Universidad de Castilla-La Mancha

Facultad de Farmacia

Departamento de Ciencia y Tecnología Agroforestal y Genética

Isolation and characterization of enzymes involved in the biosynthesis of secondary metabolites with phytotherapeutic interest

TESIS DOCTORAL

Memoria para optar al Grado de Doctor presentada por

Almudena Trapero Mozos

Dirigida por:

Dra Lourdes Gómez Gómez **Dr Oussama Ahrazem El Kadiri**

Albacete, 2013.

A mis padres y hermano

Agradecimientos

En primer lugar me gustaría dar las gracias a mis directores de tesis, por su cariño y paciencia a la hora de escribir esta tesis, sin ellos nada de esto hubiera sido posible. Además de una gran científica, la Dra. Lourdes Gómez es una gran persona, a quien le debo todo lo que soy científicamente, un ejemplo de mujer trabajadora y una gran líder. Al Dr. Oussama Ahrazem, por su enorme paciencia conmigo, por enseñarme desde lo más básico hasta lo más complicado dentro del laboratorio y por corregir mis faltas de ortografía en inglés, es una fuente de conocimiento y un ejemplo para todas las personas que hemos tenido la suerte de conocerle y trabajar con él. A la Dr. Angela Rubio Moraga, por tratarme como parte de su familia, por preocuparse por todas las facetas de mi vida, tanto personal como profesionalmente, ha sido un gran apoyo y una verdadera amiga. Y a Javier Argandoña, por su apoyo logístico y sus charlas a la hora del desayuno, el almuerzo ya no volverá a ser igual.

Ahora toca la parte formal de los agradecimientos, cuando se agradece la financiación recibida durante estos años. Durante mis años en el laboratorio de Biología Molecular y Fisiología Vegetal he recibido financiación para la realización de los trabajos que aquí se exponen. De la Junta de Comunidades de Castilla-La Mancha con cargo al proyecto PPII10-0062-7718 y PAI08-0211-6481; del Ministerio de Educación y Ciencia con cargo al proyecto BIO2009-07803 y BIO2006-00841; y también de SAFFRONOMICS FA 1101 COST ACTION para la estancia en Italia.

Grazie a tutti i ragazzi che mi hanno aiutato tantissimo in questi mesi a Roma: Marco, Giulia, Elio, Maria, Sara e Gianfranco. Grazie per avermi accolto dal primo giorno come se fosse una del gruppo, non mi sono mai sentita sola. Grazie anche al Dr Giovanni Giuliano per avermi offerto l'opportunità di lavorare con il suo team, sono tantissime le cose che ho imparato da lui.

También quiero dar las gracias a la gente del IBMCP de Valencia, al grupo de Juan Carbonell y al grupo de Toni Granell, por su ayuda en muchos de los experimentos y por recibirme con una sonrisa siempre que pisaba esas tierras.

A mis amigos: Mamen, Lucía y Alex, por estar desde el principio, dándome ánimos para continuar, a Isaac por mostrarme esta nueva forma de vida, a Rocío y Ángel por acogerme en su casa siempre que he necesitado ir a Valencia, a Jessica "Calahorra", Alicia, Vanesa, Rosa y todos los demás por los momentos de diversión. A Nathaly por

empezar conmigo esta aventura, a Carmen y Miguel por ser mis compis de master y a Rosana por la parte técnica de los primeros años y por los momentos de desconexión fuera del laboratorio.

Y finalmente quiero dar las gracias a las 3 personas más importantes de mi vida, a mi padre, por su aliento incansable, su preocupación y su coraje; a mi madre, por su inestimable ayuda y sus palabras de apoyo cuando más las necesitaba; y a mi hermano, el orgullo de esta familia.

Index

Introduction

Chapters

Chapter 1: Cloning and characterization of a glucosyltransferase from *Crocus sativus* stigmas involved in flavonoid glucosylation

Chapter 2: Genomic analysis and gene structure of the plant carotenoid dioxygenase 4 family: a deeper study in *Crocus sativus* and its allies

Chapter 3: Crocins transport in *Crocus sativus*: the long road from a senescent stigma to a newborn corm.

Chapter 4: Characterization of a glucosyltransferase enzyme involved in the formation of kaempferol and quercetin sophorosides in *Crocus sativus*

Chapter 5: Identification *in silico* of new β-glucosidases and glycosyltransferases in *Crocus sativus* stigma

Resumen

En esta Tesis Doctoral se describe el aislamiento y caracterización de diferentes enzimas involucradas en la biosíntesis y modificación de metabolitos secundarios de *Crocus sativus* que poseen interés en la industria farmacéutica.

El primer capítulo describe la caracterización y el clonaje de una glucosiltransferasa (CsGT45) presente en *Crocus sativus in vitro*, CsGT45 cataliza la transferencia de una glucosa desde UDP-Glucosa a los flavonoides kampferol y quercetina, con una marcada regioselectividad. El kaempferol es el único flavonol presente en los estigmas de *C. sativus* y los niveles de sus glucósidos cambian a lo largo del desarrollo del estigma, estos cambios, se correlacionan con los niveles de expresión de CsGT45 durante las diferentes etapas del desarrollo.

En el segundo capítulo se presenta la caracterización genómica de la familia de dioxigenasas de corte de carotenoides de tipo 4 (CCD4) presente en distintas especies vegetales y caracterizada bioquímicamente por primera vez en *Crocus sativus*. Estas CCD4s están implicadas en el corte de los dobles enlaces C_9,C_{10} y $C_{9'}$,$C_{10'}$ de compuestos carotenoides cíclicos, generando compuestos volátiles del tipo C_{13} ketonas como la β-ionona. En *C. sativus* se han aislado diferentes alelos de CDD4, que muestran diferencias en su organización génica y se han comparado sus promotores con los de genes que codifican para otras CCD4 en distintas plantas. La elaboración de plantas transgénicas de *Arabidopsis thaliana* con el promotor CsCCD4a fusionado al gen GUS ha permitido estudiar la actividad del promotor en diferentes tejidos de la planta.

En el tercer capítulo se describe la evolución de distintos apocarotenoides y de distintos compuestos volátiles durante la senescencia de las flores de *C. sativus*. Se observó que durante el proceso de senescencia de la flor tiene lugar un transporte de apocarotenoides desde los estigmas hacia los ovarios, y de estos hacia el nuevo cormo en desarrollo. Por lo tanto, en ausencia de fertilización y desarrollo de la semilla, una parte del material es reutilizado, transportándose hacía los tejidos en proceso de desarrollo.

El cuarto capítulo muestra la caracterización de una nueva glucosiltransferasa (UGT707B1) aislada de estigmas de *C. sativus*. Los transcritos de *UGT707B1* fueron detectados en el estigma de todas las especies de *Crocus* analizadas, pero el análisis de la expresión de *UGT707B1* en tépalos reveló su ausencia en determinadas especies. El

análisis de los flavonoides presentes en los tépalos de las distintas especies de *Crocus* revela que la expresión de *UGT707B1*, está asociada a la presencia de los flavonoles glucosilados: kaempferol-3-*O*-β-D-glucopiranosil-(1-2)-β-D-glucopiranósido y quercetina-3-*O*-β-D-glucopiranosil-(1-2)-β-D-glucopiranósido, El análisis de plantas transgénicas de *A. thaliana* que expresan constitutivamente *UGT707B1* bajo el control del promotor 35S del virus del mosaico de coliflor, refuerzan la implicación de UGT707B1 en la biosíntesis de flavonol-3-O-sophorosidos. Así mismo, los fenotipos de estas plantas transgénicas demuestran cómo cambios en la homeostasis de determinados flavonoides pueden provocar importantes cambios en el desarrollo vegetal, principalmente de aquellos asociados al transporte y distribución de auxinas.

El quinto capítulo se ha procedido a la búsqueda *in silico* de nuevas enzimas con actividades de glucosiltransferasa o glucosidasa presentes en el estigma de *C. sativus* utilizando herramientas bioinformáticas y de una base de datos de más de 65000 secuencias parciales obtenidas durante el proceso de secuenciación masiva de transcritos de seis estadios de desarrollo del estigma de azafrán, realizada en colaboración con el laboratorio del Dr. G. Giuliano (Green Biotechnology of ENEA, Italia). Este trabajo se ha realizado durante una estancia predoctoral de tres meses en dicho laboratorio.

Todos los resultados recogidos en esta memoria tienen como finalidad conocer mejor los mecanismos de la biosíntesis de los metabolitos secundarios de *Crocus sativus* con interés farmacológico e industrial así como determinar los genes implicados en su formación que puedan ser utilizados como herramientas para la síntesis de flavonoides con determinadas regioespecificidades.

Summary

The main goal of the present work was the isolation and characterization of some enzymes involved in the biosynthesis of secondary metabolites of *Crocus sativus* (commonly named saffron) with pharmaceutical and industrial interest.

The first chapter is dealing with the characterization and cloning of a glucosyltransferase (CsGT45) isolated from *Crocus sativus*. CsGT45 catalyses the transfer of glucose from UDP-glucose to the flavonoids kaempferol and quercetin with a specific regioselectivity. Kaempferol is the unique flavonol present in *C. sativus* stigmas and the levels of its glucosides changed during stigma development, and these changes, are correlated with the expression levels of *CsGT45* during these developmental stages.

The second chapter focus on a genomic study of carotenoid cleavage dioxygenase 4 (CCD4) family of *Crocus sativus*. The CCD4 enzymes mediate the formation of volatile C_{13} ketones, such as β-ionone, by cleaving the C_9,C_{10} and $C_{9'}$,$C_{10'}$ double bonds of cyclic carotenoids. Different CCD4 alleles have been identified. The promoter regions of CsCCD4a and CsCCD4b with available CCD4 gene promoters from other plant species have been compared. The spatial and temporal activity of the putative *CsCCD4a* promoter from *Crocus sativus* fused to the β-glucuronidase (*GUS*) gene in stably transformed by *Arabidopsis thaliana* plants, was also assessed, and the location of GUS staining was monitored in different plant tissues.

The third chapter describes the accumulation of several apocarotenoids and volatiles during flower senescence in *C. sativus*. We observed a transport of saffron apocarotenoids from the stigma tissue to the ovaries and to the developing corm. Thus, in the absence of fertilization and seed development, a part of the material invested is translocated backwards to sinks in somatic tissues, like the developing corm over the mother plant.

The fourth chapter describes the characterization of a new glucosyltranferase named UGT707B1 from C. sativus. UGT707B1 transcripts were detected in the stigma tissue of all the Crocus species analyzed, but expression analysis of UGT707B1 in tepals revealed its absence in certain species. The analysis of the glucosylated flavonoids present in Crocus tepals reveals the presence of two major flavonoid compounds in

saffron: kaempferol-3-*O*-β-D-glucopyranosyl-(1-2)-β-D-glucopyranoside and quercetin-3-*O*-β-D-glucopyranosyl-(1-2)-β-D-glucopyranoside, both of which were absent from the tepals of those *Crocus* species that did not express *UGT707B1*. Transgenic *A. thaliana* plants constitutively expressing *UGT707B1* under the control of the cauliflower mosaic virus *35S* promoter have been constructed and their phenotype analyzed. We also measured the levels of flavonoids in extracts of the transgenic plants which showed changes in the composition of flavonols when compared with wild-type plants. The data obtained showed the involvement of UGT707B1 in the biosynthesis of flavonol-3-*O*-sophorosides and how significant changes in flavonoid homeostasis can be caused by the overproduction of a flavonoid-conjugating enzyme.

The fifth chapter is dedicated to search *in silico* for new glucosyltransferases and glucosidases from *C. sativus*, using bioinformatic tools and a database of more than 65000 partial sequences which were obtained by massive sequencing of transcripts of six development stage of saffron in collaboration with Dr. G. Giuliano laboratory Green Biotechnology of ENEA, Italy. This work was realized during my short term mission scientific in that laboratory.

All the results included in this report, will help for better knowledge of biosynthesis mechanism of secondary metabolite of *Crocus sativus*, for industrial pharmaceutical interest and also to elucidate the genes involved in its formation which may be used as a tool for the synthesis of regioselective flavonoids.

Introduction

Crocus sativus

Crocus sativus L. or saffron is known for its dried stigmas and is considered the most precious and expensive spice in the world. The spice is also known as saffron and it is often called “the Golden Spice”. Nowadays, the main use of saffron is for culinary purposes. It is used for several Mediterranean dishes, often in connection with fish and seafood. Famous dishes as the Spanish rice “paella Valenciana”, the Italian “risotto alla Milanese” and the Provençal fish soup “bouillabaisse”. Furthermore, saffron appears in a few European cake recipes, where it is used for both flavour and colour. In Central Asia and Northern India, it is used extensively for rice dishes. Indian sweets (kheer, ras malai) are sometimes prepared with saffron. Saffron dyes are used today for colouring carpets, hats and traditional women’s costume in Sardinia and it is also used in the routine histological stains for dermatopathology in combination with other products such as hematoxylin and eosin [1].

However, the first uses of saffron were as a medicinal plant. Since Greek and Roman periods, saffron has been employed for medicinal purposes against coughs, flatulence, stomach disorders, colic, insomnia, feminine disorder and heart disease and also as anodyne, tranquilliser and for its emetic properties [2]. In Chinese traditional medicine, saffron has been widely used, as it is described in “Yi-Lin-Ji-Yao”, a traditional Chinese medical book composed during the Ming Dynasty (16th century) where the effects of saffron on promotion of blood circulation to remove blood stasis are prominent [3].

In the Islam culture, one of the most important books of medicine is al-Qanun or Canon of Medicine, which was written by Avicenna and who described around 800 different natural drugs. This book dedicates, a monograph dealing with saffron[4]. The traditional uses of saffron are resumed in the table 1.

Table 1. The traditional use of saffron

Systems and effects	Traditional uses (condition and effect)
Analgesic and anti-inflammatory	Earache, tooth-ache, swelling, otitis, anal pain, gout, cancer pain, gingivitis, discomfort of teething infants
Cardiovascular system	Cardiac stimulant, removes blockages of vascular
Central nervous system	Narcotic, antihysteric, CNS stimulant, hypnotic, mental disease, sedative, anticonvulsant, neurasthenia,
Eye disease	Painful eye, lacrimation, day blindness, corneal disease and cataract, purulent eye infection, pterygium, poor vision

Gastrointestinal system	Stomachic, decrease appetite, treatment of haemorrhoid, prolapse of anus, jaundice and enlargement ogf the liver, antiflatulent
Genitourinay system	Abortive, treatment of amenorrhea, aphrodisiac, impotency, contraceptive, emmenagogue, stimulate menstruation, prolapse of anus, stop mestrual periods, promote menstruation, use in puerperium period, terminate pregnancy, painful urination, diuretic, kidney stone
Infection disease	Antibacterial, antiseptic, anti-fungal, measles, smallpox, scarlet fever
Respiratory system	Astrhma, bronchitis, expectorant, pertussis, dyspnea, pleurisy, antitussive, diphtheria, disability tonslis, resulting snoring, respiratory decongestant, expectorant
Skin disease	Treatment of psoriasis, eczema, acne, wound
Miscellaneous	Immunostimulant, diaphoretic, tissue colouration, anticancer.

Currently, the increase of interest in natural products (such as spices) with low toxicity resulted in extensive studies of the biological properties and potential medical application of saffron. During these studies, the properties of saffron to prevent heart attack by decreasing cholesterol and triglyceride levels [3], to antagonise the activity of ethanol [5], [6], [7], [8] and to treat neurodegenerative disorders such as senile dementia [9] have been determined. However, most of the studies have been focused on the potential activity of saffron and its constituents in chemotherapy research, as adjuvant in chemotherapy to decrease the negative effects of the treatment of cisplatin [10], [11], [12], as a chemopreventive agent during tumorogenesis [13], [11], [14], [15] and above all as an antitumoral agent as described below.

The most important behaviour of a potential antitumoral compound is the higher activity against malignant cells than against normal cells. Saffron extract has shown this effect on different human cell lines with an IC_{50} ranged from 100 to 200 μg/mL upon typical human malignant cells, but it showed no significant effect on normal human lung cells [14], [16]. However, an important aspect for pharmacology of saffron is that high doses (10-12 g/kg) can be fatal due to the presence of poison that acts on the central nervous system and damages the kidneys [2]. Vias of administration have also been compared; intraperitoneal injection of liposome-encapsulated saffron extract enhances antitumor activity in comparison with oral administration, probably due to an increase of the drug solubility [17]. Several studies were performed to determine which active components of saffron extract were responsible for the antitumoral activity, showing that crocins are mainly responsible for such activity [18], [19], [20]. The effect of crocetin on tumoral cell lines has been the object of different criteria. There are studies where the sugars play a key role in the cytotoxic effect of crocin [18], [21], [2], while in another study the crocetin in non-toxic doses inhibited genotoxic effect and neoplastic transformation of tumoral cells [22].

García-Olmo and co-workers [23] reported that crocin increased the survival time and decreased tumour (colon adenocarcinoma) growth in female rats, without any significant effect in male animals. This suggests that the selective antitumor action of crocin in female rats might be associated to hormonal factors.

The pharmacological activities of saffron are constant at least when saffron is stored under –20ºC for at least 2 years or even longer [5]. However, temperature and humidity exert a strong influence on the degradation of the main pharmacologically active ingredients of saffron under different storage conditions [24], [25], [26], [27].

Taxonomy and Botany

The Botanic Treatise of Strasburger [28] and the description in the International Botanic Code of Nomenclature, classify Saffron as:

Super-kingdom: Eukaryota → Kingdom: Plantae → Sub-kingdom: Embryophyta → Division: Spermatophyta → Subdivisión: Magnoliophyta; Liliopsida (Angiosperma) → Class: Liliatae (monocotiledonea) → Super-Order: Lilianae → Order: Liliales → Family: Iridaceae → Genus: *Crocus* → Species: *C. sativus*.

C. sativus L has 2n=24 chromosomes and it thought to be an allopolyploid hybrid, result of crossing between diploid parents, one of which had an unreduced gamete [29] [30]. Five fertile diploid (2n=16) Crocus species are considered as possible ancestors of the sterile saffron: *C. cartwrightianus*, *C. hadriaticus*, *C. oreocreticus*, *C. pallasii* and *C. thomasii*. Preliminary RAPDs analysis has been used in order to determine the relationships among these crocuses with *C. sativus* [31], without obtaining further insights than those obtained by morphological studies. Although RAPDs have been used in many studies of genetic diversity [32], the limitations of these markers for phylogenetic analysis have been summarized by several authors [33] [34] [35], including problems associated with reproducibility and product homology. By contrast, for studies of congeneric species, noncoding sequences are often sources of phylogenetically informative characters and of the various noncoding regions, introns are being used for such purposes in plants. Recently our group have shown the utility of CsBCH1 intron sequences for phylogenetic analysis at the species level [36]. This intron has several of the attributes expected for nuclear sequences used for resolving relationships at lower taxonomic levels, such as the presence of one or more sizable noncoding regions flanked by conserved regions and single or low copy number sequences [37] [38]. BCH genes are generally members of small gene families (often only two or three loci) and maintain at least four conserved intron positions across all members. These introns are typically flanked by conserved exons that provide the opportunity to develop universal primers for direct sequencing. Traditional classification of the subgenus *Crocus* mostly relied on similarities in morphology such as floral and corm structures [39]. This subgenus is divided into two sections, which are subdivided in series using the nature of the style and corm tunic. The molecular data based on BCH1 intron sequences support such groupings. Basically, we have found two main clades: one containing sequences from all the genera included in the section Crocus and the other that correspond to the section Nudiscaspus. Furthermore, we found a good correlation at Series level. *C. sativus* BCH1 alleles grouped together with *C. cartwrightianus*, and *C. hadriaticus* accessions. Thus, *C. cartwrightianus*, *C. hadriaticus* and *C. sativus* showed a close resemblance in chromosome morphology [29], also *C. cartwrightianus*, *C. hadriaticus* and *C. sativus* showed the most similar apocarotenoid profile in the stigmas [36]. Although *C. cartwrightianus* and *C. hadriaticus* are native from Greece, they are not known to inhabit the same areas, as *C. hadriaticus* is a western Greek taxon and *C. cartwrightianus* is a south-eastern taxon. Considering different geographic origins for these species, the obtained data suggest that a human-mediated event might have been responsible for the formation of *C. sativus*, from a hybridization event between *C. cartwrightianus* and *C. hadriaticus*. However, the possibility of a natural hybridisation event could be considered assuming an overlapping distribution for both species in the past. In addition, sometimes misclassification occurs due to the fact that in many cases flowers of *C. hadriaticus* and *C. cartwrightianus* may look alike, and therefore both species could be simultaneously present in specific areas. Nevertheless, these data support the proposed allotriploid origin of *C. sativus* [30] [40] as the result of crossing between two closely related species with participation of n and 2n gametes. The re-synthesis from the presumptive parental diploids *C. hadriaticus* and *C. cartwrightianus* could help not only to demonstrate the origin of saffron, but also might provide a good tool to obtain fertile saffron plants that could be improved.[41]

Owing to the triploid genome of saffron, their beautiful flowers cannot produce any seeds and propagation is possible only via corms. The corm is a swelling stem fleshy inside and composed of slender fibrous tunics; its diameter is usually around 5 cm and is flat at the base. Roots appearing in the base of the corm are white, fine, long and numerous. The plant has from 3 to 5 white membranous cataphylls, between 5 to 11 leaves that are 1.5 to 2.5 mm wide, synanthous, green, glabrous and ciliated and from 1 to 4 fragrant autumnal flowers. Each flower presents a 3-lobed subterranean ovary and 3-lobed style. The perianth of the flower presents six deep lilac tepals. The tube of the perianth protects the style. The flower presents three stamens located below three stigmas. The stigmas have also been considered as three branches of the style [42].

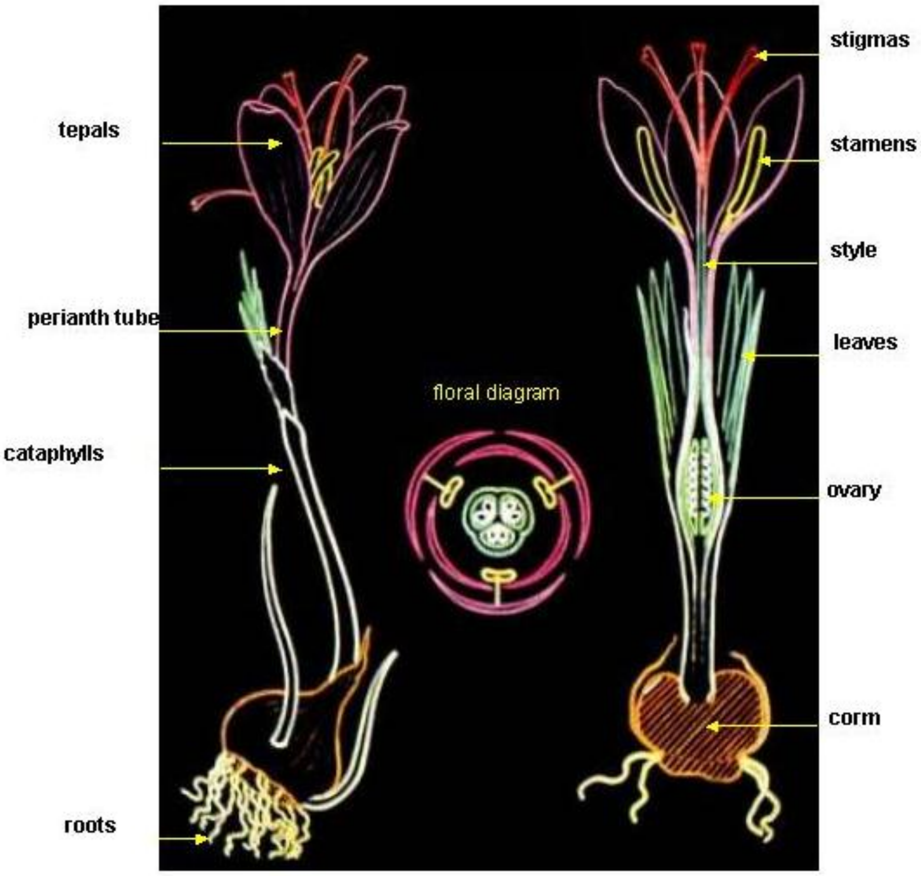


Fig. 1. Schematic representation of *C. sativus* L., transversal section and floral diagram of the plant. (Adapted from Rubio Moraga's thesis [43])

Biological cycle

The biological cycle is difficult to understand since there are moments that three corm generations co-exist at the same time [44]. The corm is a swelling stem; therefore the description of the cycle will start from when the apical buds of the mother corm develop into stems (Fig. 2).

Figure 2. Biological cycle of *C. sativus*. The sprouting stage is during the end of August or beginning of September; the flowering stage is in October; the vegetative stage is from November till April or May and the dormancy period is from June till August.

The life cycle of saffron is similar in all different countries that produce saffron, but there are wide differences in the timing of events [45]. Flowering occurs during autumn (October–November), and is followed by a vegetative stage throughout winter and the formation of replacement corms at the base of the shoots. From the moment when the plant has green leaves, the daughter corm acts as a sink organ for carbohydrates produced in photosynthesis. The accumulation of these reserves causes the swelling of the daughter corm, although the mother corm, which has roots during all this period, also maintains its growth. At the beginning of the dry season (April–May) the mother corm is completely senescent and the daughter corm becomes independent. These new corms have meristematic tissues that will generate new stems and further buds. The leaves senesce and wither, and the developed bulbs turn into dormancy. The transition from the vegetative to the reproductive stage occurs shortly afterwards in the apex of the buds of underground corms. This transition has been reported to start during March in Azerbaijan [46] [47], from March to April in Israel [48] and during July in Kashmir [49]. In Spain apical bud growth started about 40 days after the withering of the leaves. Flower formation followed shortly afterwards, and the first stages of flower initiation are already discernible by early August [45].

Secondary metabolites in plants: flavonoids and carotenoids

The sedentary lifestyle of plants has led to the evolution of a vast multitude of molecules that are synthesized for environmental adaptation and also to influence other living organisms that co-habit the biosphere, both below and above ground. Many of these molecules are products of secondary metabolism. Based on their biosynthetic origins, secondary metabolites are divided into three main groups: terpenoids, alkaloids and phenylpropanoids together with the phenolic compounds. Terpenoids are built from five-carbon isoprene units (IPP) and so can also be called "isoprenoids". Alkaloids are a group of nitrogen-containing bases, most of which are drugs. Only a few (like caffeine) are derived from purines or pyrimidines, while the large majority are produced from amino acids. The starting product of the biosynthesis of most phenolic compounds is shikimate.

Among the secondary metabolites, carotenoid and flavonoids are considered as important health promoting phytochemicals and have led to the plants that accumulated them and its products to being termed as functional food.

Carotenoids are widely distributed isoprenoid pigments fulfilling diverse functions in all taxa [50]. Due to their vital role in protecting the photosynthetic apparatus from photooxidation, carotenoids are synthesized in all photosynthetic organisms. Moreover, carotenoids represent essential structural components of the light-harvesting antenna and reaction centre complexes and are accumulated in many flowers and fruits, contributing substantially to plant-animal communication [51], [52], [53], [54]. In addition, many heterotrophic bacteria [55] and fungi [56] produce these pigments. Carotenoids also confer their colour to some crustaceans, fish and birds. However, animals do not synthesize carotenoids de novo and rely, therefore, on nutritional sources to meet their needs. In addition of the cited functions, carotenoids serve as precursors of several physiologically important compounds, synthesized through oxidative cleavage and generally known as apocarotenoids [57], [58]. Representative examples are the ubiquitous chromophore retinal, the chordate morphogen retinoic acid, the phytohormone abscisic acid, the fungal pheromone trisporic acid, and the strigolactones, essential signaling molecules attracting both symbiotic arbuscular mycorrhizal fungi and parasitic plants [59], [60] and regulators of shoot branching [61], [62].

Flavonoids represent a large class of phenylpropanoids. Besides providing beautiful pigmentation in flowers, fruits, seeds, and leaves [63], flavonoids also have key roles in signaling between plants and microbes, in male fertility of some species [64], in defense as antimicrobial agents and feeding deterrents [65], in UV protection [66], in the regulation of polar transport of auxins, and in cell cycle regulation in plants [67], [68]. There is increasing evidence suggesting that flavonoids, in particular those belonging to the class of flavonols (such as kaempferol and quercetin), are potentially health-protecting components in the human diet as a result of their high antioxidant capacity [69], [70]. Therefore, flavonoids may offer protection against major diseases such as coronary heart diseases and cancer [71], [72].

The health benefits conferred by numerous carotenoids and flavonoids have led to a higher level of inclusion in foodstuffs. The main pathways for the biosynthesis of carotenoids and flavonoids in plants have been elucidated in recent years [73], [54], [63], and some of the identified biosynthetic genes have been successfully used in metabolic engineering approaches in different plant systems [74], [75], [76], [77]. However, further elements in the complex regulatory network of these secondary compounds must be identified, such as the ones involved in the relative importance of secondary modifications, with influence in the net accumulation and availability of secondary metabolites, with the aim of influencing biosynthesis on a rational basis.

The generic structure of a flavonoid consist of two aromatic rings (ring A and ring B) linked by three carbons that are usually in an oxygenated heterocyclic ring or C ring based on differences in the generic structure of the heterocyclic C ring as well as the oxidation state and functional groups of the heterocyclic ring, they are classified as flavonols, flavones, flavanols (catechins), flavanones, anthocyanidins and isoflavonoids. Flavonoids are most frequently found in nature as conjugates in glycosylated or esterified forms.

The major source of flavonoids includes fruits and fruits products (e.g citrus fruits, rose hips, apricots, cherries, grapes, black currants, blueberry, and apples), vegetables (e.g. onions, green paper, broccoli, tomatoes, spinach), tea leaves, soybeans and herbs.

The high number and variability of secondary metabolites observed in nature is due to secondary modifications on the basic skeleton of these compounds. The widespread occurrence of glycosylated secondary metabolites, including flavonols, anthocyanins, monoterpenes, norisoprenoidic compounds and plant hormones shows that both glycoside hydrolases (GHases) and glycoside transferases (GTases) responsible for their metabolism play a central role in a large number of major biological processes, while our knowledge of the enzymes involved in these processes is still limited.

Apocarotenoids and flavonoids in *C. sativus* stigmas

The major components of saffron are derived from the oxidative cleavage of carotenoids, whose biosynthesis and accumulation during stigma development have been characterized by our group [36]. The resulting products are the apocarotenoids cis- and trans-crocins, picrocrocin (β-D-glucopyranoside of hydroxyl-β-cyclocitral), and their degradation product, the odour-active safranal [78] (Fig. 4). However, recent studies reveal a different volatile composition in unprocessed stigma tissues [79], suggesting that, as has been observed in other spices [80] degradation processes are responsible for the organoleptic properties of saffron from preformed apocarotenoid compounds [81]. Apocarotenoids have been extensively studied due to their high concentration in the stigma tissue and because of their organoleptic properties [82], [83]. It has been proposed that the biogenesis of the main colour principles, crocins, and the odour active compound, safranal, is derived by bio-oxidative cleavage of zeaxanthin [84] by a 7,8-7’,8’ cleavage reaction, and recent studies show that β-carotene also contributes to the pool of crocin and crocetin in the stigma tissue of *C. sativus*.

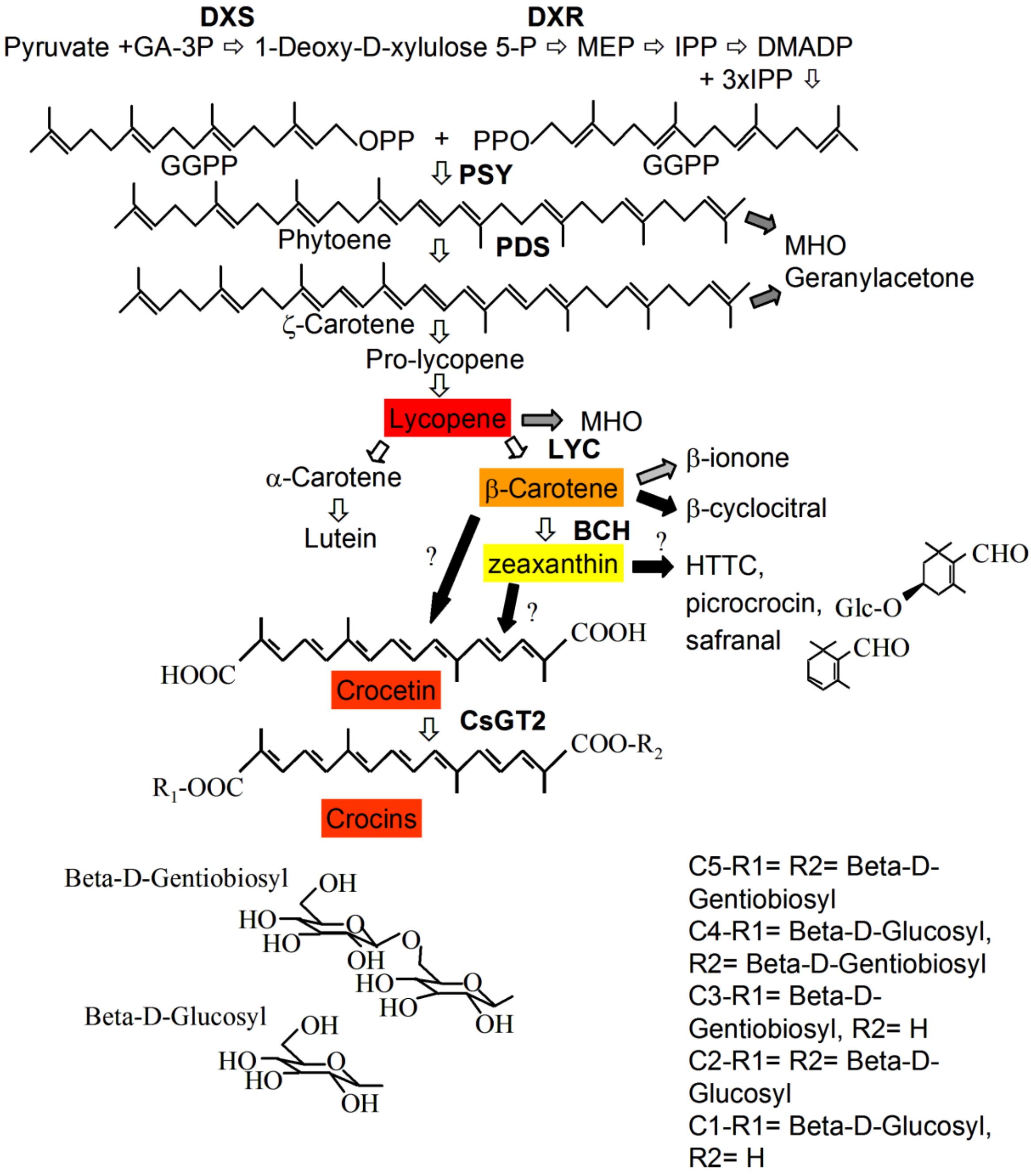


Fig 4. Biosynthesis of crocin and safranal

Other carotenoids and glucosylated apocarotenoids have been identified in *C. sativus*, but at low levels (Fig. 5), a monogentiobiosyl ester, structurally related with crocin, and that could be generated by the oxidative cleavage of zeaxanthin at the positions 7,8 and 13,14 [85], the β-D-glucopyranosides (4R)-4-hidroxy-3,5,5-trimethylcyclohex-2-enone, (4S)-4-hidroxy-3,5,5-trimethylcyclohex-2-enone, and the (4S)-4-hidroxymethyl-3,5,5-trimethylcyclohex-2-enone, and seven more apocarotenoids related to other compounds previously identified in other plant species [86]. In addition, other volatile carotenoids with C_9- or C_{10}- carbons in their skeleton were identified in 1971 by Zarghami and Heinz [87] and later by Rödel and Petrizka [88]. Using GC/MS analysis, new volatile carotenoid-derivate compounds were identified by Tarantilis and Polissiou [89].

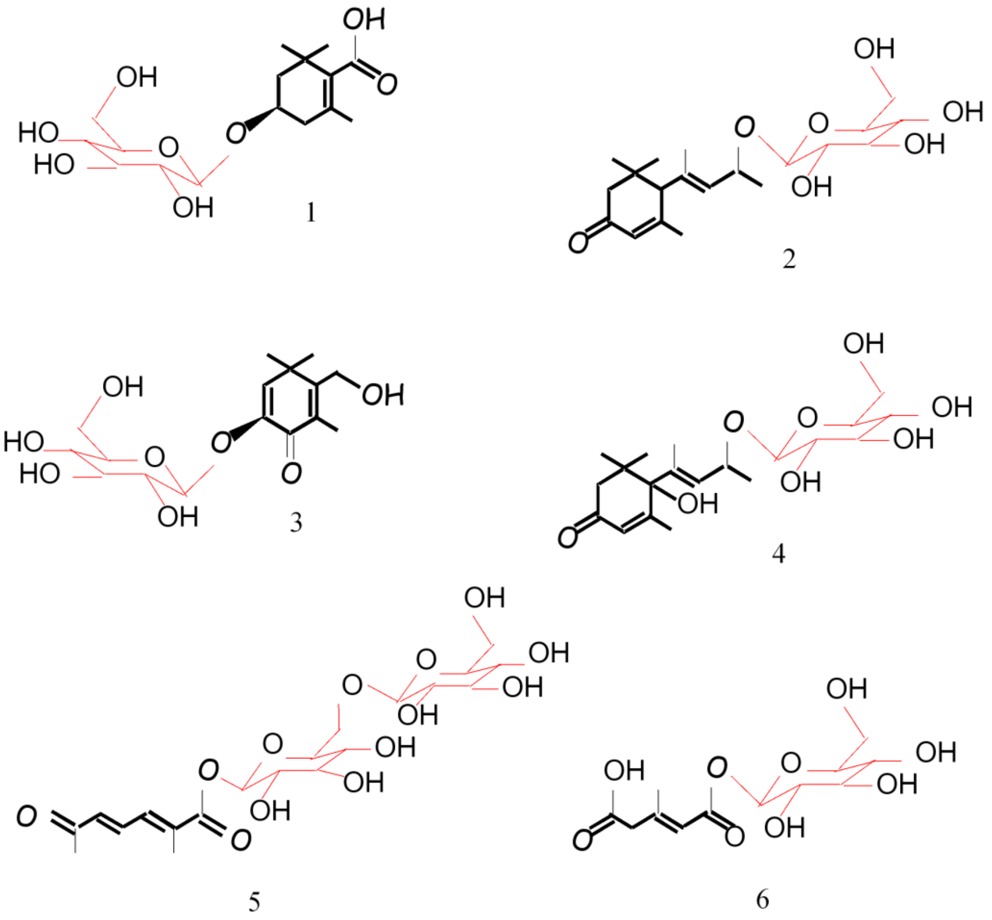


Fig 5. Other glucosylated apocarotenoid compounds identified in *C. sativus*. 1: (4R)-4-Hydroxy-2,6,6-trimethyl-1-cyclohexene-1-carboxylic acid O-β-D-glucopyranoside; **2**: (4R)-4-[(1E,3R)-3-Hydroxy-1-buten-1-yl]-3,5,5-trimethyl-2-cyclohexen-1-one β-D-glucopyranoside; **3**: 6-Hydroxy-3-(hydroxymethyl)-2,4,4-trimethyl-2,5-cyclohexadien-1-one 6-O-b-D-glucopyranoside or crocusatin-glucoside; **4**: (2R,3E)-4-[(1S)-1-Hydroxy-2,6,6-trimethyl-4-oxo-2-cyclohexen-1-yl]-3-buten-2-yl β-D-glucopyranoside or roseoside I; **5**: b-D-gentiobiosyl ester of 2-methyl-6-oxohepta-2,4-dienoic acid; **6**: (2Z)-3-Methyl-2-pentenedioic acid O-b-D-glucopyranoside.

