTACTGGAACACCCATTTAATAAAGAAGAAGATGGCTTTCAGAAGACA
ACCAGATGAAAAGGGCCAGGCCGGTAACACAAACTACACCACAAAACCTAGGGCAAGCAACATAGTGAA
GCCTCGTCCTTGGACCTTCTCCAAAAAGGCATCTGGGTTGAGGGGAAAAGCAACAACAACAACGACAGCA
ACAAGTACAGAAATTCCTGCTATTCGGACAACAACGAATGCTCAGGGTCTCAAGGATTCACCTCCGACAAT
AGAGAAGGATGATGACAACAACATGCAGTGGTGGGACAACTTGTTTCTTGACGAAGAGGGAGGTGAAGG
AACCCCGTTGGGATTTGATGCCTCGGGGGAGGGCAACGTAACTAGTTTGTGGTTCGAAGAATTGTCCCAG
ATGGTATTTAAGGAGGGTGATATTTGGACGCAGGAAAATCAAAATGGCTCGGGTTACAGCTCTACTGATG
CAGACTTGTGGGATTTTCAAATTTCCGATCACCAAACCATGCAATAGATTATTGCAGTACTGATCTTTTGCT

GTCGAATATGTTTTATACCATTTATTGGATGCATGATTACGTAGCTACGAAAAATGTTAGATTTAAGGTTTG
TATTAGTATATGTACTGATGCCAATGATTCAGTGTTGTGTACTACTTCCTGCTGGGAACTAGCTAGATATCTT
TCTCATTTATTTTGGCAAATAAATTAGTGGAAAATTACAAAAAAAAAAAAAAAAAGTCGACATCGATACGC
GTGGGTATGCGTAACACTGACGGACTAAATGCAATCTTTCCAGTGATGGATGTTCTTGCGTTAAAAAAAAA
AAAAAAA

Arabica CaMYB12

GAATTCGATTATGGGAAGAGCACCTTGCTGTGAGAAGGTAGGGTTGAGGAGAGGAAGGTGGACAGCTGA
AGAAGATGAGAAATTGATCAACTACATCAAACAAAATGGGGAAGGGTCCTGGAGATCCTTGCCCAAGAAC
GCAGGTTTACTGAGGTGTGGGAAGAGTTGCAGACTTAGATGGATTAATTACTTGAGAGCAGACTTGAAGA
GAGGAAACTTCACCAGAGAAGAGGAAGAGACAATCATCAACCTGCATCGATCCGTGGGAAATAAGTGGTC
CTTGATAGCAAGCAAATTACCAGGGAGAACAGATAACGAAGTGAAGAACTACTGGAACTCTCACCTCAGT
AGGAAAATCTATAGTTTTAGGTCCAGCGATGGTTCCTCGGTGACCACCTTAGACATGGTCAACATGCCTAG
CAAATCTAAGAGAAAAGGCGGTCGGGTGAGCCGAGCGGTTGCCAAAAAGTACAGCATGCACCCAATCACA
AAAGCTCCTCCTATTACTACTACTACTAATGAAACATCATCATCATCAAAGATTAGCAGCACAAGTGGGATT
CAAGAGGCCACAACAGCAGCGCAAAATGCACTGGTAGGTATGGTCGAAGCCACCACACATGGGGTACCTG
AGTCCACGGCTGCAGCAGAAAAAGTTTCTGATGACGCGGCGGTAGGTGCGGTTGAGGAGAGCGGAAAAT
ATCAGTTGCAGCCAATGGCCAATAATAATTATGATTACGATGAGAGCGAAGGGGGATGCATTAATAGCAT
GCACGATATTAGGTGTGGTGATGGAGGTGAAGGATGTGCGAGGGCATTATTGGTGCCCTCTCCAGAAAAG
CTGGAAGGAGAAAAGGGAGTCGCGGGACCAAATGATGAGCTAGATGATGTAACTTTGCTTCTTGAAAGTG
TTCTGGAGAGCGATTTGATGGACCTGAGCGAGATTTCAGTGCATACTGGGGACACAGAAAATGAATCAAT
GTATCCAGAAAGCGCTTCAATTAACAAGGACAGCGACTCCGGTGCCAGCAACTCAAATTCAGACACGGGG
GATGGCCTGTACGACTGTTTCGCACCGTTAGATTCTTATTTCAATCATACCTGGGATGCAGAATATGCTGTA
CCAGGATTCGGGCTTTGGGATCAAGAAGATGACGACGTACTATGGCCATGGGAGAGTTAAACCCACAAAA
ACAGACCAGTTGCAGTAAGTTCAACATCAAAATGCTTCATCGGAGGAGAAAGGCATATATCGTCAACTTTT
CTTGCTGCAGCTGCTGTAATCCCCTCTCTAAAAAGTAATAGTCGGCTAACATAAAAACTAGTAGTCCTGAAT
AAATTCACAATTTTATCTAGCAGATTGTCGCTTTATGGTAGGTAGCATGTGTAGGCAATTGATTATCAGCAT
GTACAGTATCAAGACATTTGCCAAAAATAAATAAAAACAGCACGAGTAGATGAGTATTCTAACTAAAAAAA
AAAAAAAAA

Robusta CcMYB12

atgggaagagcaccttgctgtgagaaggtagggttgaggagaggaaggtggacagctgaagaagatgagaaattgatcaactacatcaa
acaaaatggggaagggtcctggagatccttgcccaagaacgcaggtttactgaggtgtgggaagagttgcagacttagatggattaattact
tgagagcagacttgaagagaggaaacttcaccagagaagaggaagagacaatcatcaacctgcatcgatctgtgggaaataagtggtcct
tgatagcaagcaaattaccagggagaacagataacgaagtgaagaactactggaactctcacctcagtaggaaaatctacagttttaggtc
cagcgatggttcctcggtgaccaccttagacatggtcaacatgcctagcaaatctaagagaaaaggcggtcgggtgagccgagcggttgcc
aaaaagtacagcatgcacccaatcacaaaagctcctattactactgctactaatgaaacatcatcgtcatcaaagattagcagcacgagtgg
ggttcaagaggccacaacagcagcgcaaaatgcacaggtaggtatggtcgaagccaccacacatggggtacctgagtccgcggcagcagc
agaaaaagtttctgatgacgcggcggtaggtgcggttgaggagagcggaaaatatcagttgcagccaatggccaataataattatgattac
gatgagagcgaaggggatgcattaatagcatgtacgatattaggtgtggtggaggtgaaggatgtgcgagggcattattggagccctctcc
agaaaagctggaaggagaaaagggagtcgcgggaccaaatgatgagctagatgatgtaactttgcttcttgaaagtgttctggagagcga
attgatggacctgagcgagatttcagtgcatactggggacacagaaaatgaatcaataatgtatccagaaagcgcttcaattaacaaggac
agcgactccggcgccagctactcaaattcggacacgggggatggcctgtacgactgtttcgcaccgttagattcttatttcaatcatacctggg
atgcagaatatgctgtcccaggattcgggctttgggatcaagaagatgacgacgtactatggccatgggagaaatcactagtgaattcgcgg
ccgcctgcaggtcgaccatatgggagagctcccaacgcgttggatgcatag

Robusta CcMYB12b

TAGGGCGATTGGGCCCGACGTCGCATGCTCCCGGCCGCCATGGCGGCCGCGGGAATTCGATTATGGGAAG
AGCCCCTTGCTGTGAGAAAGTGGGCATAAAGAGAGGGAGGTGGACGGCTGAAGAAGATGAAATCTTGAC
CAAGTACATTCAAGCCAACGGAGAAGGTTCCTGGAGGTCACTACCAAAGAATGCAGGACTGCTTAGATGT
GGGAAGAGTTGCAGATTGAGATGGATCAACTACTTGAGGTCTGACTTGAAGCGAGGCAATATAACTGCAG
AAGAGGAAGAACTGATAATTAATCTGCATGCATCTCTCGGTAACAGGTGGTCTTTGATTGCGGGACACATG
CCAGGTAGAACAGACAATGAGATTAAGAACTACTGGAACTCTCATTTGAGCAGACAACTCCACAAATTCAA
GAAGCCAGATAGCGAAAGTAGTGTACCACCAGGTCCGCTGCAACCAGCGGCAGAGGTGATGGACTTAGTA
GTTGAGCATGGAGGCGGTGAAGTATCAAAGAAGCGAAAAGGCAGTCGAACGAATGGATCCAACATGAGG
AAGAACAGAAGGAGCACTAATCCATCGCAACCAAAAGAATTAGCGGGACCTGCTGCTGCAACGGTTCAAA
TGCCCCGTACTCCAACTCTTGAGGAAGAGACTCTGTCTAGTGCAGTCTCTTTGGAAGAGGGAGGTACGAGT
AAATTCGTGGTGTCTAGTAGTCTTATGACGGACGCATGCCCTAAACGAGAAAGAAGACAAAATTTGGCAA
GCACTGGTACTACTAGCGACGAATGCACCCCTGAAGTGGGAGAAACCAAGAGCAGCGTGATGTTGCAGCC
TGAAGGTGGATTGATGACAAGCGGTAGTGATGGAGCTGTGGGGCTGGACAGCGGGATGTTGTGTCTTGA
TGAGATGATGAATATTGGGGATCCTGATGGGATTTTGACGTTTCACGGAAATACGAGTACAAAAGAAACG
TGTAGTGAAACGGCAGCCACTAACGTGGAGCGACGATGCGATAATATGGACCCGGGCTACAACACTACAA
GAGCATCAAGTATTGTCGATGATCAGGAGGAAACTGATAAGTTGGGAAAAATGGACGTTTCAAAAGGAAG
GAATCAGATCGGCTGCACTTCTCCAGCTCATGAAGATCATCACGACGGAAAGCTTGATGACTGGGATGATT

GGCAATGGGATGAACCGGTAGTCCAAAACAATCTCACCACATTACCAGGGGAGGAAGCAGACTTGCTGTC
GTGGCTCTGGGATTCTGACAACAACAACAACTACCTGAATTCAAGTGGGAATTTTGACGATGGAGCTGGAA
TGGACGATGAAAAGCACAATGCTATGGTTGCCTGGCTTCTCTCTTGAAATCACTAGTGAATTCGCGGCCGC
CTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTGGATGCATAGCTTGAGTATTCTATA

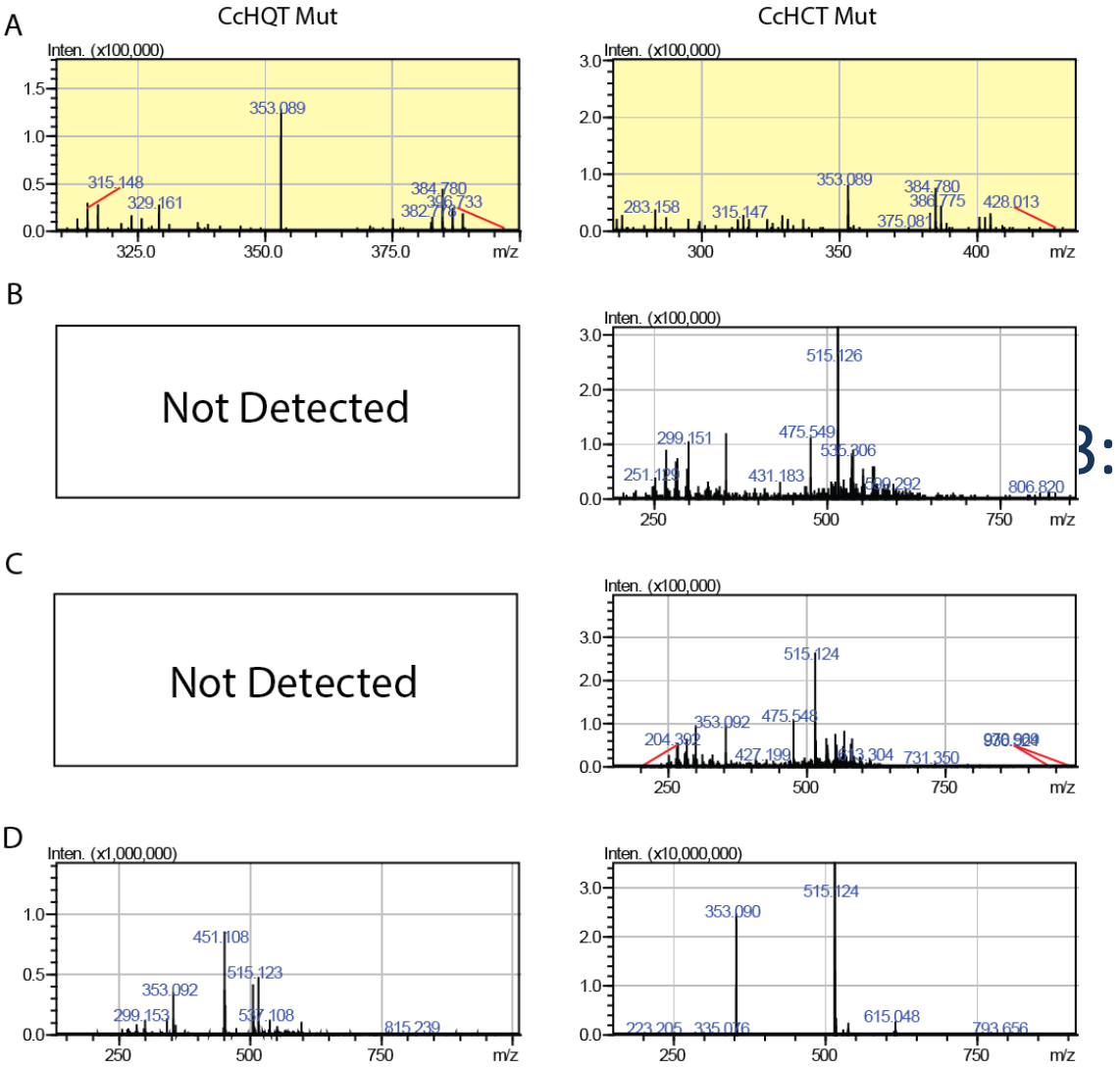


Figure 0.1 Enzyme assays performed with mutated versions of CcHQT and CcHCT. Mutated versions of plasmids were recevied from Andrew McCarthy of EMBL Grenoble Outstation, France. Unless otherwise stated all reactions were at pH 7. **A.** BAHD activty, caffeoyl CoA and quinic acid. **B**. Caffeoyl CoA with 5-CQA. **C**. Coenzyme A with 5-CQA. **D**. 5-CQA at ph5 The control forward reaction produced 5-CQA as a product (yellow) all other boxes represents the mass spectra of where diCQA fragments were detected.