Studies concerning components in saffron stigmas other than carotenoids or apocarotenoids are scarce [83], [85]; [86]; [90], as is the case for flavonoids [90], although these have been analysed in great detail in petals [91]. The majority of studies performed on *C. sativus* concentrate on the spice, and only recently the accumulation of these compounds during stigma development has been approached

The players involved: Plant Glycosyltransferases and Glycosidases

There are several hundred different β-glucosidic flavour and aroma precursors identified from plants whose aglycones are products of mevalonate (MVA), methylerythritol 4-phosphate (MEP) phenyl propanoid or shikimate pathways. Obviously, there are glucosyltransferases and β-glucosidases in source plant tissues that transfer glucose moieties or hydrolyze these flavour precursors, respectively. Thus, in each case, there is a need for isolating and characterizing the specific enzyme that transfers or hydrolyzes a β-glucoside whose aglycone moiety is of interest in food quality and processing. Such biochemical data are crucial to making practical decisions as to whether or not enzymes from host plants or other sources should be added to drinks and beverages before, during or after processing to enhance flavour, aroma and other quality factors. Likewise, such data are essential for targeting enzymes with desirable properties for overproduction in transgenic microbial or plant hosts and improvement of their catalytic properties and stability for specific uses by genetic engineering.

Glycosyltransferases from small molecules transfer sugars to a wide range of acceptors, from hormones and secondary metabolites to biotic and abiotic chemicals and toxins in the environment [92]. Parts of these metabolites are glycosylated. A broad range of different carbohydrate moieties can be added, recruiting all forms of sugars independently (monoglycosides), in parallel, or in chains (di-, tri-glycosides, etc.). This gives rise to a broad spectrum of glycosidic structures for any given aglycone. The enzymes are encoded by large multigene families and can be identified by a signature motif in their primary sequence [93], which classifies them as a subset of Family 1 glycosyltransferases (GTases). These enzymes are very attractive and are being studied for many reasons such as their wide range of substrates, their potential role in developmental and metabolic homeostasis, and their function in detoxification processes of relevance to clinical and agricultural applications. In vitro studies have shown that a single gene product can glycosylate multiple substrates of diverse origins; while multiple enzymes can also glycosylate the same substrate [94, 95].

Some flavour and colour compounds are present in plant cells as glucosides. Their hydrolysis to form volatile aglycones is catalysed by glucosidases (GHases). The number of expressed GHases involved in secondary plant metabolism and the substrate specificities of these enzymes remain largely unknown. Different glycosidic structures for any given secondary metabolite suggest the existence of different families of GHs that could hydrolyse these compounds. The O-glycoside hydrolases are a widespread group of enzymes of significant biological, biomedical and industrial importance. They catalyse the hydrolysis of the glycosidic bond between two or more carbohydrates or between a carbohydrate and a noncarbohydrate moiety. A classification of the glycoside hydrolases into 82 families, based on the amino acid sequence similarities, has been proposed and is now widely used [96]; [97]. Among the glycoside hydrolases, β-glucosidases have been extensively studied because of their important role (i) in numerous biological processes: growth and development via release of phytohormones (auxins, gibberellins, cytokinins) from their inactive glucoconjugated forms [98], host-parasite interactions [99] [100], lignification [101]; [102], cell wall degradation in the endosperm during germination [103], circadian rhythm of leaf movements [104], in floral development and pigmentation [105] and (ii) in biotechnological applications: food detoxification [106], biomass conversion [107]; [108] and, over the past decade, flavour enhancement in beverages [109]. Indeed the intensive research carried out over the past two decades has demonstrated that, in a great number of fruit and other plant tissues, important flavour compounds accumulate as non-volatile and flavourless glycoconjugates, which make up a reserve of aromas to be exploited [110].

Carotenoid cleavage dioxygenases

The first gene identified as encoding a carotenoid cleavage dioxygenase (CCD) was the maize VIVIPAROUS14 (Vp14) gene that is required for the formation of abscisic acid (ABA), an important hormone that mediates responses to drought stress and aspects of plant development such as seed and bud dormancy [111]. The VP14 enzyme cleaves at the 11,12 position of the epoxycarotenoids 9′-cis-neoxanthin and/or 9-cis-violaxanthin and is now classified as a 9-cis-epoxycarotenoid dioxygenase (NCED) [112], a subclass of the larger CCD family. Since the discovery of Vp14, many other CCDs have been shown to be involved in the production of a variety of apocarotenoids. In insects, the visual pigment retinal is formed by oxidative cleavage of β-carotene by β-carotene-15,15′-dioxygenase [113]. Retinal is produced by an orthologous enzyme in vertebrates, where it is also converted to retinoic acid, a regulator of differentiation during embryogenesis [114]. A distinct mammalian CCD is believed to cleave carotenoids asymmetrically at the 9,10 position [115] and, although its function is unclear, recent evidence suggests a role in the metabolism of dietary lycopene [116]. By contrast, in flowering plants no 15,15’ cleave activity has been detected. In addition to NCED, the CCDs from plants are distributed in four classes CCD1, CCD4, CCD7 and CCD8. CCD1 and CCD4 seem to be involved in the production of aroma volatiles. The first member of the CCD1 subfamily was identified from *Arabidopsis thaliana* [117]. Sequence homology then allowed the identification and characterization of orthologs from several plant species, such as *Crocus* [118]; [79], tomato [119], grape [120], melon [121], petunia [122], maize [123], *Fragaria* [124], *Medicago* [125] and rice [126]. CCD1 is a non-heme enzyme, which uses oxygen to cleavage a variety of carotenoid substrates symmetrically at the 9,10 and 9’,10’ positions of cyclic and acyclic carotenes, generating C_{14} dialdehydes, which are common to all carotenoid substrates, and two variable end-group-derived C_{13} ketones. The wide substrate specificity of plant CCD1s allows the production of divergent volatile C_{13} compounds, including α-ionophores, α-ionones, pseudoionone and geranylacetone [117]. The CCD1 enzymes also recognize the 5,6 or 5’,6’ bond positions of linear carotenes leading to the volatile C8 ketone 6-methyl-5-hepten-2-one [123], and the 7,8 and 7’,8’ double bonds of acyclic carotenoid ends leading to geranial [126].

The CCD4 enzymes cleave double bonds at 9,10(9’,10’) positions, and seem to be more active than the CCD1 enzymes, at least for the substrate α-carotene [79]. Several CCD4 enzymes have been characterized from different plants such apple (*Malus domestica*, MdCCD4), chrysanthemum (*Chrysanthemum morifolium*, CmCCD4a), rose (*Rosa damascena*, RdCCD4), osmanthus (*Osmanthus fragrans*, OfCCD4), and Arabidopsis, AtCCD4 [127]. CmCCD4a and MdCCD4 cleaved β-carotene to yield β-ionone, while RdCCD4, and AtCCD4 cleavage of 8'-apo-β-caroten-8'-al by to yield β-ionone, which demonstrates that all the CCD4 enzymes cleave their substrates at 9,10 and 9',10' positions. Although CCD1 and CCD4 enzymes cleave carotenoids at the same positions (9,10 and 9',10’), CCD4 enzymes seem to be more substrate specific than CCD1. CCD4s could not cleave linear carotenoids such as lycopene or carotenoids containing a hydroxyl group such as zeaxanthin and lutein. It seems that CCD4s only cleave cyclic non-polar carotenoids such as β-carotene. In addition, CCD4 seems to be targeted to the plastids, whereas CCD1 enzymes are cytosolic and lack a chloroplast transient peptide in their sequence [128]; [129]; [119] [130]. The plastid, or more exactly, the plastoglobule location of the CCD4 enzymes, as is the case of the *C. sativus* and the *Arabidopsis* enzymes [131], allows these enzymes to obtain access to plastid carotenoids, while the CCD1 activity is limited to carotenoids out of these organelles or once these organelles have lost homeostasis or are targeted for degradation.

The CCD7 and CCD8 enzymes are implicated in the generation of the apocarotenoid hormone strigolactone involved in shoot branching. CCD7 and CCD8 are conserved across angiosperm species including monocotyledons: MAX3 (more axillary shoots), RMS5 (ramosus) and HTD1(high tillering dwarf)/D17 encode (CCD7) [132] [133] [134] [135] [136]. Recombinant AtCCD7 exhibits regioselectivity for the 9,10 position similar to CCD1, yet it cleaves only once asymmetrically, resulting in C_{13} and C_{27} products [137]. Whereas MAX4, RMS1, D10 and DAD1 (decreased apical dominance) encode another subclass of CCDs designated as CCD8 [138] [139], and cleaves the product generated by CCD7 [137] [140], but is also able to act directly on carotenoid substrates [130]. Therefore, CCD7 and CCD8 might catalyse sequential carotenoid cleavage reactions.

Phytotherapy: The medicinal plants importance

Plants have been used as medicines throughout history. Indeed, studies of wild animals show that they also instinctively eat certain plants to treat themselves for certain illnesses. Medicinal plants are widely and successfully used on every continent. By definition, ‘traditional’ use of herbal medicines implies substantial historical use, and this is certainly true for many products that are available as ‘traditional herbal medicines’. About 80% of the world's people depend largely on traditional plant-derived drugs for their primary health care (PHC). Medicinal plants serve as sources of direct therapeutic agents and raw materials for the manufacture of more complex compounds, as models for new synthetic products, and as taxonomic markers.

Plants are the tremendous source for the discovery of new products with medicinal importance in drug development. Today several distinct chemicals derived from plants are important drugs, which are currently used in one or more countries in the world. Secondary metabolites are economically important drugs such atropine, codeine, morphine, digitoxin/digoxin, and quinine/artemisinin. In addition, many of the drugs sold today are simple synthetic modifications or copies of the naturally obtained substances. According to the World Health Organization ∼25% of medicines are plant-derived [141]. Discoveries of novel molecules and advances in production of plant-based products have revived interest in natural product research.

One overall goal in phytotherapy research is to genetically characterize the molecular mechanisms driving secondary metabolite biosynthesis in plant cells. Identifying key elements involved in these processes, will allow generating novel tools for metabolic engineering of plant cells to improve productivity and synthesis of valuable compounds. Even so, only a little fraction of the enormous biosynthetic potential of plant cells is being exploited. Metabolic engineering of plant cells so far has added little to the problem, since insight into the molecular mechanisms driving plant secondary metabolism is still very limited. Genetic maps of biosynthetic pathways are still far from complete and the regulation of these pathways is hardly understood [142]. Thus, the use of genomic information to discover selected pathway-specific biosynthetic genes has become a powerful means to access a desired series of compounds.

A limited number of key genes encode the enzymes that are responsible for the synthesis of the pivotal backbone structures that constitute the hallmarks of the different classes of natural products, and progress has been made in the identification of these genes[143; 144]. The subsequent decoration of the backbone structures generates the huge diversity of plant secondary products. The large majority of these decoration processes are mediated by a limited number of enzyme classes, such as glycosyl-, methyl- and acyltransferases, which are all encoded by multi gene families. The positioning of enzymes that have broad substrate specificity downstream of the conserved early pivotal enzymes of plant secondary metabolism opens the possibility of producing new secondary compounds without major re-structuring of the enzyme complement.

The ‘classical’ approach to discover enzymes is to start with a given product and to ask what enzymatic reaction is responsible for its formation, and what the substrate is.

Substrates can be hypothesized on the basis of biochemical principles and current knowledge of metabolic pathways and types of enzymes. In other words, identification of the enzymatic properties of a newly discovered protein still depends on prior biochemical knowledge relating to the family of enzymes to which the candidate protein belongs. Thus, many new methyltransferases, terpene synthases, acytl transferases, and glycosyltransferases have been discovered because we already know representative enzymes of these types and can therefore make an informed guess regarding the potential substrates with which to test candidate proteins. Currently, the rapidly advancing technologies for genome sequencing and annotation have already greatly expedited the processes for searching for biosynthetic pathways. With the ready availability of sequence data, *in silico* genome mining has now become a standard approach for the identification of biosynthetic clusters.

References

1. Cribier B, Grosshans E: **Quadri-chromic staining in routine cutaneous histopathology (hematoxylin, eosin, saffron and astra blue)**. *Annales de dermatologie et de venereologie* 1995, **122**(11-12):758-763.
2. Abdullaev FI, Riverón-Negrete L, Caballero-Ortega H, Manuel Hernández J, Pérez-López I, Pereda-Miranda R, Espinosa-Aguirre JJ: **Use of in vitro assays to assess the potential antigenotoxic and cytotoxic effects of saffron (*Crocus sativus* L.)**. *Toxicology in Vitro* 2003, **17**(5–6):731-736.
3. Abdullaev F, Frenkel G: **Saffron in Biological and Medical Research**. *Medicinal and Aromatic Plants: Industrial Profiles* 1999:103-114.
4. Hosseinzadeh H, Nassiri-Asl M: **Avicenna's (Ibn Sina) the Canon of Medicine and Saffron (*Crocus sativus*): A Review**. *Phytotherapy Research* 2012:n/a-n/a.
5. Morimoto RI, Tissieres A, Georgopoulos C: **In The biology of heat shock proteins and molecular chaperones,**. *Spring Harbor Laboratory Press* 1994:1–30.
6. Zhang Y, Sugiura M, Saito H, .Shoyama Y: **Acute effects of *Crocus sativus* L. on passive avoidance performance in mice.** *Biol Pharmacol Bull* 1994, **17**:217–221.
7. Sugiura M, Shoyama Y, Saito H, Abe K: **Crocin (crocetin di-gentiobiose ester) prevents the inhibitory effect of ethanol on long-term potentiation in the dentate gyrus in vivo**. *Journal of Pharmacology and Experimental Therapeutics* 1994, **271**(2):703-707.
8. Sugiura M, Saito H, Abe K, Shoyama Y: **Ethanol extract of *Crocus sativus* L. Antagonizes the inhibitory action of ethanol on hippocampal long-term potentiation in vivo**. *Phytotherapy Research* 1995, **9**(2):100-104.
9. Abe K, Saito H: **Effects of saffron extract and its constituent crocin on learning behaviour and long-term potentiation**. *Phytotherapy Research* 2000, **14**(3):149-152.
10. Nair SC, Salomi MJ, Panikkae B, Panikkar KR: **Modulatory effects of *Crocus sativus* and Nigella sativa extracts on cisplatin-induced toxicity in mice**. *Journal of Ethnopharmacology* 1991, **31**(1):75-83.
11. Salomi MJ, Nair SC, Panikkar KR: **Inhibitory effects of Nigella sativa and saffron (*Crocus sativus*) on chemical carcinogenesis in mice**. *Nutrition and Cancer* 1991, **16**(1):67-72.
12. el Daly ES: **Protective effect of cysteine and vitamin E, *Crocus sativus* and *Nigella sativa* extracts on cisplatin-induced toxicity in rats**. *Journal de pharmacie de Belgique* 1998, **53**(2):87-93; discussion 93-85.
13. Nair SC, Pannikar B, Panikkar KR: **Antitumour activity of saffron (*Crocus sativus*)**. *Cancer Letters* 1991, **57**(2):109-114.
14. Abdullaev FI, Frenkel GD: **The effect of saffron on intracellular DNA, RNA and protein synthesis in malignant and non-malignant human cells**. *BioFactors (Oxford, England)* 1992, **4**(1):43-45.
15. Premkumar K, Abraham SK, Santhiya ST, Ramesh A: **Protective effects of saffron (*Crocus sativus* Linn.) on genotoxins-induced oxidative stress in Swiss albino mice**. *Phytotherapy Research* 2003, **17**(6):614-617.
16. Abdullaev FI, Frenkel GD: **Effect of saffron on cell colony formation and cellular nucleic acid and protein synthesis**. *BioFactors (Oxford, England)* 1992, **3**(3):201-204.
17. Nair SC, Salomi MJ, Varghese CD, Panikkar B, Panikkar KR: **Effect of saffron on thymocyte proliferation, intracellular glutathione levels and its antitumor activity**. *BioFactors (Oxford, England)* 1992, **4**(1):51-54.
18. Escribano J, Alonso G-L, Coca-Prados M, Fernández J-A: **Crocin, safranal and picrocrocin from saffron (*Crocus sativus* L.) inhibit the growth of human cancer cells in vitro**. *Cancer Letters* 1996, **100**(1–2):23-30.
19. Morjani H, Tarantilis P, Polissiou P, Manfait M: **Growth inhibition and induction of crythroid differentiation activity by crocin, dimethylcrocetine and β-carotene on K562 tumor cells**. *Anticancer Res* 1990:1398–1406.
20. Tarantilis P, Morjani H, Polissiou P, Manfait M: **Inhibition of growth and induction of differentiation of promyelocytic leukemia (HL-60) by carotenoids from *Crocus sativus* L.** *International Institute of Anticancer Research* 1994, **14**:1913-1918.
21. Abdullaev FI, Gonzalez de Mejia E: **Antitumor effect of plant lectins**. *Natural Toxins* 1997, **5**(4):157-163.
22. Chang WC, Lin YL, Lee MJ, Shiow SJ, Wang CJ: **Inhibitory effect of crocetin on benzo(a)pyrene genotoxicity and neoplastic transformation in C3H10T1/2 cells**. *Anticancer research* 1996, **16**(6B):3603-3608.

23. Garc-Olmo DC, Riese HH, Escribano J, Onta˜´n J, Fernandez JA, Atiénzar M, Garcí-Olmo D: **Effects of Long-Term Treatment of Colon Adenocarcinoma With Crocin, a Carotenoid From Saffron (*Crocus sativus* L.): An Experimental Study in the Rat**. *Nutrition and Cancer* 1999, **35**(2):120-126.
24. Ríos JL, Recio MC, Giner RM, Máñez S: **An Update Review of Saffron and its Active Constituents**. *Phytotherapy Research* 1996, **10**(3):189-193.
25. Alonso GL, Salinas MR, Esteban-Infantes FJ, Sánchez-Fernández MA: **Determination of Safranal from Saffron (*Crocus sativus* L.) by Thermal Desorption−Gas Chromatography**. *Journal of Agricultural and Food Chemistry* 1996, **44**(1):185-188.
26. Alonso GL, Salinas MR, Garijo J: **Method to Determine the Authenticity of Aroma of Saffron (*Crocus sativus* L.)**. *Journal of Food Protection* 1998, **61**(11):1525-1528.
27. Winterhalter P, Straubinger M: **SAFFRON—Renewed interest in an ancient spice**. *Food Reviews International* 2000, **16**(1):39-59.
28. Strasburger E: **Strasburger. Tratado de Botánica**, 35 edn; 2004.
29. Brighton CA, Mathew B, Marchant CJ: **Cytology of *Crocus sativus* L. (Iridaceae) and its allies**. *Pl Syst Evol* 1977, **128**:137–157.
30. Agayev YM: **New Features in Karyotype Structure and Origin of Saffron, *Crocus sativus* L**. *CYTOLOGIA* 2002, **67**(3):245-252.
31. Grilli-Caiola MG, Canini A: **Ultrastructure of chromoplasts and other plastids in *Crocus sativus* L. (Iridaceae).** *Plant Biosyst* 2004, **138**:9.
32. KARP A, SEBERG O, BUIATTI M: **Molecular Techniques in the Assessment of Botanical Diversity**. *Annals of Botany* 1996, **78**(2):143-149.
33. Jones DT: **Successful ab initio prediction of the tertiary structure of NK-lysin using multiple sequences and recognized supersecondary structural motifs**. *Proteins: Structure, Function and Genetics* 1997, **29**(SUPPL. 1):185-191.
34. Pejic I, Ajmone-Marsan P, Morgante M, Kozumplick V, Castiglioni P, Taramino G, Motto M: **Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs, and AFLPs**. *Theoretical and Applied Genetics* 1998, **97**(8):1248-1255.
35. Harris S: **RAPDs in Systematics - a Useful Methodology?** In: *Molecular Systematics and Plant Evolution.* CRC Press; 1999: 211-228.
36. Castillo R, Fernandez JA, Gomez-Gomez L: **Implications of carotenoid biosynthetic genes in apocarotenoid formation during the stigma development of *Crocus sativus* and its closer relatives**. *Plant Physiol* 2005, **139**(2):674-689.
37. Sang T: **Utility of Low-Copy Nuclear Gene Sequences in Plant Phylogenetics**. *Critical Reviews in Biochemistry and Molecular Biology* 2002, **37**(3):121-147.
38. Soltis DE, Soltis PS: **The Role of Phylogenetics in Comparative Genetics**. *Plant Physiology* 2003, **132**(4):1790-1800.
39. Mathew B: **Botany, taxonomy and cytology of *C. sativus* L. and its allies**. *Negbi M ed(s) Saffron: **Crocus sativus** L Amsterdam,* 1999:19-30
40. Norbaek R, Nielsen K, Kondo T: **Anthocyanins from flowers of Cichorium intybus**. *Phytochemistry* 2002, **60**(4):357-359.
41. Seberg O, Petersen G: **How many loci does it take to DNA barcode a crocus?** *PloS one* 2009, **4**(2):e4598.
42. Benschop M: **Crocus**. *The Physiology of Flower Bulbs* 1993:257-283.
43. Rubio Moraga A: **Isolation, structural characterisation and pharmacological properties of bioactive saponisn from corms of *Crocus sativus***. Albacete: Castilla-La Mancha; 2004.
44. Álvarez-Ortí M: **Desarrollo y expresión génica en cormos de azafrán (*Crocus sativus* L)**. Albacete: UCLM; 2003.
45. Molina RV, Valero M, Navarro Y, Guardiola JL, García-Luis A: **Temperature effects on flower formation in saffron (*Crocus sativus* L.)**. *Scientia Horticulturae* 2005, **103**(3):361-379.
46. Azizbekova NSH, Milyaeva EL, Lobova NV, Chailakyan MKH: **Effects of gibberellic acid and kinetin on formation of flower organs in saffron crocus**. *Sov Plant Physiol* 1978, **25**:227–233.
47. Azizbekova NSH, Milyaeva EL: **Saffron cultivation in Azerbaijan**. *Saffron **Crocus sativus** L, Harwood Academic Publishers* 1999:63-71.
48. Greenberg-Kaslasi D: **Vegetative and reproductive development in the saffron crocus (*Crocus sativus* L.).** The Hebrew University of Jerusalem; 1991.
49. Koul KK, Farooq S: **Growth and differentiation in the shoot apical meristem of the saffron plant (*Crocus sativus* L.)**. *J Indian Bot* 1984, **63**:153-160.
50. Britton G: **Carotenoids Handbook**; 2004.

51. Cunningham FX, Gantt E: **Genes and Enzymes of Carotenoid Biosynthesis in Plants**. *Annu Rev Plant Physiol Plant Mol Biol* 1998, **49**:557-583.
52. Hirschberg J: **Carotenoid biosynthesis in flowering plants**. *Curr Opin Plant Biol* 2001, **4**(3):210-218.
53. Fraser PD, Bramley PM: **The biosynthesis and nutritional uses of carotenoids**. *Prog Lipid Res* 2004, **43**(3):228-265.
54. DellaPenna D, Pogson BJ: **Vitamin synthesis in plants: tocopherols and carotenoids**. *Annu Rev Plant Biol* 2006, **57**:711-738.
55. Armstrong GA: **Genetics of eubacterial carotenoid biosynthesis: a colorful tale**. *Annu Rev Microbiol* 1997, **51**:629-659.
56. Avalos FJ, Cerdá-Olmedo E: **Fungal Carotenoid Production**. In: *Handbook of Fungal Biotechnology*. Nueva York, EEUU: Marcel Dekker Inc; 2004: 367-378.
57. Giuliano G, Al-Babili S, von Lintig J: **Carotenoid oxygenases: cleave it or leave it**. *Trends Plant Sci* 2003, **8**(4):145-149.
58. Auldridge ME, McCarty DR, Klee HJ: **Plant carotenoid cleavage oxygenases and their apocarotenoid products**. *Curr Opin Plant Biol* 2006, **9**(3):315-321.
59. Akiyama K: **Chemical identification and functional analysis of apocarotenoids involved in the development of arbuscular mycorrhizal symbiosis**. *Biosci Biotechnol Biochem* 2007, **71**(6):1405-1414.
60. Bouwmeester HJ, Roux C, Lopez-Raez JA, Bécard G: **Rhizosphere communication of plants, parasitic plants and AM fungi**. *Trends in Plant Science* 2007, **12**(5):224-230.
61. Gomez-Roldan V, Fermas S, Brewer PB, Puech-Pages V, Dun EA, Pillot JP, Letisse F, Matusova R, Danoun S, Portais JC *et al*: **Strigolactone inhibition of shoot branching**. *Nature* 2008, **455**(7210):189-194.
62. Umehara M, Hanada A, Yoshida S, Akiyama K, Arite T, Takeda-Kamiya N, Magome H, Kamiya Y, Shirasu K, Yoneyama K *et al*: **Inhibition of shoot branching by new terpenoid plant hormones**. *Nature* 2008, **455**(7210):195-200.
63. Tanaka Y, Sasaki N, Ohmiya A: **Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids**. *The Plant Journal* 2008, **54**(4):733-749.
64. Taylor LP, Grotewold E: **Flavonoids as developmental regulators**. *Current Opinion in Plant Biology* 2005, **8**(3):317-323.
65. Isman MB: **Botanical insecticides, deterrents, and repellents in modern agriculture and an increasingly regulated world**. *Annual Review of Entomology* 2005, **51**(1):45-66.
66. Iwashina T: **Flavonoid Function and Activity to Plants and Other Organisms**. *Biological Sciences in Space* 2003, **17**(1):24-44.
67. Woo H-H, Faull KF, Hirsch AM, Hawes MC: **Altered Life Cycle in Arabidopsis Plants Expressing PsUGT1, a UDP-Glucuronosyltransferase-Encoding Gene from Pea**. *Plant Physiology* 2003, **133**(2):538-548.
68. Woo H-H, Jeong B, Hawes M: **Flavonoids: from cell cycle regulation to biotechnology**. *Biotechnology Letters* 2005, **27**(6):365-374.
69. Dugas N, Dereuddre-Bosquet N, Goujard C, Dormont D, Tardieu M, Delfraissy JF: **Role of nitric oxide in the promoting effect of HIV type 1 infection and of gp120 envelope glycoprotein on interleukin 4-induced IgE production by normal human mononuclear cells**. *AIDS Res Hum Retroviruses* 2000, **16**(3):251-258.
70. Ross JA, Kasum CM: **DIETARY FLAVONOIDS: Bioavailability, Metabolic Effects, and Safety**. *Annual Review of Nutrition* 2002, **22**(1):19-34.
71. Holiman PCH, Hertog MGL, Katan MB: **Analysis and health effects of flavonoids**. *Food Chemistry* 1996, **57**(1):43-46.
72. Steinmetz KA, Potter JD: **Vegetables, fruit, and cancer prevention: a review**. *Journal of the American Dietetic Association* 1996, **96**(10):1027-1039.
73. Winkel-Shirley B: **Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology**. *Plant Physiology* 2001, **126**(2):485-493.
74. Forkmann G, Martens S: **Metabolic engineering and applications of flavonoids**. *Current Opinion in Biotechnology* 2001, **12**(2):155-160.
75. Sandmann G, Romer S, Fraser PD: **Understanding carotenoid metabolism as a necessity for genetic engineering of crop plants**. *Metab Eng* 2006, **8**(4):291-302.
76. Wu S, Chappell J: **Metabolic engineering of natural products in plants; tools of the trade and challenges for the future**. *Current Opinion in Biotechnology* 2008, **19**(2):145-152.
77. Giuliano G, Tavazza R, Diretto G, Beyer P, Taylor MA: **Metabolic engineering of carotenoid biosynthesis in plants**. *Trends in Biotechnology* 2008, **26**(3):139-145.

78. Kanakis CD, Daferera DJ, Tarantilis PA, Polissiou MG: **Qualitative determination of volatile compounds and quantitative evaluation of safranal and 4-hydroxy-2,6,6-trimethyl-1-cyclohexene-1-carboxaldehyde (HTCC) in Greek saffron**. *J Agric Food Chem* 2004, **52**(14):4515-4521.
79. Rubio A, Rambla JL, Santaella M, Gomez MD, Orzaez D, Granell A, Gomez-Gomez L: **Cytosolic and plastoglobule-targeted carotenoid dioxygenases from *Crocus sativus* are both involved in beta-ionone release**. *J Biol Chem* 2008, **283**(36):24816-24825.
80. Mookherjee BD, Trenkle RW, Wilson RA: **The chemistry of flowers, fruits and spices: live vs. dead a new dimension in fragrance research** *Pure &App/ Chern* 1990, **62**:1357-1364.
81. D'Auria M, Mauriello G, Racioppi R, Rana GL: **Use of SPME-GC-MS in the study of time evolution of the constituents of saffron aroma: Modifications of the composition during storage**. *Journal of Chromatographic Science* 2006, **44**(1):18-21.
82. Pfander H, Wittwer F: **Carotenoid glycosides. II. Carotenoid composition of safran**. *Carotinoid Glykoside 2 Mitteilung Untersuchungen zur Carotinoid Zusammensetzung im safran* 1975, **58**(6):1608-1620.
83. Tarantilis PA, Tsoupras G, Polissiou M: **Determination of saffron (*Crocus sativus* L.) components in crude plant extract using high-performance liquid chromatography-UV-visible photodiode-array detection-mass spectrometry**. *J Chromatogr A* 1995, **699**(1-2):107-118.
84. Pfander H, Schurtenberger H: **Biosynthesis of C20-Carotenoids in *Crocus sativus***. *Phytochemistry* 1982, **21**:4.
85. Straubinger M, Jezussek M, Waibel R, Winterhalter P: **Novel Glycosidic Constituents from Saffron**. *Journal of Agricultural and Food Chemistry* 1997, **45**(5):1678-1681.
86. Straubinger M, Bau B, Eckstein S, Fink M, Winterhalter P: **Identification of Novel Glycosidic Aroma Precursors in Saffron (*Crocus sativus* L.)**. *Journal of Agricultural and Food Chemistry* 1998, **46**(8):3238-3243.
87. Zarghami NS, Heinz DE: **Monoterpene aldehydes and isophorone-related compounds of saffron**. *Phytochemistry* 1971, **10**(11):2755-2761.
88. Rödel W, Petrzika M: **Analysis of the volatile components of saffron**. *Journal of High Resolution Chromatography* 1991, **14**(11):771-774.
89. Tarantilis PA, Polissiou M: **Isolation and identification of the aroma constituents of saffron (Crocus sativa)**. *Journal of Agriculture and Food Chemistry* 1997, **45**:4.
90. Carmona M, Sánchez AM, Ferreres F, Zalacain A, Tomás-Barberán F, Alonso GL: **Identification of the flavonoid fraction in saffron spice by LC/DAD/MS/MS: Comparative study of samples from different geographical origins**. *Food Chemistry* 2007, **100**(2):445-450.
91. Nørbæk R, Brandt K, Kvist Nielsen J, Ørgaard M, Jacobsen N: **Flower pigment composition of Crocus species and cultivars used for a chemotaxonomic investigation**. *Biochemical Systematics and Ecology* 2002, **30**(8):28.
92. Bowles D, Lim E-K, Poppenberger B, Vaistij FE: **Glycosyltransferases of lipophilic small molecules**. *Annual Review of Plant Biology* 2006, **57**(1):567-597.
93. Hughes J, Hughes MA: **Multiple secondary plant product UDP-glucose glucosyltransferase genes expressed in cassava (Manihot esculenta Crantz) cotyledons**. *DNA Seq* 1994, **5**(1):41-49.
94. Li Y, Baldauf S, Lim EK, Bowles DJ: **Phylogenetic analysis of the UDP-glycosyltransferase multigene family of Arabidopsis thaliana**. *J Biol Chem* 2001, **276**(6):4338-4343.
95. Lim EK, Baldauf S, Li Y, Elias L, Worrall D, Spencer SP, Jackson RG, Taguchi G, Ross J, Bowles DJ: **Evolution of substrate recognition across a multigene family of glycosyltransferases in Arabidopsis**. *Glycobiology* 2003, **13**(3):139-145.
96. Henrissat B: **A classification of glycosyl hydrolases based on amino acid sequence similarities**. *Biochemical Journal* 1991, **280**(Pt 2):309.
97. Henrissat B, Davies G: **Structural and sequence-based classification of glycoside hydrolases**. *Current opinion in structural biology* 1997, **7**(5):637-644.
98. Duroux L, Delmotte F, Lancelin J, Keravis G, Jay-Allemand C: **Insight into naphthoquinone metabolism: beta-glucosidase-catalysed hydrolysis of hydrojuglone beta-D-glucopyranoside**. *Biochemical Journal* 1998, **333**(Pt 2):275.
99. Osbourn AE: **Preformed antimicrobial compounds and plant defense against fungal attack**. *The plant cell* 1996, **8**(10):1821.
100. Sue M, Yamazaki K, Yajima S, Nomura T, Matsukawa T, Iwamura H, Miyamoto T: **Molecular and structural characterization of hexameric β-D-glucosidases in wheat and rye**. *Plant Physiology* 2006, **141**(4):1237-1247.

101. Dharmawhardana D, Ellis B, Carlson I: **cDNA cloning and heterologous expression of coniferin b-glucosidase**. *Plant Mol Biol* 1999, **40**:365-372.
102. HÖSEL W, SURHOLT E, BORGMANN E: **Characterization of β-Glucosidase Isoenzymes Possibly Involved in Lignification from Chick Pea (Cicer arietinum L.) Cell Suspension Cultures**. *European Journal of Biochemistry* 1978, **84**(2):487-492.
103. Leah R, Kigel J, Svendsen I, Mundy J: **Biochemical and molecular characterization of a barley seed b-glucosidase**. *J Biol Chem* 1995, **26**:15789-15797.
104. Ueda M, Yamamura S: **Chemistry and biology of plant leaf movements**. *Angewandte Chemie International Edition* 2000, **39**(8):1400-1414.
105. Minic Z: **Physiological roles of plant glycoside hydrolases**. *Planta* 2008, **227**(4):723-740.
106. Birk R, Bravdo B, Shoseyov O: **Detoxification of cassava by Aspergillus niger B-1**. *Applied microbiology and biotechnology* 1996, **45**(3):411-414.
107. Pemberton M, Brown R, Emert G: **The role of bglucosidase in the by conversion of cellulose to ethanol**. *Canadian J Chem Eng* 1980, **58**:723-729.
108. Woodward J, Wiseman A: **Fungal and other ß-glucosidases. Their properties and applications**. *Enz Micro Technol* 1982, **4**:73-79.
109. Günata Z, Dugelay I, Sapis J, Baumes R, Bayonove C: **Role of enzymes in the use of the flavour potential from grape glycosides in winemaking**. *Progress in flavour precursor studies* 1993:219-234.
110. Winterhalter P, Skouroumounis G: **Glycoconjugated aroma compounds: occurrence, role and biotechnological transformation**. *Biotechnology of aroma compounds* 1997:73-105.
111. Zeevaart JAD, Creelman RA: **Metabolism and Physiology of Abscisic Acid**. *Annual Review of Plant Physiology and Plant Molecular Biology* 1988, **39**(1):439-473.
112. Schwartz SH, Tan BC, Gage DA, Zeevaart JA, McCarty DR: **Specific oxidative cleavage of carotenoids by VP14 of maize**. *Science* 1997, **276**(5320):1872-1874.
113. von Lintig J, Vogt K: **Filling the gap in vitamin A research. Molecular identification of an enzyme cleaving beta-carotene to retinal**. *J Biol Chem* 2000, **275**(16):11915-11920.
114. Wyss A, Wirtz G, Woggon WD, Brugger R, Wyss M, Friedlein A, Bachmann H, Hunziker W: **Cloning and Expression of β,β-Carotene 15,15′-Dioxygenase**. *Biochemical and Biophysical Research Communications* 2000, **271**(2):334-336.
115. Kiefer C, Hessel S, Lampert JM, Vogt K, Lederer MO, Breithaupt DE, von Lintig J: **Identification and characterization of a mammalian enzyme catalyzing the asymmetric oxidative cleavage of provitamin A**. *J Biol Chem* 2001, **276**(17):14110-14116.
116. Huang C-S, Fan Y-E, Lin C-Y, Hu M-L: **Lycopene inhibits matrix metalloproteinase-9 expression and down-regulates the binding activity of nuclear factor-kappa B and stimulatory protein-1**. *The Journal of Nutritional Biochemistry* 2007, **18**(7):449-456.
117. Schwartz SH, Qin X, Zeevaart JA: **Characterization of a novel carotenoid cleavage dioxygenase from plants**. *J Biol Chem* 2001, **276**(27):25208-25211.
118. Bouvier F, Suire C, Mutterer J, Camara B: **Oxidative remodeling of chromoplast carotenoids: identification of the carotenoid dioxygenase CsCCD and CsZCD genes involved in Crocus secondary metabolite biogenesis**. *Plant Cell* 2003, **15**(1):47-62.
119. Simkin AJ, Schwartz SH, Auldridge M, Taylor MG, Klee HJ: **The tomato carotenoid cleavage dioxygenase 1 genes contribute to the formation of the flavor volatiles beta-ionone, pseudoionone, and geranylacetone**. *Plant J* 2004, **40**(6):882-892.
120. Mathieu S, Terrier N, Procureur J, Bigey F, Gunata Z: **A carotenoid cleavage dioxygenase from Vitis vinifera L.: functional characterization and expression during grape berry development in relation to C13-norisoprenoid accumulation**. *J Exp Bot* 2005, **56**(420):2721-2731.
121. Ibdah M, Azulay Y, Portnoy V, Wasserman B, Bar E, Meir A, Burger Y, Hirschberg J, Schaffer AA, Katzir N *et al*: **Functional characterization of CmCCD1, a carotenoid cleavage dioxygenase from melon**. *Phytochemistry* 2006, **67**(15):1579-1589.
122. Simkin AJ, Underwood BA, Auldridge M, Loucas HM, Shibuya K, Schmelz E, Clark DG, Klee HJ: **Circadian regulation of the PhCCD1 carotenoid cleavage dioxygenase controls emission of beta-ionone, a fragrance volatile of petunia flowers**. *Plant Physiol* 2004, **136**(3):3504-3514.
123. Vogel JT, Tan BC, McCarty DR, Klee HJ: **The carotenoid cleavage dioxygenase 1 enzyme has broad substrate specificity, cleaving multiple carotenoids at two different bond positions**. *J Biol Chem* 2008, **283**(17):11364-11373.
124. Garcia-Limones C, Schnabele K, Blanco-Portales R, Luz Bellido M, Caballero JL, Schwab W, Munoz-Blanco J: **Functional characterization of FaCCD1: a carotenoid cleavage**

dioxygenase from strawberry involved in lutein degradation during fruit ripening. *J Agric Food Chem* 2008, **56**(19):9277-9285.