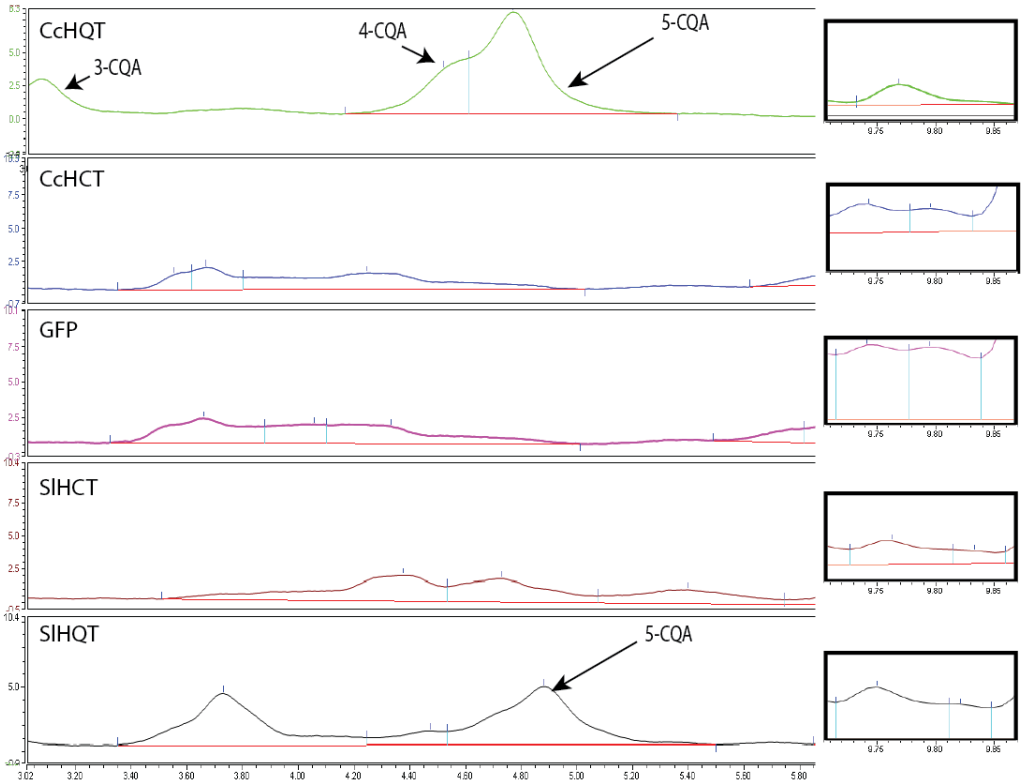


Figure 0.2 Hypertrans overexpression of HQT and HCT from coffee and tomato in HQT KO plants. CcHQT overepression caused the production of all 3 iomers of CQA. SlHQT expression caused the acumulation of 5-CQA. None of the other overexpressions caused the acumulation of CQAs. Black boxes represents area which diCQA standard is found, peaks are observed in all samples.

Chapter 4 supplementary figures

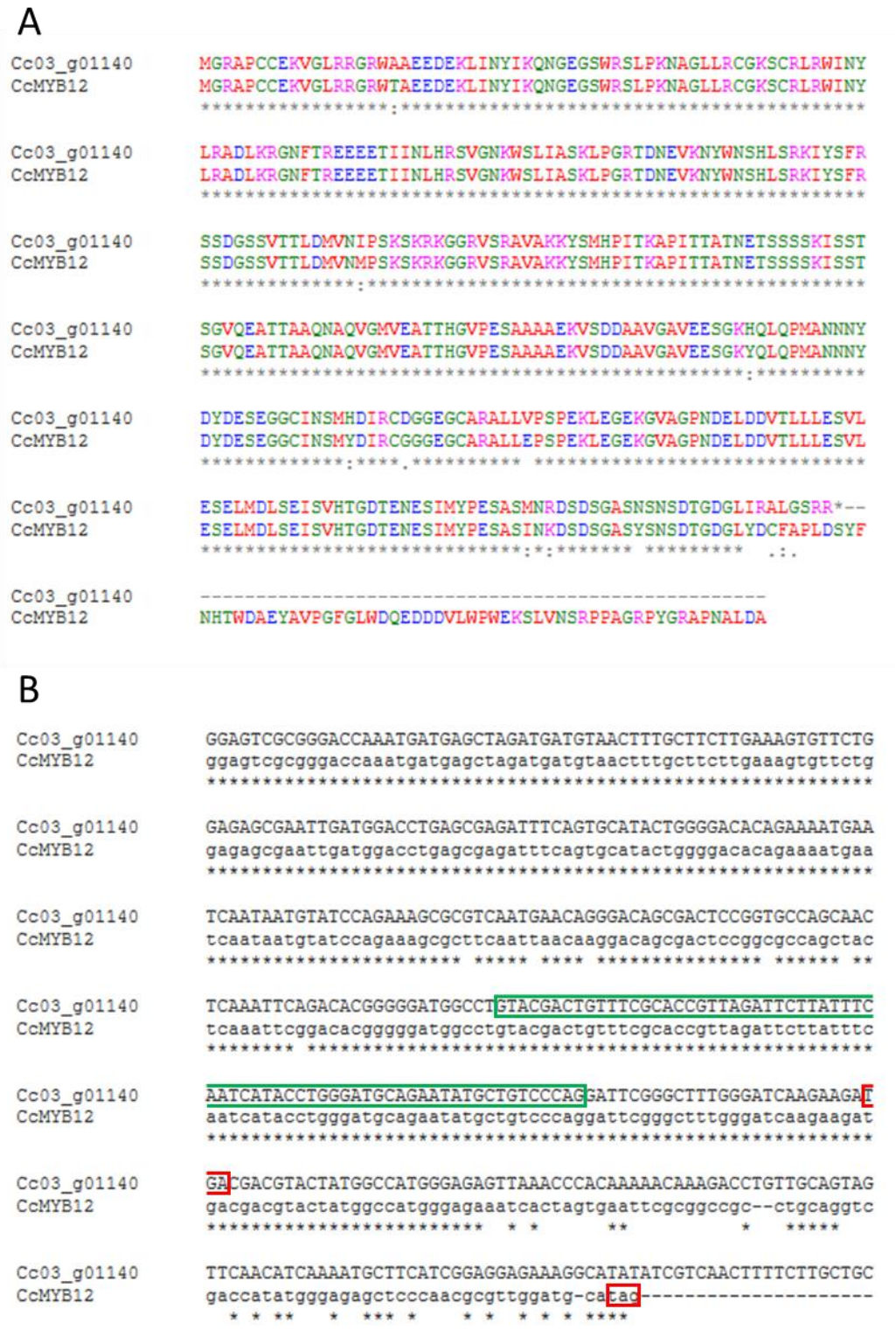


Figure 0.3 Alignments of cDNA and peptide sequences of CcMYB12 against Cc03_g01140. A. Predicted peptide alignment was performed using Clustal Omgea. CcMYB12 cDNA encodes a peptide which is longer then the predicted peptide of Cc03_g01140. **B**. Clustal Omega alignment of Cc03_g01140, of an area which includes a portion of the 3rd and 4th exon and the intron within, against the cDNA sequence of the cloned CcMYB12. The intron of Cc03_g01140 is highlighted in green and predicted stop codon in red. CcMYB12 cDNA encodes a region that is predicted to be an intron of Cc03_g01140.

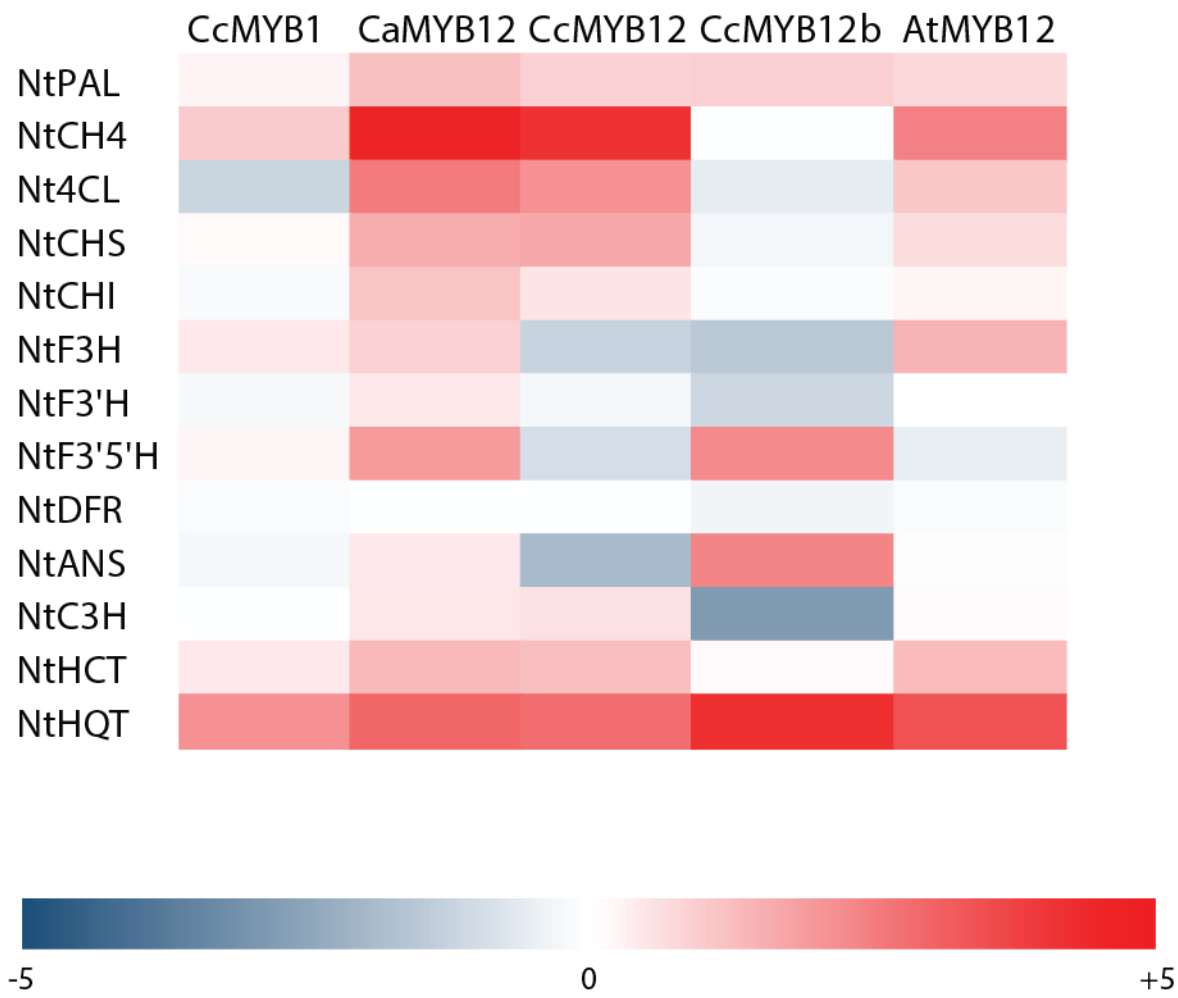


Figure 0.4 Gene expression analysis of transgenic tobacco lines expressed as a heat map. cDNA synthesised from leaf RNA using 5 biological replicates in 3 independent lines. Data shown as ΔΔCt values relative to ubiquitin represented as log10. A, phenylalanine ammonia lyase (PAL); B, cinnamate 4 hydroxylase (C4H); C, 4-coumarate:CoA ligase (4CL); D, chalcone synthase (CHS); E, chalcone isomerase (CHI); F, flavanone 3-hydroxylase (F3H); G, flavonoid 3'-hydroxylase (F3'H); H, flavonoid 3' 5' hydroxylase (F3'5'H); I, dihydroflavonol 4-reductase (DFR); J, anthocyanidin synthase (ANS); K, p-coumaroyl quinate/shikimate 3'-hydroxylase (C3'H); L, hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase (HCT); M, hydroxycinnamoyl-CoA quinate hydroxycinnamoyltransferase (HQT).

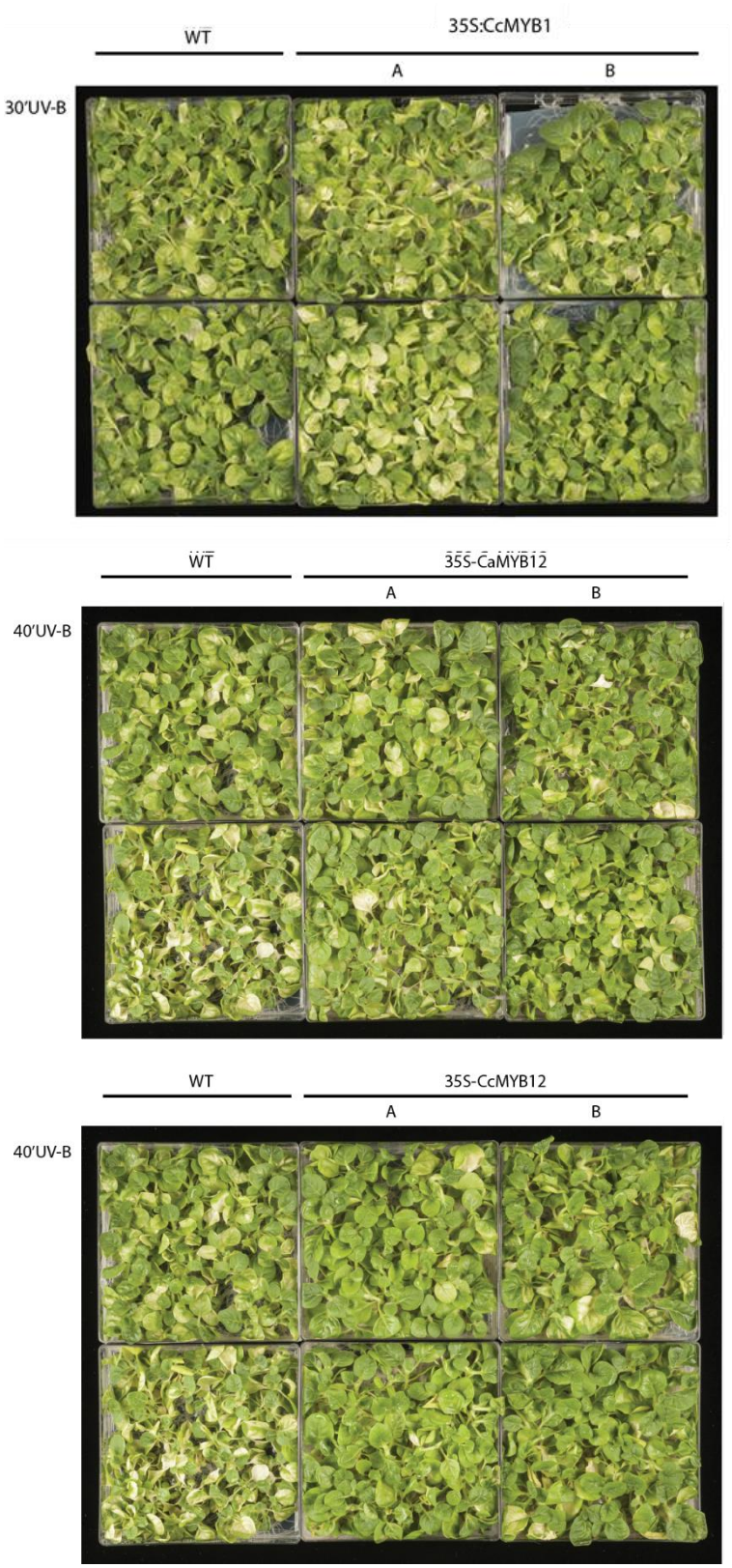


Figure 0.5 Tobacco 35S:CcMYB1 seedlings are more sensitive to UV-B light. Seedlings expressing CcMYB1 showed an enhanced sensitivity compared to WT when exposed to 30 min UV-B (3.2 mW/cm2). Two replicates of two independent transgenic lines were used for the experiment

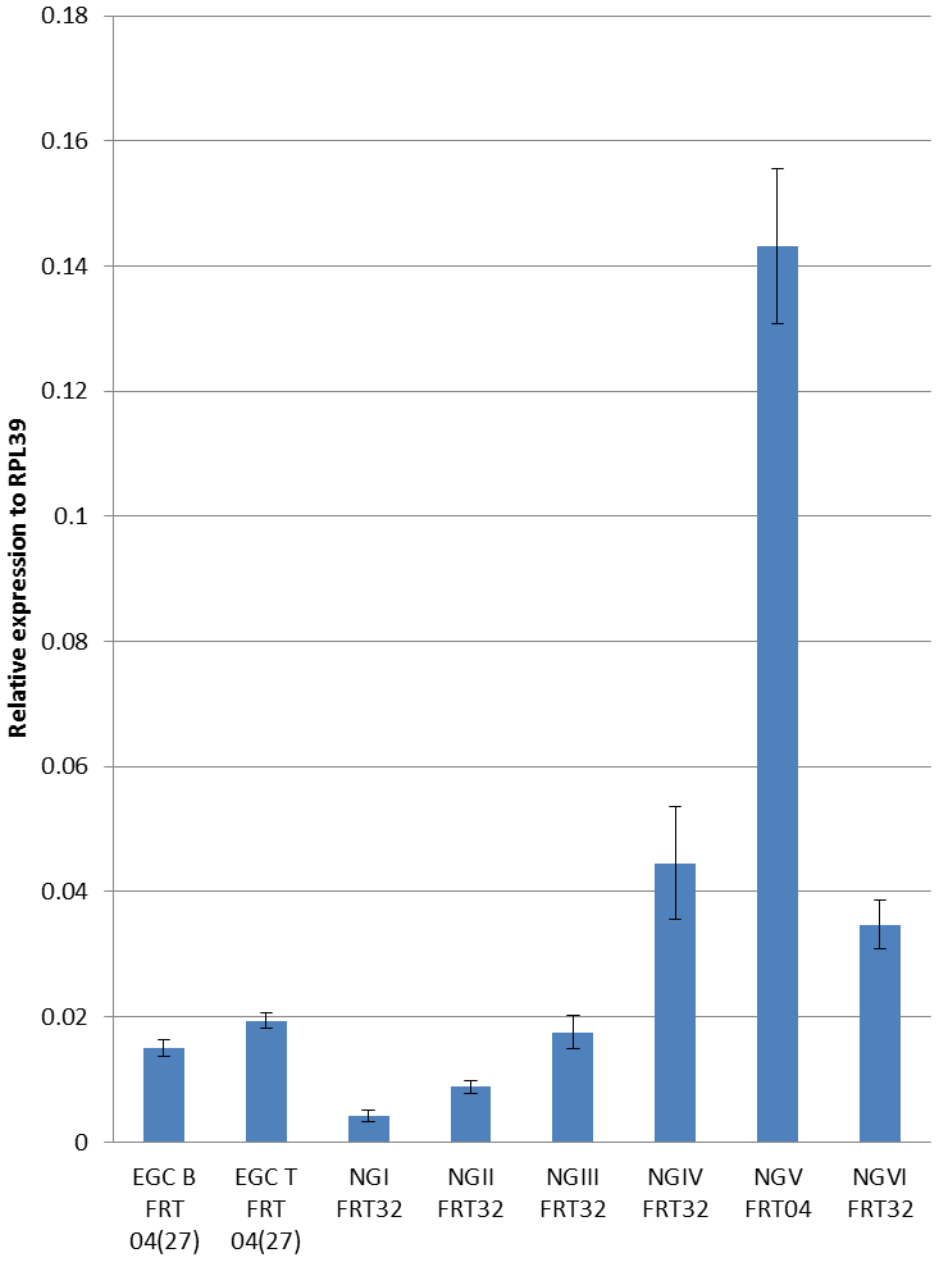


Figure 0.6 Relative transcription of MYB1 of *Robusta* (Cc) in cherry and flower developing tissues. The cDNA was synthesised from Robusta var. FRT32 and FRT04(27 cross) The data is displayed as relative quantification values to RPL39. EGC B early green cherry bottom half. EGC T, early green cherry top. NGI, very very small node; NGII very small node; NGIII small node; NGIV between node/flower; NGV nearly open flower; NGVI fully open flower. Bars = standard error

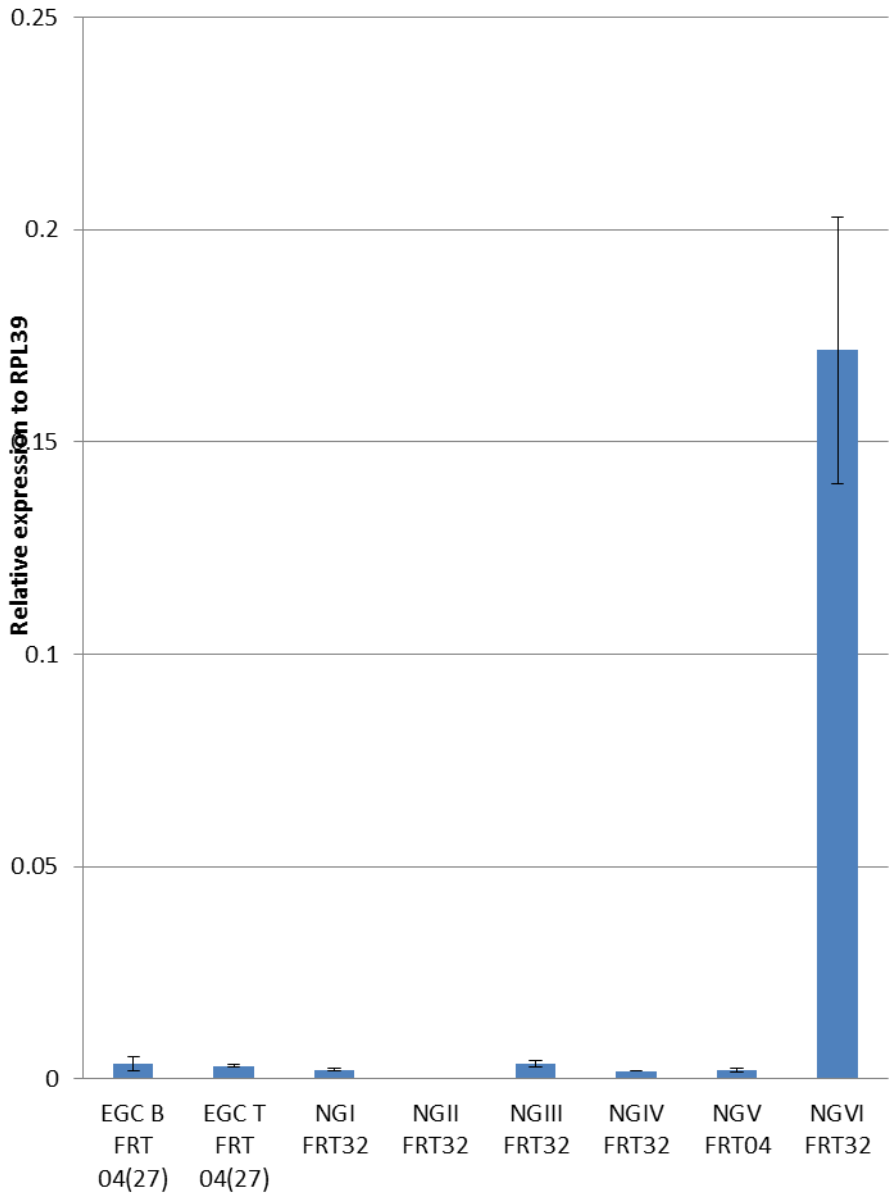


Figure 0.7 Relative transcription of MYB2 of *Robusta* (Cc) in cherry and flower developing tissues. The cDNA was synthesised from Robusta var. FRT32 and FRT04(27 cross) The data is displayed as relative quantification values to RPL 39. EGC B early green cherry bottom half. EGC T, early green cherry top. NGI, very very small node; NGII very small node; NGIII small node; NGIV between node/flower; NGV nearly open flower; NGVI fully open flower. Bars = standard error

Yeast 1 hybrid clones. Predicted MYB12 binding motif in purple. Green is the ATG start of translation. Grey is TATT box and blue is the CAAT box.

PAL UnK

AAATTAGTTTTAAAAATTACTTTATTTATTTATGGGATATGAGTTATGCATTATTAAAATAAATATTTGATTA
ATTGTGGACCCATGCTATTACACCTTGTGGAGTGTAATCCATTGTGAGTGAAAGTGTAATAAAAAAGAAGA
AAGTGGAATGCTGACGTGGCATGTGGATTACATTCCACAAAATGTAATAGCATTGTGGATGCTCTGAGTGT
ATCATGTGAAAGAGAATAAAATCAAAAGGTGGGTAGGTGTAATGACTATTTCAGGGAAACTGCCAGAAAG
CTCAAGGAAGGTATCTGAAAATTGCCAAATCTTTTCACCGGGCTACCAACCACAAATCAGTGAAAAGCCCA
TGTACCATGCCATTCCCACGTATACCCCTACCAACCGTCCAATTTATGACCGATCAATGGTCCTGATTCATAT
TCAACCAAGCCCTGCTTAGCAACGACTGCCACCATCCCCCGCCCTTACTCTTTCTCTCCCTTGTGCTCTCGAG
TCTCACCTACCAATTCTTGCATCTATATATTTCGGAGGCCTTTCGCAACTATAGCCCAGGAACTTCCTGCCCA
CCAGCATTAGCAGTTGTTCCTCTCGCGATACAAAGTAGTAGCATAGCATTACTAGTTAGTAGTCACCACTGC
TACTGCTTCAGTTCCTTTTATCACTTGCTTTACATGAATTAAGTCGATGCTCTTCCTTGAATAACTAGCGATTA
GTTTCGTGGTGACCTATCTAGCACGTCTAGCCATTTTTCTGTTTGGGTGACATCAATCCTGAACAGAGAAAG
CTGCAAGATGGAAAATGGTCATGA

PAL2

ATTGTATGGATGGGATGTTCTTTGAGTTATTGTGTATTACTGTAGAATTGTATTTAAAAAACTTATTTTTGAA
AAAATGGAGAATCCAAACGGATAACATCAAGTGCAAACTTTAGCCAAACATTTTTTTTCTAAACAAAAATGT
GTATACGACCAAACCATGTATGACTTGTGGTAGAGAGCCTTCTTCGTCATCCTAGCTGCCCATGCATTCAAA
CACGTGTCATAAATCTCAACCATCAACCGCATGCTCATCTGACGGTCGTCTACATCTTTCAACTACCCAATGC
CTGCCTCACCCAACTCTCAATGCCATCCTGTTCTCTATTTAAACAACCAAAGGTGGCTCAACCAACTCAGGA
AATTCAGCTAGAGCTATTTCCCCTTAAAGCTTTGCAAATTTCCTACTCACCCCGATCGATCAATTCCCCGCCC
CCCCTCCCCCTCCTCCCCCCAGCCGCCACCACCACCACCACTGTTTAAGCATTACAATCTTTTGGGCTAATTA
TTATTATTATTATTTTTATTACTAGCATTTACTAGTATTTTCTTGTGAGCCCGGATTTCTTGATCGTTCCAGTA
AGCTAGCTAGCCAGCCAGCCAGCAAGCAAGCAAAAATGGAGTGCGCTAATGG

PAL3

TATTATTATACCTATTAATTGCATGGTCAACAACCACGGGGGCCAACCACCACCCCATAAATCTTGACGGGT
AAAATCCAAGTTCCACGCAACTCACATTTGCACCTGACCCCCTGGGCCTACGTGTCACACATTTTATCCGTCA
GATCACATGATGATCCTTCGGCCCAAATAAATTCTCAGCTAACCACCCAACACTCTTTGAGTCGCACCCTTTA
TTTAAAGAATGCCCCAATTTACAGTTGAACCAACTCTCCATACCCTTATTCCTTCCATTGATTTTTTCCTTTCTT
CCTAAACATCATCCATTTCTAGTCTTGGCACAGAACAGAACAGAAAGCCACTCCACTTTCTTTCTTGCTTTCT

GTTTCTGTGCGGCCTTTTCTTTTAAGGGGGAAAAAAAGAGTCTCGTACCCAACATTTGTAAGCGTAACATTT
GCTTTCTTGCCCGTGACCCAAAAAAAAAAATTTTTATTTGTTCACCCAGGGAAAAAAAAGTTATATGGAGCA
CGTACATGAAAATGGAAATGTCCATCATTCCAATCGCTTCTGCGTCCAAGATGATCCCTTGAATTGGAACGC
GGCGGCGGAGTCATTGAAGGGGAGCCATTTGGATGAAGTGAAGAAAATGGTACGCGAATATCGGAAGCC
ACTAGTGAAGCTAGGTGGTGAAACGTTGACGGTAGCCCAGGTGGCAGCCGTGGCCGCCGTGCATGCAGG
TGGTGGGACGGTGGTGGAGCTGTGTGAGTCGGCCAAGGCTGGTGTGAAGGCCAGCAGTGAGTGGGTGG
TGGAGAGCATGAACAATGGTACA

C4H Ch1

TCCCACTCCTCCACACCGTCATGGATCTTCTCCTGCTAGAGAAGACCCTCTTGGGACTGTTTGCAGCCATCAT
AGTTGCCATCGTTGTTTCTAAATTACGGGGCAAGAAATTCAAGCTGCCTCCAGGCCCAATCCCAGTTCCCAT
TTTCGGAAACTGGTTACAAGTTGGTGATGATTTGAACCACCGCAACCTCACTGACTACGCCAAGAAATTTG
GAGAAATCTTCCTTCTGAGAATGGGCCAGCGCAATCTTGTGGTGGTATCGTCCCCTGAACTTGCCAAAGAC
GTCTTGCACACCCAGGGGGTGGAGTTCGGCTCCCGCACCAGAAATGTGGTGTTTGATATATTCACCGGCAA
AGGCCAGGATATGGTCTTCACCGTCTACGGCGAGCATTGGAGGAAGATGAGAAGGATTATGACTGTCCCC
TTTTTCACTAACAAAGTTGTTCAGCAGTACAGGCACGGTTGGGAGGCAGAGGTTGCCCGTGTCGTGGAGG
ATGTTAAGAAGAACCCTGAATCCTCCACCAATGGGATTGTCTTGAGGAGGAGGTTGCAGCTCATGATGTACC
AATAATATGTACCGAATCATGTTTGA

C4H Ch0-1

GCAATGCCAATGCATCTTCATGCAAACCCTTTTTGGCATCTCGGACCATATCCGTGAATATTTCGTCAAAGC
TGATTCATGACGCTGAATTATCCGTGAAGTTGTCTGTTTGAATGGCGAAGATGAGAAAGAGAGGACAGAG
ATTGAATCGATTTTAGGGTTTTTCCCACTTGAAACAACAGTAAATTTCAGATTGTTTTCTTATTTTTCTTGATT
TTGGTTATATTGGTTAAAGAGATTGAGTTTCTTTGTCCGGTTAATCGATTAATTAACTTAATATGGTCCGAAT
TGGGTTGTTTGAGGGTTTAGTTTATTAAGGGACACTTGGCGAGATTCTAAGGAGTGGGACAGGATATGGT
ACAGAAGATCCGTTGCGCGTCAATTCGGTCTGATTCTGCCACGTTTCCGCATTTATGGCATGAACATTTGAG
GCCTTACACTTTGCGTAATTATTTTGGCAAACTAATTGTAACACTTGTCCTTATTGAAAAAAGAAACCACAA
AAAAAAAAGACAATTAGACAACTGCCCCCCCTGCTATAAATTATAGTCACTCCGGCACCGATTCCACTCCAT
CACCCAACCCACTCAACCAAACCAAACCAAACCTCCCACTCTCAGTTGATTCCAACACAAACGCGCCCTCCA
CACCGTCATGGATCTTCTCCTGC

C4HCh0-2

GTAGATCGGTCATCAAAATGGTTTGGTTCTATATTAAACTCAGAAAACTAGTTTAAATCTATGTTAAACTAG
AATCTTAAAAAAACCCAATCAAAACCGGCAAAATCCAAACTACTGAAGGGGTTGTTACTTTTTTTTTTTTTG

GTAGTTCGGTAATATCTAGATTATTTGATAGTAATATTTTTAAAAAAATGTTGCATATTTGAGTTGGACTGT
GAACCTGTTGAACTGGAAGCTCATCTGGTGCACCAATCAATCTAGGGTTAAAACCATTTTTTAGCCGTGACA
GCCTTTGACATAATTTTAGACAAGTCTGCTCAATGTCTTGCTCGACTAAAATGGTTAGGTTACCTTGTCAGA
TTGCGAGAAAATTGTTACTGCCACACTGGTCTTGCATCCAAAGGTGGTAAATGATTCAGCATATTCTTAATA
GCTCTGGTGTCTTAGTTATAATTTTGCGATTCCAGGAAGCTGTTAAGAACAAAAAGCATGGATAACAACAG
TCC