125. Floss DS, Walter MH: **Role of carotenoid cleavage dioxygenase 1 (CCD1) in apocarotenoid biogenesis revisited**. *Plant Signal Behav* 2009, **4**(3):172-175.
126. Ilg A, Beyer P, Al-Babili S: **Characterization of the rice carotenoid cleavage dioxygenase 1 reveals a novel route for geranial biosynthesis**. *FEBS J* 2009, **276**(3):736-747.
127. Huang FC, Molnar P, Schwab W: **Cloning and functional characterization of carotenoid cleavage dioxygenase 4 genes**. *J Exp Bot* 2009, **60**(11):3011-3022.
128. Bouvier F, Dogbo O, Camara B: **Biosynthesis of the food and cosmetic plant pigment bixin (annatto)**. *Science* 2003, **300**(5628):2089-2091.
129. Tan BC, Joseph LM, Deng WT, Liu L, Li QB, Cline K, McCarty DR: **Molecular characterization of the Arabidopsis 9-cis epoxycarotenoid dioxygenase gene family**. *Plant J* 2003, **35**(1):44-56.
130. Auldridge ME, Block A, Vogel JT, Dabney-Smith C, Mila I, Bouzayen M, Magallanes-Lundback M, DellaPenna D, McCarty DR, Klee HJ: **Characterization of three members of the Arabidopsis carotenoid cleavage dioxygenase family demonstrates the divergent roles of this multifunctional enzyme family**. *Plant J* 2006, **45**(6):982-993.
131. Ytterberg AJ, Peltier JB, van Wijk KJ: **Protein profiling of plastoglobules in chloroplasts and chromoplasts. A surprising site for differential accumulation of metabolic enzymes**. *Plant Physiol* 2006, **140**(3):984-997.
132. Zou J, Zhang S, Zhang W, Li G, Chen Z, Zhai W, Zhao X, Pan X, Xie Q, Zhu L: **The rice HIGH-TILLERING DWARF1 encoding an ortholog of Arabidopsis MAX3 is required for negative regulation of the outgrowth of axillary buds**. *Plant J* 2006, **48**(5):687-698.
133. Johnson X, Brcich T, Dun EA, Goussot M, Haurogné K, Beveridge CA, Rameau C: **Branching Genes Are Conserved across Species. Genes Controlling a Novel Signal in Pea Are Coregulated by Other Long-Distance Signals**. *Plant Physiology* 2006, **142**(3):1014-1026.
134. Booker J, Sieberer T, Wright W, Williamson L, Willett B, Stirnberg P, Turnbull C, Srinivasan M, Goddard P, Leyser O: **MAX1 encodes a cytochrome P450 family member that acts downstream of MAX3/4 to produce a carotenoid-derived branch-inhibiting hormone**. *Developmental Cell* 2005, **8**:7.
135. Booker J, Auldridge M, Wills S, McCarty D, Klee H, Leyser C: **MAX3/CCD7 is a carotenoid cleavage dioxygenase required for the synthesis of a novel plant signaling molecule**. *Current biology* 2004, **14**:7.
136. Sorefan K, Booker J, Haurogné K, Goussot M, Bainbridge K, Foo E, Chatfield S, Ward S, Beveridge C, Rameau C *et al*: **MAX4 and RMS1 are orthologous dioxygenase-like genes that regulate shoot branching in Arabidopsis and pea**. *Genes & Development* 2003, **17**(12):1469-1474.
137. Schwartz SH, Qin X, Loewen MC: **The biochemical characterization of two carotenoid cleavage enzymes from Arabidopsis indicates that a carotenoid-derived compound inhibits lateral branching**. *J Biol Chem* 2004, **279**(45):46940-46945.
138. Snowden KC, Simkin AJ, Janssen BJ, Templeton KR, Loucas HM, Simons JL, Karunairetnam S, Gleave AP, Clark DG, Klee HJ: **The Decreased apical dominance1/Petunia hybrida CAROTENOID CLEAVAGE DIOXYGENASE8 gene affects branch production and plays a role in leaf senescence, root growth, and flower development**. *Plant Cell* 2005, **17**(3):746-759.
139. Arite T, Iwata H, Ohshima K, Maekawa M, Nakajima M, Kojima M, Sakakibara H, Kyozuka J: **DWARF10, an RMS1/MAX4/DAD1 ortholog, controls lateral bud outgrowth in rice**. *Plant J* 2007, **51**(6):1019-1029.
140. Alder A, Holdermann I, Beyer P, Al-Babili S: **Carotenoid oxygenases involved in plant branching catalyse a highly specific conserved apocarotenoid cleavage reaction**. *Biochem J* 2008, **416**(2):289-296.
141. **http://www.who.int/mediacentre/factsheets/en**
142. Carter GT: **Natural products and Pharma 2011: Strategic changes spur new opportunities**. *Natural product reports* 2011, **28**(11):1783-1789.
143. Kutchan TM: **Alkaloid Biosynthesis [mdash] The Basis for Metabolic Engineering of Medicinal Plants**. *The Plant Cell* 1995, **7**(7):1059.
144. Morant M, Bak S, Møller BL, Werck-Reichhart D: **Plant cytochromes P450: tools for pharmacology, plant protection and phytoremediation**. *Current opinion in biotechnology* 2003, **14**(2):151-162.

Objectives

Objectives

The main objective of this study is to better understand the biosynthesis of secondary metabolites in *C. sativus*, especially, those related to carotenoids and flavonoids with pharmacological and industrial interest and to identify and characterise the genes involved in its formation. To attain this purpose the following points have been designed:

- Isolation and characterization of different glycosiltransferases involve in synthesis of flavonoids in *C. sativus*.
- Genomic analysis and gene structure of the plant carotenoid dioxygenase 4 family in *C. sativus*.
- Accumulation and transport of apocarotenoids in different stage of *C. sativus*
- *In silico* identification of new glycosyltransferases and β-glucosidases in different stage of *C. sativus* stigma.

Chapters

Chapter 1

“Cloning and characterization of a glucosyltransferase from *Crocus sativus* stigmas involved in flavonoid glucosylation”

Abstract

Flavonol glucosides constitute the second group of secondary metabolites that accumulate in *Crocus sativus* stigmas. To date there are no reports of functionally characterized flavonoid glucosyltransferases in *C. sativus*, despite the importance of these compounds as antioxidant agents. Moreover, their bitter taste makes them excellent candidates for consideration as potential organoleptic agents of saffron spice, the dry stigmas of *C. sativus*.

Using degenerate primers designed to match the plant secondary product glucosyltransferase (PSPG) box, we cloned a full length cDNA encoding CsGT45 from *C. sativus* stigmas. This protein showed homology with flavonoid glucosyltransferases. *In vitro* reactions showed that CsGT45 catalyses the transfer of glucose from UDP_glucose to kaempferol and quercetin. Kaempferol is the unique flavonol present in *C. sativus* stigmas and the levels of its glucosides changed during stigma development, and these changes, are correlated with the expression levels of CsGT45 during these developmental stages.

Findings presented here, suggest that CsGT45 is an active enzyme that plays a role in the formation of flavonoid glucosides in *C. sativus*.

Keywords: *Crocus sativus*, development, flavonoids, glucosyltransferase, saffron, stigma

Introduction

Flavonols constitute a major class of plant natural products that accumulate in a wide range of conjugate structures. A large proportion of this diversity is due to the attachment of one or several sugar moieties at different positions. Besides providing beautiful pigmentation in flowers, fruits, seeds, and leaves [1], flavonoids also have key roles in signaling between plants and microbes, in male fertility of some species [2], in defence as antimicrobial agents and feeding deterrents [3], in UV protection [4], in the regulation of polar transport of auxins [5], and more recently, their role in cell cycle regulation in plants has been demonstrated [6], [7]. There is increasing evidence to suggest that flavonoids, in particular those belonging to the class of flavonols (such as kaempferol and quercetin), are potentially health-protecting components in the human diet as a result of their high antioxidant capacity [8], [9]. Therefore, flavonoids may offer protection against major diseases such as coronary heart diseases and cancer [10], [11]. Flavonoids are present at relatively high concentrations in saffron, the desiccated stigma tissue of *C. sativus* [12], [13]. Their antioxidant properties, along with their bitter taste, could qualify them as potential organoleptic agents of the spice [13], [14], [15]. In addition, they show anticonceptive and anti-inflammatory effects [16]. Nevertheless, the studies of these compounds in saffron stigma are scarce, and have only been analysed with some detail in tepals [17], [18].

Flavonoid synthesis is organ- and tissue-dependent, and is affected by environmental conditions [19]. In the early steps of flavonoid biosynthesis, phenylalanine derived from the shikimic acid pathway is converted to coumaroyl-CoA by phenylalanine ammonia-lyase, cinnamate 4-hydroxylase, and 4-coumarate: CoA ligase. Chalcone synthase, the first committed enzyme for flavonoid biosynthesis, results in the condensation of coumaroyl-CoA with three molecules of malonyl-CoA from acetyl-CoA to form naringenin chalcone, which suffers further modifications that result in the synthesis of substitute flavones, flavonols, catechins, deoxyflavonoids, and anthocyanins. The flavonoid aglycones, which have a variety of glycosylation sites, are converted into glycon by glycosyltransferases.

In higher plants, secondary metabolites are often converted to their glycoconjugates, which are then accumulated and compartmentalized in vacuoles [20], while glycosylation of phytochemicals is known to alter their regulatory properties by causing enhanced water solubility and lower chemical reactivity. Glycosylation involves a

UGT-catalysed transfer of a nucleotide diphosphate-activated sugar molecule to the acceptor aglycone [21]. The glycosylation reactions are catalysed by glycosyltransferases (GTases). Among these GTases, family 1 GTases (UGTs), commonly utilize small molecular weight compounds as acceptor molecule substrates and UDP-sugars as donors [22]. The first gene encoding a plant glycosyltransferase was isolated in *Zea mays*, during the analysis of the Bronze locus, which codes for an UDP-glucose:flavonol glucosyltransferase [23]. Since then, several clones have been characterized at a molecular level in a range of species including *Petunia hybrida* [24] [25], *Vitis vinifera* [26], *Perilla frutescens* [27], *Allium cepa* [28], *Nicotiana tabacum* [29], *Arabidopsis thaliana* [30], [31], [32], [33], [34], *Dianthus caryophyllus* [35], *Beta vulgaris* [36], *Glycine max* [37]; *Pyrus communis* [38], *Oryza sativa* [39], [40] and *Fragaria×ananassa* [41] among others.

Here the isolation of a UDP-glucose:flavonol glucosyltransferase *C. sativus* stigmas using a degenerate PCR technique is reported. The substrate specificity analyses using recombinant protein indicated that *C. sativus* flavonol GT, CsGT45, was able to catalyse glucosylation of kaempferol and quercetin. Interestingly, CsGT45 was not expressed in *Crocus* species unable to accumulate kaempferol 7-O-glucosides in stigmas, suggesting the involvement of CsGT45 in the formation of kaempferol glucosides in the stigma tissue of *C. sativus*.

Materials and Methods

Chemicals and Plant materials

Chemicals and reagents were obtained from Sigma-Aldrich unless otherwise stated. Plant tissues and stigmas from *C. sativus* grown under field conditions in Tarazona de La Mancha, Spain, were used throughout the experiments. *C. cancellatus*, *C. speciosus*, *C. niveus* and *C. cartwrightianus* were obtained from Dr. U. Jacobsen from the Agricultural University of Denmark. Stigmas were collected at the developmental stages previously described [42], and defined as follows: yellow stigma, closed bud inside the perianth tubes (around 0.3 cm in length); orange stigma, closed bud inside the perianth tubes (around 0.4 cm in length); red stigma, closed bud inside the perianth tubes (0.8 cm in length); -2da, two days before anthesis, dark red stigma in closed bud outside the perianth tubes (3 cm in length); da, day of anthesis, dark red stigma (3 cm in length); +1da, one day after anthesis, dark red stigma and +3da, three days after anthesis, dark red stigma. Tepals, style and stamens were collected from flowers at the time of anthesis and together with corms were frozen in liquid nitrogen and stored at -80°C until required. To determine stress-induced gene expression in leaves, whole leaves were collected from plants growing in fields, cut into 1 cm-long pieces and transferred to 24-well-plates containing 1 ml water supplemented with abscisic acid (ABA) (100μM), 2,4-dichlorofenoxiacetic acid (2,4D) (100μM), Giberellic acid (GA3) (100μM), 0.2 μl/ml Methyl Jasmonate (MetJa), 200 mM CaCl2, 1 mM potassium salicylate (SA), pH 6.5, and distilled water that was used as a control. All the samples were incubated under normal conditions (16 h light/8 h dark cycles at 22°C). Wounding was performed on leaves with a sterile needle and samples were then frozen immediately in liquid nitrogen and were stored at –80 ºC until used.

Cloning of *C. sativus* GTase cDNA

As a first step in identifying GTases genes expressed in saffron stigmas, total RNA and mRNA were isolated from developed saffron stigmas by using Ambion PolyAtrack and following manufacturer's protocols (Ambion Inc., Austin, TX, USA). First-strand cDNAs were synthesized by reverse transcription (RT) from 2 µg of total RNA using an 18-base pair oligo dT primer and a first-strand cDNA synthesis kit (Amersham Biosciences) according to manufacturer’s instructions. These cDNAs were used as templates for PCR using degenerate primers designed based on the conserved regions of the plants GTases [43]. The primers used were: glut-f (5’-TSNGTNGCNTAYGTNTSNTTYGG-3’) and glut-r (5’-TTCCANCCRCARTGNGTNACRAA-3’). Anchored PCR with gene-specific primers were used to analyse and identify the 5’ and 3´ ends of the glucosyltransferases. For these, 1µg of poly(A)+ RNA from stigmas was used to synthesize the 5’ and 3’ ends of the first-strand cDNA using Superscript II reverse transcriptase, using the primers 5’-CDS primer and SMARTII-A oligo for the 5’-RACE reaction and the 3’-CDS primer A for the 3’-RACE reaction supplied in the SMARTTM RACE cDNA Amplification kit (Clontech-Takara). Following dilution, the first-strand reaction product was subjected to PCR for amplification. We used the gene-specific primers CsGT45-f1 (5´-AGCTGTCGATAAGATGGATATC-3´), and CsGT45-f2 (5’-GAGTGTCTTATCGCATCCT-3’) as forward primers, and CsGT45-r1 (5’-gaagccaagctctccatcgtcga-3’) and CsGT45-r2 (5’- ctccagctggccgaatgtgttc-3’) as reverse primers in combination with the universal primer mix from the SMART RACE kit as the reverse/forward primer with the following cycling program: one cycle at 94°C for 3 min, 10 cycles at 94°C for 20 s, 66°C-0.2°C/cycle for 20 s, and 72°C for 2 min, 30 cycles at 94°C for 20 s, 64°C for 20 s and 72°C for 2 min, and a final extension at 72°C for 5 min. The amplified PCR products were analysed by electrophoresis in 1% agarose gel. The PCR products were then cloned into pGEM-T (Promega Corporation, Madison, WI, USA). The ligated DNA was transformed into *E. coli* strain JM109. The clones (20 colonies) were picked individually and amplified in 3 ml of LB medium at 37°C overnight. The plasmid DNA from each clone was extracted using a DNA plasmid Miniprep kit (Promega Corporation, Madison, WI, USA) and then analysed by EcoRI restriction digestion. One clone of each size was subjected to sequencing using an automated DNA sequencer (ABI prism 310, Perkin Elmer) with either the T7 or Sp6 sequencing primers. Computer-aided sequence similarity searches were made with the BLAST suite of programs at the National Centre for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov) Motif searches were made using PROSITE (http://expasy.hcuge.ch/sprot/prosite.html), TMPRED (http://www.isrec.isb-sib.ch/sofware/sofware.html), Signal IP (http://www.cbs.dtu.dk/services/SignalP) and PSORT II (http://psort.nibb.ac.jp). Once the 5' and 3' sequences were determined, the full-length clone CsGT45 was amplified from the cDNA and genomic DNA with the following primer sequences: the forward primer, 5’-CAGATGGACCAACATCAGCCT-3’; and the reverse primer 5’-ATTATCTCAACACCTGTGTGG -3’.

Heterologous expression

The full-length open reading frame of CsGT45 cDNA was amplified by PCR using Pfu polymerase (Promega). The oligonucleotide sequences for CsGT45 cloning were as follows: the forward primer 5’-ACCAACATCAGCCTAACATT-3’, and the reverse primer 5’-TGCGGCCGCTCCTCCTTTAAGAGGGTGA-3’. Using these primers, the generated product has a NotI site at the 3' end (underlined in the reverse primer). The PCR product was cloned directionally (SmaI-NotI) into bacterial GST expression vector pGEX-5T-3 (Amersham Biosciences/GE Healthcare) to create in-frame fusions at the 5' terminus with the GST coding sequence. The construct was sequenced to confirm that the gene was in the correct reading frame. After transformation into BL21(DE3) *E. coli* cells, colonies were selected on LB containing ampicillin (AMP) plates. Individual colonies were grown overnight in 5 ml of LB-AMP medium at 20°C, and 2.5 ml of the culture was used to inoculate 500 ml of LB-AMP fresh medium. Cells were grown at

20°C until an A600 of 0.6 was reached, after which the culture was induced with 0.5mM IPTG and allowed to grow for 16 h at 20°C. The cells were harvested by centrifugation at 5,000g for 10 min and resuspended in 20 ml PBS. Resuspended cells were sonicated with a microtip probe in ice until the viscosity disappeared. After sonication, the samples were centrifuged at 10,000g for 25 min. The supernatant and pellet were tested by PAGE (polyacrylamide gel electrophoresis)/SDS for solubility of the fusion protein by coomasie stain. The soluble proteins were applied to a glutathione Sepharose column for purification following manufacturer instructions (Amersham Biosciences/GE Healthcare). Protein concentration was determined according to the Bradford method [44], using serum albumin as standard.

Enzyme assays and analysis of reaction products

The affinity-purified enzyme was used to determine substrate specificity and enzymatic parameters. In a final assay volume of 200 μl, the reaction conditions were 50 mM Tris-HCl, pH 7.5, 14 mM 2-mercaptoethanol, 2.5 mM UDP-glucose, the recombinant enzyme (7.0 μg) and the corresponding substrates: 100 μM Quercetin, 100 μM kaempferol, 100 μM cyanidin, 1 mM trans-cinnamic acid, 1 mM sinapic acid, 1 mM indole acetic acid, 1 mM abscisic acid and 1 μM crocetin encapsulated in maltosyl-β-cyclodextrin as described [45]. All the glucosyltranferase activity assays were carried out at 30°C for 30 minutes. For the determination of the Km values of substrates, the concentrations of kaempferol and quercetin were varied from 20 to 100 μM, at a fixed UDP-glucose concentration of 2.5mM and 7.0 μg of the purified enzyme. These enzymatic reactions were performed at 30°C for 10, 15 and 20 minutes for each substrate concentration. The reactions were terminated, and the proteins precipitated, by the addition of 20 μl of trichloroacetic acid (240 mg/ml). Subsequently, samples were centrifuged at 15,000g for 5 min to collect the supernatant, and aliquots were analysed by reverse-phase HPLC as previously described [46] using a C18 Ascentis, 25 x 4.6, particle size 5 μm column (Supelco, Sigma-Aldrich). Km values were determined from Lineweaver-Burk plots of initial rate data. The assays were also analysed by HPLC DAD detector and by electrospray ionization (ESI)-mass spectrometry (MS) for the formation of glycosylated products as previously described [46] using a C18 Ascentis, 25 x 4.6, particle size 3 um column (Supelco, Sigma-Aldrich). A standard curve for peak area of quercetin and kaempferol was produced by injecting known amounts of these flavonoids. For individual quercetin glucosides, standard curves were constructed indirectly by calculating the amount of quercetin or kaempferol released after quantitative hydrolysis of the glucoside with β-glucosidase from almonds (60 min, pH 6, 30°C). The assignment of a glucosylation position to quercetin and kaempferol was determined indirectly following the method by Mabry et al. [47] for the identification of flavonoids, using published data [31], [48], [49] and by comparison with authentic flavonol standards: kaempferol 7-O-glucoside and quercetin 4'-O-glucoside (TransMIT Flavonoidforschung, DE) and quercetin 3-O-glucoside (Sigma-Aldrich).

Flavonoid analysis in stigma tissue

Stigmas at the time of anthesis were ground in liquid nitrogen. The fine powder obtained was extracted with methanol (500 µl), centrifuged and the supernatant analysed by LC-ESI-MS using a C18 Ascentis column 15 x 2.1, particle size 3 um (Supelco, Sigma-Aldrich) and following the method previously described [46]. For flavonoid analysis from stigmas at different developmental stages (yellow to +3da), three stigmas of each stage were collected and freeze-dried. The powder obtained from one stigma was extracted with 500 µl of methanol containing 0.2 mg/ml 4-methylumbelliferyl β-D-glucuronide as an internal standard. The samples were centrifuged (5,000g, 10 min), and the supernatant evaporated and treated as described [15]. Samples in triplicate were analysed by HPLC as described [46] using a C18 Ascentis, 25 x 4.6, particle size 5 um column (Supelco, Sigma-Aldrich).

Analysis of mRNA levels in different tissues and stress conditions

Reverse Transcription-PCR (RT-PCR) was used to determine the relative levels of CsGT45 and UGTCs3 messages. Total RNA was isolated from control and treated leaves, tepals, stamens, stigmas, styles and corms using the Trizol reagent (Gibco-BRL). The RNA was resuspended in 100 µl of RNase-free water and treated with RQ1 RNase-free DNase (Promega). The DNase was heat inactivated before RT-PCR. The RNA was quantified with a spectrophotometer and stored at -80°C. Various initial concentrations of mRNA, ranging over 10 fold difference, were used to demonstrate the differential accumulation of the mRNA in the tissues analysed. First-strand cDNAs were synthesized by RT from 2 µg of total RNA using a first-strand cDNA synthesis kit (Pharmacia) and random primers. Conditions for semi-quantitative RT-PCR were as follows: 65°C for 5 min, followed by 37°C for 1 h, followed by 75°C for 5 min. The cDNAs obtained were used as templates for PCR using the CsGT45 gene-specific primers: 5’-GATGGGGAGAGAGGTGTTGA-3’ and 5’-

TCCTCGCAATGCTGTCTATG-3’, and for the amplification of gene coding for the 18S ribosomal RNA (RPS18) the primers used were: 5’-AGTTTGAGGCAATAACAGGTCT-3’ and 5’-GATGAAATTTCCCAAGATTACC-3’. UGTCs3 was amplified using the primers previously described [46]. Thermal cycling parameters were 2 min at 95°C, 30 x (20 s at 95°C, 20 s at 60°C, and 30 s at 72°C). The PCR products were separated in a 2% agarose gel. The gels were photographed with the IP-010-Sd photo-documentation system (Vilber Lourmat). The PhotoCaptMw programme was used to quantify the intensity of the ethidium bromide stained DNA bands from the positive images of the gel. These experiments were repeated three times and averaged for each sample. To correct the initial mRNA levels, each intensity score was normalized to the intensity for the RPS18 gene amplification.

List of Abbreviations

C4H, cinnamate-4-hysroxylase; GTases, glycosyltransferases; HPLC-ESI-MS, HPLC-electrospray ionization-mass spectrometry; PAL, L-phenylalanine ammonia lyase;. RACE, rapid amplification of cDNA ends; RT-PCR, reverse transcription-PCR; UDP, uridine diphosphate glucose; UGT, UDP-glycosyltransferase.

Results

Profile of flavonols accumulation during stigma tissue development

In saffron, the flavonoids kaempferol 3-O-sophoroside-7-O-glucopyranoside and kaempferol 7-O-sophoroside were identified as abundant compounds [12], [13], and more recently, a kaempferol tetrahexoside and kaempferol 3,7,4’-triglucoside have been tentatively identified as minor flavonoids in saffron [15], whereas quercetin and its glucosides have not been detected. Initially the content of flavonoids present in *C. sativus* stigma at anthesis was analysed by LC-ESI-MS (Fig. 1A). In addition, six stigma developmental stages were selected and methanol extracts were analysed by HPLC. Under our experimental conditions, three significant flavonoids were evident in the HPLC chromatograms from extracts of *C. sativus* stigmas (Fig. 1A). The retention times, the UV spectra and the LC-ESI-MS analysis on stigmas at anthesis allowed us to tentatively identify these flavonoids as 3-O-sophoroside-7-O-glucopyranoside, 3,7,4’-triglucoside and 7-O-sophoroside (Fig. 1B), which was also characterized by NMR analysis and observed that the obtained structural data correspond to those found in the literature [13]. The presence of all three flavonoids increased with stigma development and the increase for the two kaempferol triglucosides was equal. The relative levels of kaempferol 7-O-sophoroside, which reached the maximum levels at anthesis, were much higher than those observed for both kaempferol 3-O-sophoroside-7-O-glucopyranoside and kaempferol 3,7,4’-triglucoside, with relative high levels in the scarlet stages (-2da to +3da) (Fig. 1C).

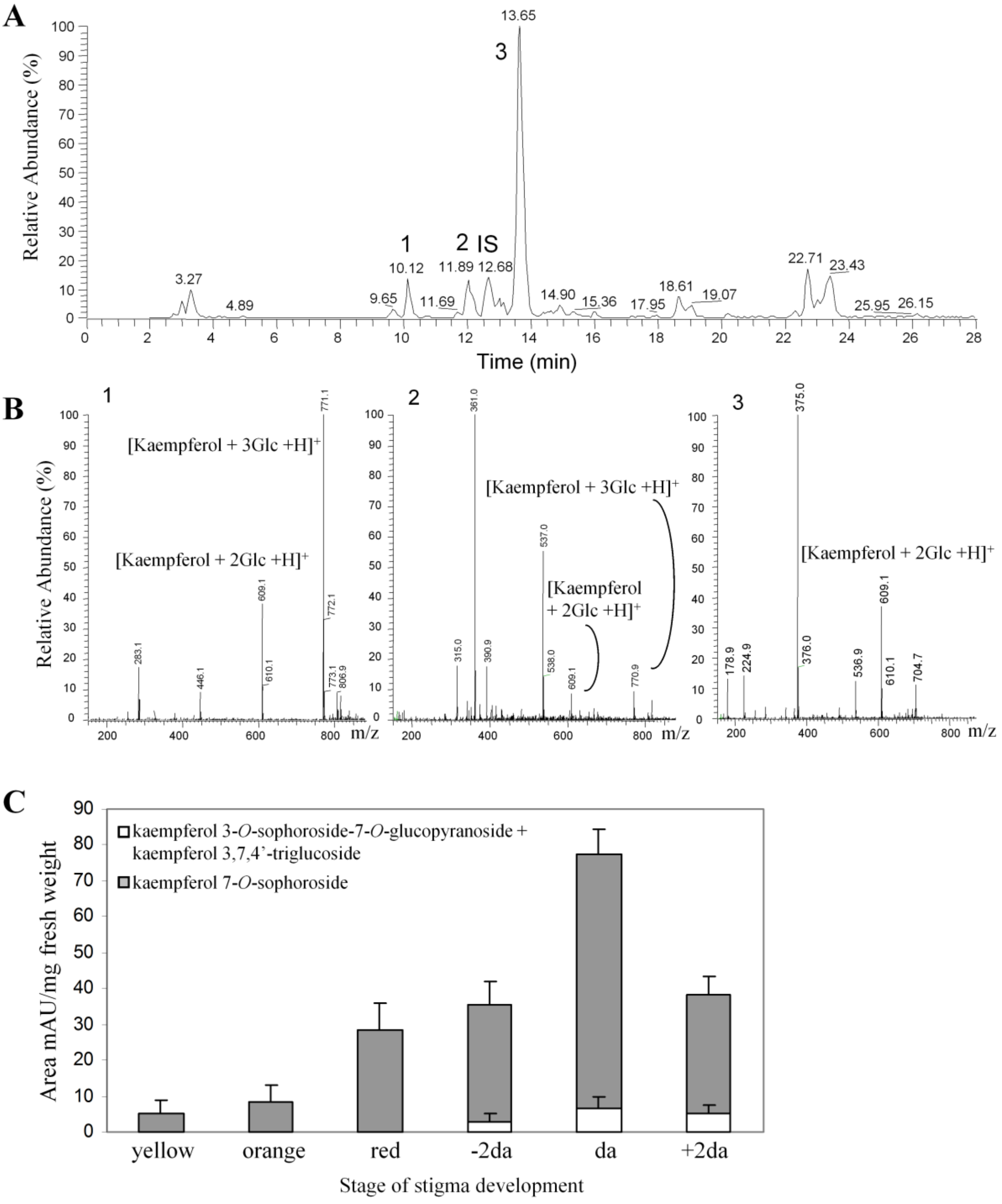


Figure 1. Presence of flavonoid glucosides in *C. sativus* stigmas. A. HPLC-ESI-MS chromatogram of a MeOH extract of *C. sativus* stigmas at anthesis. Three flavonoid peaks, 1, 2, and 3 are denoted by arrows. The compound 4-methylumbelliferyl β-D-glucuronide was used as internal standard (IS). **B**. Positive ion mass spectrum corresponding with the observed flavonoid peaks in A: 1, kaempferol 3-O-sophoroside-7-O-glucopyranoside; 2, kaempferol 3,7,4’-triglucoside, and 3, kaempferol 7-O-sophoroside acquired

during the HPLC-ESI-MS analysis. C. Relative kaempferol 3-O-sophoroside-7-O-glucopyranoside, kaempferol 3,7,4'-triglucoside and kaempferol 7-O-sophoroside levels at different stigma developmental stages.

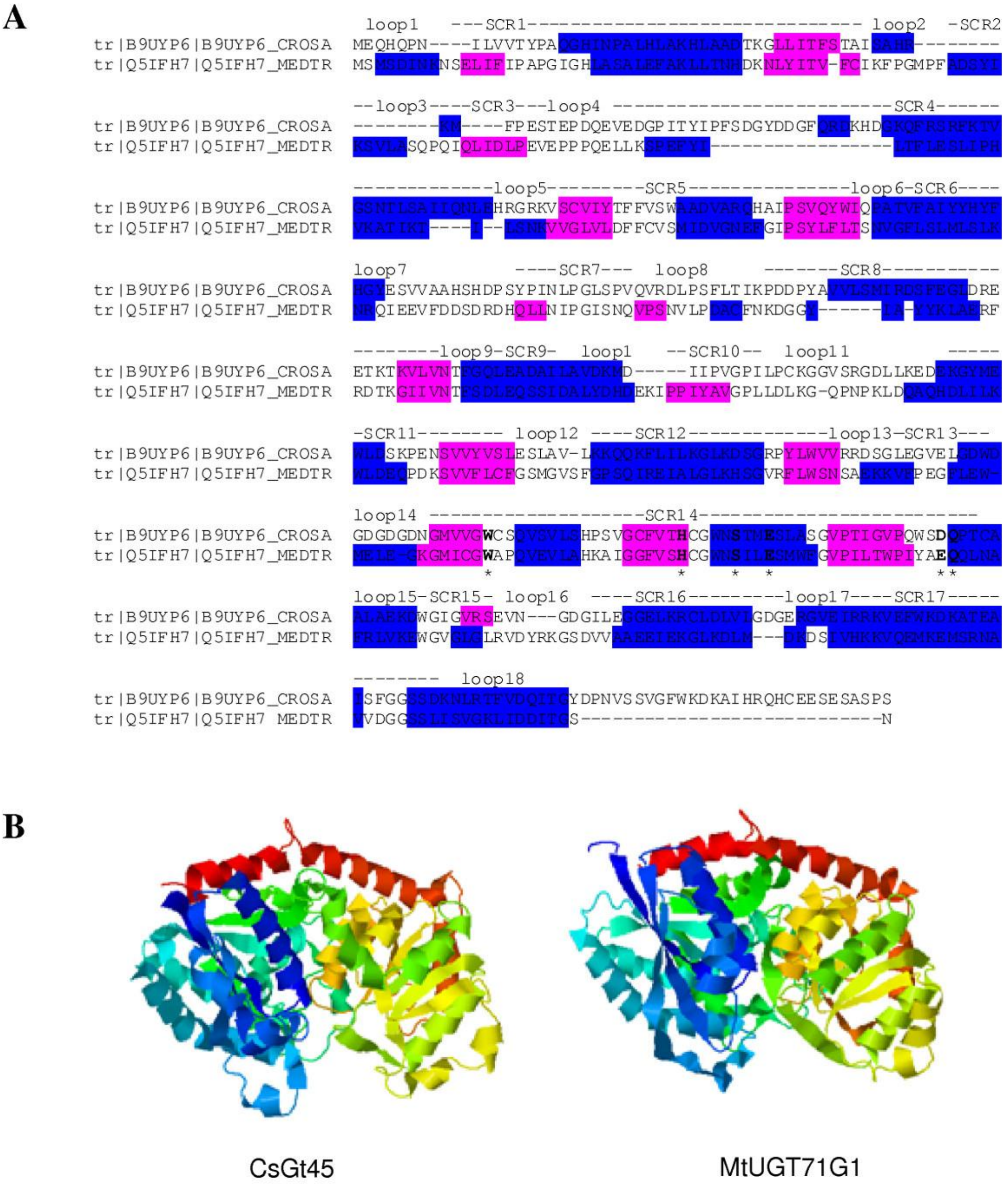
Cloning and deduced structure of CsGT45

To identify flavonoid glucosyltransferases from *C. sativus* stigmas, a homology-based strategy was used, taking advantage of specific glycosyltranferase motifs located in the C-terminus region [50]. A cDNA population was prepared by reverse transcription of poly (A)+ from total RNA isolated from *C. sativus* stigmas at anthesis, which showed the highest levels of kaempferol glucosides. DNA fragments were amplified by degenerate primers and the obtained products were cloned and analysed. Sequencing of one PCR product revealed homology to glycosyltransferases. The sequence information from this clone, CsGT45, allowed the design of PCR specific primers to obtain the full-length transcripts. We performed 5' and 3' RACE using poly(A)+ from *C. sativus* stigma as a template. The GTase gene obtained (1674 bp, Gen Bank FJ194947) was intronless, containing a putative open reading frame of 1500 bp encoding 500 amino acid residues with a calculated molecular mass of 55.42 kDa and a pI of 5.19.

Because *C. sativus* is a triploid we employed *in silico* screening of a large stigma cDNA EST database (http://www.saffrongenes.org/) [51] as an effective method for identification of potential CsGT45 alleles. We identify three EST clones with 98% identity in 611bp (EX147039.1), 98% identity in 264 bp (EX144545.1) and 84% identity in 426 bp (EX148389.1). The two first ESTs correspond to CsGT45, and the third one could correspond to a CsGT45 allele.

The carboxyl terminal of the protein contained the plant secondary product glycosyltransferase (PSPG) box signature motif. Analysis of CsGT45 sequence for N-terminal targeting signal or C- terminal membrane anchor signal using SignalP and TMpred web-based programmes predicted CsGT45 to be non-secretory with an absence of predicted signal peptides or transmembrane signals [52].

For comparative modelling CsGT45 was aligned with MtUGT71G1, whose crystal structure was recently solved [53]. CsGT45 displayed 18% overall identity to MtUGT71G1 (Fig. 2A). A molecular model of CsGT45 was constructed from the structural alignment. Structurally conserved regions of the CsGT45 model were built from crystal structure of MtUGT71G1 using the Pyre server [54](Fig. 2B). In plant GTs, the most common sugar donor is UDP-Glc. Several conserved residues, most of which are found in the PSPG motif of plant UGTs, interact with the sugar donor [22]. The conserved residues involved in the interaction with UDP-Glucose in MtUGT71G1 are also conserved in CsGT45, with the exception of the E381 residue that in CsGT45 is an aspartate residue D385, which is also found in the characterized VvGT1[22].


A

```
                         loop1    ---SCR1------------------------------ loop2  -SCR2
tr|B9UYP6|B9UYP6_CROSA   MEQHQPN----ILVVTYPAQGHINPALHLAKHLAADTKGLLITFSTAISAHR--------
tr|Q5IFH7|Q5IFH7_MEDTR   MSMSDINKNSELIFIPAPGIGHLASALEFAKLLINHDKNLYITV-FCIKFPGMPFADSYI

                         --loop3----SCR3---loop4 -------------------------SCR4------
tr|B9UYP6|B9UYP6_CROSA   --------KM----FPESTEPDQEVEDGPITYIPFSDGYDDGFQRDKHDGKQFRSRFKTV
tr|Q5IFH7|Q5IFH7_MEDTR   KSVLASQPQIQLIDLPEVEPPPQELLKSPEFYI----------------LTFLESLIPH

                         -------------loop5--------SCR5---------------loop6-SCR6----
tr|B9UYP6|B9UYP6_CROSA   GSNTLSAIIQNLEHRGRKVSCVIYTFFVSWAADVARQHAIPSVQYWIQPATVFAIYYHYF
tr|Q5IFH7|Q5IFH7_MEDTR   VKATIKT----I--LSNKVVGLVLDFFCVSMIDVGNEFGIPSYLFLTSNVGFLSLMLSLK

                         loop7          ----SCR7---  loop8      -------SCR8-----------
tr|B9UYP6|B9UYP6_CROSA   HGYESVVAAHSHDPSYPINLPGLSPVQVRDLPSFLTIKPDDPYAVVLSMIRDSFEGLDRE
tr|Q5IFH7|Q5IFH7_MEDTR   NRQIEEVFDDSDRDHQLLNIPGISNQVPSNVLPDACFNKDGGY------IA-YYKLAERF

                         --------loop9-SCR9-  loop1   --SCR10--  loop11           -----
tr|B9UYP6|B9UYP6_CROSA   ETKTKVLVNTFGQLEADAILAVDKMD-----IIPVGPILPCKGGVSRGDLLKEDEKGYME
tr|Q5IFH7|Q5IFH7_MEDTR   RDTKGIIVNTFSDLEQSSIDALYDHDEKIPPIYAVGPLLDLKG-QPNPKLDQAQHDLILK

                         -SCR11------- loop12 ----SCR12------------loop13-SCR13---
tr|B9UYP6|B9UYP6_CROSA   WLDSKPENSVVYVSLESLAV-LKKQQKFLILKGLKDSGRPYLWVVRRDSGLEGVELGDWD
tr|Q5IFH7|Q5IFH7_MEDTR   WLDEQPDKSVVFLCFGSMGVSFGPSQIREIALGLKHSGVRFLWSNSAEKKVFPEGFLEW-

                         loop14 -------------------SCR14-------------------------
tr|B9UYP6|B9UYP6_CROSA   GDGDGDNGMVVGWCSQVSVLSHPSVGCFVTHCGWNSTMESLASGVPTIGVPQWSDQPTCA
tr|Q5IFH7|Q5IFH7_MEDTR   MELE-GKGMICGWAPQVEVLAHKAIGGFVSHCGWNSILESMWFGVPILTWPIYAEQQLNA
                                     *                 *    *  *                **
                         loop15-SCR15- loop16  ----SCR16---------loop17----SCR17-----
tr|B9UYP6|B9UYP6_CROSA   ALAEKDWGIGVRSEVN---GDGILEGGELKRCLDLVIGDGERGVEIRRKVEFWKDKATEA
tr|Q5IFH7|Q5IFH7_MEDTR   FRLVKEWGVGLGLRVDYRKGSDVVAAEEIEKGLKDLM---DKDSIVHKKVQEMKEMSRNA

                         --------   loop18
tr|B9UYP6|B9UYP6_CROSA   ISFGGSSDKNLRTFVDQITGYDPNVSSVGFWKDKAIHRQHCEESESASPS
tr|Q5IFH7|Q5IFH7 MEDTR   VVDGGSSLISVGKLIDDITGS---------------------------N
```

Figure 2. Amino acids sequence alignment of CsGT45 against MtUGT71G1 and structures comparison. A. The alignment was performed guided by conservation of secondary structure, predicted for CsGT45 (B9UYP6) and observed from the solved crystal structure of MtUGT71G1 (Q5IFH7). α-helices are highlighted in blue and β-strands in pink. Structurally conserved regions (SCRs) are highlighted by dots above the alignment. Loops are numbered and named above the alignment. The

amino acids residues within the PSPG motif that interact in MtUGT71G1 with the sugar donor are marked with starts The alignment was used to model the three-dimensional structure of CsGt45. **B**. Ribbon diagrams showing the conserved secondary and tertiary structure of MtUGT71G1 (right) used as template for modelling of CsGT45 and the constructed model (left).

Comparison of the predicted amino acid sequence with that of other glycosyltransferases reveals overall positional identities of 44% with *Pyrus communis* flavonoid 7-O-glucosyltransferase (AAY27090.1), 41% with Arabidospis flavonoid 3-O-glucosyltransferase (At5g17050) and flavonoid 7-O-glucosyltransferase NtF7GT (*Nicotiana tabacum*, BAB88935). The phylogenetic tree based on deduced amino acid sequences if plant GTases is shown in Fig. 3. Currently, GTases function and specificity cannot be fully predicted based on sequence information alone. However, the phylogenetic tree of functionally characterized GTases showed several clusters, which could be characterized by the specificity of the flavonoid glycosyltransferase activities of enzymes involved therein. Cluster I is characterized by flavonoid 3-O-glycosyltransferases, cluster III mainly contains flavonoid 7-O-glycosyltransferases, and cluster IV contains broad substrate GTases. Cs45GT is included in cluster II, which contains anthocyanin 5-O-glucosyltransferases (A5GT), like VhA5GT, PfA5GT and PhA5GT which activities have been tested in vitro [25], [27] and other GTases with a broad substrate specificity that are not involved in the biosynthesis of anthocyanins, like UGT74F1 and UGT74F2 from *Arabidopsis*, which produced distinct multiple glucosides of quercetin in vitro [48], while *in vivo* act as anthranilate glycosyltransferases [55] and are also implicated in salicylic acid metabolism [56]. NtSalGT that reacts on several phenolic compounds in vitro [57], NtF7GT from *Nicotiana* that reacts on the 7-hydroxyl group of flavonol and 3-hydroxyl group of coumarin [29] and PcF7GT from *Pyrus communis* that reacts on the 7-hydroxyl group of flavonol [38]. Therefore, CsGT45 was presumed to encode a flavonoid GTase in *C. sativus* stigmas and was subjected to further analyses.

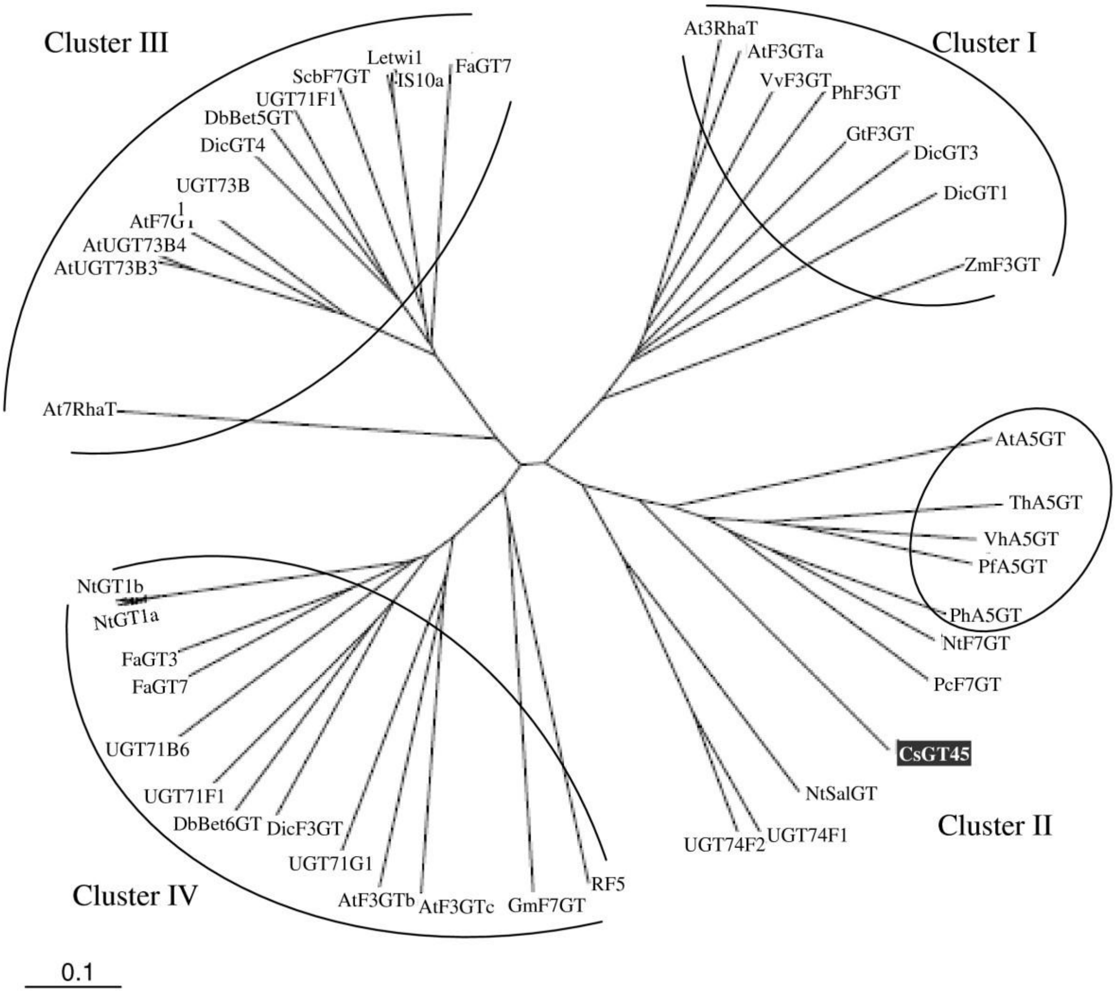


Figure 3. Unrooted phylogenetic tree of the GTases based on amino acid sequence similarity.

Only full-length members of the family are included. The predicted protein sequences were initially clustered using ClustalW. GenBank accession numbers and sources for the respective protein sequences are: CsGT45 (FJ194947) from *Crocus sativus*; ZmF3GT, flavonoid 3-O-glucosyltransferase from *Zea mays* (X13502); At GT, Flavonoid 3-O-glucosyltransferase from *Arabidopsis thaliana* (At5G17050); VvF3GT, flavonoid 3-O-glucosyltransferase from *Vitis vinifera* (AAB81682); GtF3GT, flavonoid 3-O-glucosyltransferase from *Fragaria×ananassa* (BAA12737); DicGT3 and DicGT1 flavonoid 3-O-glucosyltransferases from *Dianthus caryophyllus* (BAD52005) and (BAD52003); At3RhaT, Flavonol 3-O-rhamnosyltransferase from *Arabidopsis thaliana* (At1g30530); At7RhaT, Flavonol 7-O-rhamnosyltransferase from *Arabidopsis thaliana* (NP_563756); PcF7GT, flavonoid 7-O-glucosyltransferase from *Pyrus communis* (AAY27090); NtF7GT, flavonoid 7-O-glucosyltransferase from *Nicotiana tabacum* (BAB88935); NtSalGT, salicylic acid glucosyltransferase from *Nicotiana tabacum* (AAF61647); AtUGT73B3, pathogen-responsive glucosyltransferase from *Arabidopsis thaliana* (AAD17393); AtUGT73B4, flavonoid 3-O-glucosyltransferase from *Arabidopsis thaliana* (AAD17392); AtF7GT, flavonoid 7-O-glucosyltransferase from *Arabidopsis thaliana* (AAR01231); DicGT4, chalcononaringenin 2'-O-glucosyltransferase (BAD52006) from *Dianthus caryophyllus*; DbBet5GT, betanidin-5-O-glucosyltransferase from *Dorotheanthus bellidiformis* (CAB56231); UGT74F1, UGT74F2, and UGT73B1, flavonoid glucosyltransferases from *Arabidopsis thaliana* (AAB64022.1), (AAB64024.1) and (At4g34138); ScbF7GT, flavonoid 7-O-glucosyltransferase from *Scutellaria baicalensis* (BAA83484); Letwi1, wound-inducible glucosyltransferase from *Solanum lycopersicum* (CAA59450); NtIS5a, immediate-early salicylate-induced glucosyltransferase from *Nicotiana tabacum* (AAB36653); FaGT7, multi-substrate flavonol-O-glucosyltransferase (ABB92749); AtF3GTb, putative flavonol 3-O-glucosyltransferase from *Arabidopsis thaliana* (NP_180535.1); AtF3GTc, putative flavonol 3-O-glucosyltransferase from *Arabidopsis thaliana* (NP_180534.1); DicF3GT, flavonol 3-O-glucosyltransferase from *Dianthus caryophyllus* (BAD52004); DbBET6GT, betanidin 6-O-glucosyltransferase from *Dorotheanthus bellidiformis* (AAL57240); UGT71B6, glucosyltransferase from *Arabidopsis thaliana* (AB025634); FaGT3 and FaGT7, flavonol-O-glucosyltransferases from *Fragaria×ananassa* (AAU09444) and (ABB92748); NtGT1a and NtGT1b, broad substrate specificity glucosyltransferases from *Nicotiana tabacum* (BAB60720) and (BAB60721); PhF3GT, flavonol 3-O-glucosyltransferase from *Petunia hybrida* (AAD55985); GtF3GT, flavonol 3-O-glucosyltransferase from Gentiana triflora (BAA12737); AtA5GT, glucosyltransferase from *Arabidopsis thaliana* (AAM91686); ThA5GT, anthocyanin 5-O-glucosyltransferase from *Torenia hybrida* (BAC54093); VhA5GT, anthocyanin 5-O-glucosyltransferase from *Verbena hybrida* (BAA36423); PfA5GT, anthocyanin 5-O-glucosyltransferase from *Perilla frutescens* (BAA36421); PhA5Gt, anthocyanin 5-O-glucosyltransferase from *Petunia hybrida* (BAA89009); UGT71F1, regioselective 3,7 flavonoid glucosyltransferase from *Beta vulgaris* (AY526081); UGT73A4, regioselective 4’,7 flavonoid glucosyltransferases from *Beta vulgaris* (AY526080); UGT71G1, triterpene glucosyltransferase from *Medicago truncatula* (AAW56092). The horizontal scale shows the number of differences per 100 residues derived from the ClustalW alignment.

Biochemical characterization

To identify the function of CsGT45, the full-length open reading frame was cloned into the expression vector pGEX-5T-3 for heterologous protein expression in *E. coli*. The recombinant protein was affinity purified on a glutathion sepharose column that binds the protein’s N-terminal GST-tag (Fig. 4A). Due to its homology with other flavonoid glycosyltransferases, CsGT45 was expected to glucosylate flavonoids. Activity tests were performed with UDP-Glucose and the flavonols quercetin and kaempferol (Fig. 4B). CsGT45 forms monoglucosides on the 7-, hydroxyl group of kaempferol (Fig. 4C and E), whereas over quercetin forms monoglucosides on the 7-, 3’-, and 4’-hydroxyl groups of quercetin (Fig. 4D and F). Glucosylation positions of the kaempferol and quercetin reaction products were assigned based on the hypsochromic shift data [47], comparison with published data [31], [48], [49] and when available, using authentic reference compounds. Flavonols have two absorption maxima: Band I (350-380) and Band II (240-280) corresponding to the B- and A-ring, respectively. Conjugation of 3-, 5-, or 4’-hydroxyl groups causes a Band I hypsochromic shift, which is larger for a 3-substitution (12-17nm) than a 4'-conjugation (3-5 nm). The maximum absorbances of kaempferol were 266 and 368, and those of the kaempferol reaction product were 268 and 368 nm. The lack of a hypsochromic shift between substrate and reaction product strongly suggests that glycosylation occurred at the hydroxyl group of C-7, which was confirmed by comparison with an authentic reference standard (Fig. 4C). For quercetin (256, 372) only the product P1 (256, 372) did not show a hypsochromic shift (Fig. 4D) suggesting conjugation at the 7-hydroxyl group. P2 (254, 368) showed a Band I hypsochromic shift of 4 nm suggesting conjugation at the 4'-hydroxyl group, which was confirmed by comparison with an authentic reference standard (Fig. 4D).

The product P3 (252, 370) was tentatively assigned to quercetin 3'-O-glucoside based on comparison of related flavonoid product elution profiles [31], [48], and by the lack of coincidence with the quercetin 3-O-glucoside standard regarding spectral data and elution time (Fig 4D). When longer incubation times (60 min) and higher substrate concentration (100 mM) of kaempferol or quercetin were used the formation of one diglucoside was observed for each flavonoid (data not shown). Other compounds, i.e. trans-cinnamic acid, sinapic acid, crocin, IAA and abscisic acid were assayed, but no activity was detected with any of these substrates. The results obtained suggest that CsGT45 acts on flavonols *in vivo*. The kinetic parameters for the individual glucosides formed were determined at variable concentrations of quercetin and kaempferol. The K_{cat} and K_m values are described in Table 1. The V_{max}/K_m ratios clearly demonstrate that CsGT45 exhibits the highest specificity towards 7-OH of kaempferol (100%), followed by the 7-OH and 4'-OH of quercetin (20.5 and 9.1%, respectively), and low affinity toward the 3'-OH (3.1%).

The kinetic constants for UDP-glucose were also calculated. Different concentrations of UDP-glucose were assayed keeping the level of kaempferol constant. UDP-glucose showed a Km of 0.6 mM and a V_{max} of 2.9 nkat/mg. Suggesting that glucose is a good substrate for CsGT45.

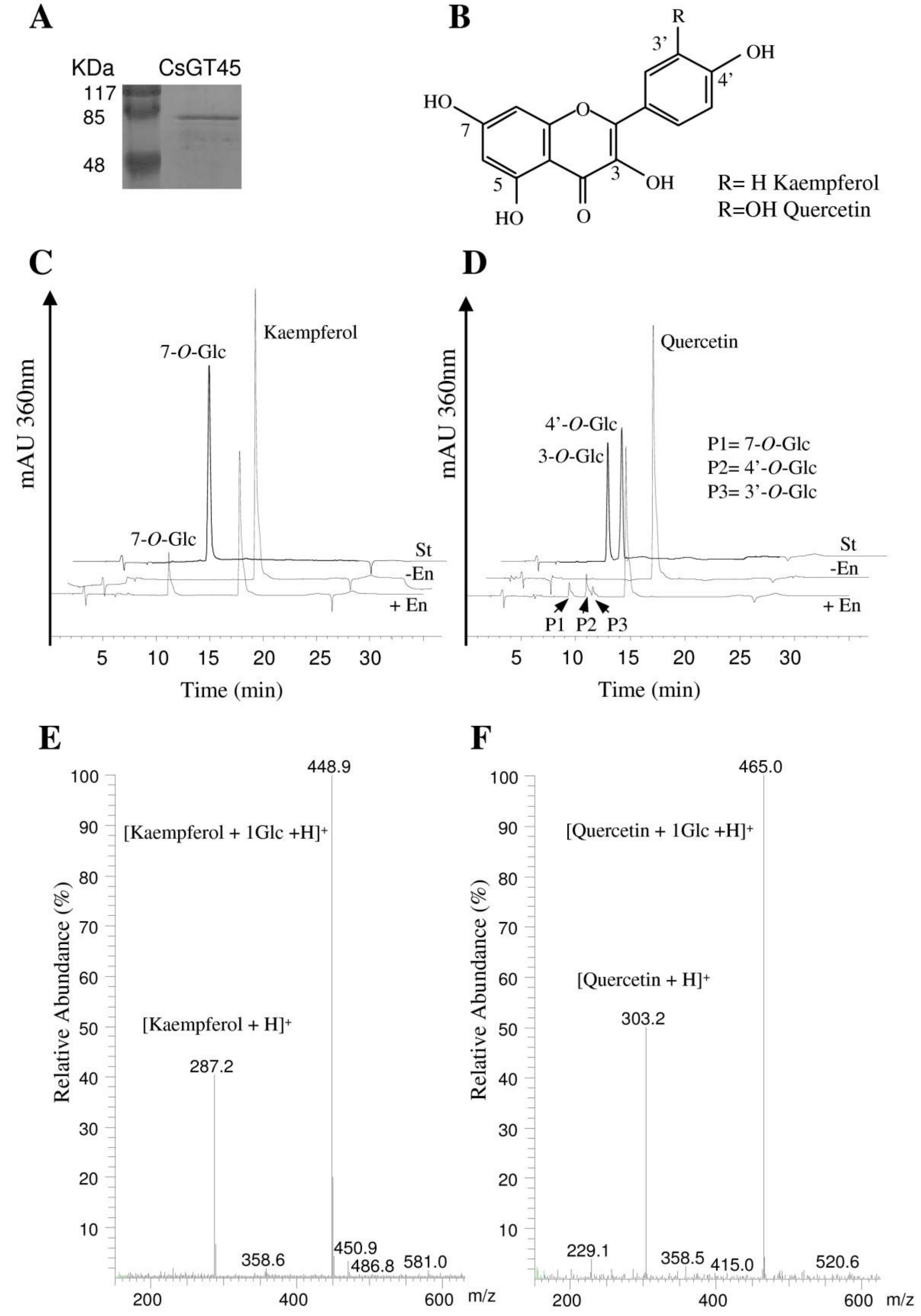


Figure 4. The glutathione S-transferase-CsGT45 fusion protein shows activity toward flavonoids.

A. The recombinant CsGT45 was purified from *E. coli* using glutathione-coupled Sepharose, analyzed using 10% (w/v) SDS-PAGE, and visualized with Coomassie staining. **B**. Chemical structures of the flavonoids kaempferol and quercetin. **C**. HPLC analysis of CsGT45 activity toward kaempferol. **D**. HPLC analysis of CsGT45 activity toward quercetin. The obtained products, P1, P2, and P3 are denoted by arrows. **E** Positive ion mass spectrum of kaempferol 7-O-glucoside acquired during the HPLC-ESI-MS analysis. **F**. Representative positive ion mass spectrum obtained for quercetin 7-O-glucoside, quercetin 4’-O-glucoside and 3’-O-glucoside acquired during the HPLC-ESI-MS analysis of each reaction product. Abbreviations: St, flavonol standard; -E, minus enzyme; and +E plus enzyme.

Table 1. The kinetic parameters K_m and V_{max} and the relation (V_{max}/K_m) of CsGT45, toward kaempferol and quercetin with a fixed UDPG concentration.

Substrate	K_m (μM)	V_{max} (pkat/mg protein)	V_{max}/K_m
Kaempferol 7-OH	15.6 ± 1.2	366 ± 19.8	23.46
Quercetin 4’-OH	86.95 ± 8.4	186 ± 12.4	2.14
Quercetin 3’-OH	30.3 ± 3.6	22.9 ± 2.6	0.75
Quercetin 7-OH	21.50 ± 2.3	104 ± 5.46	4.83

Enzyme assays were carried out using purified CsGT45 (7 μg), substrate (20 to 100 μM) and UDP-glucose (2.5 mM). Reactions mixtures were incubated at 30° C, and performed in triplicate.

Spatial and developmental expression

The spatial and temporal expression pattern of CsGT45 was studied by RT-PCR throughout stigma development. Analyses were performed with RNA isolated from different stages of stigma development, i.e. flowers containing yellow, orange and red stigmas, which are characterized by the presence of immature anthers, and small tepals that do not show the characteristic purple coloration of *C. sativus*. These immature flowers are contained inside perianth tubes that elongate as flowers develop inside. Only when flowers are completely developed do they emerge from the perianthesis tubes and open when anthesis (da) occurs a few days later. Upon emerging, all flowers exhibited purple tepals and scarlet stigmas (-2da to +3da). The RT-PCR analysis revealed that CsGT45 expression is developmentally regulated. The CsGT45 transcript level in the yellow and orange stages was low, but increased from the red stage, and reached a peak at anthesis (Fig. 5). The expression of the CsGT45 was also examined in different tissues. The expression in flower tissues showed that CsGT45 transcripts were present in pollen, tepals and styles at low levels whereas expression in corms was practically undetectable under these conditions (Fig. 5).

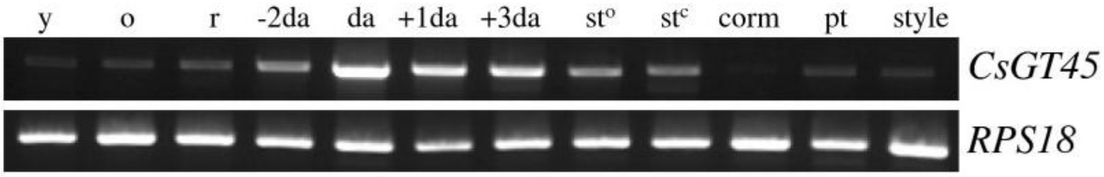


Figure 5. Expression analysis of CsGT45 in plant tissues. The level of CsGT45 was analysed in the stigma tissue of *C. sativus* in different developmental stages: yellow (y), orange (o), red (r), two days before anthesis (-2da), anthesis (da), one day after anthesis (+1da), and three days after anthesis (+3da). Expression levels for CsGT45 were also analysed in closed and open stamen (stc and sto), corm, tepals (pt) and style. Equal amounts of total RNA were used in each reaction. The levels of the constitutively expressed RPS18 coding gene were assayed as controls. The PCR products were separated by 2% (w/v) agarose gel electrophoresis and visualized by ethidium bromide staining.

The high expression levels of CsGT45 transcripts in the stigma tissue and its *in vitro* activity suggested that CsGT45 was associated with the observed kaempferol glucosylation in the stigma tissue. To investigate further such correlation the expression levels of CsGT45 were investigated in the stigma tissue of *Crocus* species in which kaempferol with substitutions in the 7-OH position were not detected (Fig. 6A-B).

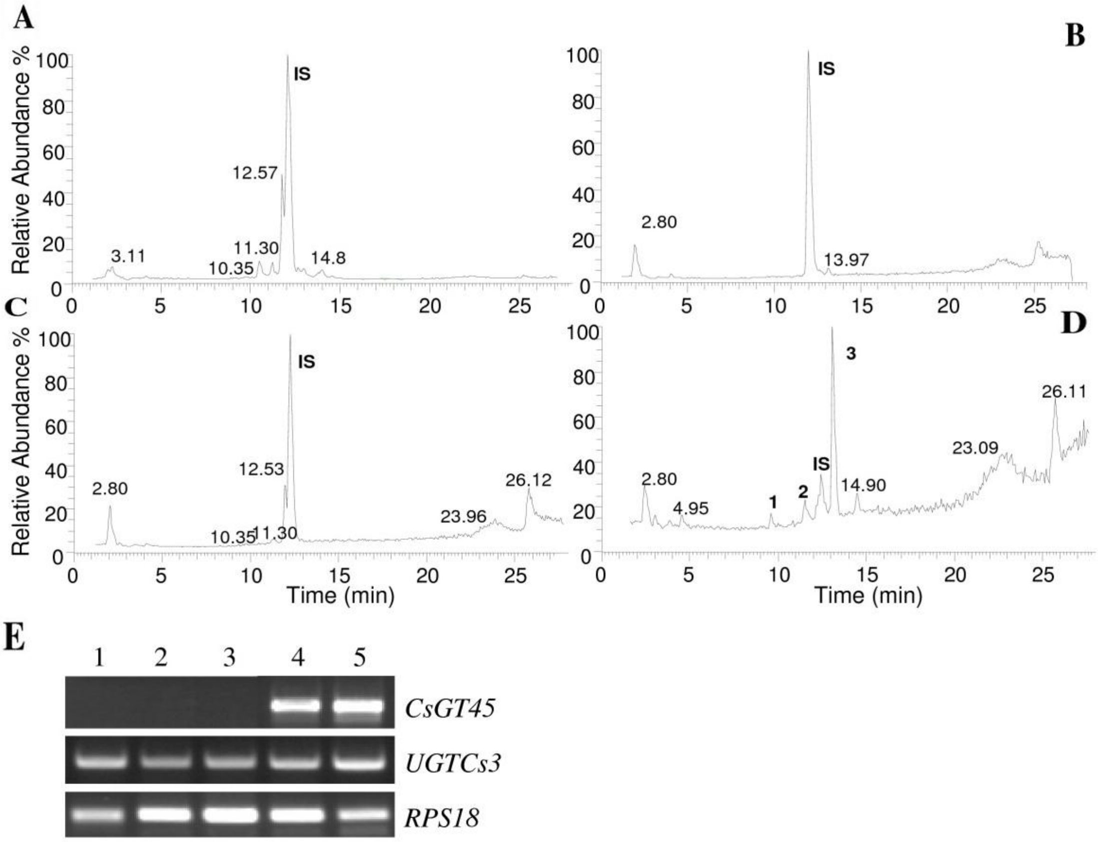


Figure 6. HPLC-ESI-MS chromatograms and expression analyses from three *Crocus* species.

A. HPLC-ESI-MS chromatograms of MeOH extract of *C. cancellatus* stigmas at anthesis. **B**. HPLC-ESI-MS chromatograms of MeOH extract of *C. niveus* at anthesis. **C**. HPLC-ESI-MS chromatograms of MeOH extract of *C. speciosus* stigmas at anthesis. **D**. HPLC-ESI-MS chromatograms of MeOH extract of *C. cartwrightianus* stigmas at anthesis. The peaks 1, kaempferol 3-*O*-sophoroside-7-*O*-glucopyranoside; 2, kaempferol 3,7,4'-triglucoside; and 3, kaempferol 7-*O*-sophorosid. The compound 4-methylumbelliferyl β-D-glucuronide was used as internal standard (IS). **E**. Transcript levels of *CsGT45* in the stigma tissue of different *Crocus* species: 1, *C. niveus*; 2, *C. cancellatus*; 3, *C. speciosus*; 4, *C. sativus* and 5, *C. cartwrightianus*. To ensure the detection of the transcripts, 40 PCR cycles were carried out.

These three *Crocus* species showed reduced flavonoid levels in comparison with *C. sativus*. In *C. niveus* we were unable could to detect kaempferol glucosides, in *C. speciosus* and *C. cancellatus* (Fig. 6A and C) a kaempferol tretahexoside was identified at position 10.35. This compound, substituted at position 3, has been also identified in *C. sativus* as a minor flavonoid [15]. The expression of CsGT45 was not detected in the stigma tissue of *C. niveus*, *C. speciosus* and *C. cancellatus* (Fig. 6E), while was present in the stigma tissue of *C. sativus* and *C. cartwrightianus* that accumulate kaempferol with substitutions in the 7-OH position (Fig. 1A and Fig. 4E). By contrast, the expression of UGTCs3, a GTase previously identified in *C. sativus* stigmas [46], was detected in all the species (Fig. 6E). The absence of CsGT45 expression in the stigma tissue of of *C. niveus*, *C. speciosus* and *C. cancellatus* suggests a role of CsGT45 in the accumulation of specific kaempferol glucosides in the stigma of *Crocus* species.

Unaltered Expression of CsGT45 under stress conditions

Several studies have shown that GTases are induced by a variety of stresses, including: salicylic acid [57] [58], auxin [59], methyl jasmonate [60] and wounding [61]. To determine whether the gene expression levels of CsGT45 were influenced by exogenous hormones or by other stimuli such as drought stress and wounding, total RNA was isolated from treated leaves and used as template in the RT-PCR reactions. The expression of the gene was not altered 24 hours after the treatments (Fig. 7). Shorter times were also tested with the same results (data not shown). Exogenous JA, ABA, GA3, or 2,4D did not significantly promote the expression of the genes (Fig. 7). Drought, wounding and SA failed to affect the expression levels of CsGT45 (Fig. 7).

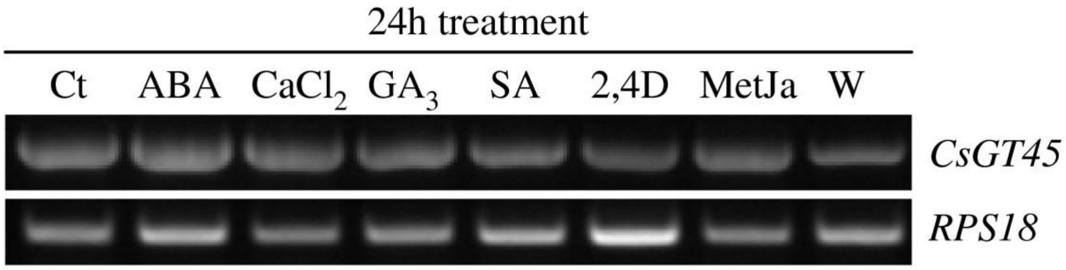


Figure 7. Transcript levels of CsGT45 in response to different treatments.

Untreated control leaves (Ct), treated leaves with abscisic acid (ABA), calcium chloride ($CaCl_2$), giberellic acid (GA_3), potassium salicylate (SA), 2,4-dichlorofenoxiacetic acid (2,4D), methyl jasmonate (MetJa) and wounded leaves of *C. sativus* were collected 24 hr after treatment and total RNA extracted for CsGT45 expression analysis.

Discussion

In general, GTases that use secondary metabolites as substrates are minor constituents in plant cells [21]. Although many of these enzymes have been isolated from several plant species and assayed in vitro, in many cases their roles in the secondary metabolism of these plants are still unknown.

The saffron CsGT45 protein belongs to glucosyltransferase family 1, as do most of the UGTs involved in plant secondary metabolism. This protein possessed a PSPG box with a conserved sequence of 45 amino acid residues and showed specificity towards flavonoid aglycones. This protein has no signal sequence, nor any clear membrane-spanning or targeting signals, as the plant glycosyltransferases identified to date [62]. This suggests that these enzymes function in the cytosol, although within that compartment the proteins may associate as peripheral components of the endomembrane system, as previously suggested [63]. Sequence analysis showed CsGT45 as being most closely related to the *Pyrus communis* flavonoid 7-O-glucosyltransferase and belonging to the same clade of the phylogenetic tree, in which other glucosyltransferases of flavonoids attach sugars without high regiospecificity. The presence in this clade of A5GT enzymes suggest a common ancestral gene for all these GTases, where the A5GTs enzymes showed a strict substrate specificity [25], [27], [64], and seem to have evolved to a more specific function.

Plant secondary product glycosyltransferases have been reported to exhibit a rather strict regioselectivity towards the position of the sugar attachment [21]. The most common site on the flavonol molecule for glycosyl addition is carbon 3 of the C-ring, although other sites, especially the hydroxyl at carbon 7, are often substitutes [65]. However, in proportion, there are few studies on the enzyme activity and genes implicated in the catalysis of the 7-O-glucoside reaction. Many plant GTases recognize quercetin as an acceptor when assayed in vitro, and some others can glucosylate multiple hydroxyl groups of the aglycone and even form diglucosides in some cases [28], [29], [30], [31], [36], [41], [61]. In *Arabidopsis*, from ninety one GTases analyzed for their activity toward quercetin, 29 enzymes showed catalytic activity, and four recognize three sites [31]. Analysing the activity of some enzymes related to CsGT45, the *Arabidopsis* enzyme UGT74F1, glycosylated the 3’-OH, 4’-OH and 7-OH positions of quercetin [48]. We have observed similar activity for CsGT45 toward quercetin, but with a preference for the 7-OH position (K_m 21.5 μM). However, CsGT45 showed high regioselectivity toward kaempferol, and the same was reported for NtGT7 [29], present in the CsGT45 cluster. This feature is characteristic for several 7-O-glucosyltransferases present in cluster III and IV and distantly related to CsGT45 [28], [37], [39], [40], [61]. The K_m value of kaempferol with CsGT45 is 15.6 μM, indicating its sufficient affinity to the substrate. These Km values were the same as other plant GTases reported [29], [61], [65], suggesting that these substrates are reasonable acceptors for CsGT45. Moreover, the CsGT45 enzyme did not react on other OHs of kaempferol, indicating that in this case the regiospecificity of the glucosylation is strictly determined. It has recently been reported that the hydroxylation pattern in ring B of the acceptor molecule can influence product specificity. UGT73A4 from *Beta vulgaris* accepts the positions 4′ and 7 of flavonols. If a hydroxy group is present at position 3′ (e.g. quercetin), 4′-O-glucosides are preferentially formed. If the hydroxy group is missing (e.g. kaempferol), the enzyme produced 7-O-glucosides [36]. Perhaps the differences observed for CsGT45 towards quercetin or kaempferol are due to this fact.

Because quercetin and kaempferol are substrates for CsGT45 in vitro, it is reasonable to propose that this enzyme is involved in the glucosylation of phenolics in the stigma tissue. Analysis of the flavonoid fraction from saffron revealed the unique presence of kaempferol [15]. The gene expression pattern of CsGT45 correlates with high levels of kaempferol glucosides in the stigma tissue. Interestingly the three main flavonoid glycosides detected in the stigma at anthesis were kaempferol 3-O-sophoroside-7-O-glucopyranoside, kaempferol 3,7,4’-triglucoside and kaempferol 7-O-sophoroside. CsGT45 was found to be active on the C-7 position of kaempferol, with the production of a monoglucoside and probably a diglucoside under the experimental conditions tested. Because kaempferol-7-O-sophoroside was the main flavonoid in the stigma tissue. We can speculate whether CsGT45 use UDP-sophoroside as a sugar donor, although UDP-Glc seems to be a good donor and the sugar donor preference of a specific GT is often very narrow, showing little or no activity with alternative sugars [22]. By a molecular modeling approach, we observed that CsGT45 and MtUGT71G1 share higher structural similarity, indicative of similar binding modes with sugar donors. In addition, the residues that interact with the UDP-Glc molecule were conserved in CsGT45. To our knowledge, none GTase has been shown to use UDP-sophoroside as sugar donor. Furthermore, the bright blue or red flowers in the Japanese morning glory (*Ipomoea nil*) contain anthocyanidin 3-O-sophoroside derivatives, and the UDP-glucose:anthocyanidin 3-O-glucoside-2''-O-glucosyltransferase (3GGT) enzyme mediates the glucosylation of anthocyanidin 3-O-glucosides to yield anthocyanidin 3-O-sophorosides.Whereas another glucosyltransferase catalyzes the addition of a glucose molecule to the 3 position [66]. Thus, most probably CsGT45 is responsible for the production of Kaempferol 7-O-glucoside and other GGT could be responsible for the generation of kaempferol 7-O-sophoroside.

The absence of CsGT45 activity toward the C-3 or the C-4’ positions of kaempferol indicates that other glucosyltransferases are implicated in the flavonoid glucosylation in *C. sativus* stigma. In fact, enzymes that catalyzed the transfer of glycosyl groups to the flavonol C-3 position are included in a different cluster than CsGT45.

The increase of flavonoids in the stigma tissue could be associated with the role of flavonoids in protection against abiotic and biotic stresses [67]. The ultra-violet (UV)-absorbing characteristics of flavonoids have long been considered to be evidence for the role of flavonoids in UV protection. The mechanism of protection by flavonoids could be suppression of free radicals formed upon exposure to UV light [68]. Flavonoids are often present in the epidermal cell layers of leaves and in tissues that are susceptible to UV light, such as pollen and the apical meristem [69]. In *C. sativus* the flavonoid levels are specially high in scarlet stigmas (-2da to +3da) that are characterized for being exposed to sunlight, whereas the stigmas for earlier developmental stages are under the soil and enclosed inside the perianth tubes, and therefore protected from the light.

Another well-documented property of flavonoids is their antimicrobial effect [70], [71]. High levels of kaempferol are even reported to inhibit the growth of viruses [72]. The mechanism of kaempferol toxicity is not known, but kaempferol can promote radical formation that might interfere with vital functions of pathogens [73]. The stigma appears to offer a hostile environment to bacteria and fungi since growth of these organisms on the stigma is rare [74]. The presence of flavonoids in the stigma tissue could be associated with the protection of stigmas from pathogen attack.

In addition to these possible functions of flavonoids in the stigma of *C. sativus*, flavonoids are also involved in the control of polar auxin transport [5], [75], [76], which determines plant organ morphology, such as leaf and flower shape [77]. In flowers, stigma and style growth is due to cell elongation, but not to cell division. The transition from the red stage of stigma development in *C. sativus* to the fully developed stigma (-2da to +3da) is accompanied by a rapid increase in size. When the stigmas of *C. sativus* are fully developed they are slender at the base and wider at the apex where they fold to give a trumpet-like structure [78]. The typical morphology of *C. sativus* stigmas could be the result of different auxin distribution controlled by flavonoid signals.

Expression analysis showed low levels of CsGT45 in tepals. The tepals of *C. sativus* are characterized by high levels of anthocyanins, up to 90% of total flavonoids [17]. Nevertheless, a total of nine quercetin and kaempferol glycosides have been identified in minor amounts in *C. sativus* tepals, and a flavonoid-7-O-glucosyltransferase was predicted to be responsible for the formation of two of them [17].

References

1. Tanaka Y, Sasaki N, Ohmiya A: **Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids**. *Plant J* 2008, **54**:733-749.
2. Taylor LP, Grotewold E: **Flavonoids as developmental regulators**. *Curr Opin Plant Biol* 2005, **8**(3):317-323.
3. Isma M: **Botanical insecticides, deterrents, and repellents in modern agriculture and increasingly regulated world**. *Annu Rev Entomol* 2006, **51**:46-66.
4. Iwashina T: **Flavonoid function and activity to plants and other organisms**. *Biol Sci Space* 2003, **17**:24-44.
5. Murphy A, Peer W, Taiz L: **Regulation of auxin transport by aminopeptidases and endogenous flavonoids**. *Planta* 2000, **212**:315-324.
6. Woo H-H, Faull K, Hirsch A, Hawes M: **Altered life cycle in *Arabidopsis thaliana* plants expressing PsUGT1, a UDP-glucuronosyltransferase encoding gene from Pisum sativum**. *Plant Physiol* 2003, **133**:538-548.
7. Woo H-H, Jeong B, Hawes M: **Flavonoids: from cell cycle regulation to biotechnology**. *Biotechnol Lett* 2005, **27**:365-374.
8. Dugas A, Castaneda-Acosta J, Bonin G, Price K, Fischer N, Winston G: **Evaluation of the total peroxyl radical-scavenging capacity of flavonoids: Structure-activity relationships**. *J Nat Prod* 2000, **63**:327-331.
9. Ross JA, Kasum CM: **Dietary flavonoids: bioavailability, metabolic effects, and safety**. *Annu Rev Nutr* 2002, **22**:19-34.
10. Hertog M, Hollman P: **Potential health effects of the dietary flavonol quercetin**. *Eur J Clin Nutr* 1996, **50**:63-71.
11. Steinmetz K, Potter J: **Vegetables, fruit, and cancer prevention**. *J Am Diet Assoc* 1996, **96**:1027-1039.
12. Tarantilis P, Tsoupras G, Polissiou M: **Determination of saffron (*Crocus sativus* L.) components in crude plant extract using high-performance liquid chromatography-UV-visible photodiode-array detection-mass spectrometry**. *J Chromatogr* 1995, **699**:107-118.
13. Straubinger M, ezussek J, Waibel R, Winterhalter P: **Two kaempferol sophorosides from Crocus sativus**. *Natural Products Letter* 1997, **10**:4.
14. Tarantilis PA, Polissiou M: **Isolation and identification of the aroma constituents of saffron (Crocus sativa)**. *Journal of Agriculture and Food Chemistry* 1997, **45**:4.
15. Carmona M, Sánchez A, Ferreres F, Zalacaín A, Tomás-Barberán F, Alonso G: **Identification of the flavonoid fraction in saffron spice by LC/DAD/MS/MS: Comparative study of samples from different geographical origins**. *Food Chem* 2007, **100**:445-450.
16. Hosseinzadeh H, Younesi HM: **Antinociceptive and anti-inflammatory effects of Crocus sativus L. stigma and petal extracts in mice**. *BMC Pharmacology* 2002, **2**(7).
17. Nørbæk R, Brandt K, Nielsen J, Orgaard M, Jacobsen N: **Flower pigment composition of Crocus species and cultivars used for a chemotaxonomic investigation**. *Biochemical Systematics and Ecology* 2002, **30**:28.
18. Hadizadeh F, Khalili N, Hosseinzadeh H, Khair-Aldine R: **Kaempferol from saffron tepals**. *I Pharma Res* 2003, **2**:251-252.
19. Coronado C, Zuanazzi J, Sallaud C, Quirion J, R.Esnault, Husson H, Kondorosi A, Ratet P: **Alfalfa root flavonoid production is nitrogen regulated**. *Plant Physiol* 1995, **108**:533-542.
20. Lim E-K, Bowles D: **A class of plant glycosyltransferases involved in cellular homeostasis**. *EMBO J* 2004, **23**:2915-2922.
21. Vogt T, Jones P: **Glycosyltransferases in plant natural product synthesis: characterization of a supergene family**. *Trends Plant Sci* 2000, **5**(9):380-386.
22. Bowles D, Lim EK, Poppenberger B, Vaistij FE: **Glycosyltransferases of lipophilic small molecules**. *Annu Rev Plant Biol* 2006, **57**:567-597.
23. Ralston E, English J, Dooner H: **Sequence of three bronze alleles of maize and correlation with the genetic fine structure**. *Genetics* 1988, **119**:185-197.
24. Kroon J, Souer E, de Graaff A, Xue Y, Mol J, Koes R: **Cloning and structural analysis of the anthocyanin pigmentation locus Rt of Petunia hybrida: characterization of insertion sequences in two mutant alleles**. *Plant J* 1994, **5**(1):69-80.
25. Yamazaki M, Yamagishi E, Gong Z, Fukuchi-Mizutani M, Fukui Y, Tanaka Y, Kusumi T, Yamaguchi M, Saito K: **Two flavonoid glucosyltransferases from Petunia hybrida: molecular cloning, biochemical properties and developmentally regulated expression**. *Plant Mol Biol* 2002, **48**(4):401-411.