C4H Ch0-3

TTCTTGTGTTCAACCATGTCATGCATAGTTTTTGCTGTGGTATAGGTGTAATACTCTCTTGCCATTGACTTGA
TATCAACAATATTCAATCCATATCCATCACACACTGAGAATTAGAGTTGGTATAAAATGTGATTCAAGCCCA
TGGAACAAAAGACCAACATATGATATGAGCATCCAATACATATCTGATTGTTGATCTCAGTTTCACAAGTAC
TTAAATTTATTTTCTCCACCAAAGCCAAATTAAGGTAAAAACCGGCAAGAAATTATCTAGTTCCTCGTGTGA
TTTCCACACAAATTGTAGAAAGGCAAGTTTGAATTGTTTCAAGACTTTGTACAAAGATGGATGGCCTTGGA
GAGGCACTTCAACTACCAAACCTGCAAGTGCAACCACCAAATTTGACTTGAGAAAATGGTATTTCATTCTCC
TAACCAGATTCATAAGCAAGCCCGTGGATTTTCCTAAAACCAGCGGATACCTACCTTCCAAACCCACTTACA
TTAGTCATATATAAACTCAATGTCCTTACCTTTCCTTCCTCGTACTGATCAGTTTCCTATCCCTCTTACAACAT
GGGCAAACTTCTCAAG

4CL

AAAGACTGCTCCCCTAAGCTTATCATCACTGTCCCTGAATTGATTTCCAAAGTTTCTGGCTTTAATTTGCCGA
TTTTGCTCCTGGGGTCCGATAAAAATCAAGAATTTGGCAGTAAATTATCTGGGAAAGTCATGTTTTTTGCTG
ACTTGATCAAGATAGTTGGGCCAATGAACTGGGATATCTCTAGTACTGTTATTAAGCAATCGGACACTGCT
GCATTGTTGTACTCTTCTGGGACAACTGGTATAAGCAAAGGTGTAGTCTTGACGCATGGGAATTTTATTGCT
TCCTCATTGATGGTGACTGCTGATCAAGAACTTGCAGGGGAATTGCATAATGTGTATTTGTGTGTTCTGCCT
ATGTTTCATGTTTTTGGGCTTGCTGTTATCATGTATGCACAGCTGCAGAGGGGTAATGCTGTTGTTTCAATG
GGAAAGTTTGATCTGGAAATGATTTTCAAAGCTGTTGAAAAGCACAGGATCACTCACATGTGGGTTGTGCC
ACCAATCATATTAGCTTTGGCTAAGAATAGTGTGGTTAAGAAGTATGATTTGTCATCCTTGAAGCAGATTG
GATCAGGGGCTGCACCACTGGGGAGGGAATTGATGCAAGAATG

CHS Ch3

CTAACTCATAGATTTTCAACCTTGAATTCTTATTAGACCAAGACTTATTTAAACAATATTAAGCAATTAAACC
ATTAATGAAATACTAGTTAAAAAAATAGTAGTAATACTAGTGAACCAAAATTTTTTTCTTTTTTGTCCTTCCT
CTTCTCTTCACACCCTTTTTTTCCTCAAAATTCGGACCGTACACCGGACAGATAGCATGAAAAACTAGTAAAT
CAATTTTTTCTCAAAAAACCACCATATCGAAATAGGTTTATTATAGCTTTATTACTAGCTTGTGTGTTGATTG

ATGACACTAAGGTCCAAACATGGCAACAAAAATTGAGAGTGTACGCAGGGAAATTGGAACTCGGTGTTAG
GTACCAAATTTCGGGGTCATGTAGATGGTTTGAGGGGCAAGTGCAAACCACACGAAACTATCACGTGCGG
CCGGCAACCTTACGTTCGCTAGGCATGTGATTCTAAGCTACCATTATAGTAATTACCCCATCTTCGAGTACT
ATATATACCAAACCATTTTGCAGACTTGCACAGCCACAACTTCTTCCGGAGCTACAAAGCTAGCAGCTACCT
GCAAAACTCTACGTACCTTAATTTTTCTCTTTTTCCGAGCAAAAATGGTTACC

CHS Ch0-1

CTACCATCACATTAGTAAACATTCTAAGGACTTTATCAGTTATATCCAATATTTAACAGATAAATAAGAAATT
ATTAAAATACAAAACATATAATTTAGGAATGTATAAAATAAAAACAAATACATTGTCTTAGGGCATACATTC
CAACACGTTGATGCAGTAAAAGAGACAAAATGTATGCTTTCAAGAGTGAAAATACCAAAGGATAAATGGA
AGTTTTGTTTTTGGCCTTCCCTGGATTACAAAAGTAGGAACATGACTCAAAGGAAAAAGACAAAAAAAGAA
AACTATCACATGATTTATGAAATATGTTTTTGGAAACTCCAGATTAATTGAAGAATTAAAAAGTCAGATTCT
TTCTTGGTTAAAGGTTCATGCTAAGATGTGTGTTAATTTCTTAATTCTTTTCTTTGCAGATAGATAGGGTAGA
CACATATGTAAGCATACTTTTCTGATGGGAATTAGATATTGCCCCAAGTGGATAAAACTCAACTTTCTTGCA
TTAGAAAACGTATTCAAACTGTTCAACCATGATGACTTGATAATTCCCAATTCATCATCTAAGTAAGCAGCT
GGATGATCCAAATTGGTCTAATGAACTTTTTTTTTTTTTTGAATTTCATTTCAGGTGACAAATCCATGATTAA
GAAGCG

CHS Ch0-2

CCGCCCTTATAAGAACATTTCAACACTATCAGCATCTGAGTCCGTTGATGCAGTAAAAGAGACAAAATGCA
TGCTTTCAAGAGTGAAAATACCAAAGGATATAAGAACATTTCAACACTATCAGCATCTGAGACAAAATGTA
TGCTTTCAAGAGTGAAAATACCAGAGGATAAATGGAAGTTTTGTTTTTGGCCTTCCCTGGATTACAAAAGT
AGGAACATGACTCAAAGGAAAAAGACAAAAAGAGAAAACTATCACATGATTTATGAAATATGTTTATGGA
AGCTCCAGATTAATTGAAGAATTAAAAAGTCAGATTCTTTCTTGGTTAAAGGTTCATGCTAAGATGTGTGTT
AATTTCTTAATTCTTTTCTTTGCAGAAAGATAGGGTAGACACATATGTAAGCATACTTTTCTGATGGAAATTA
GATATTGCCCCAAGTGGATAAAACTCAACTTTCTTGCATTAGAAAACGTATTCAAACTGTTCAACCATGATG
ACTTGATGATTCCTAAAATTCATCATCTAAGTAAGCAGCTGGATGATCCAAATTGGTTCTAATGATTTTTTTT
TTTGAATTTCATTTCAGGTGACAAATCCATGATTAAGAAG

CHS Ch2

CCGCCCTTATAAGAACATTTCAACACTATCAGCATCTGAGTCCGTTGATGCAGTAAAAGAGACAAAATGCA
TGCTTTCAAGAGTGAAAATACCAAAGGATATAAGAACATTTCAACACTATCAGCATCTGAGACAAAATGTA
TGCTTTCAAGAGTGAAAATACCAGAGGATAAATGGAAGTTTTGTTTTTGGCCTTCCCTGGATTACAAAAGT
AGGAACATGACTCAAAGGAAAAAGACAAAAAGAGAAAACTATCACATGATTTATGAAATATGTTTATGGA

AGCTCCAGATTAATTGAAGAATTAAAAAGTCAGATTCTTTCTTGGTTAAAGGTTCATGCTAAGATGTGTGTT
AATTTCTTAATTCTTTTCTTTGCAGAAAGATAGGGTAGACACATATGTAAGCATACTTTTCTGATGGAAATTA
GATATTGCCCCAAGTGGATAAAACTCAACTTTCTTGCATTAGAAAACGTATTCAAACTGTTCAACCATGATG
ACTTGATGATTCCTAAAATTCATCATCTAAGTAAGCAGCTGGATGATCCAAATTGGTTCTAATGATTTTTTTT
TTTGAATTTCATTTCAGGTGACAAATCCATGATTA

CHI Ch2

TTAAACCCCCTGCAACAGCTGAATCTACTCGTCACTCTCACCTTCATTAATTTCCCCATCAAAATTTCCAAGCT
CCTACAGCGAGTTTTCGTGTGTTTGCATAACTTCGCCGGCAGTTTAGCGTCGGTGTCATGGTACTAGGCCAT
CCTACTCTCTGGGGAAGTACTCTACTGCATTTTTTTTTGGTCCATACTCCCCAGCTATATATATATATATAAA
AGCCATAACCTGACCCTGCTTTTTTCCAATAACAAGAAGAACATTTTATTGATAACAACAGATATCGTCCATT
CCTCCTACTTTCTCACTAACCAGTCCACCCATCTCCAATTCTCCCTCCACCACCCTTCGGTCAATTTTGCTTCTC
TACCCCCCAAAAAAAAAAAAAAGTTTCAATCTCCAAGGCTCCAAATTCTGCCCATCCCAACTAACCCCTTCG
TTTTAAACGACTCTATATATAGCCTCATACCACCACTACCACTGCTACAAACTGCAAAACTCCCAGACTGCCA
ACTCTTCAAAGGTGCGCGAAATCTCTGAGTGCTAGTTACAATATGTCCCTCTCG

CHI Ch3

AAGGATAACTTGTTCTGCCGCTTGCAAAAGGTGTACGACAAGCAACAAATATCAATTCAGAATTGCTGAAA
TCCATGTTTCAGGGAAATGGTAGTTTGTTTGAATTCTAGTTCTTTTCTTCGGCAAACATTAACTTTGAAGCTA
CGAGCTGGAATGAATTCCCTTACTATTATATATTTCTAGGGAAGACTTAGATTCCGAATCTCTGATTCCTTTT
TGGGGGGTTCTATGGATGCTGCACATAATCCATCTGGCCACCAAAGATAAACTTGTGAAATGGGGTACAAT
AACACGTTCTTTTTTTGATATTATGATACAATTTGTCGTATTTACCAGACTTAGATACCAATAATGCCTTGAG
GTCTTCCTCTATAAATAAACATCTTACAGTTCCATGTAAGAAGCAAAGCCAACAATCCGTTCAGAAAATTTT
ACTCCTACCTCTTCTTCCACCTCTACAAATTACTACCAGACTTCCAGAGTTGAGGTTGTATGATACTAATGGC
GCCGGCAT

F3H

AACAATTCTTATATAAAGGCACGTTAATGAGAAGTAAGATTTGAACCCTTTACATCTCACCCCATAATTAAT
GTGATGGCCTTTACATCCCACCCCATAATTAATGTAGTGGCCAGCTCCTCTAGAGGGAGTTAACAACATCTG
TATGGAGAATGCATTTCTGGTCGGGAATCCGATAATATGCAGACGCATCACTTTGCTCAAATTAGTCATCAT
TTGCTTTCTTTGATGTAGAATCTCAATTTCCAAGCCAAGCCAAGTAGTGAGAATTAGTACCAAGCAAACCAT
TAACGCAAGCAAAGATCATATGCTAATTCAAGTATCACAAAGCCGGCCATGCTGTATCCGGATGTCCACGT
GTTGGTAGTTGTCTCCAAGAAGGGTAGCTAGCTGCCACGCCATTGCATTTCCCCTCCTCCTCTTGCTACAAA
CCTGAGCCGTTCAATCAATCTAACCTTCCCAAACCAAAATTCCTAATAGCCGTTCAATATAGACTTACCAACC

ACTCTCTATATATATGAATCAAAAAACATTATTGTTGTTCTTCATCACAAGAATCGTGTTGGTCCGCCGATAA
TTAATTCATCGTATCGAGTGTGGGCAAAACAAGAAAGAAAGCTGACAAACTGAAACAGTGAGACAGATGG
CTCCCTCAACTCTCACA

HCT

TCGCTAGTCCCTGCTCGCACAACAACCATCACTTCGGCAGCAGAATTTGAACAATGACTGTTTGTAAGAAAA
GTTATTCGTTTTTATCAATTGAACCAACTCGTGTTAGCCCCTTTAGCGTGTGCATGTGTGTTTTTTTCTTTCTT
TTCTTGTTTTTTAATTAATTTTTCTTTGGTTACGGCCACTTTCTCCGTCACTCACAACCATGTTGTCTTTCTTTT
TCTCTCTATCGTTGACGGACTTTAGGAGTTAGTACTAAAAAACTATTTTCCCATTTAATTTATGCAATCTCTTA
TAGTTAGGCAATTGATTTGTTGTGCGGCGAACCCGTCAATAACTTTCTTGTAGTGGGGTTACAAGTAAGCTT
TCTTAGGCGCGGTAGGTAAGTCCCCCAAAACAAATTATATAGTCCCCCTGCGGTCTCGCACTCTTTAACATT
TTTTTTCCCGTTTTGTTAGGACGCAGATCAGATTCATTTCGATTAAAGTCCAACCCACTTCGAACAACACCAT
ACCTACAACATCATTGTCCTTACCACTTCTCCCATTAGCCAGCCTACCACCTCCTCCTCCTCCTCCTCTTCC
TTTTTGAATTTCTTATCCCACCACCCAACATATTCTTCTTCCTTCTCAACTCTTCAGGTACGTTTAATGCAAATT
TCACGTAATATTTCACGTTCGAAATCCATTCTTCATCACCATTAAACTAATCAAAGTACGTTTATTGCACCATT
TTCTTTTATGTTTTTTGGGGTACAGTTAGTTGTTGTACGTGTCAAGCAAGGATCATACAACTTTCAACCATGA
AAATCGAGGTGAA

HQT

TTCTATCGTGTTTGAATTGCTATTTTTAGGAGTTTTTTTTTAAAAAATATATTATAGCGATTCGATAAATATAT
ATGTGATAAAAGGGTGGTTGAGAAATGTGTTCATAAAAAATATAAAAAATTTTTTGGCTCAATTGCATTTG
AATGAGACCCACTCTTAAATATTAATGGCTCAAATTGGACATTAATGCAATTTCATTCGAAATCTGATTATTA
GGCTAACCATTTTATTTAATGGCTCGCATTTGACCCTTTTATTCATCTAATAAACAAAAAAAAAAAAAAAATA
TAAAATAAAAAGGAAGAGAGAGAAAATTCAACGTAGCTGTACTATAAATATGTACAGCCAGCGGCAATTA
CAGTCGTTGCCTGGAAAAGAGAAGCCTCCCTCTTTCACAAGACACTTTTGTGGTCATCTTTCAATAGCCGCT
CTTCCTCATCACATTCCACCAAGGACAAAGGTACGGGTACTTATAGCCAAGCAAAATTGCACAAAGTCATTT
TCCAATTTGCTTCTGAAACGTCAGGACAAAAAAAATCTGTAACTCTAACACATAAATTTTCAATTGACTTCCA
ATTCATCTCTCTCATCTGCCCTGCACTTCTATTTATCTAGACCATTTGTCCGTCTACACACCTTTCTTCTCCATC
CCAGCTCGTTTCTGACAAATCCCTTTCGTCAATCAACAAAACTCTTTACATATTTTCTGCTAAATCTCAGCTTA
TTTCATCTTGTTTGATCAGCGTGGGAAGCGAGAAAATGAAGATAACC

FLS

AAGTGGAATAGAGAGTAATACAAGATTTGTGAGGTGATATGCACGGCCTGACAAAACAAAATGGTGATTA
TATTTGTGTATCATCACATAATTCGAATGTATTCAATCAATTCTTCACTTTTCATCACATAATTCGAATGTATT

CAATCAATTCTTCACTTTTCTAAAAATATTTTTAACAGTGATAACAATTCTTATACATTCGGTAGGATGTAAC
ATATTGGTTCAAAACTTTTTCACAAATTTATGGAGTTCAGGAGGAGTTCGAGAACTACGAGGCTTCAAATG
CTTAACAGGTTTTATAAACAGAAAATAGAAAAGAAAACGATAAATCCTTACTCAAATTGATCTTTAGGTTGT
TATTATAAAAAGTCATGTTTACAAATTGTAATTTTATTTAACTTGCACACTCAACTCATGTGGTCATGTGCTA
TGCATCAATAAGTTTAGGTACCAAGCCACCAACCATCCTGGATCCAATGTTCTCTAACTGTCGCATTCAATT
GAACAACATAGCGTATGCATATAATAATAAATGTAGTCACACTAAAATAAAAAAGGTAGAAAAACACTAAG
AATTTTGCCCATAATTAAATGTTGTATTACGTAACCAATATAATTTGTTCGATTTCAAATCAACTTCATCTGA
GCCGTCAATAGTAATTTATAGAGGCAGAGAATTTTAACCCAACTATCCTATCAACATATTCTTCGTATGATG
CAAGATGTAGGCTACTATAGCGGTGCATGCTGTTGAGGTTTGGCATGATCTATCAAATGAACAGGACTCAC
CCACGTAACCCCTTTTTTATCTTTTCATAGAAAATTAATGACATCCACAATGCATATTATTGCTGGTGTTGAT
GAATTTGGTTCCAGCTTTTGTCTTGCCTCCCTTTATATATGCTGCAAAGAGGCAAGACTCTAGCATAAAGAA
AGAAAATCATCAAAAGTTTTTCTGTTGGAGAGTGAGGGAAGGAAAAAAAAAAAAGAAATGGAGCTGGAA
AGAGTGCAACAAA

Appendix 4: Medium recipes for yeast, bacterial and plant culture

Lysogeny Broth (LB)

To prepare 1L LB liquid, add 10 g tryptone, 5 g yeast extract, and 10 g NaCl Adjust the pH of the medium to 7.0 using 1N NaOH and bring volume up to 1 liter. To prepare LB agar medium, Prepare LB liquid medium as above, but add 15 g/L agar before autoclaving.

YPD (glucose)

Amount to add (per 1 L final volume)

Agar plates

Bacto agar -- 24 g, Bacto peptone 20 g 20 g, Yeast extract 10 g, Water 950 mL 950 mL. Autoclave the mixture.

For liquid media

Add 50 mL of sterile 40% (w/v) glucose. Mix.

Allow to cool before use.

SD (glucose)

Add a Minimal SD Base Medium and the appropriate dropout supplement in the amounts shown in Table I to 1 L of deionized H2O and stir to dissolve.

Verify that the medium has a pH of 5.8; adjust if necessary.

Autoclave at 121°C for 15 min.

Store autoclaved, liquid SD medium at 4° in subdued light.

Minimal Plating Media with Agar:

Add a Minimal SD Agar Base and the appropriate dropout supplement(s) in the amounts shown in Table I to 1 L of deionized H2O. Mix to dissolve. The agar will not dissolve until it is autoclaved.

Verify that the medium has a pH of 5.8; adjust if necessary.

Autoclave at 121°C for 15 min. Cool to ~50°C before pouring plates.

Allow plated medium to harden at room temperature. Store plates inverted, in a plastic sleeve at 4°C.

MS (Murashige and Skoog) medium

Formula for 1 litre of de-ionised water.

Murashige and Skoog medium

(micro and macro elements including vitamins) 4.41g

Sucrose 30.0g

Adjust pH to 5.8 with 1M NaOH

MS 0.8%

Formula for 1 litre of de-ionised water.

Murashige and Skoog medium (including vitamins) 4.41g

Sucrose 30.0g

Adjust pH to 5.8 with 1M NaOH

Add per litre

Formedium agar (AGA03) 8.0g

SOC

Formula for 1 litre of de-ionised water

Tryptone 20.0g

Yeast Extract 5.0g

NaCl 0.58g

KCl 0.19g

MgCl2 2.03g

MgSO4 7H2O 2.46g

Glucose 3.6g

2 x TY medium

Formula for 1litre of de-ionised water

2 X TY broth for medium YDB0102 31g

pH 7.4 with 5M NaOH

Mix contains

Formula for 1 litre of de-ionised water.

Tryptone 16.0g

Yeast Extract 10.0g

NaCl 5.0g

Adjust pH to 7.4 with 5M NaOH

If solid medium is required add 10g for medium agar

Appendix 5: Dual catalytic activity of hydroxycinnamoyl-coenzyme A quinate transferase from tomato allows it to moonlight in the synthesis of both mono- and dicaffeoylquinic acids.

Dual Catalytic Activity of Hydroxycinnamoyl-Coenzyme A Quinate Transferase from Tomato Allows It to Moonlight in the Synthesis of Both Mono- and Dicaffeoylquinic Acids[1][W][OPEN]

Andrea Moglia, Sergio Lanteri, Cinzia Comino, Lionel Hill, Daniel Knevitt, Cecilia Cagliero, Patrizia Rubiolo, Stephen Bornemann, and Cathie Martin*

Department of Agriculture, Forest and Food Sciences, Plant Genetics and Breeding, University of Torino, 10095 Grugliasco, Italy (A.M., S.L., C.Co.); John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom (L.H., D.K., S.B., C.M.); and Department of Drug Science and Technology, University of Torino, 10125 Turin, Italy (C.Ca., P.R.)

ORCID ID: 0000-0002-3640-5080 (C.M.).

Tomato (*Solanum lycopersicum*), like other Solanaceous species, accumulates high levels of antioxidant caffeoylquinic acids, which are strong bioactive molecules and protect plants against biotic and abiotic stresses. Among these compounds, the monocaffeoylquinic acids (e.g. chlorogenic acid [CGA]) and the dicaffeoylquinic acids (diCQAs) have been found to possess marked antioxidative properties. Thus, they are of therapeutic interest both as phytonutrients in foods and as pharmaceuticals. Strategies to increase diCQA content in plants have been hampered by the modest understanding of their biosynthesis and whether the same pathway exists in different plant species. Incubation of CGA with crude extracts of tomato fruits led to the formation of two new products, which were identified by liquid chromatography-mass spectrometry as diCQAs. This chlorogenate:chlorogenate transferase activity was partially purified from ripe fruit. The final protein fraction resulted in 388-fold enrichment of activity and was subjected to trypsin digestion and mass spectrometric sequencing: a hydroxycinnamoyl-Coenzyme A:quinate hydroxycinnamoyl transferase (HQT) was selected as a candidate protein. Assay of recombinant HQT protein expressed in *Escherichia coli* confirmed its ability to synthesize diCQAs in vitro. This second activity (chlorogenate:chlorogenate transferase) of HQT had a low pH optimum and a high K_m for its substrate, CGA. High concentrations of CGA and relatively low pH occur in the vacuoles of plant cells. Transient assays demonstrated that tomato HQT localizes to the vacuole as well as to the cytoplasm of plant cells, supporting the idea that in this species, the enzyme catalyzes different reactions in two subcellular compartments.

The importance of plant-based foods in preventing or reducing the risk of chronic disease has been widely demonstrated (Martin et al., 2011, 2013). In addition to vitamins, a large number of other nutrients in plant-based foods promote health and reduce the risk of chronic diseases; these are often referred to as phytonutrients. The presence of phytonutrients in fruit and vegetables is of significant nutritional and therapeutic importance, as many have been found to possess strong antioxidant activity (Rice-Evans et al., 1997). Phenolics are the most widespread dietary antioxidants and caffeoylquinic acids, such as chlorogenic acid (CGA), dicaffeoylquinic acids (diCQAs), and tricaffeoylquinic acids (triCQAs), play important roles in promoting health (Clifford, 1999; Niggeweg et al., 2004). CGA limits low density lipid oxidation (Meyer et al., 1998), diCQAs possess anti-hepatotoxic activity (Choi et al., 2005), and triCQAs reduce the blood Glc levels of diabetic rats (Islam, 2006). diCQA derivatives have been shown to protect humans from various kinds of diseases; diCQAs suppress melanogenesis effectively (Kaul and Khanduja, 1998), show anti-inflammatory activity in vitro (Peluso et al., 1995), and exhibit a selective inhibition of HIV replication (McDougall et al., 1998). The physiological effects of caffeoylquinic acid derivatives with multiple caffeoyl groups are generally greater than those of monocaffeoylquinic acids, perhaps because the antioxidant activity is largely determined by the number of hydroxyl groups present on the aromatic rings (Wang et al., 2003; Islam, 2006). Furthermore, both diCQAs and triCQAs may function as inhibitors of the activity of HIV integrase, which catalyzes the insertion of viral DNA into the genome of host cells (McDougall et al., 1998; Slanina et al., 2001; Gu et al., 2007).

CGA is the major soluble phenolic in Solanaceous crops (Clifford, 1999) and the major antioxidant in the average U.S. diet (Luo et al., 2008), while different isomers of diCQAs have been identified in many crops

[1] This work was supported by a postdoctoral fellowship from the Piedmont Region of Italy (to A.M.), the core strategic grant of the Biological and Biotechnological Science Research Council to the John Innes Centre (to C.M.), and the Institute Strategic Program Understanding and Exploiting Plant and Microbial Secondary Metabolism from the Biological and Biotechnological Science Research Council (grant no. BB/J004596/1 to C.M., S.B., and D.K.).

* Address correspondence to cathie.martin@jic.ac.uk.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Cathie Martin (cathie.martin@jic.ac.uk).

[W] The online version of this article contains Web-only data.

[OPEN] Articles can be viewed online without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.114.251371

such as coffee (*Coffea canephora*), globe artichoke (*Cynara cardunculus*), tomato (*Solanum lycopersicum*), lettuce (*Lactuca sativa*), and sweet potato (*Ipomoea batatas*; Clifford, 1999; Islam, 2006; Moco et al., 2006, 2007; Moglia et al., 2008). In tomato, CGA accounts for 75% and 35% of the total phenolics in mature green and ripe fruit, respectively, amounting to 2 to 40 mg 100 g^{-1} dry weight (DW), although levels decline after ripening and during postharvest storage (Slimestad and Verheul, 2009). diCQAs and triCQAs also accumulate in tomato fruit (diCQAs, approximately 2 mg 100 g^{-1} DW; and triCQAs, 1–2 mg 100 g^{-1} DW; Chanforan et al., 2012).

Three pathways (Villegas and Kojima, 1986; Hoffmann et al., 2003; Niggeweg et al., 2004) have been proposed for the synthesis of CGA: (1) the direct pathway involving caffeoyl-CoA transesterification with quinic acid by hydroxycinnamoyl-Coenzyme A:quinate hydroxycinnamoyl transferase (HQT; Niggeweg et al., 2004; Comino et al., 2009; Menin et al., 2010; Sonnante et al., 2010); (2) the route by which *p*-coumaroyl-CoA is first transesterified with quinic acid via hydroxycinnamoyl-Coenzyme A transferase (HCT) acyltransferase (Hoffmann et al., 2003; Comino et al., 2007), followed by the hydroxylation of *p*-coumaroyl quinate to 5-caffeoylquinic acid, catalyzed by C3′H (*p*-coumaroyl-3-hydroxylase; Schoch et al., 2001; Mahesh et al., 2007; Moglia et al., 2009); and (3) the use of caffeoyl-glucoside as the acyl-donor (Villegas and Kojima, 1986). In tomato, the synthesis of CGA involves transesterification of caffeoyl-CoA with quinic acid by HQT (Niggeweg et al., 2004).

To date, it is not clear whether diCQAs are derived directly from the monocaffeoylquinic acids (such as CGA) through a second acyltransferase reaction involving an acyl-CoA or not, although their structural similarity provides good a priori evidence supporting this hypothesis. Recently the in vitro synthesis of 3,5-diCQA from CGA and CoA by HCT from coffee has been reported (Lallemand et al., 2012). By contrast, in sweet potato, an enzyme that catalyzes the transfer of the caffeoyl moiety of CGA to another molecule of CGA, leading to the synthesis of isochlorogenate (3,5-di-*O*-caffeoylquinate), has been described, but the corresponding gene has not been identified (Villegas and Kojima, 1986).

We report a chlorogenate:chlorogenate transferase (CCT) activity leading to the synthesis of diCQAs in tomato fruits and describe how alternative catalysis, by a single enzyme, leads to the production of both CGA and diCQA in different cellular compartments.

RESULTS

Identification of an Enzyme Activity Involved in diCQA Synthesis in Tomato

We incubated crude extracts of tomato fruit with CGA and detected the formation of three new products, absent in the negative control reactions (Fig. 1). The products were identified by liquid chromatography (LC)-photodiode array detector (PDA)-tandem mass spectrometry (MS/MS) analysis (Supplemental Fig. S1) comparing their UV profiles and mass fragmentation patterns to those obtained from the analysis of diCQA standard mixture (Fig. 2; Supplemental Fig. S2). The three new peaks (retention times: 9.60, 9.75, and 10.40 min) were identified as isomers of diCQAs; their precursor ions (mass-to-charge ratio [m/z] 515 and 517 in negative electrospray ionization mode [ESI^-] and positive electrospray ionization mode [ESI^+], respectively) corresponded to that of diCQA standard (Fig. 2, C and D), while in ESI^-, the observed fragment ions of m/z 179 and 135 matched well with the presence of a caffeoyl moiety in the structure (Fig. 2, E and F) and the fragment ions at m/z 191 and 173 are indicative of a quinic acid moiety in the molecule.

The three isomers were identified respectively as 1,5-diCQA, 3,5-diCQA, and 4,5-diCQA by comparison with authentic standards. This showed that the synthesis of dicaffeoyl quinates in tomato occurs through CCT activity. When tomato crude extracts were tested in the presence of CGA and caffeoyl-CoA (crude plus caffeoyl-CoA; Fig. 1) for acyl transferase activity, no increase in diCQA content was detected.

Optimization of the Enzyme Assay

The effect of pH on the catalytic activity of the enzyme was tested from pH 3 to 7. CCT activity was highest at

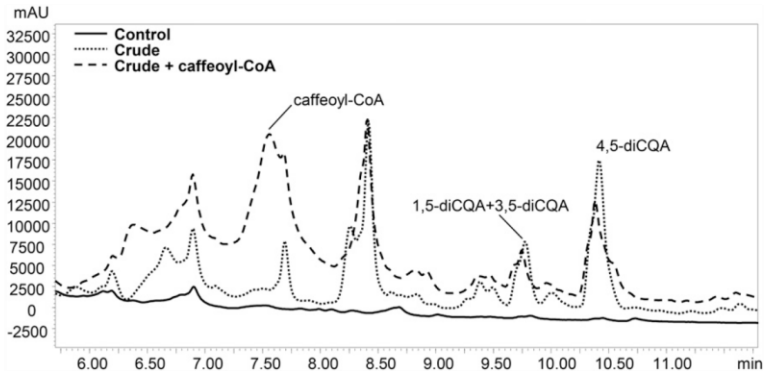


Figure 1. Identification of CCT enzyme activity giving synthesis of diCQAs in tomato. LC-PDA (at 325 nm) profiles of CCT activity with inactive (boiled) extract of ripe tomatoes (Control) and crude extract of ripe tomatoes (Crude). Incubations with crude enzyme extract and CGA were performed with or without caffeoyl-CoA. mAU, Milliabsorbance units.