26. Ford CM, Boss PK, Hoj PB: **Cloning and characterization of *Vitis vinifera* UDP-glucose:flavonoid 3-O-glucosyltransferase, a homologue of the enzyme encoded by the maize Bronze-1 locus that may primarily serve to glucosylate anthocyanidins in vivo**. *J Biol Chem* 1998, **273**(15):9224-9233.
27. Yamazaki M, Gong Z, Fukuchi-Mizutani M, Fukui Y, Tanaka Y, Kusumi T, Saito K: **Molecular cloning and biochemical characterization of a novel anthocyanin 5-O-glucosyltransferase by mRNA differential display for plant forms regarding anthocyanin**. *Biol Chem* 1999, **274**:7405-7411.
28. Kramer CM, Prata RT, Willits MG, De Luca V, Steffens JC, Graser G: **Cloning and regiospecificity studies of two flavonoid glucosyltransferases from Allium cepa**. *Phytochemistry* 2003, **64**(6):1069-1076.
29. Taguchi G, Ubukata T, Hayashida N, Yamamoto H, Okazaki M: **Cloning and characterization of a glucosyltransferase that reacts on 7-hydroxyl group of flavonol and 3-hydroxyl group of coumarin from tobacco cells**. *Arch Biochem Biophys* 2003, **420**(1):95-102.
30. Jones P, Messner B, Nakajima J, Schaffner AR, Saito K: **UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in *Arabidopsis thaliana***. *J Biol Chem* 2003, **278**(45):43910-43918.
31. Lim EK, Ashford DA, Hou B, Jackson RG, Bowles DJ: **Arabidopsis glycosyltransferases as biocatalysts in fermentation for regioselective synthesis of diverse quercetin glucosides**. *Biotechnol Bioeng* 2004, **87**(5):623-631.
32. Lee D, Polisensky D, Braam J: **Genome-wide identification of touch- and darkness-regulated Arabidopsis genes: a focus on calmodulin-like and XTH genes**. *New Phytol* 2005, **165**:429-444.
33. Kim JH, Kim BG, Park Y, Ko JH, Lim CE, Lim J, Lim Y, Ahn JH: **Characterization of flavonoid 7-O-glucosyltransferase from *Arabidopsis thaliana***. *Biosci Biotechnol Biochem* 2006, **70**(6):1471-1477.
34. Kubo H, Nawa N, Lupsea SA: **Anthocyaninless1 gene of Arabidopsis thaliana encodes a UDP-glucose:flavonoid-3-O-glucosyltransferase**. *J Plant Res* 2007, **120**(3):445-449.
35. Ogata J, Itoh Y, Ishida M, Yoshida H, Ozeki Y: **Cloning and heterologous expression of cDNAs encoding flavonoid glucosyltransferases from *Dianthus caryophyllus***. *Plant Biotechnol* 2004, **21**:367-375.
36. Isayenkova J, Wray V, Nimtz M, Strack D, T.Vogt: **Cloning and functional characterisation of two regioselective flavonoid glucosyltransferases from *Beta vulgaris***. *Phytochem* 2006, **67**:1598-1612.
37. Noguchi A, Saito A, Homma Y, Nakao M, Sasaki N, Nishino T, Takahashi S, Nakayama T: **A UDP-glucose:isoflavone 7-O-glucosyltransferase from the roots of soybean (*Glycine max*) seedlings. Purification, gene cloning, phylogenetics, and an implication for an alternative strategy of enzyme catalysis**. *J Biol Chem* 2007, **282**(32):23581-23590.
38. Fischer T, Gosch C, Pfeiffer J, Halbwirth H, Halle C, Stich K, Forkmann G: **Flavonoid genes of pear (*Pyrus communis*)**. *Trees* 2007, **21**:521-529.
39. Kim J, Shin K, Ko J, Ahn J: **Glucosylation of flavonols by Escherichia coli expressing glucosyltransferase from rice (*Oryza sativa*)**. *J Biosci Bioeng* 2006, **102**:135-137.
40. Ko JH, Kim BG, Kim JH, Kim H, Lim CE, Lim J, Lee C, Lim Y, Ahn JH: **Four glucosyltransferases from rice: cDNA cloning, expression, and characterization**. *J Plant Physiol* 2008, **165**(4):435-444.
41. Griesser M, Vitzthum F, Fink B, Bellido ML, Raasch C, Munoz-Blanco J, Schwab W: **Multi-substrate flavonol O-glucosyltransferases from strawberry (*Fragaria x ananassa*) achene and receptacle**. *J Exp Bot* 2008, **59**(10):2611-2625.
42. Rubio A, Rambla JL, Santaella M, Gomez MD, Orzaez D, Granell A, Gomez-Gomez L: **Cytosolic and plastoglobule-targeted carotenoid dioxygenases from *Crocus sativus* are both involved in beta-ionone release**. *J Biol Chem* 2008, **283**(36):24816-24825.
43. Ross J, Li Y, Lim E, Bowles DJ: **Higher plant glycosyltransferases**. *Genome Biol* 2001, **2**(2):REVIEWS3004.
44. Winterhalter P, M.Straubinger: **Saffron-renewed interest in an ancient spice**. *Food Rev Int* 2000, **16**:39-59.
45. Cormier F, Dufresne C, Dorion S: **Enhanced crocetin glucosylation by means of maltosyl-β-cyclodextrin encapsulation**. *Biotechnol Tech* 1995, **9**:553-556.
46. Pollack P, Vogt T, Mo Y, Taylor L: **Chalcone synthase and flavonols accumulation in stigmas and anthers of *Petunia hybrida***. *Plant Physiol* 1993, **102**:925-932.

47. Mabry TJ, Markham KR, Thomas MB: **The systematic identification of flavonoids** New York Springer-Verlag 1970.
48. Kelley L, Sternberg M: **Protein structure prediction on the web: a case study using the Phyre server**. *Nature Protocols* 2009, **4**:363-371.
49. Vogt T, Grimm R, D.Strack: **Cloning and expression of a cDNA encoding betanidin 5-O-glucosyltransferase, a betanidin- and flavonoid- specific enzyme with high homology to inducible glucosyltransferases from the Solanaceae**. *Plant J* 1999, **19**:509-519.
50. Hughes J, Hughes MA: **Multiple secondary plant product UDP-glucose glucosyltransferase genes expressed in cassava (Manihot esculenta Crantz) cotyledons**. *DNA Seq* 1994, **5**(1):41-49.
51. D'Agostino N, Pizzichini D, Chiusano M, G.Giuliano: **An EST database from saffron stigmas**. *BMC Plant Biol* 2007, **7**:53.
52. Nielsen H, Engelbrencht J, Brunak S, Heijne Gv: **Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites**. *Prot Eng* 1997, **10**:1-6.
53. Cartwright AM, Lim EK, Kleanthous C, Bowles DJ: **A kinetic analysis of regiospecific glucosylation by two glycosyltransferases of Arabidopsis thaliana: domain swapping to introduce new activities**. *J Biol Chem* 2008, **283**(23):15724-15731.
54. Shao H, He X, Achnine L, Blount JW, Dixon RA, Wang X: **Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from Medicago truncatula**. *Plant Cell* 2005, **17**(11):3141-3154.
55. Quiel J, Bender J: **Glucose conjugation of anthranilate by the Arabidopsis UGT74F2 glucosyltransferase is required for tryptophan mutant blue fluorescence**. *J Biol Chem* 2003, **278**:6275-6281.
56. Dean J, Delaney S: **Metabolism of salicylic acid in wild-type, ugt74f1 and ugt74f2 glucosyltransferase mutants of Arabidopsis thaliana**. *Physiol Plant* 2008, **132**:417-425.
57. Lee HI, Raskin I: **Purification, cloning, and expression of a pathogen inducible UDP-glucose:Salicylic acid glucosyltransferase from tobacco**. *J Biol Chem* 1999, **274**(51):36637-36642.
58. Fraissinet-Tachet L, Baltz R, Chong J, Kauffmann S, Fritig B, Saindrenan P: **Two tobacco genes induced by infection, elicitor and salicylic acid encode glucosyltransferases acting on phenylpropanoids and benzoic acid derivatives, including salicylic acid**. *FEBS Lett* 1998, **437**:319-323.
59. Taguchi G, Yazawa T, Hayashida N, Okazaki M: **Molecular cloning and heterologous expression of novel glucosyltransferases from tobacco cultured cells that have broad substrate specificity and are induced by salicylic acid and auxin**. *Eur J Biochem* 2001, **268**(14):4086-4094.
60. Imanishi S, Hashizume K, Kojima H, Ichihara A, Nakamura K: **An mRNA of tobacco cell, which is rapidly inducible by methyljasmonate in the presence of cycloheximide, codes for a putative glycosyltransferase**. *Plant cell Physiol* 1998, **39**:202-211.
61. Hirotani M, Kuroda R, Suzuki H, Yoshikawa T: **Cloning and expression of UDP-glucose: flavonoid 7-O-glucosyltransferase from hairy root cultures of Scutellaria baicalensis**. *Planta* 2000, **210**(6):1006-1013.
62. Li Y, Baldauf S, Lim EK, Bowles DJ: **Phylogenetic analysis of the UDP-glycosyltransferase multigene family of Arabidopsis thaliana**. *J Biol Chem* 2001, **276**(6):4338-4343.
63. Winkel-Shirley B: **Evidence for enzyme complexes in the phenylpropanoid and flavonoid pathways**. *Physiol Plant* 1999, **107**:142-149.
64. Nakatsuka T, Sato K, Takahashi H, Yamamura S, Nishihara M: **Cloning and characterization of the UDP-glucose:anthocyanin 5-O-glucosyltransferase gene from blue-flowered gentian**. *J Exp Bot* 2008, **59**(6):1241-1252.
65. Williams C, Harbone J: **In The Flavonoids: Advances in Research since 1986**. London; 1994.
66. von Rad U, Huttl R, Lottspeich F, Gierl A, Frey M: **Two glucosyltransferases are involved in detoxification of benzoxazinoids in maize**. *Plant J* 2001, **28**(6):633-642.
67. Morita Y, Hoshino A, Kikuchi Y, Okuhara H, Ono E, Tanaka Y, Fukui Y, Saito N, Nitasaka E, Noguchi H *et al*: **Japanese morning glory dusky mutants displaying reddish-brown or purplish-gray flowers are deficient in a novel glycosylation enzyme for anthocyanin biosynthesis, UDP-glucose:anthocyanidin 3-O-glucoside-2''-O-glucosyltransferase, due to 4-bp insertions in the gene**. *Plant J* 2005, **42**(3):353-363.
68. Schmelzer E, Jahen W, Hahlbrock K: **In situ localization of light-induced chalcone synthase mRNA, chalcone synthase, and flavonoid end products in epidermal cells of parsley leaves**. *Proc Natl Acad Sci U S A* 1988, **85**:2989-2993.

69. Winkel-ShirleyI B: **The Biosynthesis of flavonoids**. In: *The Regulation of Flavonoid Biosynthesis.* Edited by Grotewold E. New York: Springer; 2006: 71-96.
70. Lawton M, Lamb C: **Transcriptional activation of plant defence genes by fungal elicitor, wounding and infection**. *Mol Cell Biol* 1987, **7**:335-341.
71. Ververidis F, Trantas E, Douglas C, Vollmer G, Kretzschmar G, Panopoulos N: **Biotechnology of flavonoids and other phenylpropanoid-derived natural products. Part I: Chemical diversity, impacts on plant biology and human health**. *Biotechnol J* 2007, **2**:1214-1234.
72. French C, Towers G: **Inhibition of infectivity of potato virus X by flavonoids**. *Phytochem* 1992, **31**:3017-3020.
73. Ahmad S: **Biochemical defense of prooxidant plant allelochemicals by herbivore insects**. *Biochem System Ecol* 1992, **20**:269-296.
74. Jung J: **Sind Narbe und Griffel Eintrittspforten für Pilzinfektionen?** *Phytopathol* 1956, **27**:405-426.
75. Peer WA, Murphy AS: **Flavonoids and auxin transport: modulators or regulators?** *Trends Plant Sci* 2007, **12**(12):556-563.
76. Woo H-H, Kuleck G, Hirsch A, Hawes M: **Flavonoids: signal molecules in plant development**, vol. 505. New York: Kluwer Academic/Plenum; 2002.
77. Kesler S, Sinha N: **Shaping up: the genetic control of leaf shape**. *Curr Opin Plant Biol* 2004, **7**:65-72.
78. Grilli M, Canini A: **Ultrastructure of chromoplast and other plastids in *Crocus sativus* L. (Iridaceae)**. *Plant Biosystems* 2004, **138**(1):10.

Chapter 2

“Genomic analysis and gene structure of the plant carotenoid dioxygenase 4 family: a deeper study in *Crocus sativus* and its allies”

Abstract

The plastoglobule-targeted enzyme carotenoid cleavage dioxygenase (CCD4) mediates the formation of volatile C_{13} ketones, such as β-ionone, by cleaving the C_9-C_{10} and $C_{9'}$-$C_{10'}$ double bonds of cyclic carotenoids. Here, we report the isolation and analysis of CCD4 genomic DNA regions in *Crocus sativus*. Different *CCD4* alleles have been identified: *CsCCD4a* which is found with and without an intron and *CsCCD4b* that showed the presence of a unique intron. The presence of different *CCD4* alleles was also observed in other *Crocus* species. Furthermore, comparison of the locations of *CCD4* introns within the coding region with *CCD4* genes from other plant species suggests that independent gain/losses have occurred. The comparison of the promoter region of *CsCCD4a* and *CsCCD4b* with available *CCD4* gene promoters from other plant species highlighted the conservation of *cis*-elements involved in light response, heat stress, as well as the absence and unique presence of *cis*-elements involved in circadian regulation and low temperature responses, respectively. Functional characterization of the *Crocus sativus CCD4a* promoter using *Arabidopsis* plants stably transformed with a DNA fragment of 1,400 base pairs (P-CsCCD4a) fused to the β-glucuronidase (*GUS*) reporter gene showed that this sequence was sufficient to drive *GUS* expression in the flower, in particular high levels were detected in pollen.

Keywords: *Crocus sativus*, carotenoids, intron, promoter, stigma, volatiles

Introduction

Carotenoids are isoprenoid pigments synthesized by all photosynthetic organisms and some nonphotosynthetic bacteria and fungi. In plants, carotenoids are essential in protecting the photosynthetic apparatus from photo-oxidation, and represent essential constituents of the light-harvesting and reaction centre complexes. The oxidative cleavage of carotenoids occurs in plants, animals, and micro-organisms and leads to the release of a range of apocarotenoids that function as signalling molecules with diverse functions [1-2], including the ubiquitous chromophore retinal, plant hormone abscisic acid and strigolactones. Other apocarotenoids with unknown functions in plants but with high economic value are bixin in *Bixa orellana* and saffron in *Crocus sativus*[3], used as a flavouring and colouring agent. *Crocus sativus* is a triploid sterile plant, most probably originated by a human-mediated event, which is propagated by corms. Among all known *Crocus* species, *Crocus sativus* is particularly appreciated due to the high levels of apocarotenoids on the stigmas that determine the commercial value of this plant. Hence, studying apocarotenoid formation in all these species will help to determine the mechanisms underlying apocarotenoid accumulation.

The synthesis of apocarotenoids is initiated by the oxidative cleavage of double bonds in in its stigmas carotenoid backbones, catalyzed by carotenoid oxygenases, which are nonheme iron enzymes present in all taxa [1, 4]. The first gene identified as encoding a carotenoid cleavage dioxygenase was the Vp14 maize gene required for the formation of abscisic acid (ABA) [5]. On the basis of their substrate specificity, VP14 and its orthologous have been termed 9-cis epoxy-carotenoid dioxygenases (NCEDs). Plants possess a second group of carotenoid oxygenases, carotenoid cleavage dioxygenases (CCDs), which act on different carotenoid substrates [1]. These include CCD1, CCD4, CCD7 and CCD8.

Plants release volatile apocarotenoids, including C_{13} ketones such as β-ionone and damascone, which constitute an essential aroma note in tea, grapes, roses, tobacco and wine. Two classes of CCD enzymes have been implicated in plant volatile production, CCD1 and CCD4. Both plant enzymes cleave C(40) carotenoids at the C_9-C_{10} and $C_{9'}$-$C_{10'}$ double bonds into C_{14}dialdehydes, which are common to all carotenoid substrates, and two variable end-group-derived C_{13} ketones [1]. The CCD1 enzymes act in the cytosol, where most probably are involved in apocarotenoid cleavage, whereas CCD4 enzymes have been shown to reside in plastids, where their substrates are localized, suggesting a direct involvement in volatile formation. The first member of the CCD4 subfamily was identified in *Chrysantemum morifolium*[6] and the enzymatic activity has been recently characterized in *Crocus sativus*[7], *C. morifolium*, *Arabidopsis thaliana*, *Rosa damascena*, and *Malus domestica*[8].

The different CCD families are characterized by the heterogeneity in their gene structures, with genes containing no introns or containing as many as 10 introns, as in the human gene β,β-carotene-15,15-dioxygenase (BCMO) [9] or 11 and 13 introns for the case of the CCD1 genes of rice and *Arabidopsis*[1], respectively. The plant group of CCDs containing CCD7 and CCD8 is characterized by the presence of multiple intron sequences in all the available genes. The CCD7 genes from petunia, rice and pea are characterized by the presence of 6 introns [10-11], whereas 5 introns are present in the CCD7 gene of *Arabidopsis* [12], while four of these introns are well conserved among these plant species [11]. The CCD8 of *Arabidopsis* and petunia also contains 5 intron sequences [13].

By contrast, compared with the CCD1 and the CCD7/8 groups, the CCD4 group seems to contain genes without intron sequences or with one or two introns [6, 8, 14].

The genomic structure of a gene determines its regulation. Gene regulation is mainly determined by the promoter region, but several other types of gene regulation, both positive and negative, involve plant introns. Some introns contain enhancer elements or alternative promoters, while many others elevate mRNA accumulation by a different process that has been named intron-mediated enhancement (IME), which is thought to result from synergistic interactions between the factors involved in the various steps of gene expression from transcription to translation [15]. Furthermore, conservation of exon-intron structure in clades of orthologous genes, as well as in families of paralogous genes and protein superfamilies, support the use of gene features as sources for phylogenetic inference [16]. Thus, the knowledge of the genomic structure is important in order to characterize gene families and for the establishment of evolutionary relationships.

In this report, we have isolated and compared the genomic structure of the CCD4 genes of *Crocus sativus* and compared it with its allies and with other plant species, and determined the intron presence and conservation of intron arrangement within this CCD family. In addition, we have isolated and analysed the genomic DNA sequence upstream from the CCD4 genes of *C. sativus*, which were analysed along with other CCD4 promoters (*Arabidopsis*, rice, tomato, poplar, papaya, *Vitis*, *Medicago* and

Brassica) in order to identify common cis-regulatory motifs and compare the distribution patterns of these motifs. The spatial and temporal activity of the putative CsCCD4a promoter from *C. sativus* fused to the β-glucuronidase (GUS) gene in stably transformed *Arabidopsis* plants was also assessed, and the localization of GUS staining was monitored in different plant tissues.

Materials and Methods

Plant material

For this study, we included 14 species of *Crocus* with a total of 37 different populations. Specimens were obtained from saffron growers in Tarazona de La-Mancha (Spain), from Dr. U. Jacobsen from the Agricultural University of Denmark and from private collections in the UK (Potterton Nursery) (Table 1).

Plant tissues were independently harvested and frozen in liquid nitrogen and stored at -80 °C until required.

Seeds from *Arabidopsis* wild type Columbia ecotype (Col-0), and transgenic lines were sown in pots and watered with nutrient solution under a controlled environment with 16 h light/8 h dark cycles at 22 °C. Seeds from transformed *Arabidopsis* plants were surface sterilized by rinsing them in 70% (v/v) ethanol for 1 min, followed by a 15 min treatment in 10% (v/v) bleach + 0.05% (v/v) Triton X-100 and three rinses in sterile distilled water.

Table 1. *Crocus* species and populations sampled for *CCD4a* and *CCD4b* genomic sequences: a+: *CCD4a* plus intron; b+: *CCD4b* plus intron; a-: *CCD4a* minus intron;b-: *CCD4b* minus intron

Series	Nomenclature	Origin of plant material	*CCD4* alleles			
			a+	b+	a-	b-
Crocus	*C. sativus* L.	Spain, Tarazona	+	+	+	-
	C. sativus L.	China, Yunan (C 395)	+	+	+	-
	C. sativus L.	Greece, Kozani, Krokos	+	+	+	-
	C. sativus L.	Iran, Mashhad,	+	+	+	-
	C. sativus L	(UK)	+	+	+	-
	*C. oreocreticus*B.L. Burtt (GKØ00-32)	Greece, Creta, Mt Dikti	+	+	-	-
	C. oreocreticus B.L. Burtt (G94-15)	Greece, Creta, Rethimno	+	+	-	-
	C. cartwrightianus Bory and Chaub (12629)	Greece, NomosAtiki, Plaka	+	+	-	-
	C. cartwrightianus Bory and Chaub (GJLL 01-56)	Greece, NomosAtiki, Mt. Imitos	+	+	-	-
	C. cartwrightianus Bory and Chaub (NJG 98-112)	Greece, NomosAtiki, Keratea	+	+	+	-
	C. cartwrightianus Bory and Chaub (12627)	Greece, NomosAtiki, Ch. Taxiarchiso	+	+	-	-
	C. cartwrightianus Bory and Chaub (GKØ00-8)	Greece, Creta, Hania	+	+	+	-
	C. cartwrightianus Bory and Chaub (GKØ00-44)	Greece, Creta, Hania	+	+	+	-
	C. hadriaticus Herbert (Ø01-124)	Greece, Kefalonia, Mt Etnos	+	+	-	-
	C. hadriaticus Herbert (J01-37)	Greece, Peloponnese, W of Pelei	+	+	-	-
	C. hadriaticus Herbert (ØJ01-32)	Greece, Peloponnese, N of Vasilios	+	+	-	-
	C. hadriaticus Herbert (GNJ01-141)	Greece, Peloponnese, NomosArkadia	+	+	-	-
	C. hadriaticus Herbert (ØJ01-40)	Greece, Peloponnese, Mt Didima	++	+	-	-
	C. hadriaticus Herbert (G98-31)	Greece, Peloponnese, Mt Taigetos	++	+	-	-
	C. pallasiiGoldbssppallasii(GKØ01-64)	Greece, Lesbos, above Agra	-	+	+	-
	C. pallasiiGoldbssppallasii (GKØ01-36)	Greece, Lesbos, Mt Olymbos	+	+	-	-
	C. pallasiiGoldbssppallasii (GKØ01-45)	Greece, Lesbos, Petra toKalloni	-	++	-	-
	C. pallasiiGoldbssppallasii (GKØ01-61)	Greece, Lesbos, above Agra	+	+	-	-
Longiflori	*C. medius*Balbis (C 430)		-	-	+	-
	C. mediuss (UK)		-	-	+	-
	C. goulimyi (UK)		-	-	+	-
	*C. goulimyi*Turrill (Ø01-12)	Greece, Peloponnese, N PirgosDirou	-	-	+	-
	C. niveus Bowles (00-154)	Greece, Peloponnese, NomosArkadia	-	-	+	-
	C. niveus(UK)		-	-	+	-
	C. nudiflorus (UK)		-	-	+	-
Versicolores	*C. imperati* (UK)		-	-	+	-
Kotschyani	*C. ochloroleucus* (UK)		-	-	+	+
	C. kostcyanus(UK)		-	-	+	-
Speciosi	*C. pulchellus* Herbert (98-81)	Greece, Mt Pangaeon	-	-	+	-
	C. pulchellus (I-19)	UK, Kew Gardens	-	-	+	-
	C. speciosus M. (UK)		-	-	+	+
	*C. speciosus*M. Bieb (C12)	without origin	-	-	+	+

+: indicates +: indicates the presence of one copy of that gene; ++: indicates the presence of two copies.the presence of one copy of that gene; ++: indicates the presence of two copies.

Isolation of CsCCD4a and CsCCD4b genomic clones

Genomic DNA samples were prepared from *Crocus* leaves by using a CTAB (hexadecyltrimethylammonium bromide) method. The isolated DNA was quantified and amplified via PCR in a BioRad MJ Mini Thermal Cycler. In order to isolate the genomic clones of CCD4 in different *Crocus* species, the primers CsCCD4a-f, CsCCD4b-f and CsCCD4a/b-r (Table 2) were used. PCR products were purified from 0.8% agarose gels with the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI). The PCR products were then cloned into the pGEM-T Easy plasmid vector using a TA cloning ligase kit (Promega, Madison, WI). For each sample, 12-20 transformed colonies that contained the PCR insert were sequenced with Sp6 and T7 primers using an automated DNA sequencer (ABI PRISM 3730xl, Perkin Elmer) from Macrogen Inc. (Seoul, Korea).

Isolation of putative CCD4 promoter from *C. sativus*

Genomic DNA was prepared from *C. sativus* leaves by using a CTAB method. The *C. sativus* CCD4 upstream flanking sequences were isolated with the Genome Walker Universal Kit (BD Biosciences, Palo Alto, CA, USA) using two nonoverlapping gene specific primers: CCD4-P-r1 and CCD4-P-r2 (Table 2), based on the CCD4 cDNA sequences obtained (Genbank accession numbers ACD62476 and ACD62477). All PCR reactions were performed using Advantage 2 Polymerase mix (BD Biosciences, Palo Alto, CA, USA).

Table 2.*CCD4* genes, proteins and gene structure in several plant species.

Organism	protein	Chromosome and location	Genomicsequence	No. introns	Introns length
Aquilegia coerulea	AcCCD4a		V1.002207mg	0	
	AcCCD4b		V1.005639mg	0	
Arabidopsis thaliana	AtCCD4 Np_193652.1	IV, 10481786-10483857	AL161550.2	0	
Arabidopsis lyrata	AlCCD4 XP_002869997	IV	XM_002869951	0	
Artemia annua	AaCCD4		EZ188260	0	
Brachypodium distachyon	BdCCD4	III, 155125-153498	ADDN01001089	0	
Brassica rapa	BrCCD4a	VI	AC232508.1	0	
Carica papaya	CpCCD4		ABIM01016290	0	
Cucumis sativus	CcsCCD4		ACHR01009958	0	
Chrysanthemum x morifolium	CmCCD4a ABY60885			1	105
	CmCCD4b BAF36656.2			unknown	
Citrus clementina	CcCCD4a ABC26011			unknown	
	CcCCD4b ABC26012			unknown	
Crocus sativus	CsCCD4a ACD62476.1			1 and 0	666
	CsCCD4b ACD62477.1			1	525
Eucaliptus grandis	EgCCD4			1	119
Glycine max	GmCCD4		ACUP01000498	1	669
Hordeum vulgare	HvCCD4 AK248229.1			unknown	
Lactuca sativa	LsCCD4 AB120111			unknown	
Malus domestica	MdCCD4 EU327777			1	1285
Medicago truncatula	MtCCD4a	V, 23071-21329	CT025839.21	0	
	MtCCD4b	V, 37397-39133	CT025839.21	0	
	MtCCD4d	V, 46180-47913	CT025839.21	0	
Mimulus guttatus	MgCCD4a		MGV1a003497m	2	134, 150
	MgCCD4b*		MGV1a023345m	1	493
	MgCCD4c*		MGV1a025390m	1	
Osmanthus fragans	OfCCD4 EU334434			2	1942, 491
Oryza sativa	OsCCD4a NP_001047858.1	II, 29046669-29044592	AP008208.1	0	
	OsCCD4b ABA97976.1	XII, 14279169-14277220	AP008218.1	0	
Picea sitchensis	PcsCCD4 ABK24456			unknown	
Pisum sativum	PsCCD4 BAC10552			unknown	

Table 2 (*continued*).*CCD4* genes, proteins and gene structure in several plant species.

Organism	protein	Chromosome and location	Genomicsequence	No. introns	Introns length
Populus trichocarpa	PtCCD4a XP_002307055	V, 685642-684119	NC_008471.7	0	
	PtCCD4b XP_002326037	XIX, 10102690-10100921	NC_008485.1	0	
	PtCCD4c XP_002312876	IX, 990213-992488	NC_008475	1	696
	PtCCD4d XP_002312877	IX, 1004373-1006130	NC_008475	0	
	PtCCD4e XP_002312878	IX, 1015994-1017778	NC_008475	0	
Ricinus communis	RcCCD4 XP_002519944		NW_002994348	0	
Rosa x damascena	RdCCD4 EU334433			0	
Setaria italica	SiCCD4		Siprov004032	0	
Shorghum bicolor	SbCCD4a	IV, 10190662-10191240	NC_012873.1	0	
	SbCCD4b	IV, 60615854-60617290	NC_012873.1	0	
Solanum lycopersicum	SlCCD4a	VIII, 48149-46341	AP009393.1	1	78
	SlCCD4b	VIII, 65635-63288	AP009393.1	1	566
Thellungiella halophyla	ThCCD4		BAJ34120	0	
Vitis vinifera	VvCCD4a XP_002268404.1	II, 17383078-17385173	NC_012008.2	0	
	VvCCD4b XP_002270161.1	II, 17412105-17414090	NC_012008.2	0	
	VvCCD4c XP_002269538.1	XVI, 792204-802594	NC_012022.2	1	8649
	VvCCD4d XP_002269309	XVI, 721247-720579*	NC_012022.2	0	
Zea mays	ZmCCD4a	IV	AC190588	0	
	ZmCCD4b	V	AC194862	0	

*****Truncated gene

Two rounds of PCR amplification using the CCD4-P-r1 primer with an adapter-specific primer and the nested PCR amplification with CCD4-P-r2 and another adapter-specific primer gave fragments around 1,400 base pairs. All PCR products were ligated to pGEM-T with the TA Cloning Kit (Promega, Madison, USA). The ligated DNA was transformed into E. coli strain JM109. The clones (50 colonies) were picked individually and amplified in 3 ml of LB medium at 37 °C overnight. The plasmid DNA from each clone was extracted using a DNA Plasmid Miniprep Kit (Promega, Madison, WI, USA). Plasmids were sequenced using an automated DNA sequencer (ABI PRISM 3730xl, Perkin Elmer) from Macrogen Inc. (Seoul, Korea).

Promoters were analysed by the FOOTPRINTER programme for the identification of consensus patterns in unaligned DNA sequences [17].This tool takes into account the evolutionary relationships and distances between the genes compared (based on a phylogenetic tree).

Sequence analyses

Sequence analyses performed in this study used publicly available genomic sequences of CCD4 (Table 3) CCD1, CCD7 and CCD8. The gene prediction programme GenScan was used to predict the gene structures for the CCD genes, together with GenBank using NCBI BLAST programmes. Putative cis-regulatory elements shown in this study and their positions were identified with the use of PlantCare software [18].

Phylogenetic analysis

A phylogenetic analysis was initially performed with representative members of the different carotenoid cleavage enzymes obtained from Genbank. All the CCD proteins that fell into the same group were named accordingly as CCD1, CCD4, CCD7 and CCD8. The CCD4 proteins were extracted and used for a more exhaustive identification of additional CCD4 enzymes to be used for a CCD4 phylogenetic analysis. To construct the phylogenetic trees, the amino acid sequences were aligned using the BLOSUM62 matrix with the ClustalW (http://www.clustal.org) algorithm-based AlignX module from MEGA Version 4.0 (http://www.megasoftware.net/mega.html). The alignments were saved and executed by MEGA Version 4.0 to generate Neighbour Joining Tree with bootstrapping (500 replicates) analysis and handling gaps with pairwise deletion.

Arabidopsis transformation

To produce transgenic plants in which the GUS protein was expressed under the control of the CsCCD4a promoter, the vector pGWB3 was used [19]. The strategy followed for cloning the CsCCD4a promoter in this vector was based on Gateway Technology, and the oligonucleotides used were attb1_ccd4prom_f3 and attb2_CDD4prom-r2 (Table 3) present in the coding sequence.

Table 3. Primer sequences used for *CsCCD4* gene analysis and promoter isolation

Primer	*sequence*	*orientation*
CsCCD4af	5’-CAATCTCAAGTATTAGCATTC-3’	sense
CsCCD4bf	5’-CACTACCCATCTCATCAAGA-3’	sense
CsCCD4a/b-r	5’- CTGCTGTGACAGCAGCTCAGC-3’	antisense
CCD4-P-r1	5’-CTTGTTGATACTGATACTCTTCT-3’	antisense
CCD4-P-r2	5’-TTGCCGGGTGAGGGAAAGTGGA-3’	antisense
att_ccd4prom_f3	5’-GGGGACAAGTTTGTACAAAAAAGCAGGCTCCGTTGTCCATCTTCTGTTCCTCT-3’	sense
att_ccd4prom_r2	5’-GGGGACCACTTTGTACAAGAAAGCTGGGTCGAGAAGGGAGTGGAGCATGCCGTCG-3’	antisense

The amplified fragment was introduced into the vector pDONR™221 (Invitrogen, Carlsbad, CA) by a BP recombination reaction, and from this vector to the destination vector pGWB3 by a LR recombination reaction using Gateway Technology (Invitrogen, Carlsbad, CA). The recombinant CCD4a::pGWB3 vector was transferred into Agrobacterium tumefaciens strain GV3101 by electroporation and bacteria were selected on YEB agar with 50 µg mL^{-1} kanamycin and hygromicin and 100 µg mL^{-1} rifampicin and 25 µg mL^{-1} gentamicin. *Arabidopsis* plants were transformed by floral dipping [20] and transformants selected on Murashige and Skoog agar with 50 µg mL^{-1} kanamycin and hygromicin.

Histochemical β-glucuronidase (GUS) staining

The GUS activity of transgenic T3 *Arabidopsis* plants was analysed by histochemical staining using 5-bromo-4-chloro-3-indolyl glucuronic acid (X-gluc) as described [21]. Briefly, whole plants at various stages were immersed in an X-gluc solution (1 mM X-gluc, 50 mM sodium phosphate (pH 7.0), 0.1% Triton X-100), and after applying vacuum for 5 min, were incubated at 37°C until satisfactory staining was observed. For better visualization of stained tissue, samples were rinsed, at room temperature, with an ethanol series for at least 1 h to remove chlorophyll. Images were taken using a Leica MZ16F stereomicroscope.

Results

The CCD4 protein family

BLAST was utilized to search the Genbank database with reported CCD4 protein sequences as queries. In model species for which the genome is completely sequenced, 26 CCD4 genes were identified (Table 4): one in *A. thaliana, Artemisia annua, Brachypodium distachyon, Brassica rapa, Carica papaya* and *Glycine maxima*, 2 in *Oryza sativa, Shorghum bicolour, Solannum lycopersicum* and *Zea mays*, 3 in *Medicago truncatula*, 4 in *Vitis vinifera* and 5 in *Populus trichocarpa*. Table 4also includes other genomic sequences previously identified in Citrus [22], *Crocus sativus*[7], *Lactuca sativa*[23], *M. domestica, O. fragans, R. damascena*, and *C. morifolium*[6, 8]. In addition, 5 more CCD4 genes were identified in 5 other plant species.

The physical gene distribution was analyzed in the species with more than one CCD4 gene present in the same chromosome (Fig. 1). The *Vitis* a and b genes located in chromosme II and separated 26.9 kbp showed 73% identity to each other. By contrast, the *Vitis* c and d genes in chromosome XIV, separated 71.6 kbp showed 97% identity in 677 nucleotides, but along the region where this truncated gene is located, two additional fragments of 164 and 125 bp at 4826 and 4950 pb from the 5’ end of the truncated clone are present (Fig. 1). The *M. truncatula, S. lycopersicum* and *P. trichocarpa* genomes showed that the CCD4 genes consecutively located in the same chromosome showed high identities to each other, suggesting that they were generated via tandem duplications from one of the genes. *S. lycopersicum* genes, located in chromosome VIII and separated 14.4 kbp showed 82% identity. The *M. truncatula* CCD4 genes located in the same chromosome showed identities over 82% and were separated 14.3 kbp (a and b) and 7.0 kbp (b and c). The *P. trichocarpa* genes PtCCD4c, PtCCD4d and PtCCD4e, located in chromosome IX, and separated 9.8 Mb (c and d) and 11. 9 Mb (d and e) showed over 92% identity. Tandemly duplicated genes are considered to be relatively younger than the genes duplicated via whole-genome duplication [24-25], and show high identity sequence levels. By contrast, in *Z. mays* and *O. sativa* the two genes were located in different chromosomes. Rice genes showed 62% identity, which is similar to the identities observed between CCD4 genes from different plant species, and the *Vitis* genes located in different chromosomes showed identities around 54 and 53%. Although located in different chromosomes the poplar genes PtCCD4a and PtCCD4b showed 96% identity, this high similarity could be due to the fact that duplication events, nucleotide substitutions, tandem gene duplications, and gross chromosomal rearrangement appear to proceed substantially more slowly in poplar than in other plant species [26].

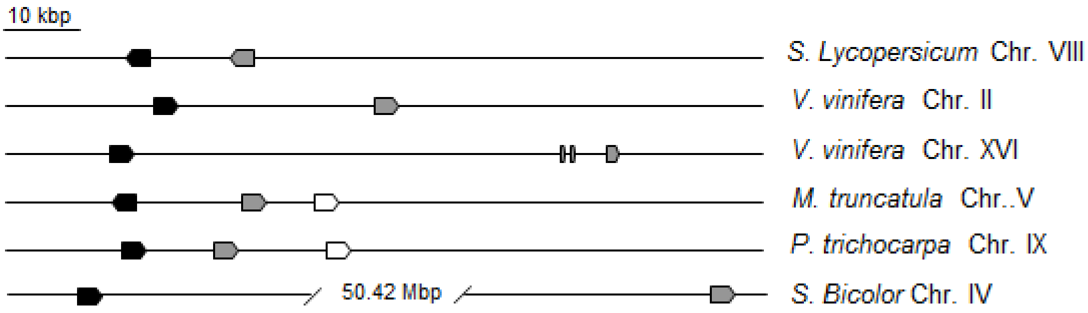


Figure 1.Schematic representation of the genomic organization of the *CCD4* genes found in several plant species. The CCD4 duplicates and genes fragments are shown in different gray scale colours. Arrows indicate gene orientation.

The CCD4 proteins were used to study the phylogenetic relationships. The phylogenetic tree with representative members of the different carotenoid cleavage dioxygenases and the CCD4 phylogenetc tree suggested that the CCD4 plant proteins could derive from a common ancestor, with the monocotyledon group later being separated (Fig 2 and Fig. 3). All the deduced amino acid sequences of CCD4 proteins contain a plastid-targeting transit peptide at the N-terminus and four highly conserved His residues (Fig. 4) that may be involved in coordinating a nonheme iron that is required for enzymatic activity [5, 27].

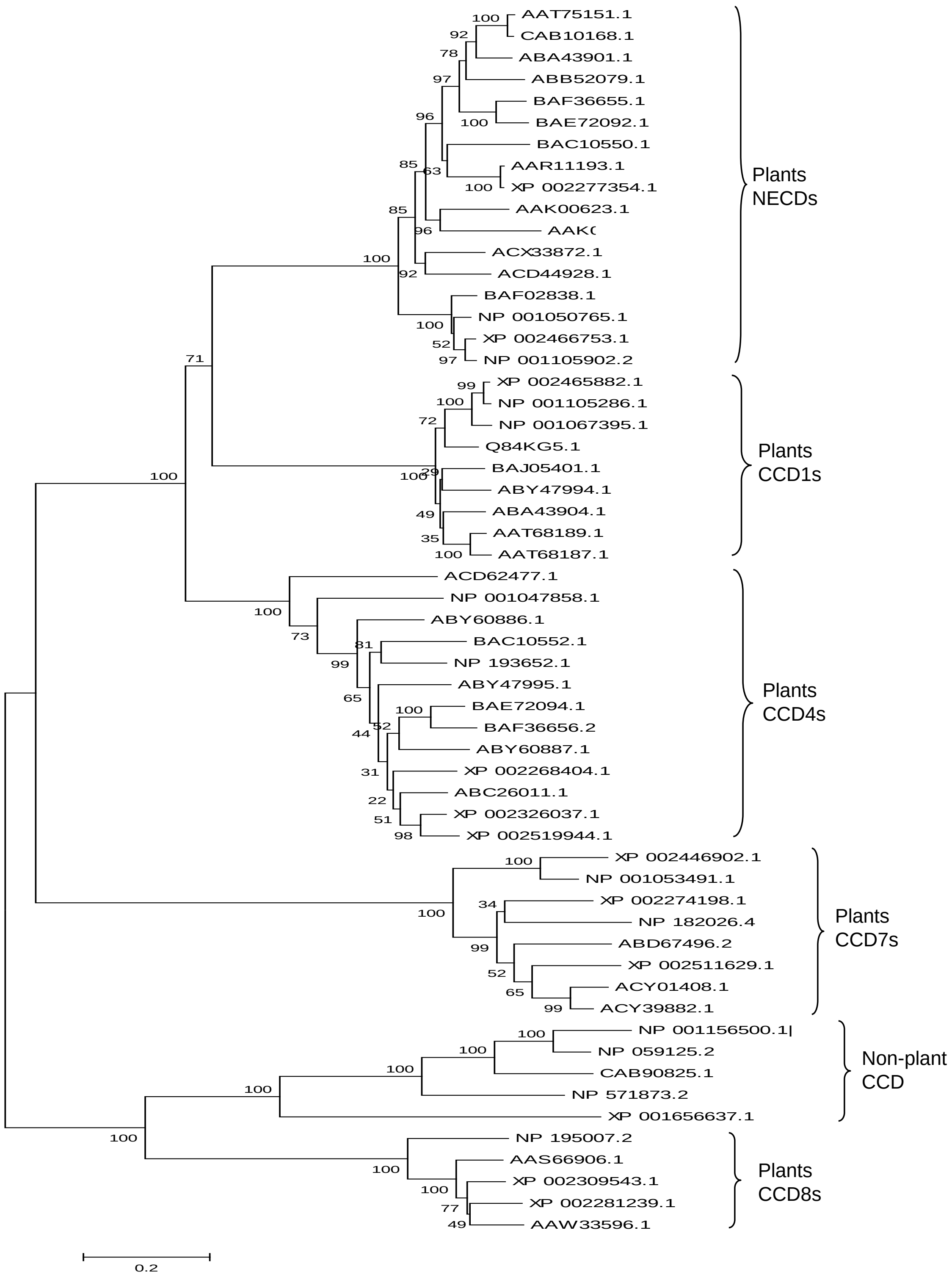


Figure 2. Phylogenetic tree of plant carotenoid cleavage enzymes and related proteins with their GenBank Accession numbers. The tree was inferred by the neighbor-joining method after alignment of the amino acid sequences of 57 carotenoid cleavage proteins.