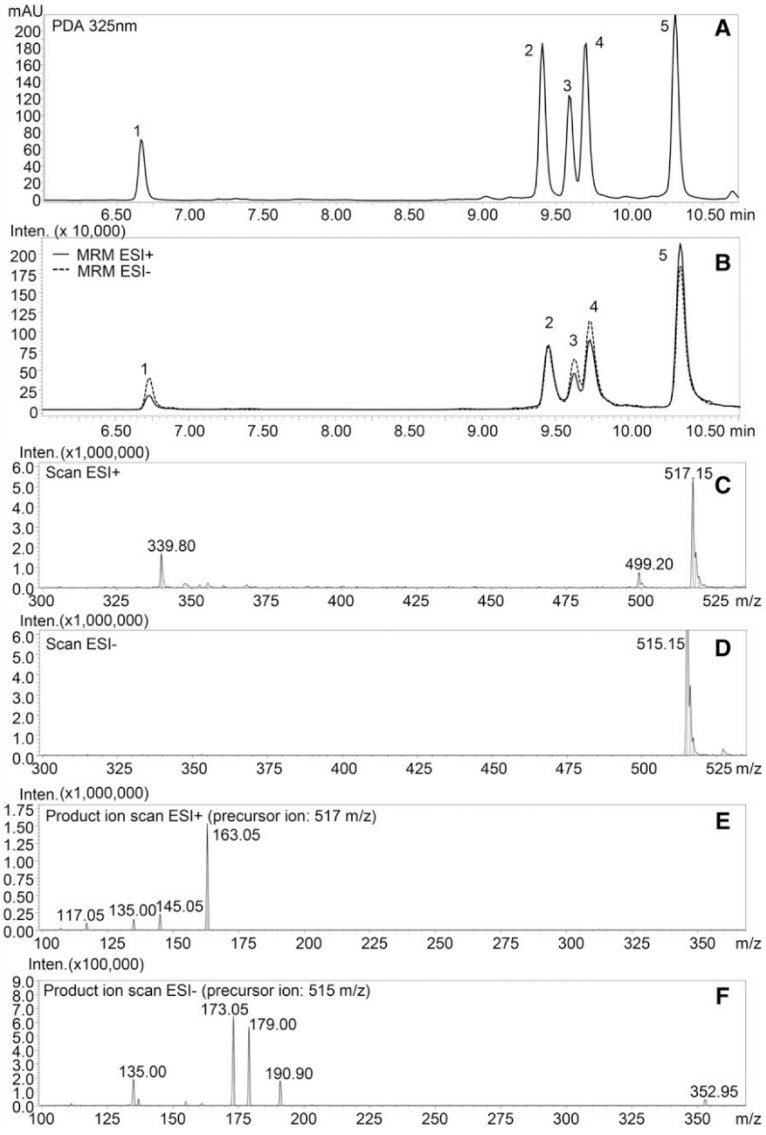


Figure 2. LC-PDA-MS/MS profile of diCQA standard mixture. A, PDA profile at 325 nm. B, MS profile obtained in MRM acquisition mode (for transitions, see "Materials and Methods"). C, ESI^+ full-scan spectra of 4,5-diCQA. D, ESI^- full-scan spectra of 4,5-diCQA. E, ESI^+ product ion scan of 4,5-diCQA (precursor ion, 517 *m/z*). F, ESI^- product ion scan of 4,5-diCQA (precursor ion, 515 *m/z*). For peak numbers, 1 indicates 1,3-diCQA, 2 indicates 3,4-diCQA, 3 indicates 1,5-diCQA, 4 indicates 3,5-diCQA, and 5 indicates 4,5-diCQA. mAU, Milliabsorbance units; Inten., ion intensity.

pH 4.0 (0.043 nkat mg^{-1} protein) in sodium phosphate buffer (Fig. 3A) without any cofactor requirements, although addition of $MgCl_2$ (2.5 mM) and dithiothreitol (DTT; 1 mM) increased activity slightly. The reaction showed typical Michaelis-Menten kinetics with increasing CGA concentrations and gave a K_m for CGA of 16 ± 3 mM (Fig. 3B). The activity was irreversible because when crude extracts were incubated in the presence of diCQA (3,5-, 4,5-, or 1,5-diCQA) and quinic acid, no substrate was consumed or product formed.

CCT activity and diCQA levels were determined at different stages during fruit ripening (green, breaker, turning, pink, and red). An increase in CCT activity was observed during ripening, reaching a maximum of 0.036 nkat mg^{-1} protein in the peel of red fruit (Fig. 4A). The levels of diCQAs also increased during ripening, reaching a maximum in fully ripe tomatoes (Fig. 4B). Peel tissue had slightly higher levels of CCT activity (0.034 ± 0.0034 nkat mg^{-1} total protein) than flesh tissue (0.027 ± 0.002 nkat mg^{-1} total protein) and correspondingly higher levels of diCQAs (Fig. 4C).

Purification of CCT Activity

Protein extracts from red fruit exhibiting high enzymatic activity were used for enzyme purification (Supplemental Table S1). A 45% to 60% ammonium sulfate-precipitated

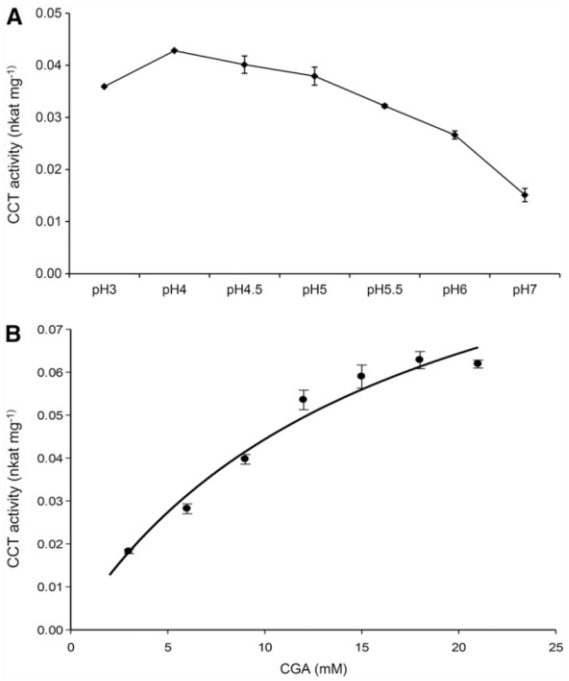


Figure 3. CCT activity in red tomato fruit. A, CCT activity under different pH conditions (3–7). B, CCT activity in crude extract of ripe tomato with increasing concentrations of CGA. Values on the *y* axis are nkat mg^{-1} of total protein. The curve represents the best fit of the Michaelis-Menten equation to the data using SigmaPlot (Systat Software). Error bars indicate SD calculated from three biological replicates.

protein fraction (21.7-fold purification) was separated by ion-exchange chromatography, resulting in a fraction with a specific activity of 1.34 nkat mg^{-1} protein, corresponding to a purification of 388-fold (Supplemental Table S1). This protein fraction was subjected to trypsin digestion and mass spectrometric sequencing. The list of peptide masses was aligned with predicted tryptic digests of known proteins by MASCOT (Supplemental Table S2). These peptide sequences were used to interrogate the EST database of tomato to search for potential candidates for CCT. Among the peptides obtained from mass spectrometric analysis, an HQT protein (Q70G32) was selected as a good candidate for further investigation as two unique peptides from the tryptic digest (IWSSNLDLIVGR and SALDYLELQPDLSTLIR) matched exactly the predicted peptide sequence of this gene.

CCT Activity of Enzyme Expressed in *Escherichia coli*

As previously described, the tomato BAHD acyl transferase HQT (SlHQT) can use quinic acid as an acceptor with caffeoyl-CoA and *p*-coumaroyl-CoA donors for the synthesis of CGA and coumaroyl quinate, respectively (Niggeweg et al., 2004). (The acyl-CoA-dependent BAHD superfamily is defined by the names of the first four acyltransferase enzymes of the family isolated from plant species; D'Auria, 2006.)

The crude enzyme extract of *E. coli* expressing SlHQT was incubated in the presence of high quantities of CGA, according to the optimum conditions determined for the CCT activity in crude extracts of tomato fruit. In the presence of SlHQT, we detected the formation of three new products, absent in control reactions. These were identified as 1,5-, 3,5-, and 4,5-diCQA acid on the basis of UV spectra, the deprotonated $[M-1]^-$ and protonated $[M+1]^+$ molecular ions and by comparison to authentic standards (Fig. 5, wild-type protein; Supplemental Fig. S3). When SlHQT was assayed in the reverse direction (using, alternatively, 1,5-, 3,5-, 4,5-diCQA acids and quinic acid as

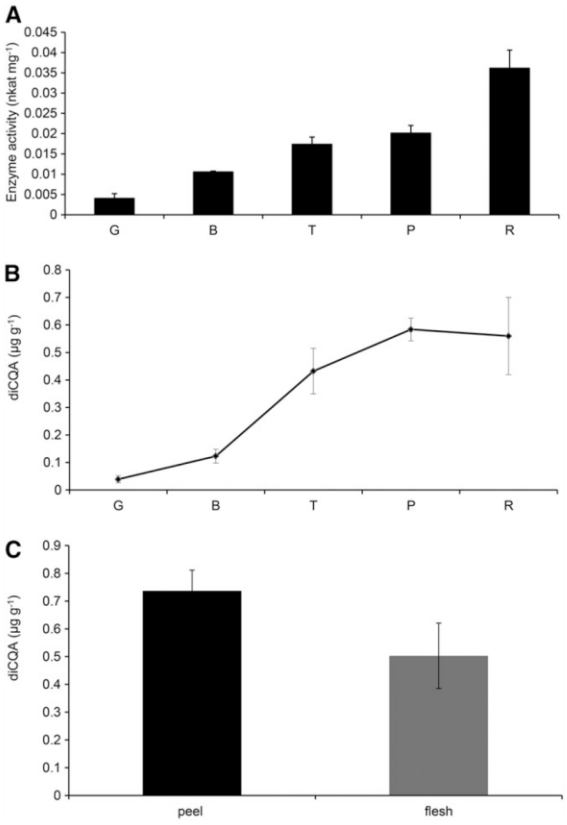


Figure 4. CCT activity and levels of diCQAs during ripening and in different tissues of tomato fruit. A, Changes in CCT activity during ripening of tomato. CCT activity was measured in tissue at different stages of fruit ripening and expressed as specific activity (nkat mg^{-1} of total protein). Error bars indicate SD calculated from three biological replicates. G, Green; B, breaker; T, turning; P, pink; R, red stage. B, Content ($\mu g\ g^{-1}$ fresh weight) of diCQAs during fruit ripening. Error bars indicate SD calculated from three biological replicates. C, Content ($\mu g\ g^{-1}$ fresh weight) of diCQAs in different tissues of red fruit. Error bars indicate SD calculated from three biological replicates.

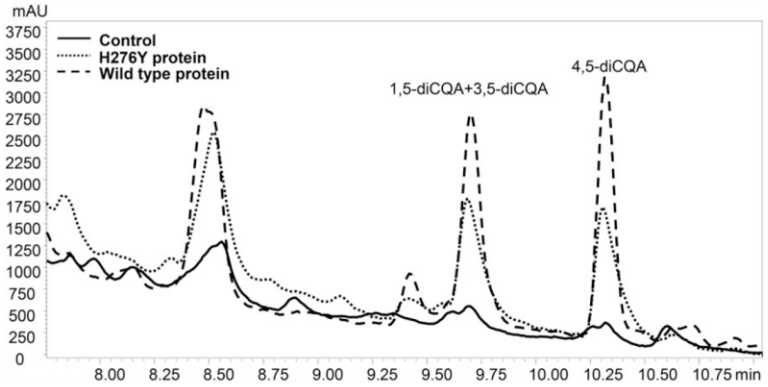


Figure 5. LC-PDA profiles (at 325 nm) of CCT activity with recombinant wild-type SlHQT enzyme, His-mutated SlHQT enzyme (H276Y), and inactive enzyme (Control). mAU, Milliabsorbance units.

substrates), no activity was detected. The affinity of SlHQT as CCT for CGA as an acceptor (K_m = 19 ± 2 mM) was very similar to that observed earlier in crude extracts of tomato (K_m = 16 ± 3 mM).

To gain insight into the involvement of SlHQT in diCQA synthesis, the levels of transcripts were measured by means of real time quantitative PCR, using ubiquitin to normalize expression. The levels of expression were 5.9-, 8.7-, 12.0-, and 24.5-fold higher in breaker, turning, pink, and red fruits, respectively, compared with green-stage tissue, indicating an increase of gene expression levels during ripening of tomato fruits, with a peak at the red ripe stage (Supplemental Fig. S4).

Subcellular Localization of SlHQT

The subcellular localization of SlHQT was investigated using two vectors expressing the SlHQT complementary DNA (cDNA) under the control of the 35S promoter, such that the SlHQT protein was fused at its C terminus to either enhanced yellow fluorescent protein (EYFP) or monomeric red fluorescent protein (mRFP). Plasmid DNA was bombarded into onion (*Allium cepa*) epidermal tissue, because these cells have no chloroplasts and fluorescence could be observed free from interference from chlorophyll (Fig. 6). As a control, a vector for expression of EYFP or mRFP alone, under the control of the 35S promoter, was also bombarded into onion epidermis.

Both SlHQT:EYFP- and EYFP-transformed cells showed clear cytoplasmic localization, indicated by strong fluorescence at the periphery of the cells and in cytoplasmic strands (Fig. 6, A–D). However, in contrast to 35S:EYFP, a faint fluorescence was also observed with SlHQT:EYFP in vacuolar regions, particularly in epidermal cells incubated for 96 h in the dark, implying that there might be low levels of SlHQT in vacuoles as well as in the cytoplasm (Fig. 6, B and D). Because mRFP is particularly stable at low pH (Hunter et al., 2007), transient assays were repeated with 35S:SlHQT:mRFP compared with 35S:mRFP. Again, the predominant fluorescence was cytoplasmic, as evidenced by the strong fluorescence from the zone delimiting the inner edge of the cells and the occasional cytoplasmic strands. The nucleus also fluoresced strongly, similar to 35S:mRFP. However, fainter fluorescence from SlHQT:mRFP was clearly visible in the vacuolar region of the cells, particularly in samples incubated for 96 h in the dark (Fig. 6, E and F), compared with cells bombarded with 35S:mRFP, where the fluorescence remained cytoplasmic (Fig. 6, G and H). Finally, cells were cobombarded with 35S:SlHQT:mRFP (Fig. 6I) and 35S:EYFP (Fig. 6J). Cells expressing both plasmids showed that SlHQT:mRFP was in the vacuole while EYFP was completely cytoplasmic (Fig. 6K; Supplemental Fig. S5). This demonstrated that SlHQT is localized in vacuoles as well as in the cytoplasm of plant cells.

Structural Features Determining the Duel Catalytic Activity of SlHQT/CCT

SlHQT in tomato appears to be capable of using two types of acyl donor. It uses caffeoyl-CoA or *p*-coumaroyl-CoA as acyl donors and quinate as the acyl acceptor in the synthesis of CGA (Niggeweg et al., 2004), but it is also able to use CGA as both acceptor and acyl donor in the synthesis of diCQAs (Fig. 7A).

The 3,5-diCQA isomer could be made in one step by a CCT, but the 4,5 isomer would likely require an additional spontaneous acyl migration from the 3 to the 4 position (Fig. 7A), as suggested by Lallemand et al. (2012). We generated a homology model of HQT with Swiss model (Guex and Peitsch, 1997) using the 4KEC (a ternary complex form) structure of sorghum (*Sorghum bicolor*) HCT (SbHCT) bound to *p*-coumaroyl-shikimate described by Walker et al. (2013). The Arg-371 residue of SbHCT approaches the shikimate ring from the same face as the transferred acyl group and forms an H-bonding interaction with the carboxyl group of shikimate with a distance of 2.7 Å (Fig. 7B). In SlHQT, the equivalent Arg-352 would be expected to form a similar interaction with quinate, except that its carboxyl group is out of the plane of the ring unlike with shikimate. On the other face of the acceptor substrate and opposite the Arg residues, His-276 in SlHQT is equivalent to Tyr-296 in SbHCT. For the CCT reaction to occur, CGA would have to bind with its ring in

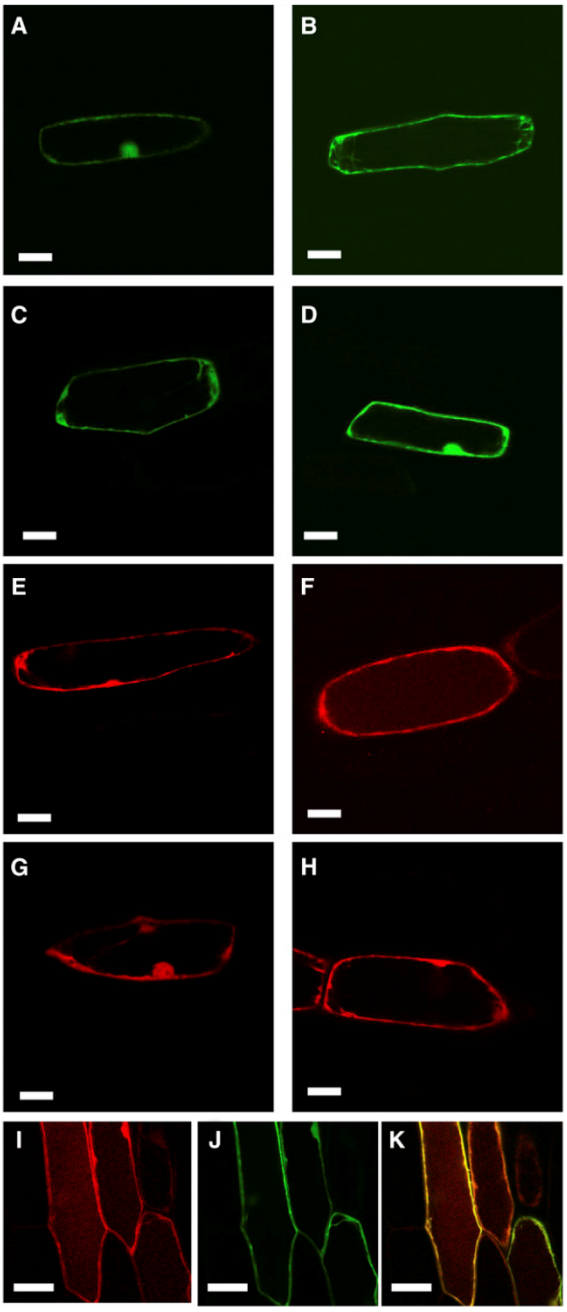


Figure 6. Subcellular localization of SlHQT. Onion epidermis was bombarded with 35S:SlHQT fused to EYFP or mRFP with 35S:EYFP or 35S:mRFP as controls for cytoplasmic localization. A, Fluorescence from SlHQT:EYFP after 24 h in the dark. B, Fluorescence from SlHQT:EYFP after 96 h in the dark. Fluorescence in the vacuole is faintly visible. C, Fluorescence from EYFP after 24 h in the dark. D, Fluorescence from EYFP after 96 h in the dark. E, Fluorescence from SlHQT:mRFP after 24 h in the dark. F, Fluorescence from SlHQT:mRFP after 96 h in the dark. Fluorescence in the vacuole is clearly visible. G, Fluorescence from mRFP after 24 h in the dark. H, Fluorescence from mRFP after 96 h in the dark. Fluorescence remains cytoplasmic. I, Fluorescence from SlHQT:mRFP cobombarded with EYFP after 96 h in the dark. J, Fluorescence from EYFP from the same cell. K, Merged image showing fluorescence from SlHQT:mRFP in the vacuole as well as the cytoplasm, while the EYFP remains entirely cytoplasmic. Bars = 50 μm.

the opposite orientation such that its carboxyl group would more likely engage with His-276 of SlHQT rather than Arg-352. The low pH optimum of the CCT reaction implies that His-276 would be positively charged and able to form a favorable interaction with CGA in this orientation, only at low pH. The absence of a potentially positively charged residue at this location in its active site would suggest that SbHCT does not have CCT activity.

To examine the role of this specific amino acid on the enzymatic activity of SlHQT, we selectively changed His-276 to Tyr-276. The activity of the H276Y mutant was compared with that of the unmutated enzyme in the CCT reaction. The SlHQT H276Y mutant showed significantly lower production of diCQA in the CCT reaction when compared with wild-type SlHQT protein (58% of wild-type activity; 0.0021 $\pm$ 0.0001 nkat mg^{-1} compared with 0.0036 $\pm$ 0.0001 nkat mg^{-1}, $P < 0.01$ Student's t test; Fig. 5; Supplemental Fig. S3). The H276Y mutation barely affected the activity of the enzyme as HQT in CGA synthesis, as monitored by activity with caffeoyl-CoA and quinic acid assayed using the spectrophotometric method (Niggeweg et al., 2004) on the same extracts (92% of wild-type activity; 11.16 $\pm$ 0.3 nkat mg^{-1} for H276Y protein compared with 12.16 $\pm$ 0.4 nkat mg^{-1} for wild-type SlHQT).

Alignment of sequences of enzymes shown to encode HCT from different plants species indicated that in all HCT proteins, the residue equivalent to Tyr-296 in SbHCT was also Tyr (Supplemental Fig. S6). Alignment of sequences of enzymes with HQT activity from Solanaceous species (tomato, tobacco (*Nicotiana tabacum*), and potato (*Solanum tuberosum*); Niggeweg et al., 2004), coffee (Lepelley et al., 2007), and globe artichoke (Comino et al., 2009; Menin et al., 2010; Sonnante et al., 2010) showed that in HQT from Solanaceous species, the residue equivalent to Tyr-296 was His (His-276 in SlHQT), but in HQT from other species, this residue was Tyr (Fig. 8). We tested whether the HQT from globe artichoke (possessing Tyr-276 instead of His-276) had CCT activity following expression in *E. coli*, but only traces of diCQAs were detected following incubation with CGA at pH 5 (Supplemental Fig. S7). These data confirmed our structural analysis, suggesting that the His at position 276 in SlHQT plays a key role in CCT activity and also suggesting that although CCT activity is likely responsible for diCQA formation in Solanaceous species, diCQAs may form by other routes in other plant species.

DISCUSSION

Assays of crude extracts of tomato confirmed an early report of the synthesis of isochlorogenic acid (3,5-diCQA)

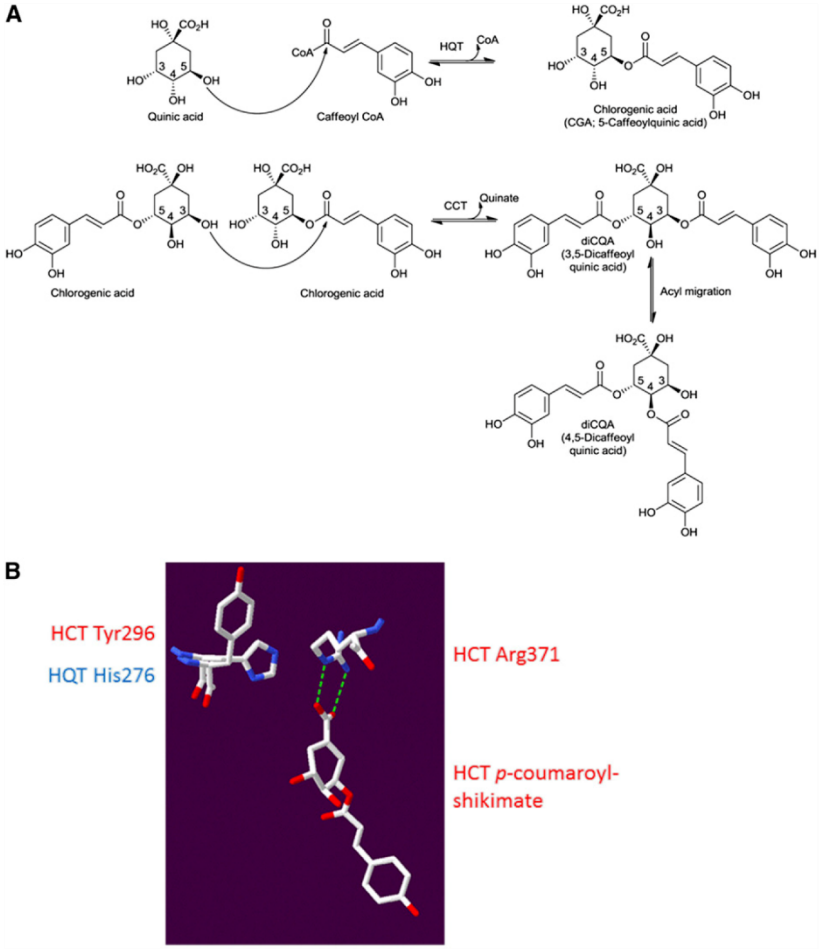


Figure 7. Structural analysis of SlHQT modeled on sorghum HCT (Protein Data Bank code 4KEC; Walker et al., 2013). A, Proposed reaction scheme for diCQA synthesis by CCT using CGA as the acyl donor as well as the acyl acceptor compared with the BAHD activity of HQT and HCT using caffeoyl-CoA as the acyl donor and quinate as the acyl acceptor, respectively. B, Structure of sorghum HCT (Protein Data Bank code 4KEC; Walker et al., 2013) with a superposed homology model of SlHQT based upon this structure. CGA would have to bind to SlHQT with its carboxyl group facing away from the Arg residue, allowing a favorable electrostatic interaction between His-276 and CGA that would promote CCT activity.

in sweet potato through an enzyme reaction in which CGA acts both as acyl donor and acceptor (Villegas and Kojima, 1986; Fig. 7A). These data suggest the presence, in tomato, of a CCT, which catalyzes the formation of diCQAs using CGA as acyl donor. The increase in CCT activity correlated well with increases in levels of diCQAs with advancing fruit ripening (Fig. 4A). As previously observed by Moco et al. (2007), we detected increases in diCQA isomers in tomato epidermis upon ripening; their increase was lower in flesh (Fig. 4C), probably as a consequence of a lower CCT activity.

Partial purification of CCT activity and identification of proteins in the enriched fraction suggested HQT (an enzyme already known to bind CGA) as a candidate for CCT activity. In most cases, acyltransferases involved in hydroxycinnamate synthesis use CoA thioesters (for example, caffeoyl-CoA) as acyl donors (D'Auria, 2006). There is one reported case in which CGA is used as an acyl donor in tomato: chlorogenate:glucarate caffeoyltransferase, a lipase-like protein that catalyzes the transfer of the caffeoyl moiety of CGA to the acceptor molecule glucarate (Strack and Gross, 1990; Teutschbein et al., 2010).

Tomato CCT activity is optimal at pH 4 (Fig. 3A) and seems likely to be biologically relevant only when the HQT enzyme is located within the vacuole of tomato cells, where an acid pH prevails. Vacuolar localization of CCT activity would favor diCQA biosynthesis, because high vacuolar concentrations of CGA could promote diCQA biosynthesis despite the relatively low affinity of CCT for CGA. In the cytoplasm, SlHQT could act as a BAHD acyl transferase at pH 6 to 7 in the presence of aromatic acyl-CoA donors (K_m, 0.3 ± 0.1 mM) with a high affinity for CGA (K_m, 0.05 mM) and V_{max} of 119 and 58 nkat mg^{-1} protein (forward and reverse reactions, respectively; Niggeweg et al., 2004). By contrast, it could act as a CCT in the vacuole at pH 4 to 5 in the absence of aromatic acyl-CoA donors and presence of relatively high concentrations of CGA allowing activity despite a low affinity for CGA (K_m, 16 ± 3 mM) and a V_{max} of 0.12 ± 0.01 nkat mg^{-1} protein. The low V_{max} of the CCT activity could, in other circumstances, limit the biological significance of this

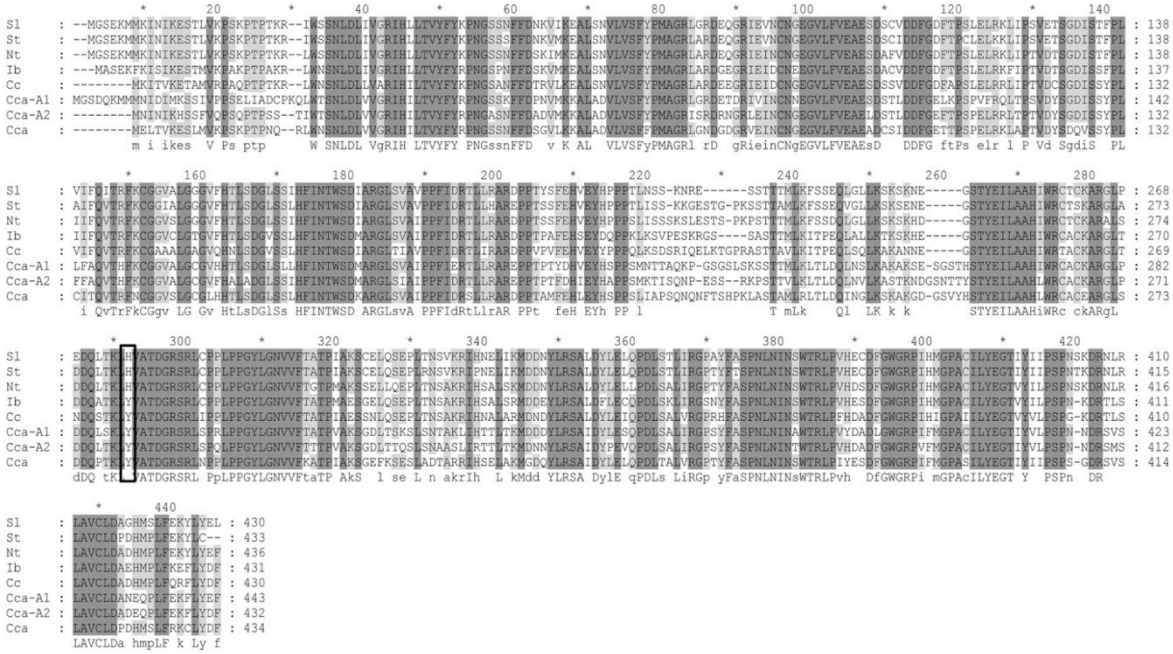


Figure 8. Sequence alignment of sequences belonging to the plant HQT family. NP_001234850 from tomato (Sl), ABA46756 from potato (St), CAE46932 from tobacco (Nt), BAA87043 from sweet potato, ABO77957 from coffee (Cc), ABK79689 from globe artichoke (Cca), ADL62854 from globe artichoke (Cca-A1), and ADL62855 from globe artichoke (Cca-A2). Black box indicates His-276 in SlHQT and equivalent residues in HQT enzymes from other plant species.

activity, but in the context of fruit ripening, over a period of days, this level of activity is sufficient. Thus, SlHQT likely performs two distinct functions in the very different cytoplasmic and vacuolar subcellular compartments. SlHQT was consistently localized in the vacuole lumen as well as in the cytoplasm when monitored by C-terminal fluorescent tags in transient assays (Fig. 6). This dual localization was distinct from mRFP or EYFP expressed without the fusion protein, which were exclusively cytoplasmically localized. We suggest that once significant amounts of CGA have been synthesized in the cytoplasm and transported to the vacuole for storage, tomato HQT moonlights by participating in diCQA synthesis in the vacuole.

The BAHD activity of SlHQT is freely reversible (Niggeweg et al., 2004), whereas that of CCT is irreversible. Consequently, CGA can accumulate only if it is sequestered in the vacuoles of cells. There, in high concentrations, at low pH, and in the absence of acyl-CoA donors, SlHQT can form diCQAs through its CCT activity.