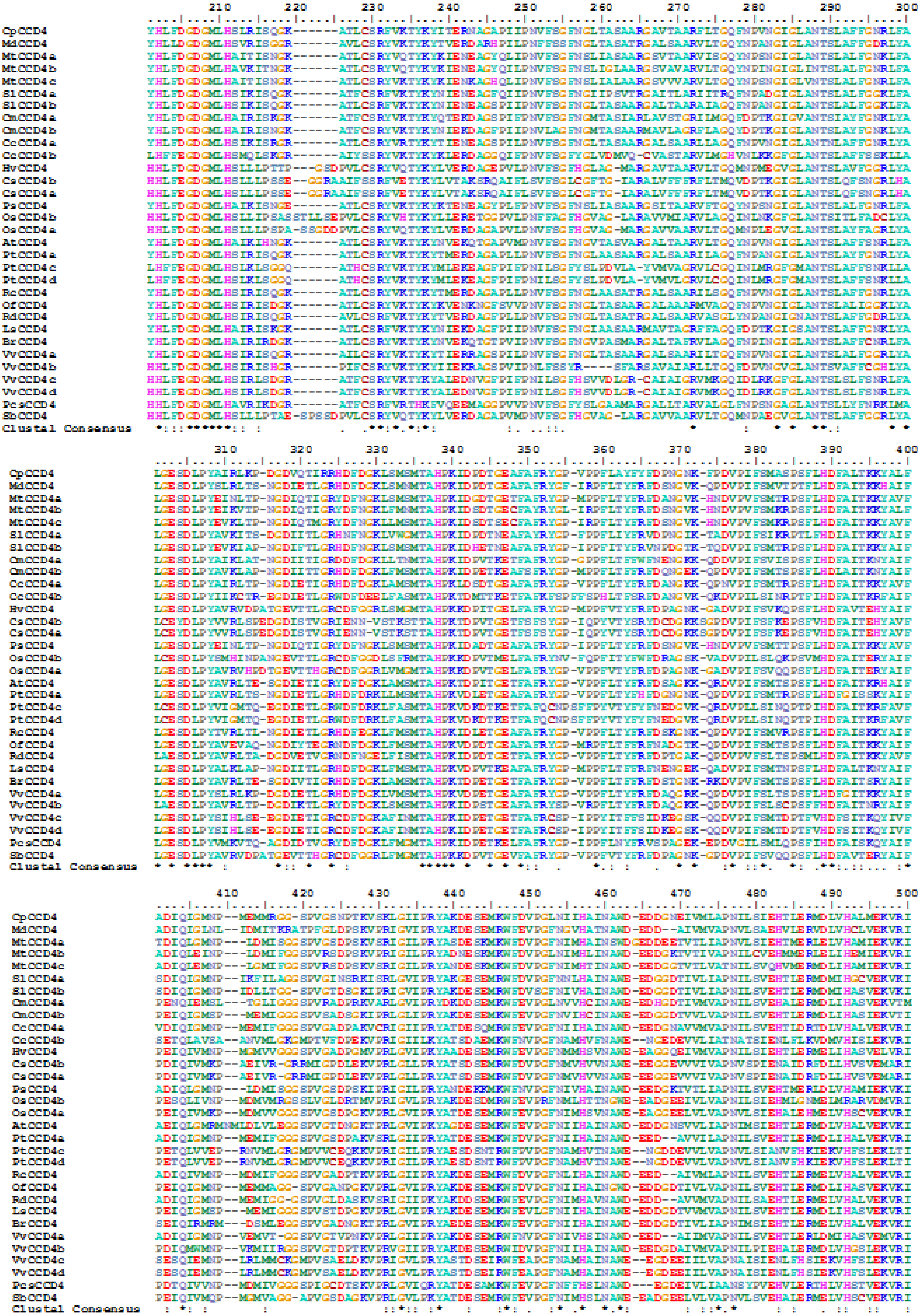


Figure 3. Alignment of the CCD4 amino acid sequences. Alignment of the CCD4 amino acid sequences showing the presence of four highly conserved His residues that may be involved in coordinating a nonheme iron that is required for enzymatic activity. Histidines are in pink.

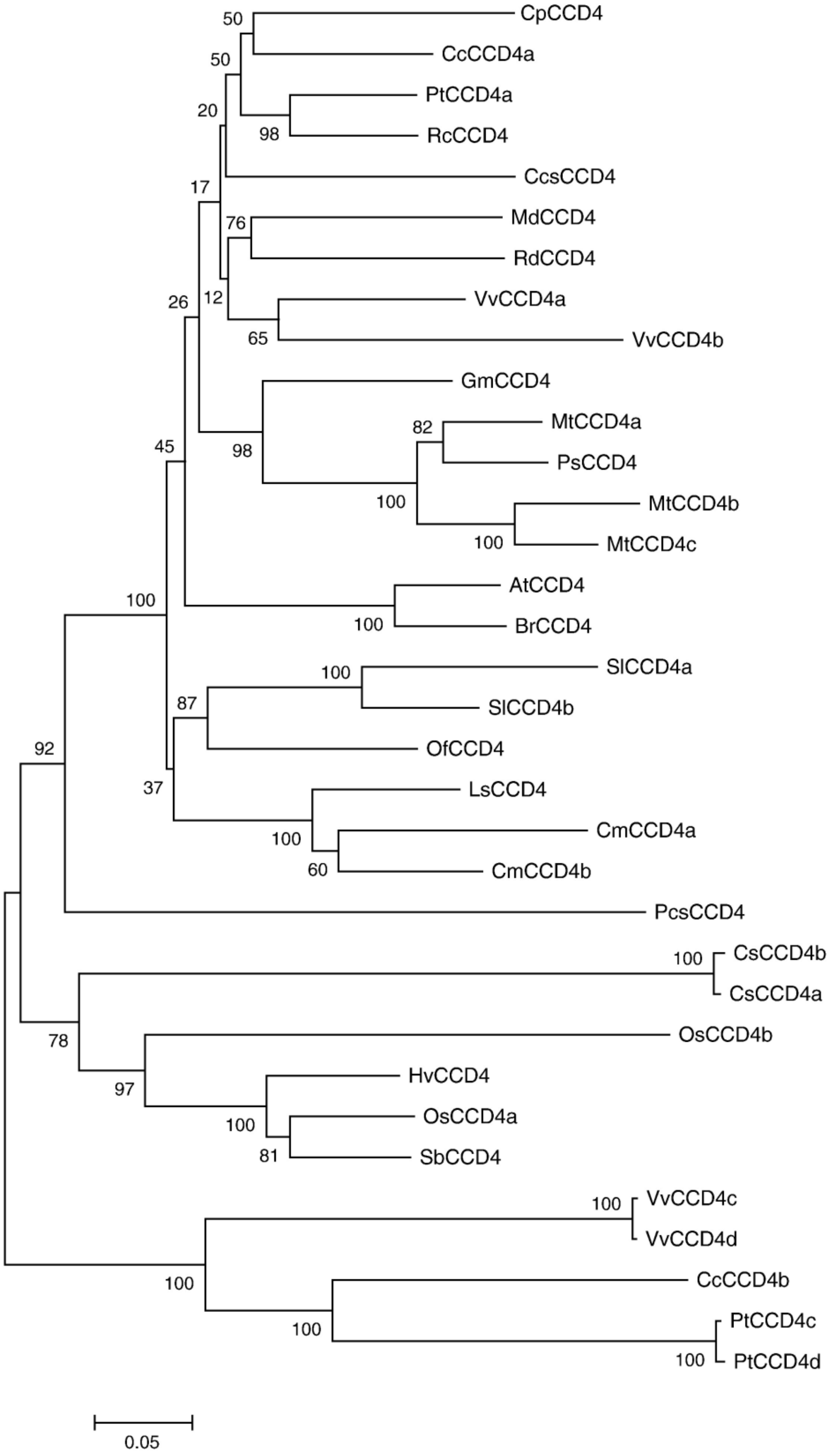


Figure 4. Phylogenetic tree of the CCD4 proteins based on amino acid sequence similarity. Bootstrapping values (500 replicates) are shown on the tree. Only full-length members of the family are

included. The predicted protein sequences were initially clustered using ClustalW. Accession numbers are in Table 3.

Genomic organization in *Crocus sativus*

Using specific oligonucleotides for CsCCD4a and CsCCD4b, we isolated four genomic clones from *C. sativus* leaves (Fig. 5A and B). Two copies presented one intron, other copy contained no intron sequences and a third copy was a truncated clone previously described [7], which has been only detected in *C. sativus*. CsCCD4a and CsCCD4b are considered to be alleles and their presence could be due to the hybrid origin of *Crocus sativus* and its triploid nature. Two related *Crocus* species, *C. cartwrigthianus* and *C. hadriaticus* have been suggested as possible parentals of *Crocus sativus*[28]. Therefore, the presence of CCD4 was investigated in other *Crocus* species (Fig. 5A). Both CCD4a and CCD4b homologous genes were found to be polymorphic among and in some of the analysed species. Interestingly, major differences were found for CCD4a genes. Differences in the number of copies could be associated with the differences observed in chromosome numbers [29]. Sequence analysis showed that the differences in size observed among the amplified products were due to intron absence or presence and intron length, although intron position was always maintained among the sequences analysed.

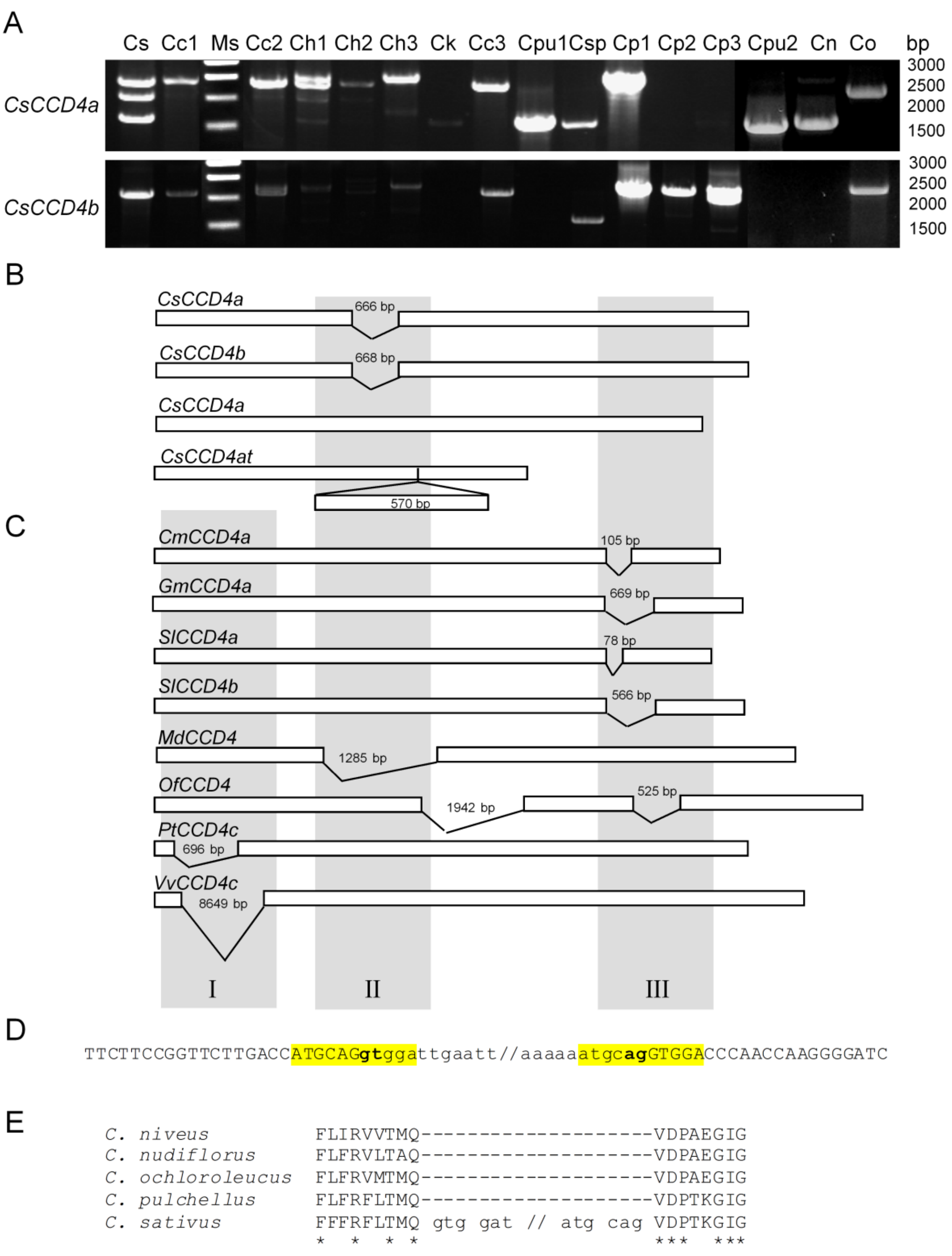


Figure 5. Gene structure of *CCD4*. A. A representative result of an ethidium bromide gel showing PCR amplifications with the primer sets for *CCD4a* and *CCD4b* from genomic DNA of different *Crocus* species: Cs, *C. sativus*; Cc1, *C. cartwrightianus* (12629); Cc2, *C. cartwrightianus* (12627); Cc3, *C. cartwrightianus* (G94-15); Ch1, *C. hadriaticus* (98-31); Ch2, *C. hadriaticus* (01-37); Ch3, *C. hadriaticus* (01-32); Ck, *C. kostchyanus*; Cpu1, *C. pulchellus* (98-81); Cpu2, *C. pulchellus* (I-19); Csp, *C. speciosus* (UK); Cp1, *C.*

pallasii (01-61); Cp2, *C. pallasii* (01-64); Cp3, *C. pallasii* (01-45); Cn, *C. niveus*(UK); Co, *C. oreocreticus*(G94-15). **B**. Diagrammatic representation of the *CsCCD4* genes of *Crocus sativus*. Boxes and lines denote coding regions and introns, respectively. **C**. Diagrammatic representation of *CCD4* genes of *Crysanthemun morifolium* (*CmCCD4*), *Glycine max* (*GmCCD4*), *Solanum lycopersicum*(*SlCCD4a* and *SlCCD4b*), *Malus domestica*(*MdCCD4*), *Osmanthus fragans* (*OfCCD4*), *Populu strichocarpa* (*PtCCD4c*) and *Vitis vinifera* (*VvCCD4c*). Boxes and lines denote coding regions and introns, respectively. Introns sizes are shown by numbers. Three grey columns indicate the three introns positions. **D**. Representation of the direct inverted repeats present in the exon-intron boundaries of *CsCCD4a* and *CsCCD4b* introns. Intron sequence (lower case) with the repeat (highlighted in yellow) and splice sites (bold). The remaining intronic sequence finds no significant BLAST hit within NCBI. **E** Sequence alignment between five species showing amino acid substitutions around the intron site. Ms corresponds to the molecular size marker, and Msbp corresponds to the lengths in bp of Ms.

Intron distribution in CCD families

To further study the genomic structure of the *C.s sativus* CCD4 genes, the genomic clones were compared with the CCD4 genes found in the *A. thaliana*, *C. papaya*, *O. sativa*, *S. lycopersicum*, *S. bicolor*, *V. vinifera*, *P. trichocarpa*, *M. truncatula*and *B. rapa* genomes and the genomic clones isolated from *M. domestica*, *O. fragans*, *R. damascena*, and *C. morifolium*. Table 2 shows the characteristics of the CCD4 genes, including a comparative analysis of the intron position in the nucleotide sequence among CsCCD4 and other CCD4 genes. The absence or presence and the pattern position of the introns in the different CCD4 genes were observed (Table 1 and 3 and Fig. 5C). Introns were not widely distributed among the sequences analysed; from the available genomic sequences, only 25% of the sequences contain at least one intron. The *G. maxima* gene showed an intron of 669 bp, located at 1,140bp from the start codon (Fig. 5C). The two *S. lycopersicum* CCD4 genes showed the presence of one intron, located at 1,446 bp from the start codon but with different lengths, as observed for two of the five CsCCD4 genes identified (Fig. 2C). The *M. domestica* CCD4 gene has an intron of 1,285 bp at 535 bp from the start codon [8] (Fig. 5C). For the CCD4a gene from Chrysanthemum, a 105 bp intron at 1,442 bp from the start codon has been reported [6] (Fig. 5C). Two introns were observed in *O. fragans* CCD4, 1,942 bp and 491 bp long, and inserted at 878 bp and 1470 bp from the start codon, respectively [8] (Fig. 5C). And finally, one intron located close to the 5’ end was detected in CCD4c from poplar and CCD4c from *Vitis*, 696 and 8649 bp long, respectively an located at 62 and 87 bp from the start codon, respectively (Fig. 5C). The global analysis of intron positions reveals the presence of three main insertion sites for introns in the CCD4 sequences (Fig. 5C). Analysis of the insertion positions reveals that the first and second positions do not appear to be conserved, whereas introns present close to the 3’end of the sequence can be considered as shared since they were found in the same amino acid position in the sequence alignments of the FASTA formatted files containing the CCD4 protein.

The intronic sequences of CsCCD4a (GenBank accession no. FN908206) and CsCCD4b (GenBank accession no. FN908207) showed a 94% identity. The AT content of both introns (62% and 60.6%) conforms to the AT-rich bias of introns from other plant genes [30]. Both introns are in phase zero and are inserted in a G|G proto-splice site at 638 and 605 bp from the transcription start codon of CsCCD4a and CsCCD4b, respectively. Both genes showed practically conserved the consensus sequences (A/C)AG|GU(A/G)AGU and the CAG|G, at the donor and acceptor splice sites, respectively (Fig. 5D). In addition, the acceptor splice sequence was preceded by a polypirimidine tract, typical of the acceptor splice site, which is recognised by the U5 snRNA of the splicesoma[31]. We observed that short direct repeats of 11 base pairs (ATGCAGGTGA) flank both introns, with one repeat positioned within the end of an adjacent exon and the other repeat near the opposite end of the intronic sequence (Fig. 5D). We searched BlastN and BlastX Genbank and found no significant match to these introns, which do not possess any detectable long inverted repeats that are often associated with transposable elements nor do they contain ORFs longer than 30 aa.

It has been proposed that the functionally important (A/C)AG|G exon sequences flanking introns are relics of recognition signals for the insertion of introns or protosplice sites [31], and that introns, indeed, predominantly insert into or are preferentially fixed in specific protosplice sites, which have this consensus sequence. Due to the observed difference in intron presence among the *Crocus* species studied, the sequences flanking the intron insertion were analysed (Fig. 5E). The sequence CAAG|Gt was conserved in the species that show introns in the CCD4 genes and in the species that lack introns in the CCD4 genes, suggesting that these consensus nucleotides flanking the splice junctions were present in the original protosplice site and did not evolve convergently after the intron insertion. Intron number and position have been determined for some CCD1, CCD7 and CCD8 genes. We also analysed the number of intron sequences in the available genomic sequences for these gene families that had not been previously analysed. For CCD1, more than 10 introns were detected for all the available genomic sequences. *Cucumis sativus*, *V. vinifera*, *R. communis*, *G. maxima*, and *S. lycopersicum* with 12 introns; and *P. trichocarpa*, *Z. mays*, *B. rapa* and *S. bicolour* containing 11 introns (Fig. 7).

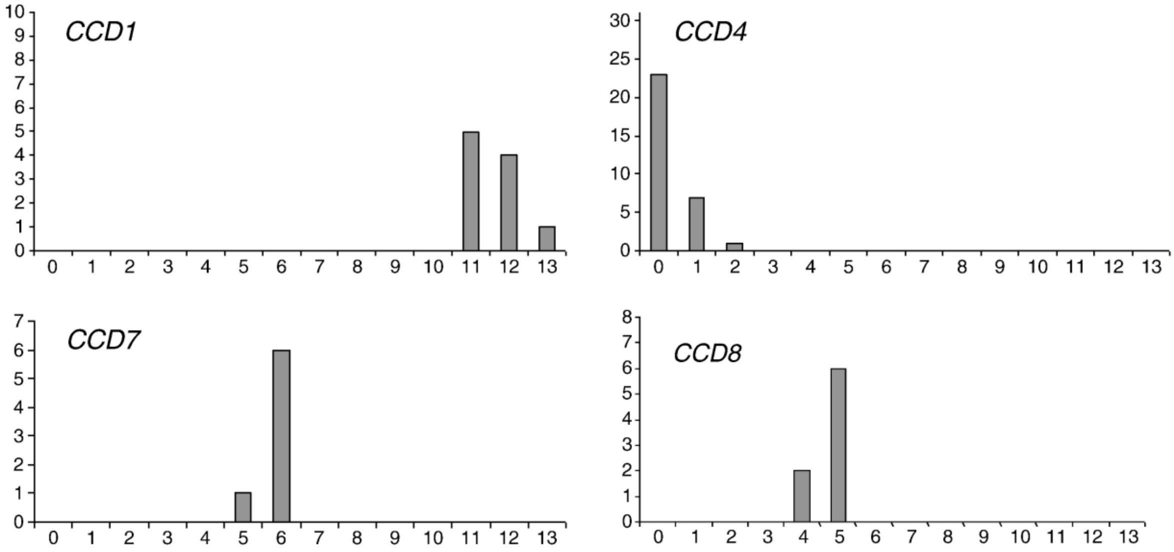


Figure 6. Frequency histograms of intron numbers in all the *CCD* genes examined.
Abscissa, intron numbers; ordinate, number of genes, which showed numbers of intron insertions in the coding regions as indicated on the abscissa.

In addition to the already characterized CCD7 genes, with four well conserved introns in terms of position, six introns have been identified in *S. lycopersicum*, *A. annua*, and *S. bicolour* (Fig. 7). For CCD8, a total of 5 introns were identified in *Musa balbisiana*, *P. trichocarpa*, *V. vinifera*, and *B. rapa*, whereas 4 introns were detected in *Z. mays* and *S. bicolour* (Fig. 7). The results show that in contrast to CCD1, CCD7 and CCD8 genes where intron presence and conservation seems to be characteristic of these genes (Fig. 7), in the CCD4 genes the intron-exon structure is highly dynamic.

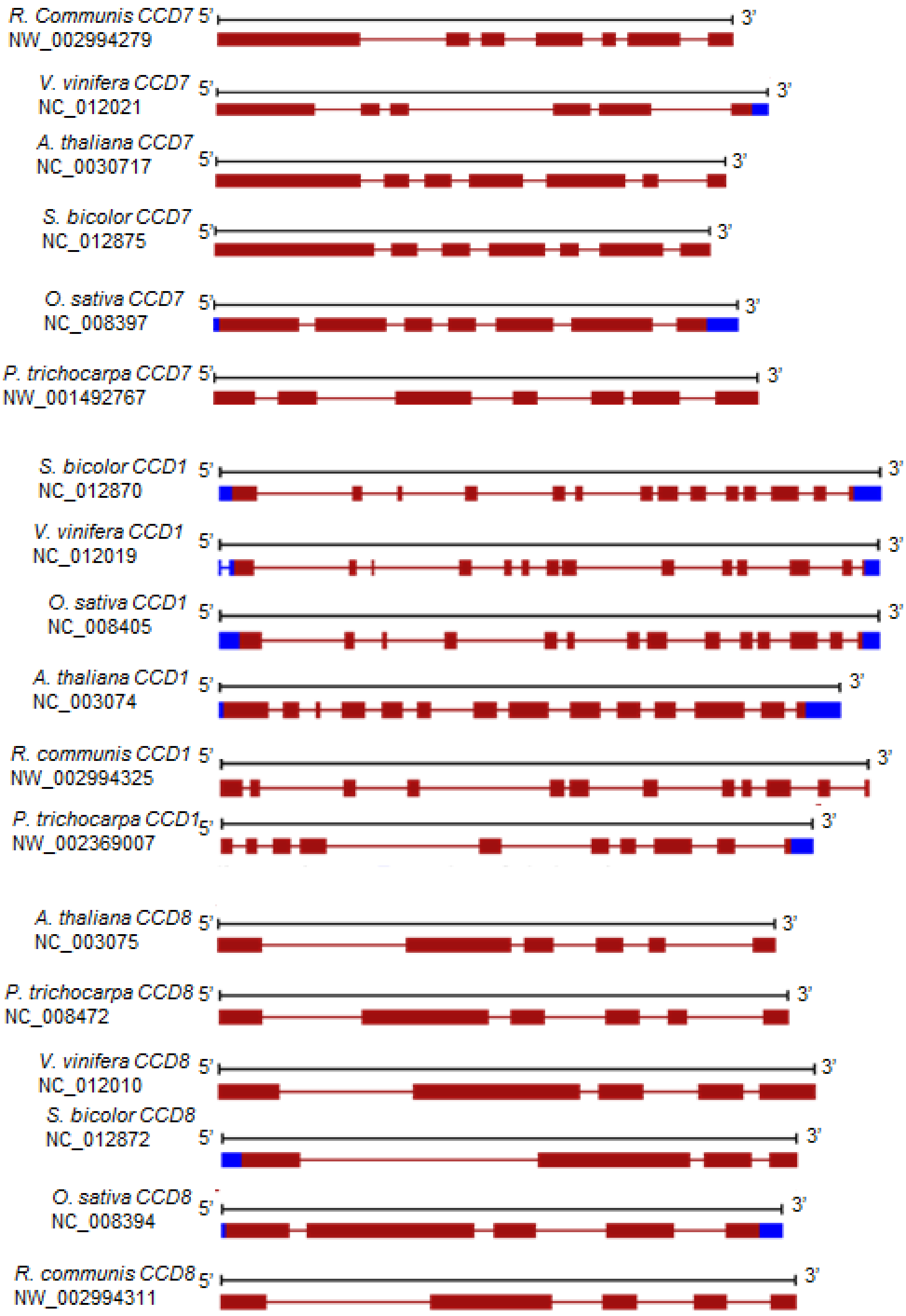


Figure 7. Analysis of gene structure among *CCD7*, *CCD8* and *CCD1* homologues.

Exons and introns are in red, exons are represented as closed boxes and introns as solid lines. In blue are shown the untranslated regions.

Sequence analysis of CCD4 promoters

The genomic DNA fragment upstream from the CCD4 genes was isolated by using a PCR based genome walking procedure for *C. sativus*. Comparison of the CsCCD4a (GenBank accession no. FN908208) and CsCCD4b (GenBank accession no. FN908209) promoter regions showed a 93% identity (Fig. 8). This high conservation at the promoter level, together with the high conservation at the level of the coding sequences, confirmed the allelic relationships of CsCCD4a and CsCCD4b genes. Both were compared with other promoters, and the highest identity was obtained with the CCD4a and CCD4b rice promoter sequences. Pair-wise alignment with these sequences gave 48% identity. Both *S. lycopersicum* promoter regions showed a 54% identity, the *M. truncatula* sequences showed a 52% identity among each other, and a 51% identity was found for *V. vinifera* sequences from VvCCD4a, VvCCD4b and VvCCD4c/d. The *P. trichocarpa* promoters of PtCCD4c, PtCCD4d and PtCCD4e showed 59% identity and 60% identity between PtCCD4a and PtCCD4b. All the promoters were analysed for conserved motifs, and overall we observed a poor degree of conservation among all the promoters analysed (Fig. 10). The promoters of rice and saffron were analysed separately. Although several conserved motifs were shared between both promoters, the spatial organization of these motifs clearly diverged (Fig. 8).

In silico promoter analysis showed that for CsCCD4a the most proximal putative TATA-boxes were located at -45, -59 and at -82 positions. In CsCCD4b the most proximal TATA-box was located at -23. Conserved CAAT-box sequences were located farther upstream in the promoter sequence of CsCCD4a from -460 to -785 and at -70, -252 and -587 in CsCCD4b. Potential regulatory elements were identified upstream from the CCD4 gene using the Plant-Care databases [18], and included consensus sequences for the binding of transcription factor proteins like Myb (five different consensus sequences), several light response elements, heat stress elements, low temperature response elements, and a salicylic acid-responsive-element among others (Table 4). These elements were also identified in the promoter regions from *M. truncatula*, *A. thaliana*, *B. rapa*, *C. papaya*, *V. vinifera*, *S. lycopersicum*, *P. trichocarpa* and *Oryza sativa*.(Table 4).

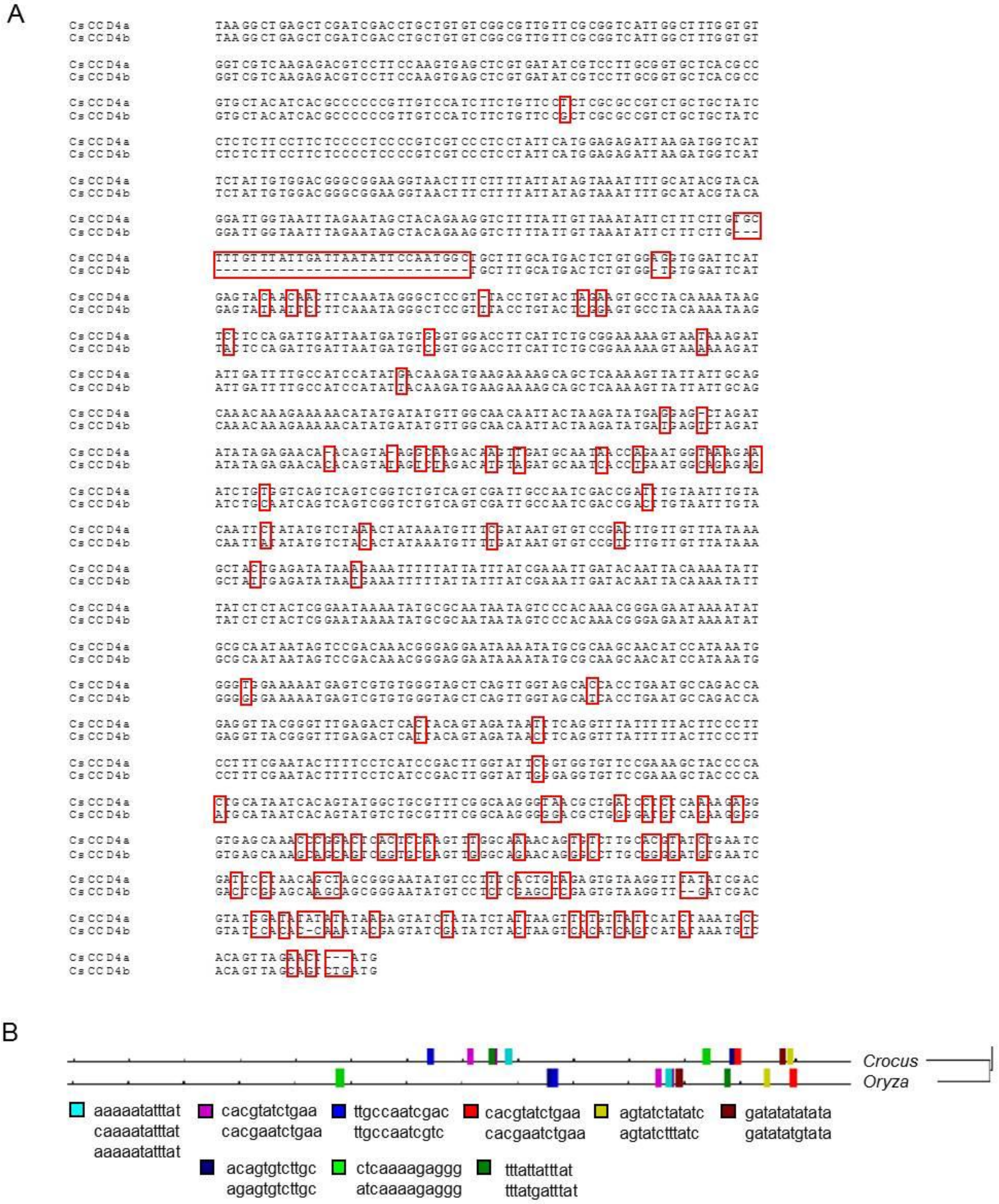


Figure8. Promoter analysis of *CsCCD4*. A. Sequence alignment of *CsCCD4a* and *CsCCD4b* promoter regions. The differences between both promoters are highlighted by red boxes. **B**. Visual representation of motifs detected by FOOTPRINTER in the promoters of rice and saffron *CCD4* genes (FOOTPRINTER parameters: motif size: 12, allowed mutations 2, losses).

Table 4. Promoter elements identified in the promoter region of the *CCD4* genes.

Positions of the cis-acting elements are indicated for the *CCD4* promoters. These positions are indicated with respect to the ATG initiation codon. Only the elements that have been localized in *CsCCD4a* and *CsCCD4b* promoters are shown.

Sequence	1	2	3	**4**	**5**	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
GGTAAAGAAA	902			**1104** **744**	**1075**		863 821										982 788			
TACGTG	700		807	**149**			111			1285 1064										
ATTAAT		11818 28 806	1297 947 132	**10809** **60**	**962**					1020 616 669 244	1097 327 276	1048 1028	1095 857 621 567	1036 296 68	609			1145		
CACGTA	776	1350	807	**149**	**420**		211			1283 1064	669			668					1196	751
CACGAC	157	1097		**421** **149**							217		260	641	1074					
AAGGAAGA	1375		158	**1268**	**1239**					1193		1213						584	681	1133
GCATTC	704 1315			**388**	**388** **778**	417 139 854		738		980		1231	1045				205			285 157 38
CCGAAA				**270** **231**	**269** **230**						990 920 900									
AAAAAATTTC	389 452 1011	760 111	109	**599**	**598**		880	1284	864 884			599	983	281	1088	945 1334	1000	841 269 687	1189 1051	519
ATTTTCTCCA		730 921		**885**	**887**	396 1238	319 639			1012	780	1280	1138 872	1175		1214	1219	528		
CCATCTTTTT	943			**205**	**204**	869	786	310	908		596 445		276			343	348	337	434	1159
TGACG					**1359**					1044	201	22	1182	579						
CGTCA				**1388**	**1354**					604	21		158	419						
CGGTCA	1020	608	226	**1406** **406** **14**	**13814** **03** **16**			355	472		945	143		930	17	919	924	11401 098		336
CAANNNNATC	1183	10841 068	404	**254**		1079 633	247 753	511	834	1107	1290	1070 916	880 681	564	800	628 718	798	442	235	424 157

Spatial and temporal expression patterns of P-CsCCD4a::GUS in transgenic *Arabidopsis* plants

To determine the spatial and temporal expression patterns of CsCCD4a, we examined GUS staining in various tissues at different developmental stages in 3-day-old, and 7-week-old flowering transgenic *Arabidopsis* plants. CsCCD4a Prom::GUS expression was only observed in flowers (Fig. 9A and B) and no signal was detected in vegetative tissues (Fig. 9C and D) including stem and roots (data not shown). GUS expression was not detected in immature flower buds before opening (data not shown) and expression increased when flowers were fully open (Fig. 9A). CsCCD4a Prom::GUS expression was mainly detected in pollen and low staining was observed in stigmas, style surface and sepals (Fig. 9B). The analysis of the CsCCD4a promoter using the PLACE programme revealed the presence of elements required for flower expression: CARGATCONSENSUS with the sequence CCWWWWWWGG; two different elements for pollen expression: GTGANTG10, with the sequence GTGA repeated 8 times, and POLLEN1LELAT52 with the sequence AGAAA repeated 6 times; and for vascular expression: BS1EGCCR, with the sequence AGCGGG.

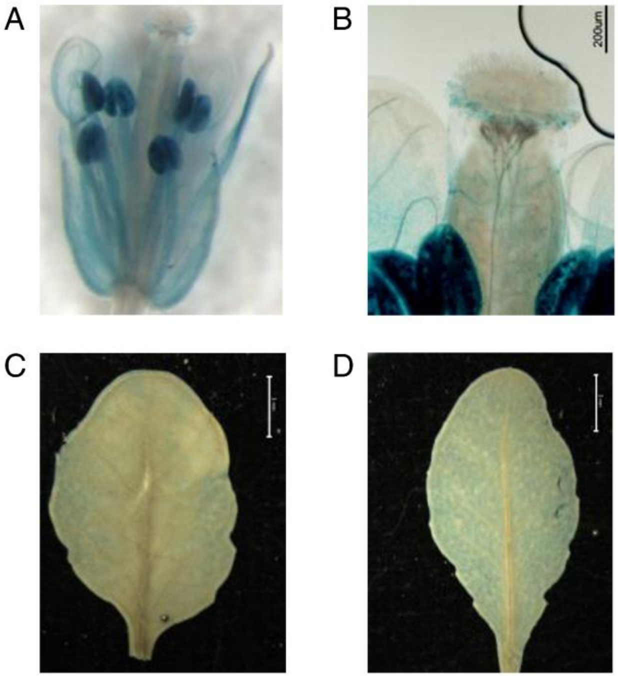


Figure 9. Histochemical assay of *GUS* reporter gene in transgenic plants of *Arabidopsis* driven by the CsCCD4a promoter. A Opening flower. **B.** Detail of the stigma surface. **C** and **D**. Leaves from a 35 d-old transgenic plant.

Discussion

In this work, we have analysed CCD4 gene structure in relation to CCD4 genes from related *Crocus* species and also from other plant species, revealing a history of intron loss and gain during the evolution of these genes. In addition, we report the sequences of the putative promoter regions of CsCCD4a and CsCCD4b genes, and a functional evaluation of a 1,400-bp genomic DNA fragment from the CsCCD4a promoter.

Several CCD4 copies were found in more than 70% of the analysed plant genomes. In addition to the CCD4 gene fragment found in *Vitis*, gene fragments of CCD4 were also identified in the same chromosome in *S. bicolor*, and *P. trichocarpa*, indicating that these CCD4 duplicates were pseudogenized (data not shown). The presence of a truncate CCD4 gene in *C. sativus* suggest that this gene, probably originally duplicated from CsCCD4a, suffered the same process.

Gene and genome duplications have been shown to be particularly prominent in plant genomes and have greatly influenced their organization and evolution [32-34]. Gene duplication is achieved via whole-genome duplication or local tandem duplication. Tandem gene duplication is mainly caused by unequal crossing over between homologous chromosomes or through unequal crossing over. Gene duplicates may get lost over time or if maintained these duplication events have profound effects on gene function and regulation [35]. The analysis of CCD4 gene sequence identity and chromosome location suggested that mainly tandem and whole genome duplication are responsible for the presence of CCD4 copies in different plant species. Furthermore, although the coding sequences of the CCD4 tandemly duplicate genes, as do those found in *M. truncatula* and poplar are relatively well conserved, the upstream regulatory regions among these CCD4 genes in each plant are highly divergent, except for the core promoter near the translation start site of poplar and *M. truncatula* genes (Fig. 10). The data suggest that the expression of these duplicated genes is likely to be regulated differently. Tandemly duplicated genes tend to be rapidly diverged in expression [36], and duplicated genes are easily released from original selection pressures, providing adaptive and developmental novelty.

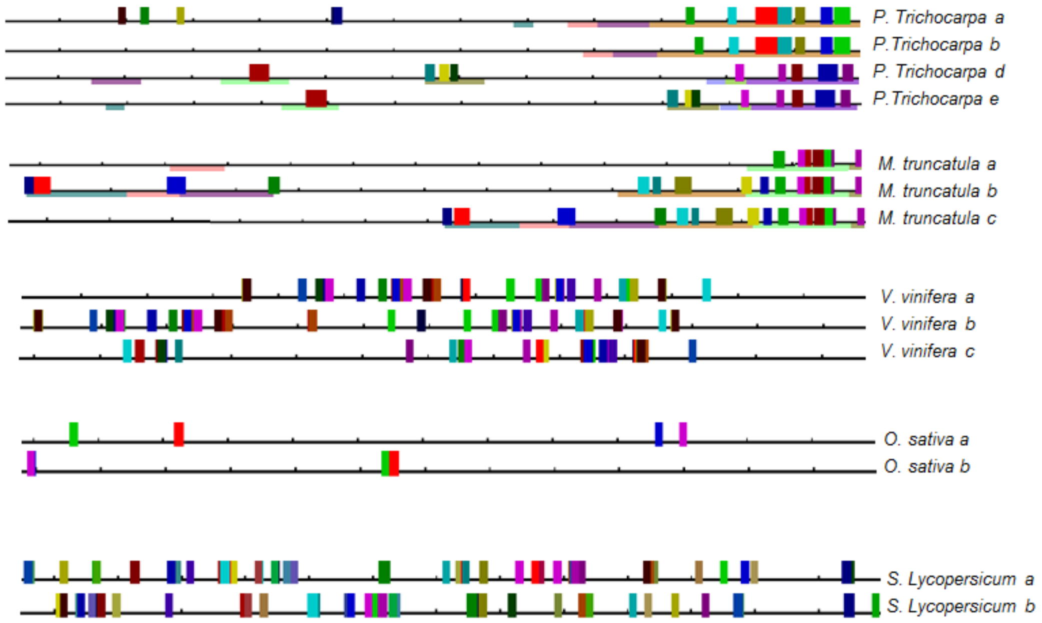


Figure 10. Visual representation of motifs detected by FOOTPRINTER in the promoters of *CCD4* genes from different plant species (FOOTPRINTER parameters: motif size: 12, allowed mutations 2, losses).

Promoter analysis showed the presence of several potential cis-regulatory elements in CsCCD4a and CsCCD4b promoters and many were found to be shared with other CCD4 gene promoter sequences from *Arabidopsis*, *O. sativa*, *S. lycopersicum*, *M. truncatula*, *V. vinifera* and *B. rapa*.

Genomic analysis among different *Crocus* species showed differences in the presence of CCD4a and CCD4b alleles, as well as differences in intron presence. Absence of one of the two alleles or the presence of both does not follow a species-specific pattern, as can be observed from the analysis of different *C. pallasi* and *C. hadriaticus* populations. In these populations, in addition to *C. oreocreticus*, *C. sativus* and *C. cartwrightianus*, all belonging to the same series *Crocus*, we observed the presence of the CCD4 alleles with the intron in the genomic sequences. Interestingly, for the other *Crocus* species analysed, which belong to the Longiflori, Kotschyani, Versicolores and Speciosi series, no introns were detected in any of the populations analysed.

The early eukaryotic progenitor has been assumed to be intron-rich on the basis of the presence of introns in homologous positions of orthologous genes of widely divergent eukaryotes [37]. In addition, the presence of several introns in other plant and animal CCDs suggests that the CCD gene ancestor was intron rich. The genes included in CCD1, CCD7 and CCD8 gene families are all characterized by the presence of several intronic sequences, and only in the related NCED group are all the analysed genes intron-less [38]. In the intron-containing CCD4 sequences, intron number and position were not conserved among the CsCCD4 alleles. Taking into account that a wide variety of green plants from vascular plants to enslaved algae show an excess of intron loss over intron gain [39], the results obtained suggest loss of introns in CCD4 sequences. Furthermore, the analysis of intron number and position suggest the presence of an ancestral intron in CCD4, lost in more than 89% of the available genomic sequences, but still present in *S. lycopersicum*, *C. morifolium*, *G. maxima* and *O. fragans*. However, with the available data it cannot be determined if that insertion occurred before or after the divergence of monocots and dicots, although it has not been detected in any monocot sequences.

Massive losses of ancestral introns have been postulated to have occurred in select lineages, and it has been suggested that rates of intron gain and loss have been declining in the past 1.3 billion years in most eukaryotes, with a greater decline of gains than losses [40]. Furthermore, intron gain has been argued to be a rare event, with a rate on the order of $<4 \times 10^{-6}$ per coding site per million years, which is orders of magnitude lower than estimated rates of loss [41]. However, the other introns present in CsCCD4a and b, MdCCD4, along with the first intron of OfCCD4 seem not to be really conserved, if assuming that an intron is conserved in two or more genes when it occupies the same position in these genes [37]. In this context these introns can be considered as newly gained introns. Further support for this possibility came from the absence of this intron in CCD4 sequences from *Crocus* species that do not belong to the *Crocus* series, in agreement with the fact that the rate of intron gain is so variable between species [42].

There are at least three global mechanisms for the de novo origin of intron positions [43]: (i) transposition, which would include duplication of preexisting introns, even though this process is extremely rare; (ii) insertion of intron-like transposons; and (iii) novel introns which develop gradually via the optimisation of previously non-intronic sequence. These mechanisms assume that (i) every new intron position originates from a “formative” intron, (ii) formative introns derive from intron donors elsewhere in the genome (including introns, transposons, and exons), and (iii) formation of a novel intron position is instantaneous. The CsCCD4 intronic sequence finds no significant BLAST hit within NCBI, excluding the second mechanism of intron gain elements. Furthermore, a manual inspection of the sequence flanking this novel intron identified a single event reminiscent of tandem duplication. Several recently discovered novel introns are flanked by short direct repeats [42, 44]. These repeats may represent the signature of non-homologous end joining (NHEJ) after uneven double-stranded breaks (DSBs), a process known to generate insertions flanked by direct repeats. In consideration of this, we note that the duplication observed here may also be explained by a direct repeat flanking a sequence of unknown origin, suggesting the third mechanism as responsible for the generation of this new intron in the *Crocus* series.

Recent studies have shown that intron number and average/total intron length per gene have a significant negative correlation with expression levels in monocot and dicotyledonous plants [45-46]. Various hypotheses have been proposed to explain the detrimental nature of introns with respect to gene expression. This could be associated to the relative costs of introns, in terms of the kinetics of transcription and splicing, which occur concurrently in the nucleus, where the splicing of two or more introns requires more time than transcription and becomes rate limiting. Thus, transcripts that are modulated during stress, either in increasing or decreasing levels under different conditions, are significantly under enriched for introns [47], as seems to be the case for the stress-regulated NCEDs genes [38]. In addition, the CsCCD4a and CsCCD4b genes have been shown to be regulated in leaves under different stress conditions, in contrast to the CsCCD1 gene [7].

Several expression studies showed that CCD4 genes are preferentially expressed in floral organs [6, 48]. In *C. sativus*, the CsCCD4 genes have been detected in different flower tissues, but both genes were preferentially expressed in stigma, with this expression developmentally regulated and associated to the accumulation of carotenoids and apocarotenoids along the style and stigma [7]. In transgenic *Arabidopsis* plants, GUS staining was mainly detected in pollen and weak expression was observed in sepals, style and stigma tissues. Although expression was restricted to flower organs, the expression was not as expected from the previously CsCCD4 expression in *C. sativus*. The differences could be due to the different flower organization of *C. sativus* compared with *A. thaliana*. The flower of *C. sativus* consists of six petaloid three tepals in whorl 1 (outer tepals) and three tepals in whorl 2 (inner tepals), 3 stamens and a unique style divided into 3 stigmas. A modified ABC model has been proposed to explain this flower morphology [49]. Therefore, the differences in the expression pattern may be due to a different set of transcription regulatory factors connected with the determinate architecture of *C. sativus*. Nevertheless, promoters derived from monocot species often fail to exhibit a regulated pattern of expression in transgenic dicots, due in many cases to the absence of conserved trans-acting factors and cis-acting elements that enable these promoters to be regulated in a tissue-specific fashion in both dicots and monocots. Interestingly, the phylogenetic analysis showed that the promoters of rice and saffron CCD4 genes grouped together, and several conserved motifs were clearly detected showing a one-to-one relationship, although changes in spacing were observed. However, these motifs were not shared with the dicotyledonous CCD4 gene promoters, suggesting the presence of specific cis-acting elements in these monocot promoters.

References

1. Auldridge ME, McCarty DR, Klee HJ: **Plant carotenoid cleavage oxygenases and their apocarotenoid products**. *Curr Opin Plant Biol* 2006, **9**(3):315-321.
2. Camara B, Bouvier F: **Oxidative remodeling of plastid carotenoids**. *Arch Biochem Biophys* 2004, **430**(1):16-21.
3. Pfander H, Wittwer F: **[Carotenoid composition in safran (author's transl)]**. *Helv Chim Acta* 1975, **58**(7):2233-2236.
4. Kloer DP, Schulz GE: **Structural and biological aspects of carotenoid cleavage**. *Cell Mol Life Sci* 2006, **63**(19-20):2291-2303.
5. Tan BC, Schwartz SH, Zeevaart JA, McCarty DR: **Genetic control of abscisic acid biosynthesis in maize**. *Proc Natl Acad Sci U S A* 1997, **94**(22):12235-12240.
6. Ohmiya A, Kishimoto S, Aida R, Yoshioka S, Sumitomo K: **Carotenoid cleavage dioxygenase (CmCCD4a) contributes to white color formation in chrysanthemum petals**. *Plant Physiol* 2006, **142**(3):1193-1201.
7. Rubio A, Rambla JL, Santaella M, Gomez MD, Orzaez D, Granell A, Gomez-Gomez L: **Cytosolic and plastoglobule-targeted carotenoid dioxygenases from *Crocus sativus* are both involved in beta-ionone release**. *J Biol Chem* 2008, **283**(36):24816-24825.
8. Huang FC, Molnar P, Schwab W: **Cloning and functional characterization of carotenoid cleavage dioxygenase 4 genes**. *J Exp Bot* 2009, **60**(11):3011-3022.
9. Yan W, Jang GF, Haeseleer F, Esumi N, Chang J, Kerrigan M, Campochiaro M, Campochiaro P, Palczewski K, Zack DJ: **Cloning and characterization of a human beta,beta-carotene-15,15'-dioxygenase that is highly expressed in the retinal pigment epithelium**. *Genomics* 2001, **72**(2):193-202.
10. Johnson X, Brcich T, Dun EA, Goussot M, Haurogne K, Beveridge CA, Rameau C: **Branching genes are conserved across species. Genes controlling a novel signal in pea are coregulated by other long-distance signals**. *Plant Physiol* 2006, **142**(3):1014-1026.
11. Drummond RS, Martinez-Sanchez NM, Janssen BJ, Templeton KR, Simons JL, Quinn BD, Karunairetnam S, Snowden KC: ***Petunia hybrida* carotenoid cleavage dioxygenase7 is involved in the production of negative and positive branching signals in petunia**. *Plant Physiol* 2009, **151**(4):1867-1877.
12. Auldridge ME, Block A, Vogel JT, Dabney-Smith C, Mila I, Bouzayen M, Magallanes-Lundback M, DellaPenna D, McCarty DR, Klee HJ: **Characterization of three members of the Arabidopsis carotenoid cleavage dioxygenase family demonstrates the divergent roles of this multifunctional enzyme family**. *Plant J* 2006, **45**(6):982-993.
13. Snowden KC, Simkin AJ, Janssen BJ, Templeton KR, Loucas HM, Simons JL, Karunairetnam S, Gleave AP, Clark DG, Klee HJ: **The Decreased apical dominance1/Petunia hybrida carotenoid cleavage dioxygenase8 gene affects branch production and plays a role in leaf senescence, root growth, and flower development**. *Plant Cell* 2005, **17**(3):746-759.
14. Zwanenburg B, Mwakaboko AS, Reizelman A, Anilkumar G, Sethumadhavan D: **Structure and function of natural and synthetic signalling molecules in parasitic weed germination**. *Pest Manag Sci* 2009, **65**(5):478-491.
15. Rose AB: **Intron-mediated regulation of gene expression**. *Curr Top Microbiol Immunol* 2008, **326**:277-290.
16. Rokas A, Holland PW: **Rare genomic changes as a tool for phylogenetics**. *Trends Ecol Evol* 2000, **15**(11):454-459.
17. Blanchette M, Tompa M: **FootPrinter: A program designed for phylogenetic footprinting**. *Nucleic Acids Res* 2003, **31**(13):3840-3842.
18. Rombauts S, Dehais P, Van Montagu M, Rouze P: **PlantCARE, a plant cis-acting regulatory element database**. *Nucleic Acids Res* 1999, **27**(1):295-296.
19. Nakagawa T, Kurose T, Hino T, Tanaka K, Kawamukai M, Niwa Y, Toyooka K, Matsuoka K, Jinbo T, Kimura T: **Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation**. *J Biosci Bioeng* 2007, **104**(1):34-41.
20. Clough SJ, Bent AF: **Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana**. *Plant J* 1998, **16**(6):735-743.
21. Jefferson RA, Kavanagh TA, Bevan MW: **GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants**. *EMBO J* 1987, **6**(13):3901-3907.

22. Agusti J, Zapater, M., Iglesias, DJ., Cercos, M., Tadeo, FR. Talon, M. : **Differential expression of putative 9-cis-epoxycarotenoid dioxygenases and abscisic acid accumulation in water stressed vegetative and reproductive tissues of citrus**. *Plant Sciences* 2007, **172**(1):10.
23. Sawada Y, Aoki M, Nakaminami K, Mitsuhashi W, Tatematsu K, Kushiro T, Koshiba T, Kamiya Y, Inoue Y, Nambara E *et al*: **Phytochrome- and gibberellin-mediated regulation of abscisic acid metabolism during germination of photoblastic lettuce seeds**. *Plant Physiol* 2008, **146**(3):1386-1396.
24. Rizzon C, Ponger, L., Gaut, B.S.: **Strikimg similarities in the genomic distribution of tandemly arrayed genes in arabidopsis and rice**. *PLoS Computational Biology* 2006, **2**:e115.
25. Kliebenstein D: **A role for geneduplication and natural variation of gene expression in the evolution of metabolism**. *PloS ONE* 2008, **3**:e1838.
26. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A *et al*: **The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray)**. *Science* 2006, **313**(5793):1596-1604.
27. Schwartz SH, Tan BC, Gage DA, Zeevaart JA, McCarty DR: **Specific oxidative cleavage of carotenoids by VP14 of maize**. *Science* 1997, **276**(5320):1872-1874.
28. Castillo R, Fernandez JA, Gomez-Gomez L: **Implications of carotenoid biosynthetic genes in apocarotenoid formation during the stigma development of *Crocus sativus* and its closer relatives**. *Plant Physiol* 2005, **139**(2):674-689.
29. Ozhatay N: **Diversity of bulbous monocots in turkey with special reference. Chromosome numbers**. *Pure and Applied Chemistry* 2002, **74**:9.
30. Goodall GJ, Filipowicz W: **The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing**. *Cell* 1989, **58**(3):473-483.
31. Rogozin IB, Sverdlov AV, Babenko VN, Koonin EV: **Analysis of evolution of exon-intron structure of eukaryotic genes**. *Brief Bioinform* 2005, **6**(2):118-134.
32. Otto SP, Whitton J: **Polyploid incidence and evolution**. *Annu Rev Genet* 2000, **34**:401-437.
33. Wendel JF: **Genome evolution in polyploids**. *Plant Mol Biol* 2000, **42**(1):225-249.
34. Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y: **Modeling gene and genome duplications in eukaryotes**. *Proc Natl Acad Sci U S A* 2005, **102**(15):5454-5459.
35. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J: **Preservation of duplicate genes by complementary, degenerative mutations**. *Genetics* 1999, **151**(4):1531-1545.
36. Casneuf T, De Bodt S, Raes J, Maere S, Van de Peer Y: **Nonrandom divergence of gene expression following gene and genome duplications in the flowering plant Arabidopsis thaliana**. *Genome Biol* 2006, **7**(2):R13.
37. Fedorov A, Merican AF, Gilbert W: **Large-scale comparison of intron positions among animal, plant, and fungal genes**. *Proc Natl Acad Sci U S A* 2002, **99**(25):16128-16133.
38. Tan BC, Joseph LM, Deng WT, Liu L, Li QB, Cline K, McCarty DR: **Molecular characterization of the Arabidopsis 9-cis epoxycarotenoid dioxygenase gene family**. *Plant J* 2003, **35**(1):44-56.
39. Roy SW, Penny D: **Patterns of intron loss and gain in plants: intron loss-dominated evolution and genome-wide comparison of O. sativa and A. thaliana**. *Mol Biol Evol* 2007, **24**(1):171-181.
40. Carmel L, Rogozin IB, Wolf YI, Koonin EV: **Evolutionarily conserved genes preferentially accumulate introns**. *Genome Res* 2007, **17**(7):1045-1050.
41. Carmel L, Rogozin IB, Wolf YI, Koonin EV: **Patterns of intron gain and conservation in eukaryotic genes**. *BMC Evol Biol* 2007, **7**:192.
42. Farlow A, Meduri E, Dolezal M, Hua L, Schlotterer C: **Nonsense-mediated decay enables intron gain in Drosophila**. *PLoS Genet* 2010, **6**(1):e1000819.
43. Tarrio R, Ayala FJ, Rodriguez-Trelles F: **Alternative splicing: a missing piece in the puzzle of intron gain**. *Proc Natl Acad Sci U S A* 2008, **105**(20):7223-7228.
44. Li W, Tucker AE, Sung W, Thomas WK, Lynch M: **Extensive, recent intron gains in *Daphnia populations***. *Science* 2009, **326**(5957):1260-1262.
45. Ingvarsson PK: **Gene expression and protein length influence codon usage and rates of sequence evolution in *Populus tremula***. *Mol Biol Evol* 2007, **24**(3):836-844.
46. Yang H: **In plants, expression breadth and expression level distinctly and non-linearly correlate with gene structure**. *Biology Direct* 2009, **4**.
47. Jeffares DC, Penkett CJ, Bahler J: **Rapidly regulated genes are intron poor**. *Trends Genet* 2008, **24**(8):375-378.

48. Yamamizo C, Kishimoto S, Ohmiya A: **Carotenoid composition and carotenogenic gene expression during Ipomoea petal development**. *J Exp Bot* 2009.
49. Kanno A, Nakada M, Akita Y, Hirai M: **Class B gene expression and the modified ABC model in nongrass monocots**. *ScientificWorldJournal* 2007, **7**:268-279.

Chapter 3

“Crocins transport in *Crocus sativus*: the long road from a senescent stigma to a newborn corm”

Abstract

Saffron, the desiccated stigmas of *Crocus sativus*, is highly appreciated for its peculiar colour, flavour and aroma. The main compounds that accumulate throughout stigma development in *C. sativus*, formed from the oxidative cleavage of β-carotene and zeaxanthin, are crocetin, its glucoside derivatives, crocins, and picrocrocin, all of which increase as stigmas reach a fully developed stage. After anthesis, and in the absence of fertilization, the flower enters into a senescence programme. Senescence represents the ultimate stage of floral development and results in the wilting of entire flower. The programmed senescence of flowers allows for the removal of metabolically active tissue. Senescence is an active process and governed by a well-defined cell death programme. During the senescence of *C. sativus* flowers, changes in the composition of saffron apocarotenoids occurs and transport of crocins from the senescent stigma to the ovaries and the developing corm takes place. Afterwards, deglucosylation of crocins in these tissues results in crocetin accumulation. This mobilization mimics the export to storage cells (resorbed) of different compounds during leaf senescence, thus avoiding loss of nutrients in leaves that would otherwise be cycled back into the soil system through leaf litter decomposition. In *C. sativus*, the resorbed apocarotenoids are stored within the developing corm, and used for its active metabolism during early and active phases of corm development, where the glucose molecules from crocins stimulate cell initiation and elongation.

Keywords: apocarotenoids, corm, flower senescence, saffron, stigma, transport

Introduction

Carotenoids are widely distributed isoprenoid pigments fulfilling diverse functions in all taxa [1]. The oxidative cleavage of carotenoids leads to the release of a range of apocarotenoids that function as signalling molecules with diverse functions [2-4], including the ubiquitous chromophore retinal, phytohormone abscisic acid, strigolactones, a group of C_{15} apocarotenoids attracting both symbiotic arbuscular mycorrhizal fungi and parasitic plants [5, 6] and, as recently shown, exerting functions as novel plant hormones regulating shoot branching [7-9]. Other apocarotenoids with unknown functions in plants but with high economic value are bixin in *Bixa orellana*[10] and crocetin in *Crocus sativus*[11]. However, the ability to synthesize crocetin is not restricted to *C. sativus* and other related species, but is also present in fruits of *Artocarpus heterophyllus*[12] the flower extract of *Buddleja*[13], in *Jacquinia angustifolia*[14], in *Coleus forskolii*[15], in the fruits of *Gardenia*[16], and even in the cyanobacterium *Microcystis*[17].

In *C. sativus*, these apocarotenoids have only been detected in the stigma tissue. The accumulation of these compounds in the stigma begins early in the development, crocetin is already detected in the earlier stages, with levels rapidly increasing to a maximum in the red-immature stage and decreasing thereafter in the mature stigma and at anthesis. The total content in crocins follows the same behavior as crocetin, but remains high in the mature stages until the time of anthesis[18]. The stigmas of *C. sativus* are collected from saffron preparation when flowers emerge, or at anthesis. After the flowers wither, new corms are produced asexually above the old corm and continue to develop until leaves enter in senescence and dry out: while at that time the mother corm have deteriorated completely [19]. Corms are essential for *C. sativus* propagation, which is unknown in the wild state [20], and depending on corm weight, each corm includes 1-4 apical bud(s) which are origins of flowers and leaves or just leaves, thus suggesting that the corms should have a threshold age to be capable of producing flowers [20, 21].

C. sativus is an autumn flowering geophyte, and intensive differentiation of *C. sativus* flower organs occurs when daughter corms are “dormant”, with temperature being the key factor controlling flowering [22]. The female organs of *C. sativus* consist of a trilocular ovary, a long style, and 3 long stylar branches forming the stigmas [23]. Flower pollination never results in fertilization due to the male sterility of saffron.

Although the pollen grains stick to the papillae and undergo rapid hybridization, a high number of pollen tubes fail to grow and the few fertilized ovules do not succeeded in setting seeds [24, 25]. Therefore, either in the presence or in the absence of pollination, the next step in the developmental process of saffron flowers is senescence (Fig. 1). The senescence of floral organs which have lost their role in the post-pollination development of the flower, such as sepals, stamens, styles and stigmas, is connected with the remobilization of reserve substances from dying parts of the flower to the ovary or in the absence of fertilization to other sinks in the mother plant [26]. However, relatively small attention has been given to this process.

The process of flower senescence and the changes in the apocarotenoid profile during senescence have not been studied in *C. sativus*, mainly due to the fact that the real value of saffron lies in the fresh stigmas, which are removed from flowers collected at the time of anthesis. In this paper, we have studied the accumulation of several apocarotenoids and volatiles during flower senescence and observed a transport of saffron apocarotenoids from the stigma tissue to the ovaries and to the developing corm. Thus, in the absence of fertilization and seed development, a part of the material invested is translocated backwards to sinks in somatic tissues, like the developing corm over the mother plant.

Material and methods

Plant material

Flower tissues from *C. sativus* grown under field conditions in Tarazona de la Mancha, Spain, were used throughout the experiments. Stigmas, style, ovaries, peduncle and developing corms were collected at different developmental stages. All tissues were frozen in liquid nitrogen and stored at -80 °C until required.

Extraction and analysis of apocarotenoid compounds by HPLC

Apocarotenoids from different flower parts were extracted in a 2 mL microcentrifuge tube by grinding 30 mg of material with a micropestle in 700 µL Tris-HCl (50 mM, pH 7.5) (containing 1 M NaCl), and incubated for 10 min on ice. Chloroform was then added (700 µl), mixed and the extract incubated on ice for an additional 10 min. The phases were separated by centrifugation at 3,000g for 5 min at 4 °C. The lower chloroform phase was evaporated under vacuum and the dried residues were stored together with the upper aqueous phases at -80 °C until high-performance liquid chromatography analysis. The assays were all performed in triplicate.

Apocarotenoids were analysed by reverse phase HPLC as previously described [27] using a Hewlett Packard 1100 HPLC (Palo Alto, CA) equipped with a SugerlaborInertsil ODS-2.5-µm C18 column (250 x 4.6 mm) and connected on line to a photodiode array detector, with a dynamic range from ultraviolet to visible region (190-700nm). The different carotenoid derivatives were identified on the basis of HPLC retention times and UV-visible light spectra [27, 28], and their concentrations calculated as previously described [29].

Sugar analysis and aglycon determination

Acid hydrolysis of one of the fractions of the senescent extract was performed to determine the sugar moiety and the aglycon nature of the apocarotenoids detected. The sample was incubated at 85 °C with 1 M HCl for 120 min. Samples were cooled after hydrolysis and neutralized with NaOH. After appropriate dilution of the neutralized samples, the neutral sugars were converted into their corresponding alditol acetates [30] and identified and quantified by GLC as described previously [31]. Another aliquot was analysed by LC-ESI-MS using a C18 Ascentis column (15 x 2.1 mm; particle size 3 um) (Supelco, Sigma-Aldrich) and following the method previously described [18].

Results

Apocarotenoid profile in senescent stigma

The presence of the apocarotenoids crocetin, picrocrocin and the different crocins in senescent stigmas was investigated by extracting senescent stigmas (3 weeks after anthesis) as previously described [32] and analysing the extracts by reverse HPLC. These stigmas were characterized by a low apocarotenoid content, with C4 (crocetin di-(β-D-gentibiosyl) ester) (442-466 nm) and C3 (crocetinβ-D-gentibiosyl-β-D-glucosyl ester) (440-464 nm) were present at ± 1μmol/fresh weight (Fig. 1A) and by the presence of several apocarotenoid compounds (retention time from 14-20 min) showing an UV/Vis spectra similar to crocins but with unidentified structure (Fig. 2A). These compounds were hydrolysed and the resulting products were analysed by GLC (data not shown) and LC-ESI-MS (Fig. 1B and C). Glucose was detected as the unique sugar compound and crocetin was detected as the aglycon of these compounds and was found in cis and trans configurations (Fig. 1B-E). In contrast, the hydrolysis of crocins obtained during the preanthesis and anthesis stage showed only trans-crocetin (data not shown). The picrocrocin precursor 4-hydroxy-2,6,6-trimethyl-1-cyclohexen-1-carboxaldehyde (HTCC) was also detected in this senescent stigma, but the levels were around 20 times less compared with the stigma at anthesis[33], while picrocrocin was not detected at this stage.

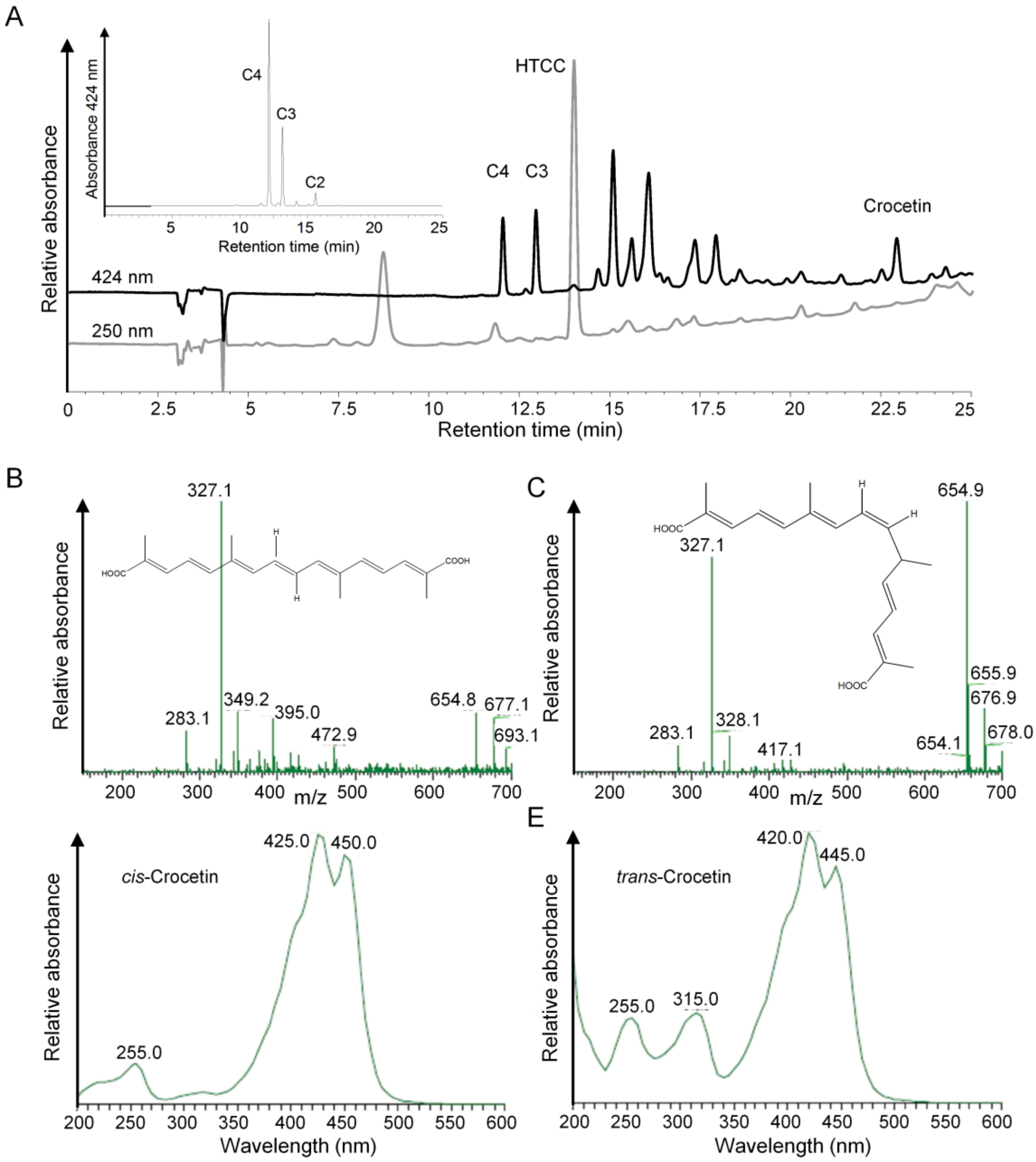


Figure 1 HPLC elution profiles of apocarotenoids from saffron stigmas, three weeks after anthesis. **A**. In black is shown the HPLC elution profile of the aqueous extract of *C. sativus* stigmas measured at 424 nm. The apocarotenoids C4, C3 and crocetin were detected and are indicated. In grey is shown the HPLC elution profile of the chloroform extract of *C. sativus* stigmas measured at 250 nm. The main peak detected corresponds to HTCC and the second non-labelled peak corresponds to flavonoids. Inset a representative HPLC elution profile of extracts of stigmas at anthesis recorded at 424 nm is shown. **B**. Negative ion mass spectrum of transcrocetin acquired during the HPLC-ESI-MS analysis and its chemical structure. **C**. Negative ion mass spectrum of ciscrocetin acquired during the HPLC-ESI-MS analysis and its chemical structure. **D**. Online diode-array spectrum of transcrocetin. **E**. Online diode-array spectrum of ciscrocetin.

Apocarotenoid transport

In addition to be male sterile, *C. sativus* flowers showed self-incompatibility [34]. Hence, in absence of genetically suitable pollen the unfertilised flowers of *C. sativus* begin to wilt (Fig. 2A-C). In an intent to study the process of flower senescence in *C. sativus*, visual colour changes during this process where observed as a progressive increase in colour in apex to base direction along the style tissue, which further progressed through the ovary and the peduncle until reaching the new developing corm (Fig. 2D). A dissection of these corms showed the accumulation of orange compounds close to the base, where the new developing corms are attached to the mother corms (Fig. 2E). When stigmas were removed from the flowers at anthesis and the developing corms were dissected, it was not possible to detect the presence of those orange compounds, which were also absent in the developing corms from non-producing flowers shoots (Fig 2F and G). In order to determine the nature and origin of these pigments, extracts from the stigma, style, ovary, peduncle and the developing corm from preanthesis (2 days before anthesis) flowers (Fig. 2H) and senescent flowers (10 days after anthesis) were analysed by HPLC, in addition to developing corms from non-flowering plants.

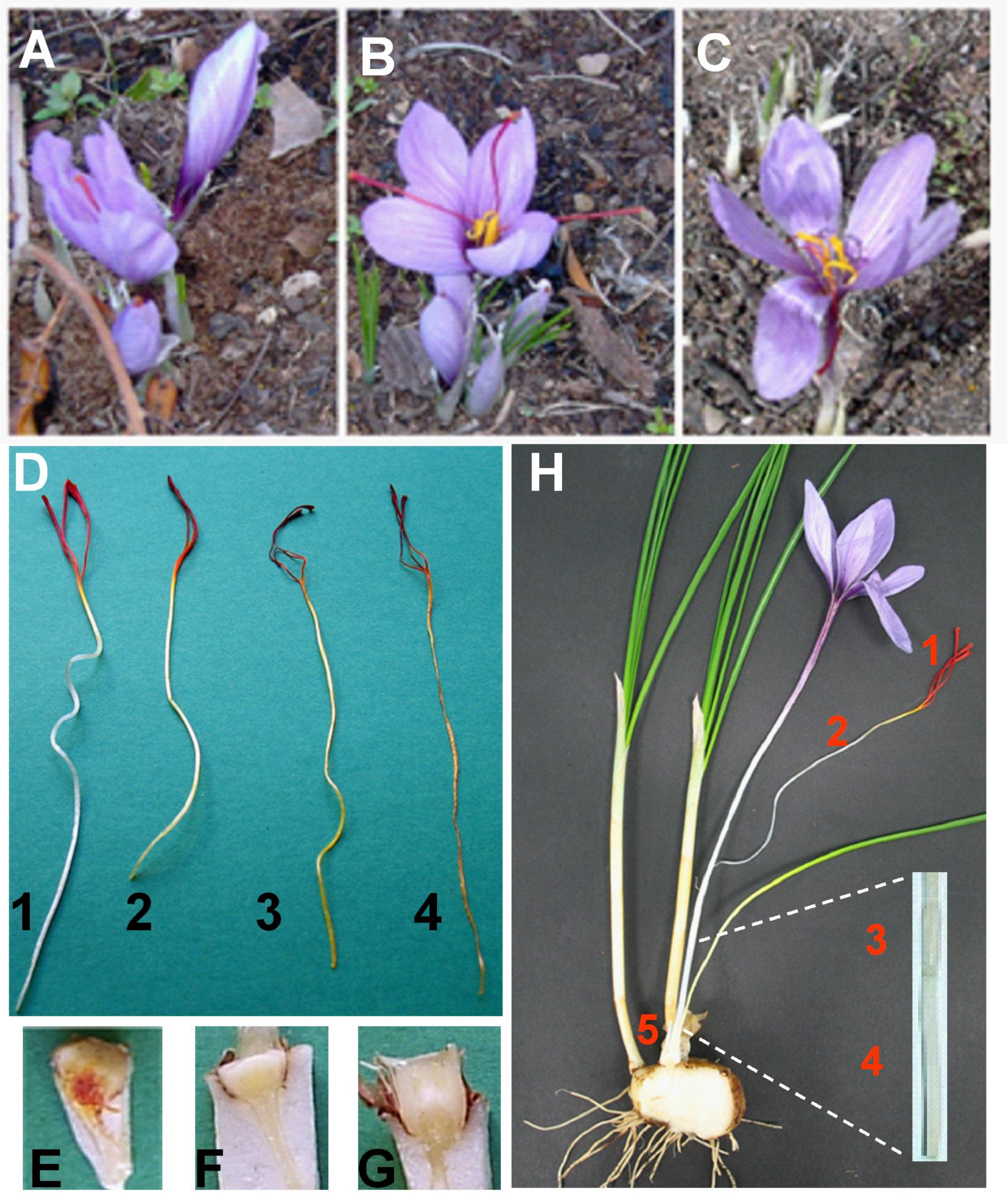


Figure 2. Visual changes associated with the senescent process in flowers of *Crocus sativus*.

A. Flowers in preanthesis (close flowers) and anthesis. **B** Flowers in preanthesis and one day after anthesis showing turgent petals and withered stigmas. **C**. Flowers in postanthesis (7 days) showing translucency of the petals and withered stigma. **D**. Colour changes associated with the senescent process in stigma and style: 1, from flowers at anthesis; 2, from flowers five days after anthesis; 3, from flowers ten days after anthesis; 4, from flowers 20 days after anthesis. **E**. Dissection of a developing corm from shoots with flowers ten days after anthesis. **F**. Dissection of a developing corm from shoots with flowers without stigmas ten days after anthesis. **G**. Dissection of a developing corm from shoots that do not produce flowers. **H**. An open and dissected flower in preanthesis showing the different parts analysed for saffron apocarotenoids: 1, stigma; 2, style; 3, ovary; 4, peduncle; 5, developing corm attached to the mother corm.

In the stigma from postanthesis flowers the major apocarotenoids detected were the trans crocins C5 (crocetin three-(β-D-glucosyl)-β-D-gentibiosyl) ester) (444-468 nm), C4, C3, C2 (crocetinβ-D-gentibiosyl ester or di-(β-D-glucosyl) ester) (434-460 nm) and C1 (crocetinβ-D-glucosyl ester) (434-458 nm) were identified in addition to crocetin (428-452 nm) and picrocrocin (250 nm) (Fig. 3A). The analysis of the organic extract showed the presence of HTCC, the picrocrocin precursor, which was not detected in the preanthesis stigma (data not shown). The stigma from preanthesis flowers did not differ so much from the senescent stigma, although minor levels of apocarotenoids were observed in terms of fresh weight, as previously described for stigmas at this developmental stage [18]. However, picrocrocin levels were higher at this stage than in senescent stigmas (Fig. 3A), which was also found for the stigma at anthesis (data not shown). In the style from postanthesis flowers the major apocarotenoids detected were the trans-crocins C5 (444-468 nm), C4 (442-466 nm), C3 (440-464 nm), C2 (434-460 nm) and C1 (434-458 nm) and crocetin (C) (428-452 nm), among them the highest level was found for the C2 crocin (Fig. 3 A and 4 A). By contrast, in the preanthesis style, C1 was not detected and C2 was at lower levels compared with C4 and C3 crocins (Fig. 3 A and 4 A), and the style was characterized by the presence of high levels of flavonoids in comparison with the other tissues analysed (data not shown). In the preanthesis ovary C4 and C2 were found at similar levels and C1 and crocetin were also detected at lower levels (Fig. 3 A). All these compounds were found in the senescent ovary, but at higher levels (Fig. 3 A and 4 A), with the exception of C1 (Fig. 4 A). Major changes were observed between the peduncle of senescent flowers and the peduncle of preanthesis flowers. Very low levels of C4, C2 and C1 could be detected in the preanthesis stage (Fig. 3 A), whereas the peduncle from senescent flowers accumulated 40 fold levels of C2 and 150 fold levels of C4 (Fig. 3 A and 4 A).

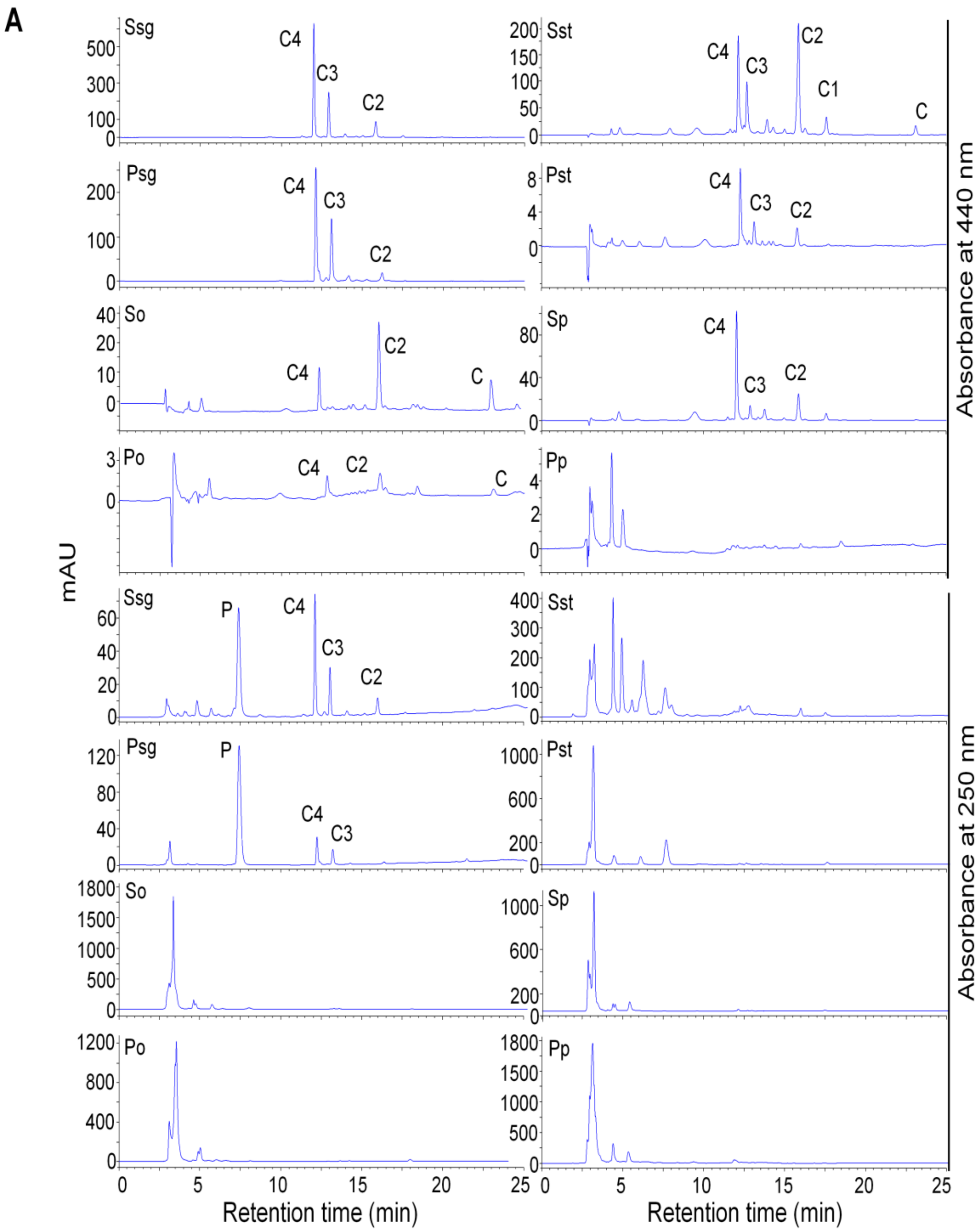


Figure 3A HPLC profiles of saffron apocarotenoids in different tissues and in different developmental stages. HPLC elution profiles measured at 424 and 250 nm for the detection of crocins, crocetin, picrocrocin and HTCC on: Ssg, senescent stigma; Psg, preanthesis stigma; Sst, senescent style; Pst, preanthesis style; So, senescent ovary; Po, preanthesis ovary; Sp, senescent pedancle; Pp, preanthesis peduncle.