This possibility is supported by the close association between levels of CGA and levels of diCQAs in tomato under different cultivation conditions (Vallverdú-Queralt et al., 2012) and between HQT expression levels and CCT activity observed during fruit ripening (Fig. 4; Supplemental Fig. S4). Although our results indicate that HQT is localized in vacuoles as well as in the cytoplasm, we do not yet know whether transport to the vacuole involves an active transport mechanism. The fact that SlHQT is present in the vacuole gives significance to the CCT activity despite its high K_m and low pH optimum.

In coffee, the synthesis of diCQAs has been proposed to occur via an uncharacterized activity of HCT on CGA and CoA (Lallemand et al., 2012), while in tomato, diCQAs appear to be synthesized by a different route involving the CCT activity of SlHQT.

Site-directed mutagenesis of SlHQT confirmed the role of His-276 in the CCT reaction (Fig. 5). Interestingly, HQT from globe artichoke, which has a Tyr residue rather than a His residue at the position equivalent to His-276 in SlHQT, had very low CCT activity when assayed under optimized conditions in vitro. This suggests that the ability to moonlight in the production of diCQAs may be limited to HQT enzymes from Solanaceous species that carry a His residue at an appropriate position for interaction with the carboxyl group of the quinate moiety of the CGA donor molecule, under conditions of relatively low pH (<5). Presumably, species lacking HQT or HCT enzymes with His in this position may use different mechanisms to form diCQAs, and alternative routes may have arisen by convergent evolution in different families of plants.

Our results broaden the view of one enzyme-one activity and suggest that changing subcellular environments may diversify the range of products of secondary metabolism. This concept could operate in acyl decoration of

The identification of the components was based on their UV spectra and mass spectral information in Multiple Reaction Monitoring (MRM) mode in both positive and negative ionization mode (respectively, ESI^+ and ESI^-). MS operative conditions were heat block temperature, 400°C; nebulizing gas (nitrogen) flow rate, 3 min^{-1}; drying gas (nitrogen) flow rate, 15 min^{-1}; and desolvation line temperature, 250°C. Collision gas was argon (230 kPa). Transitions monitored included ESI^+, m/z 517 → 163, m/z 517 → 145, and m/z 517 → 135 (dwell time, 20 ms; collision energy, –35 V; and event time, 0.096 s), and ESI^-, m/z 515 → 179, m/z 515 → 191, and m/z 515 → 135 (dwell time, 20 ms; collision energy, 35 V; and event time, 0.096 s). The MRM transitions were selected on the basis of the fragments obtained by analyzing the diCQA standards in full-scan mode in both ESI^+ and ESI^- in the range of 300 to 1,200 m/z, with a scan speed of 1,000 D s^{-1}, and then in product ion scan mode in both ESI^+ and ESI^- in the range of 100 to 550 m/z, with a scan speed of 1,000 D s^{-1} and using, as precursor ions, 517 m/z protonated molecular ion ($[M+H]^+$) for ESI^+ and 515.00 m/z deprotonated molecular ion ($[M–H]^-$) for ESI^-.

The quantification of diCQAs was undertaken on the PDA profiles (at 325 nm) using the external calibration method. A three-point calibration curve was built analyzing, in triplicate, the pure standard of 4,5-diCQA in the range of 100 to 500 ppb. The determination coefficient was 0.9951.

Subcellular Localization of SlHQT

Tomato *HQT* was amplified from cDNA using primers for Gateway recombination (HQTF, 5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGGAAGTGAAA-AAATGATGAAAATTAATATC-3′; and HQTR, GGGGACCACTTTGTACAAG-AAAGCTGGGTATAATTCATATAAATATTTTTCAAATAG-3′) and recombined into entry vector pDONR207. The resulting clone was recombined into pB7YFWG2.0 for EYFP fusions and pB7RWG2.0 for mRFP fusions (Karimi et al., 2002). To construct the 35S:EYFP and 35S:mRFP control vectors, pB7YWG2.0 and pB7RWG2.0 were cut with *Eco*RV and *Spe*I (to remove the Gateway destination cassette), filled in with the Klenow fragment of DNA polymerase, and self-ligated to yield 35S:EYFP and 35S:mRFP, respectively. Plasmid DNA was prepared using a Qiagen mini-prep kit. Bombardment was conducted using a particle inflow helium gun based on the design of Vain et al. (1993). Onion (*Allium cepa*) epidermal tissue was prepared freshly from large onions and rinsed in sterilized and distilled water. Plasmid DNA (5 μg) was precipitated onto 2-mg gold particles (1.0-μm diameter) through the addition of 50 μL of 2.5 M CaCl2 and 20 μL of 100 mM spermidine. After precipitation, 90 μL of supernatant was discarded, and particles were washed twice in ethanol and resuspended in 50 μL of ethanol. Gold particles were prepared immediately before use. For bombardment, tissue was placed on Murashige and Skoog (1962) medium plus 8% (w/v) agar (MS medium) in a petri dish, within the desiccator in the gun range of 120 to 160 mm. Tissue was bombarded with 7 μL of gold suspension using a 50-ms burst of helium at a pressure of 7,580 kPa within a vacuum of −98 kPa. After bombardment, tissue was incubated on MS medium at 20°C in the dark for 24 or 96 h. Bombarded tissue was imaged using a Zeiss 510 Meta Confocal microscope. For EYFP, a 488-nm laser line was used, and emission at 500 to 550 nm was imaged. For mRFP, a 561-nm laser line was used, and emission at 570 to 615 nm was imaged using a ×25, 0.7 multiimmersion objective. Images were processed using Fiji (http://fiji.sc/wiki/index.php/Fiji).

Bioinformatic Analysis of Tomato HQT

Homology model of SlHQT was performed with Swiss model (Guex and Peitsch, 1997) using the 4KEC structure of SbHCT bound to *p*-coumaroyl-shikimate described by Walker et al. (2013). Alignment of sequences of enzymes shown to encode HCT and HQT from different plants species was performed by means of ClustalW set with standard parameters.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. CCT activity in crude extracts of tomato.

Supplemental Figure S2. MS/MS profiles of diCQA standards.

Supplemental Figure S3. CCT activity of recombinant wild-type SlHQT enzyme and H276Y-mutated SlHQT enzyme.

Supplemental Figure S4. Expression of *SlHQT* in tomato.

Supplemental Figure S5. Subcellular localization of SlHQT.

Supplemental Figure S6. Sequence alignment of sequences belonging to the plant hydroxycinnamoyl CoA shikimate/HCT family.

Supplemental Figure S7. CCT activity of wild-type SlHQT enzyme and globe artichoke HQT.

Supplemental Table S1. Purification scheme for chlorogenate:chlorogenate transferase activity from tomato.

Supplemental Table S2. List of peptide masses present in purified enzyme fraction.

ACKNOWLEDGMENTS

We thank Gerhard Saalbach (John Innes Centre) for technical assistance with protein MS, Silke Robatzek (Sainsbury Laboratory) and Grant Calder (John Innes Centre) for invaluable advice and help with subcellular imaging, and Anna Maria Milani (University of Torino) for excellent technical assistance.

Received October 1, 2014; accepted October 7, 2014; published October 9, 2014.

LITERATURE CITED

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem **72:** 248–254

Chanforan C, Loonis M, Mora N, Caris-Veyrat C, Dufour C (2012) The impact of industrial processing on health-beneficial tomato microconstituents. Food Chem **134:** 1786–1795

Choi J, Park JK, Lee KT, Park KK, Kim WB, Lee JH, Jung HJ, Park HJ (2005) In vivo antihepatotoxic effects of *Ligularia fischeri* var. spiciformis and the identification of the active component, 3,4-dicaffeoylquinic acid. J Med Food **8:** 348–352

Clifford M (1999) Chlorogenic acids and other cinnamates: nature, occurrence and dietary burden. J Sci Food Agric **79:** 362–372

Comino C, Hehn A, Moglia A, Menin B, Bourgaud F, Lanteri S, Portis E (2009) The isolation and mapping of a novel hydroxycinnamoyltransferase in the globe artichoke chlorogenic acid pathway. BMC Plant Biol **9:** 30

Comino C, Lanteri S, Portis E, Acquadro A, Romani A, Hehn A, Larbat R, Bourgaud F (2007) Isolation and functional characterization of a cDNA coding a hydroxycinnamoyltransferase involved in phenylpropanoid biosynthesis in *Cynara cardunculus* L. BMC Plant Biol **7:** 14

D'Auria JC (2006) Acyltransferases in plants: a good time to be BAHD. Curr Opin Plant Biol **9:** 331–340

Gu R, Dou G, Wang J, Dong J, Meng Z (2007) Simultaneous determination of 1,5-dicaffeoylquinic acid and its active metabolites in human plasma by liquid chromatography-tandem mass spectrometry for pharmacokinetic studies. J Chromatogr B Analyt Technol Biomed Life Sci **852:** 85–91

Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis **18:** 2714–2723

Hoffmann L, Maury S, Martz F, Geoffroy P, Legrand M (2003) Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. J Biol Chem **278:** 95–103

Hunter PR, Craddock CP, Di Benedetto S, Roberts LM, Frigerio L (2007) Fluorescent reporter proteins for the tonoplast and the vacuolar lumen identify a single vacuolar compartment in Arabidopsis cells. Plant Physiol **145:** 1371–1382

Islam S (2006) Sweetpotato (*Ipomoea batatas* L.) leaf: its potential effect on human health and nutrition. J Food Sci **71:** R13–R21

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

Kaul A, Khanduja KL (1998) Polyphenols inhibit promotional phase of tumorigenesis: relevance of superoxide radicals. Nutr Cancer **32:** 81–85

Lallemand LA, Zubieta C, Lee SG, Wang Y, Acajjaoui S, Timmins J, McSweeney S, Jez JM, McCarthy JG, McCarthy AA (2012) A structural basis for the biosynthesis of the major chlorogenic acids found in coffee. Plant Physiol **160:** 249–260

Lepelley M, Cheminade G, Tremillon N, Simkin A, Caillet V, McCarthy J (2007) Chlorogenic acid synthesis in coffee: an analysis of CGA content

and real-time RT-PCR expression of HCT, HQT, C3H1, and CCoAOMT1 genes during grain development in *C. canephora*. Plant Sci **172:** 978–996

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

Mahesh V, Million-Rousseau R, Ullmann P, Chabrillange N, Bustamante J, Mondolot L, Morant M, Noirot M, Hamon S, de Kochko A, et al (2007) Functional characterization of two *p*-coumaroyl ester 3′-hydroxylase genes from coffee tree: evidence of a candidate for chlorogenic acid biosynthesis. Plant Mol Biol **64:** 145–159

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

Martin C, Zhang Y, Tonelli C, Petroni K (2013) Plants, diet, and health. Annu Rev Plant Biol **64:** 19–46

McDougall B, King PJ, Wu BW, Hostomsky Z, Reinecke MG, Robinson WE Jr (1998) Dicaffeoylquinic and dicaffeoyltartaric acids are selective inhibitors of human immunodeficiency virus type 1 integrase. Antimicrob Agents Chemother **42:** 140–146

Menin B, Comino C, Moglia A, Dolzhenko Y, Portis E, Lanteri S (2010) Identification and mapping of genes related to caffeoylquinic acid synthesis in *Cynara cardunculus* L. Plant Sci **179:** 338–347

Meyer A, Donovan J, Pearson D, Waterhouse A, Frankel E (1998) Fruit hydroxycinnamic acids inhibit human low-density lipoprotein oxidation in vitro. J Agric Food Chem **46:** 1783–1787

Moco S, Bino RJ, Vorst O, Verhoeven HA, de Groot J, van Beek TA, Vervoort J, de Vos CH (2006) A liquid chromatography-mass spectrometry-based metabolome database for tomato. Plant Physiol **141:** 1205–1218

Moco S, Capanoglu E, Tikunov Y, Bino RJ, Boyacioglu D, Hall RD, Vervoort J, De Vos RC (2007) Tissue specialization at the metabolite level is perceived during the development of tomato fruit. J Exp Bot **58:** 4131–4146

Moglia A, Comino C, Portis E, Acquadro A, De Vos RC, Beekwilder J, Lanteri S (2009) Isolation and mapping of a C3′H gene (CYP98A49) from globe artichoke, and its expression upon UV-C stress. Plant Cell Rep **28:** 963–974

Moglia A, Lanteri S, Comino C, Acquadro A, de Vos R, Beekwilder J (2008) Stress-induced biosynthesis of dicaffeoylquinic acids in globe artichoke. J Agric Food Chem **56:** 8641–8649

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue culture. Physiol Plant **15:** 473–478

Niggeweg R, Michael AJ, Martin C (2004) Engineering plants with increased levels of the antioxidant chlorogenic acid. Nat Biotechnol **22:** 746–754

Peluso G, De Feo V, De Simone F, Bresciano E, Vuotto ML (1995) Studies on the inhibitory effects of caffeoylquinic acids on monocyte migration and superoxide ion production. J Nat Prod **58:** 639–646

Rice-Evans C, Miller NJ, Paganga G (1997) Antioxidant properties of phenolic compounds. Trends Plant Sci **2:** 152–159

Schoch G, Goepfert S, Morant M, Hehn A, Meyer D, Ullmann P, Werck-Reichhart D (2001) CYP98A3 from *Arabidopsis thaliana* is a 3′-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. J Biol Chem **276:** 36566–36574

Slanina J, Taborska E, Bochorakova H, Slaninova I, Humpa O, Robinson W, Schram K (2001) New and facile method of preparation of the anti-HIV-1 agent, 1,3-dicaffeoylquinic acid. Tetrahedron Lett **42:** 3383–3385

Slimestad R, Verheul M (2009) Review of flavonoids and other phenolics from fruits of different tomato (*Lycopersicon esculentum* Mill.) cultivars. J Sci Food Agric **89:** 1255–1270

Sonnante G, D'Amore R, Blanco E, Pierri CL, De Palma M, Luo J, Tucci M, Martin C (2010) Novel hydroxycinnamoyl-coenzyme A quinate transferase genes from artichoke are involved in the synthesis of chlorogenic acid. Plant Physiol **153:** 1224–1238

Strack D, Gross W (1990) Properties and activity changes of chlorogenic acid:glucaric acid caffeoyltransferase from tomato (*Lycopersicon esculentum*). Plant Physiol **92:** 41–47

Teutschbein J, Gross W, Nimtz M, Milkowski C, Hause B, Strack D (2010) Identification and localization of a lipase-like acyltransferase in phenylpropanoid metabolism of tomato (*Solanum lycopersicum*). J Biol Chem **285:** 38374–38381

Vain P, Keen N, Murillo J, Rathus C, Nemes C, Finer J (1993) Development of the particle inflow gun. Plant Cell Tissue Organ Cult **33:** 237–246

Vallverdú-Queralt A, Odriozola-Serrano I, Oms-Oliu G, Lamuela-Raventós RM, Elez-Martínez P, Martín-Belloso O (2012) Changes in the polyphenol profile of tomato juices processed by pulsed electric fields. J Agric Food Chem **60:** 9667–9672

Villegas RJ, Kojima M (1986) Purification and characterization of hydroxycinnamoyl D-glucose. Quinate hydroxycinnamoyl transferase in the root of sweet potato, *Ipomoea batatas* Lam. J Biol Chem **261:** 8729–8733

Walker AM, Hayes RP, Youn B, Vermerris W, Sattler SE, Kang C (2013) Elucidation of the structure and reaction mechanism of sorghum hydroxycinnamoyltransferase and its structural relationship to other Coenzyme A-dependent transferases and synthases. Plant Physiol **162:** 640–651

Wang M, Simon JE, Aviles IF, He K, Zheng QY, Tadmor Y (2003) Analysis of antioxidative phenolic compounds in artichoke (*Cynara scolymus* L.). J Agric Food Chem **51:** 601–608