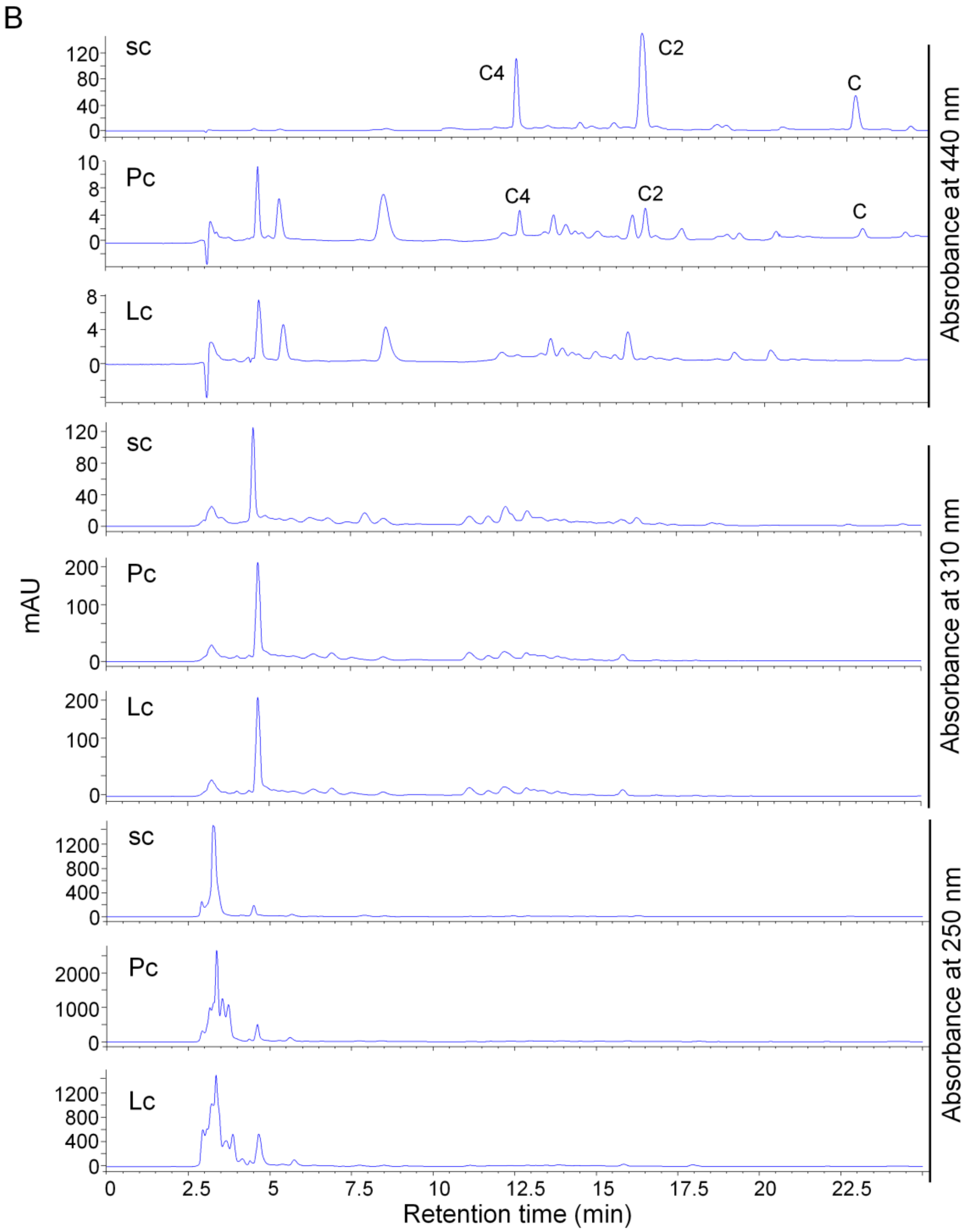


Figure 3B. HPLC profiles of saffron apocarotenoids in different tissues and in different developmental stages. HPLC elution profiles measured at 424, 310 and 250 nm for the detection of crocins, crocetin, picrocrocin and HTCC. Sc, corm from senescent flower shoot; Pc corm from preanthesis flower shoot; Lc, corm from non-flowering shoot.

The extracts of corms from not flowering shoots (Lc), flowers during senescence (Sc) and in preanthesis flowers (Pc), were analysed at three different wavelengths: 250 nm, 310 nm, and 440 nm. The chromatograms recorded at 310 nm did not show appreciable changes among the three corms types (Fig. 3B). At 250 nm the chromatogram of the extracts of Lc corms and the one from Pc corms were also very very similar (Fig. 3B). By contrast, major differences were observed at 440 nm. The peaks detected in the Lc corms (Fig. 3B) were also present in the Pc corms (Fig. 3B), and the additional peaks detected correspond to crocins C4 and C2 and to crocetin that were found at high levels in the Sc corm (Fig. 3B and 4A).

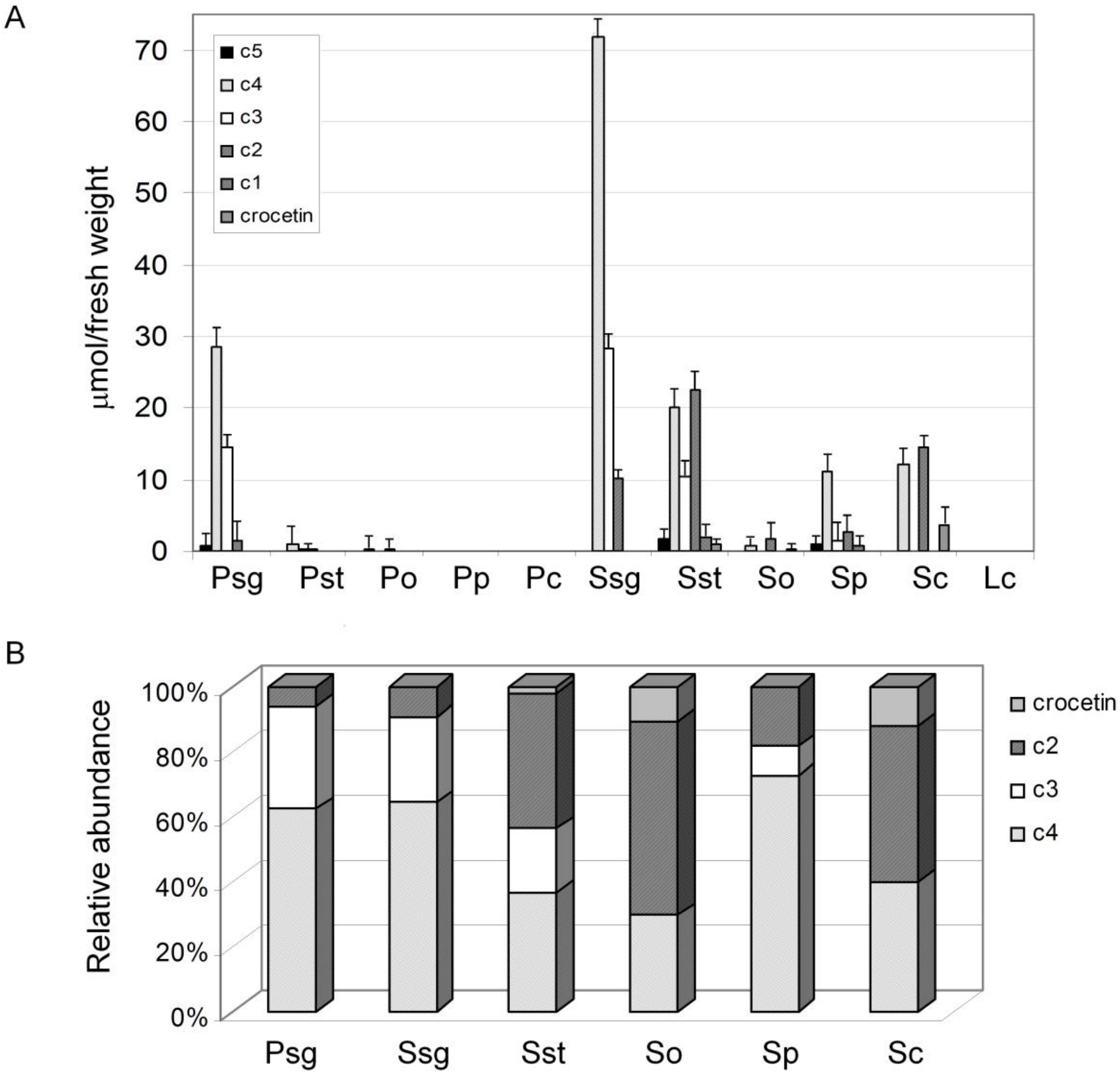


Figure 4. Evolution of saffron apocarotenoids in different tissues during flower senescence.

A. Levels of each apocarotenoid compound, crocin: C5, C4, C3, C2, C1, and crocetin in tissues at different developmental stages. **B.** Relative abundance in different tissues in preantheis and senescence of the main apocarotenoids detected in the stigma.Psg, preanthesis stigma; Pst, preanthesis style; Po, preanthesis ovary; Pp, preanthesis peduncle; Pc, corm from preanthesis flowers shoot; Ssg, senescent stigma; Sst, senescent style; So, senescent ovary; Sp, senescent peduncle; Sc, corm from senescent flowers shoot; Lc, corm from non-flowering shoot.

The overall results showed that picrocrocin and its precursor HTCC were not transported. In addition, the accumulation of crocetin observed in the ovaries and in the developing corm from senescence flowers (Fig. 4B) suggest that crocetin was originated by crocin deglucosylation, which can be also responsible for the higher percentage of C2 crocin detected in the ovaries and in the corm (Fig. 4B) in comparison with the percentage level of this crocin in the preanthesis and senescence stigmas, suggesting a transformation of C3 and C4 crocins in C2 crocin.

Apocarotenoid degradation in the developing corm

In order to test the presence of these transported compounds in the developing corms. Corms developed from non-flowering shoots, corms developed from flowering shoots but with the stigmas removed in preanthesis, and those developed from flowering shoots without stigmas removal, were collected 1, 2 and 4 months after flower senescence, extracted and analysed by HPLC. In all the samples tested it was not possible to detect the presence of crocin or crocetin. Interestingly, we observed that 1 month corms developed from flowering shoots without stigmas were smaller in size than the corms developed from flowering shoots (Fig. 5 A and B), although these differences in size were not significant three months later (Fig. 5 C and D). The corms developed from non-flowering shoots were always smaller in size at the beginning and at the end of the growing period (data not shown).

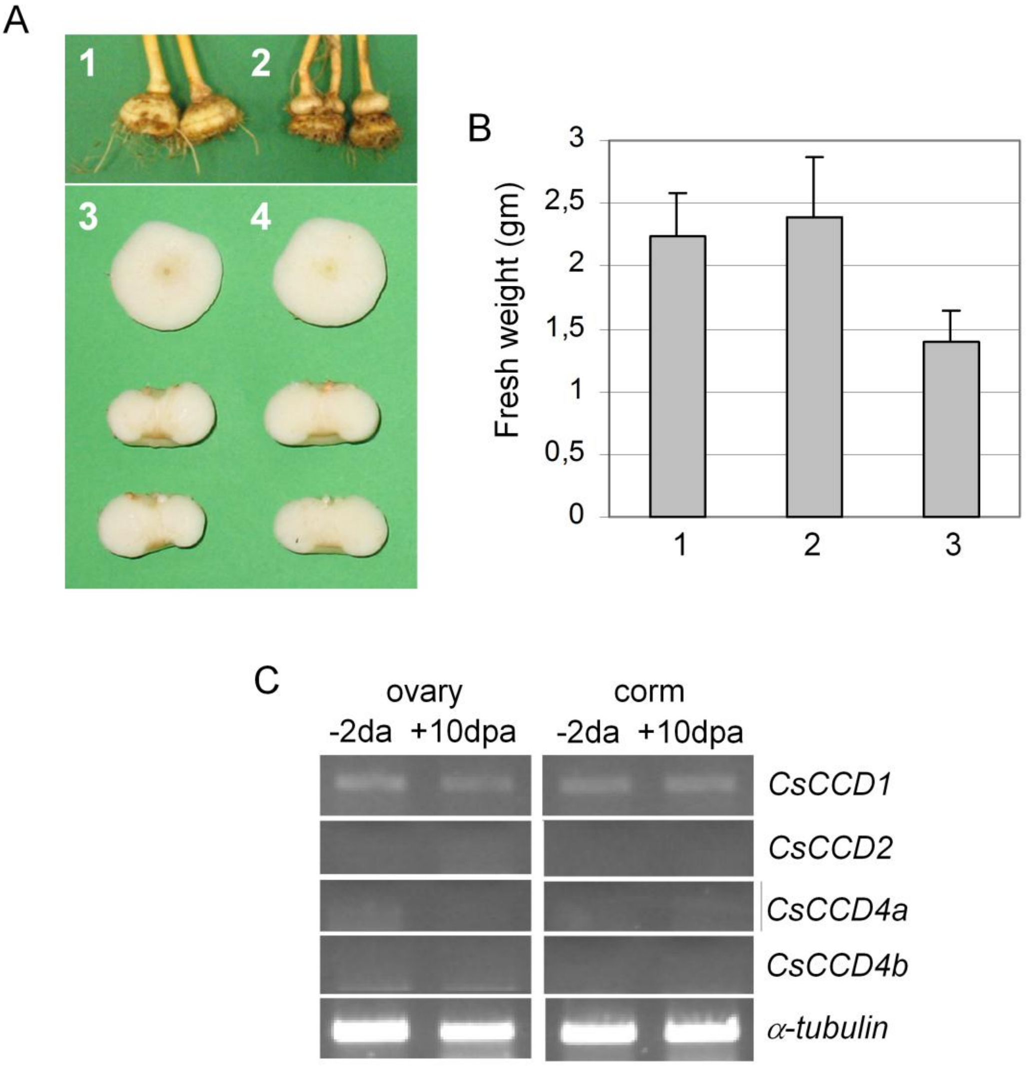


Figure 5. Apocarotenoid degradation during ovary senescence and corm development. A. Developing corms of *C. sativus*. 1, Corm from flowering shoot, with stigmas removed at preanthesis, growing on the mother corm, one month after flower senescence; 2, corm from flowering shoots growing on the mother corm, one month after flower senescence; 3, transversal and longitudinal sections of corms from flowering shoots, with stigmas removed at preanthesis, growing on the mother corm, two months later; 4, transversal and longitudinal sections of corms from flowering shoots, growing on the mother corm, three months later. **B**. Fresh weight of corms three months later. Each bar represent the mean + SE. 1, corms from flowering shoots with stigmas removed at preanthesis (*n* = 21); 2, corms from flowering shoots (*n* = 38); 3, corms from non-flowering shoots (*n* = 27). **C**. Expression analysis of *CsCCD*1, *CsCCD*2, *CsCCD*4*a* and *CsCCD*4*b* by RT-PCR in ovaries and corms two days before anthesis (nce); 2, corm from flowering shoots growing on the mother corm, one month after flowertrol. The PCR products were separated by 1% (w/v) agarose gel electrophoresis and visualized by ethidium bromide staining..

Discussion

C. sativus is a geophyte with interesting ecophysiological properties with respect to redistribution of macroelements between above and underground parts at the beginning and the end of their growing season [19, 35]. This redistribution is highly important for the economical use of nutrients and has evolved to survive adverse environmental conditions. In this study, we showed that the remobilization of substances from unfertilised flower parts could be considered as part of this life strategy in *C. sativus*.

Senescence metabolism has been widely studied in leaves of several plant species [36] and in the case of flowers, petal senescence has concentrated the attention due to its impact on the cut flower market [37, 38]. A key feature of senescence in leaves and flowers is remobilization of resources such as starch and proteins, which are broken down into smaller, soluble ones, such as sugars and amino acids, and then exported to storage cells [39, 40] or pigments transport [26], and this is reflected in some of the major classes of genes whose expression is up-regulated in both organs during senescence [40, 41]. To cope with the transport of dissolved substances to the sinks in mother organs functional vascular bundles are necessary, in fact, the functioning of vascular bundles especially their phloem parts has been characterized as long persisting during senescence [26, 39, 42, 43]. During the senescence of *C. sativus* flowers, it is possible to follow the transport of remobilized compounds through the increasing colour of the style, ovary and peduncle, thus crocins may serve as naturals indicators of substance remobilisation and movement connected with flower senescence. Crocins C4, C3 and C2 were transported out from the senescent stigma, which in the latest stage of senescence presented low levels of these compounds, while an important accumulation of C4 and C2 was found in the developing corm, where can be used to support its intense growth [19]. The relatively higher levels of C2, and the absence of C3 suggest that C4 and C3 crocins are substrates of glucosidase activities in the corm. By contrast, picrocrocin and its precursor HTCC were not transported during flower senescence and the absence of picrocrocin and the reduced concentration of HTCC in the more senescent stigma suggest that most probably both are degraded during the senescent process and released as volatiles, which happens during the drying process of stigmas collected for saffron production [44, 45]. Furthermore, release of volatiles has been reported to occur late during leaf senescence, after cessation of phloem transport [46, 47], and the chemical composition of the fragrance signature in senesced flowers proved to be very different from that for non-senescent flowers [48, 49].

We observed a similar pattern of relative apocarotenoid composition and accumulation in the ovary and in the corm, although it was in this latter organ where a significant accumulation was observed. Nutrients in pollinated flowers are likely to be remobilized to the developing ovary, while in unpollinated flowers they are being transported to sink tissues outside of the flower [50]. Although pollinated, the flower of *C. sativus* never became fertilized, thus if initially after pollination the induction and coordination of ovary and ovule development in preparation for fertilization and the remobilization of nutrients out of the perianth to these developing structures start, after the lack of fertilization this process is most probably interrupted and the compounds are redirected to other sinks organs. Therefore, the low levels of apocarotenoids detected in the ovary could be the result of this process, with the developing corm acting as the sink organ for these compounds in absence of fertilization.

During the growth of the corm subtle changes, such as modifications of the ratio of hexoses to sucrose, influence its growth [51]. High hexose to sucrose ratios are reported during the early stages of corm development and would stimulate cell initiation and elongation, whereas cell differentiation would be triggered by lower ratios of hexoses to sucrose [52, 53]. Hexoses are known to stimulate cell division and elongation [52]. The glucose molecules present in the transported crocins could contribute to increase the hexose pool in the developing corm; in fact the high levels of crocetin in the corm suggest crocin deglucosylation. These cells in the corm would thus enter into the differentiation phase much earlier in the season. This model is consistent with the fact that final cell size was reached earlier in plants with stigmas than in plants without stigmas, although at the end of the growth period the final sizes of corms with or without stigmas were not significant.

In summary, resorbing and storing of the mobilized compounds permit the plant to shed its flowers while avoiding loss of their nutrients. This, in turn, allows the plant to avoid having to compete with other plants and soil microbes for the resorbed nutrients that would otherwise be cycled back into the soil system through flower litter decomposition. Given the role of the stigmas of *C. sativus* as visual and odour cues of pollination, senescence has likely evolved as a mechanism to conserve resources by focusing the attention of pollinators on those flowers that are receptive to pollination. Furthermore, in contrast to leaves they are a sink rather than a source of photosynthates, a feature of stigma senescence that supports their role in the remobilization of cellular constituents from the senescing organs to other structures that continue to develop.

References

1. Britton G: **Overview on carotenoid biosynthesis**, vol. 3. Basel: Birkenhauser; 1998.
2. Auldridge ME, McCarty DR, Klee HJ: **Plant carotenoid cleavage oxygenases and their apocarotenoid products**. *Curr Opin Plant Biol* 2006, **9**(3):315-321.
3. Camara B, Bouvier F: **Oxidative remodeling of plastid carotenoids**. *Arch Biochem Biophys* 2004, **430**(1):16-21.
4. Agusti J, Zapater, M., Iglesias, DJ., Cercos, M., Tadeo, FR. Talon, M. : **Differential expression of putative 9-cis-epoxycarotenoid dioxygenases and abscisic acid accumulation in water stressed vegetative and reproductive tissues of citrus**. *Plant Sciences* 2007, **172**(1):10.
5. Akiyama K: **Chemical identification and functional analysis of apocarotenoids involved in the development of arbuscular mycorrhizal symbiosis**. *Biosci Biotechnol Biochem* 2007, **71**(6):1405-1414.
6. Bouwmeester HJ, Roux C, Lopez-Raez JA, Becard G: **Rhizosphere communication of plants, parasitic plants and AM fungi**. *Trends Plant Sci* 2007, **12**(5):224-230.
7. Gomez-Roldan V, Roux C, Girard D, Becard G, Puech-Pages V: **Strigolactones: promising plant signals**. *Plant Signal Behav* 2007, **2**(3):163-164.
8. Gomez-Roldan V, Fermas S, Brewer PB, Puech-Pages V, Dun EA, Pillot JP, Letisse F, Matusova R, Danoun S, Portais JC *et al*: **Strigolactone inhibition of shoot branching**. *Nature* 2008, **455**(7210):189-194.
9. Umehara M, Hanada A, Yoshida S, Akiyama K, Arite T, Takeda-Kamiya N, Magome H, Kamiya Y, Shirasu K, Yoneyama K *et al*: **Inhibition of shoot branching by new terpenoid plant hormones**. *Nature* 2008, **455**(7210):195-200.
10. Rouseff RL: **Liquid chromatographic determination of naringin and neohesperidin as a detector of grapefruit juice in orange juice**. *J Assoc Off Anal Chem* 1988, **71**(4):798-802.
11. Pfander H, Wittwer F: **[Carotenoid composition in safran (author's transl)]**. *Helv Chim Acta* 1975, **58**(7):2233-2236.
12. A.M. Buddhika Priyadarshani ERJaHP: **Studies on the carotenoids of jakfruit (Artocarpus heterophyllusLam.) from Matale and Kurunegala Districts**. *J Natn Sci Foundation Sri Lanka* 2007, **35** (4):4.
13. Liao Y-H, Houghton PJ, Hoult JRS: **Novel and Known Constituents from Buddleja Species and Their Activity against Leukocyte Eicosanoid Generation**. *Journal of Natural Products* 1999, **62**(9):1241-1245.
14. Eugster CH HH, Leuenberger HJ **Crocetindyaldehyd und crocetinhaldaldehyb als blutenfarbstoffe von Jacquinia angustifolia.** . *Helv Chim Acta* 1969, **52**:2.
15. Tandon JS KS, Ruedi P, Eugster CH: **Crocetin-dialdehyde from Coleus forskohlii Briq., labiatae.** . *Helv Chim Acta* 1979, **274**:2.
16. Pfister S MP, Steck A, Pfander H. : **Isolation and structure elucidation of carotenoid glycoslyesters on gardenia fruits (Gardenia jasminoides) and saffron (Crocus sativus)**. *Journal of Agriculture Food Chemistry* 1996, **44**:4.
17. **Jütner F HB**: **Evidence of β-carotene 7,8(7,8) oxygenase (β-cyclocitral,crocetindial generating) in *Microcystis***. *Arch Microbiol* 1985., **141**:7.
18. Moraga AR, Rambla JL, Ahrazem O, Granell A, Gomez-Gomez L: **Metabolite and target transcript analyses during Crocus sativus stigma development**. *Phytochemistry* 2009, **70**(8):1009-1016.
19. Milyaeva EL AN: **Cytophysiological changes in the course of development of stem apex of saffron**. *Soviet Plant Physiol* 1978, **25**:7.
20. De Mastro G RC: **Relations between corm size and saffron (Crocus sativus L.) flowering**. *Acta Horticulturae* 1993, **344**:6.
21. Kaushal SKaRGU: **Studies on variation in corm size and its effect on cormel production and flowering in Crocus sativus L. under midd hill conditions of HP** *Research on crops* 2002, **3**(1):3.
22. R.V. Molina MV, Y. Navarro1, J.L. Guardiola, A. García-Luis: **Temperature effects on flower formation in saffron (Crocus sativus L.)** *Scientia Horticulturae* 2004, **103**(3):19.
23. Canini MGCA: **Ultrastructure of chromoplasts and other plastids in Crocus sativus L. (Iridaceae)** *Plant Biosystems* 2004, **138**(1):10.
24. CHICHIRICCO G. GCM: **Crocus Sativus pollen germination and pollentube growth in vitro and after intraspecific and interspecific pollination**.*Canadian Journal of Botany* 1986, **64**(11):4.

25. M. CGaGC: **Crocus sativus pollen tube growth in intra- and interspecific pollinations.***Caryologia* 1984, **37**(1-2):11.
26. Erdelská O OM: **Senescence of unfertilised flowers in Epiphyllum hybrids**. *Biology Plantarum* 2004, **48**:8.
27. Moraga AR, Nohales PF, Perez JA, Gomez-Gomez L: **Glucosylation of the saffron apocarotenoid crocetin by a glucosyltransferase isolated from Crocus sativus stigmas**. *Planta* 2004, **219**(6):955-966.
28. Tarantilis PA TG, Polissiou MG. : **1995. Determination of saffron (Crocus sativus L.) components in crude plant extract using high-performance liquid chromatography-UV–visible photodiode-array detection-mass spectrometry**. *Journal of Chromatography* 1995, **699**:12.
29. Himeno H SK: **Synthesis of crocin, picrocrocin and sdafranal by saffron stigma-like structures proliferate din vitro.** . *Agric Biol Chem* 1987, **51**:5.
30. Laine RA, Esselman, W. J., Sweeley, C. C. : **Gas-liquid chromatography of carbohydrates**. *Methods Enzymol* 1972, **28**:9.
31. Gómez-Miranda B, Rupérez, P. , Leal, J. A. : **Changes in chemical composition during germination of Botrytis cinerea sclerotia.** . *Curr Microbiol* 1981, **6**:4.
32. Castillo R, Fernandez JA, Gomez-Gomez L: **Implications of carotenoid biosynthetic genes in apocarotenoid formation during the stigma development of Crocus sativus and its closer relatives**. *Plant Physiol* 2005, **139**(2):674-689.
33. Rubio-Moraga A, Castillo-Lopez R, Gomez-Gomez L, Ahrazem O: **Saffron is a monomorphic species as revealed by RAPD, ISSR and microsatellite analyses**. *BMC Res Notes* 2009, **2**:189.
34. Caiola MG: **Embryo origin and development in Crocus sativus L. (Iridaceae)**. *Plant Biosystems* 2005, **139**(3):9.
35. Chrungoo NH FS: **Correlative changes in carbohydrate content and starch hydrolyzing enzyme in corm of saffron crocus (Crocus sativus L.) during dormancy and sprouting**. *Biochem Physiol der Pflan* 1985, **180**:7.
36. Lim PO KH, Nam HG. : **Leaf senescence**. *Annu Rev Plant Biol* 2007, **58**:22.
37. Reid M: **Senescence in flowers of geophytes**. *Acta Hort* 1997, **430**:11.
38. Woodson WR JM: **In Search of Eternal Youth: The Delay of Postharvest Senescence in Flowers**. *Acta Hort* 2003, **624,**:10.
39. Bieleski RL: **Onset of Phloem Export from Senescent Petals of Daylily**. *Plant Physiol* 1995, **109**(2):557-565.
40. van Doorn WG, Woltering EJ: **Physiology and molecular biology of petal senescence**. *J Exp Bot* 2008, **59**(3):453-480.
41. Price AM, Aros Orellana DF, Salleh FM, Stevens R, Acock R, Buchanan-Wollaston V, Stead AD, Rogers HJ: **A comparison of leaf and petal senescence in wallflower reveals common and distinct patterns of gene expression and physiology**. *Plant Physiol* 2008, **147**(4):1898-1912.
42. Matile P WF: **Function of lysosomes and lysosomal enzymes in the senescing corolla of the morning glory (Ipomoea purpurea)**. *J Exp Bot* 1971, **22**:13.
43. Bieleski RL RM: **Physiological changes accompanying senescence in the ephemeral daylily flower** *Plant Physiol* 1992, **98**:8.
44. Kanakis CD, Daferera DJ, Tarantilis PA, Polissiou MG: **Qualitative determination of volatile compounds and quantitative evaluation of safranal and 4-hydroxy-2,6,6-trimethyl-1-cyclohexene-1-carboxaldehyde (HTCC) in Greek saffron**. *J Agric Food Chem* 2004, **52**(14):4515-4521.
45. Bolandi M, Ghoddusi, H.B.: **Flavour and colour changes during processing and storage of saffron (Crocus sativus L.)**. *Dev Food Sci* 2006, **43**:4.
46. Keskitalo J, Bergquist G, Gardestrom P, Jansson S: **A cellular timetable of autumn senescence**. *Plant Physiol* 2005, **139**(4):1635-1648.
47. Husted S, Schjoerring JK: **Ammonia Flux between Oilseed Rape Plants and the Atmosphere in Response to Changes in Leaf Temperature, Light Intensity, and Air Humidity (Interactions with Leaf Conductance and Apoplastic NH4+ and H+ Concentrations)**. *Plant Physiol* 1996, **112**(1):67-74.
48. Schade F, Legge RL, Thompson JE: **Fragrance volatiles of developing and senescing carnation flowers**. *Phytochemistry* 2001, **56**(7):703-710.
49. Ackermann IE BD, Fordham WD, Kinder JP, Poots I.: **b-Glucosides of aroma components from petals of Rosa species : assay, occurrence and biosynthetic capabilities. 1**. *Journal of Plant Physiology* 1989, **134**:6.

50. Chapin LaJ, M. : **Nutrient remobilization during pollination-induced corolla senescence in Petunia**. *Acta Hort* 2007, **755**:10.
51. Lundmark M, Hurry V, Lapointe L: **Low temperature maximizes growth of Crocus vernus (L.) Hill via changes in carbon partitioning and corm development**. *J Exp Bot* 2009, **60**(7):2203-2213.
52. Kock K: **Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development.** . *Current Opinion in Plant Biology* 2004, **7**:12.
53. Gibson S: **Control of plant development and gene expression by sugar signalling** *Current Opinion in Plant Biology* 2005, **8**:10.

Chapter 4

“Characterization of a glucosyltransferase enzyme involved in the formation of kaempferol and quercetin sophorosides in *Crocus sativus*”

Abstract

UGT707B1 is a new glucosyltransferase isolated from saffron that localizes to the cytoplasm and the nucleus of stigma and tepal cells. UGT707B1 transcripts were detected in the stigma tissue of all the *Crocus* species analysed, but expression analysis of UGT707B1 in tepals revealed its absence in certain species. The analysis of the glucosylated flavonoids present in *Crocus* tepals reveals the presence of two major flavonoid compounds in *C. sativus*: kaempferol-3-*O*-β-D-glucopyranosyl-(1-2)-β-D-glucopyranoside and quercetin-3-*O*-β-D-glucopyranosyl-(1-2)-β-D-glucopyranoside, both of which were absent from the tepals of those *Crocus* that did not express UGT707B1. Transgenic *Arabidopsis* plants constitutively expressing UGT707B1 under the control of the CaMV35S promoter have been constructed and their phenotype analysed. The transgenic lines displayed a number of changes that resembled those described previously in lines where flavonoid levels had been altered. The plants showed hyponastic leaves, a reduced number of trichomes, thicker stems and flowering delay. Levels of flavonoids measured in extracts of the transgenic plants showed changes in the composition of flavonols when compared to wild-type plants. The major differences were observed in the extracts from stems and flowers, with an increase in 3-sophoroside flavonol glucosides. Furthermore, a new compound not detected in Col-0 wild-type plants was detected in all the tissues and identified as kaempferol-3-*O*-sophoroside-7-*O*-rhamnoside. The data reveal the involvement of UGT707B1 in the biosynthesis of flavonol-3-*O*-sophorosides, and how significant changes in flavonoid homeostasis can be caused by overproduction of a flavonoid-conjugating enzyme.

Keywords: saffron, stigmas, tepals, flavonoids, glucosyltransferases, expression, sophoroside

Introduction

Flavonoids encompass a very large family of phenolic compounds, with >9000 of these known [1-3], including anthocyanin, proanthocyanidin and phlobaphene pigments, as well as the flavonol, flavone and isoflavone subfamilies that possess specific biological functions in the species in which they accumulate [4-7]. In plants, flavonoids have many diverse functions including defence, UV protection, auxin transport inhibition, allelopathy, plant-microorganism communication, regulation of reactive oxygen species and flower colouring to attract pollinators, while in many species they are required for pollen viability [2]. A number of biological processes, such as transcriptional regulation, signal transduction, and cell-to-cell communication are also influenced by flavonoids. In addition, flavonols play a crucial role in human nutrition and have pharmaceutical potential, as the intake of flavonols from vegetables, fruit, berries and beverages has been favour ably linked with reduced risks of a number of diseases in large population studies [1, 8-10].

Flavonoids are synthesized in plants via the flavonoid branch of the phenyl propanoid and acetate-malonate metabolic pathway. Their biosynthesis has been investigated extensively with nearly all the enzymes isolated and functionally characterized [3, 11]. However, the pathways for sequential modification, such as glycosylation, acylation, and methylation, are still relatively unexplored even though modification produces a huge chemical diversity and is essential for the stable accumulation of flavonoids. Glycosylation renders them more water-soluble and less toxic, and it may enable flavonol transport and compartmentation[12]. Glycosylation is achieved by UDP-carbohydrate dependent glycosyltransferases (UGTs) that catalyze the transfer of a carbohydrate from an activated donor sugar onto small molecule acceptors by the formation of a glycosidic bond [13]. These enzymes exhibit a rather broad acceptor tolerance, but have been shown to display strict regioselectivity in many cases [14]. Moreover, the differentiation of the sugar donor specificity of UGTs appears to have occurred in a lineage-specific manner after establishment of general regiospecificity for the glycosyl acceptor [15-16]. UGTs also catalyse the attachment of additional sugars to flavonoid glycosides. With respect to flavonoid glycosides, cDNAs encoding anthocyanidin 3-O-glucoside: 2''-glucosyltransferase from *Ipomoea nil* [17] and *Veronica persica* [18], anthocyanidine 3-O-glucoside: 2''-glucuronyltransferase from *Berris perennis*[19], flavanone 7-O-glucoside: 2''-rhamnosyltransferase (yielding

7-O-neohesperidoside) from *Citrus maxima* [20], flavonoid 3-O-glucoside: 6"-glucosyltransferase from *Catharanthus roseus* [21], cyanidin 3-O-galactoside: 2"-xylisyltransferase from *Actinidia chinensis*[22], andanthocyanin 3-O-glucoside: 2"-O-xylosyltransferase [16] have been thus far cloned and functionally characterized. Although various flavonol 1,2"-sophorosides have been isolated from various sources of higher plants [23] none have been reported for flavonol glucoside 1,2"-glucosyltransferases.

Saffron (*Crocus sativus*) is one of the most valuable crops worldwide due to its stigmas, widely used as spice, medicinal drugs, and food additives. These stigmas are characterized by the presence of apocarotenoids and flavonoids. Their antioxidant properties, along with their bitter taste, could qualify them as potential organoleptic agents of the spice [24-25]. In addition, they show anticonceptive and anti-inflammatory effects [26]. In *Crocus* species, flavonols are usually glycosylated at their 3-OH, 7-OH and 4′-OH positions [27-28], generating a quite complex pattern of flavonols. Three main glucosides of kaempferol alone have been identified in *C. sativus* stigmas: kaempferol 7-O-sophoroside, kaempferol 3-O-sophoroside-7-O-glucopyranoside and kaempferol 3,7,4’-triglucoside [25, 28] and 21 different glycosides of kaempferol, quercetin, myricetin, naringenin, taxifolin, tamarixetin and isorhamnetin have been identified in tepals [29]. With regard to flavonoid aglycones in tepals, kaempferol dominates as kaempferol glycosides constitute between 70 and 90% of the total contents of flavonoids; quercetin glycosides vary from 5 to 10% and glycosides of dihydrokaempferol, isorhamnetin, naringenin, taxifolin, tamarixetin, and myricetin are only minor components [27].

In order to broaden our understanding on the enzymes responsible for flavonol glucosylation in saffron, we focused attention on the correlation of the accumulation of specific flavonol glycoside end-products with the expression of UGT707B1 in the flower tissues from different *Crocus* species. In addition, to determine the activity of UGT707B1 in plants, we developed transgenic *Arabidopsis* lines over-expressing UGT707B1 and then we carried out liquid chromatography-mass spectrometry (LC-MS)-based metabolite profiling to allow for a more comprehensive identification of flavonol metabolites in the transgenic lines. Finally, integration of comprehensive flavonol identification/annotation from the transgenic *Arabidopsis* lines and *Crocus* species with UGT707B1 expression analysis identified UGT707B1 as a gene encoding a flavonol 2"-O-glucosyltransferase, responsible for the formation of flavonoid sophorosides, which interfere with specific developmental plant processes.

Materials and methods

Chemicals and Plant materials

Chemicals and reagents were obtained from Sigma-Aldrich unless otherwise stated. Plant tissues and stigmas from *C. sativus* grown under field conditions in Tarazona de La Mancha, Spain, were used throughout the experiments. *C. goulimyi*, *C. medious*, *C. cartwrightianus* “albus”, *C. speciosus*, *C. niveus*, *C. pulchellus*, *C. chrysanthus*, *C. ochroleucus* and *C. cartwrightianus* were obtained from Pottertons Nursery (Nettleton, UK). Stigmas were collected at the developmental stages previously described [30], and defined as follows: yellow stigma, closed bud inside the perianth tubes (around 0.3 cm in length); orange stigma, closed bud inside the perianth tubes (around 0.4 cm in length); red stigma, closed bud inside the perianth tubes (0.8 cm in length); -2da, two days before anthesis, dark red stigma in closed bud outside the perianth tubes (3 cm in length); da, day of anthesis, dark red stigma (3 cm in length); +1da, one day after anthesis, dark red stigma and +3da, three days after anthesis, dark red stigma. Tepals, style and stamens were collected from flowers at the time of anthesis and together with corms were frozen in liquid nitrogen and stored at -80°C until required.

Seeds of *Arabidopsis* wild type ecotype Columbia (Col-0), and transgenic lines were sown in pots containing vermiculite and watered with nutrient solution under a controlled environment with 16 h light/8 h dark cycles at 22 °C/18 °C.

Cloning of *C. sativus* UGT707B1 cDNA

As a first step in identifying glycosyltransferase genes expressed in saffron stigmas, total RNA and mRNA were isolated from developed saffron stigmas by using AmbionPolyAtrack and following manufacturer’s protocols (Ambion Inc., Austin, TX, USA). First-strand cDNAs were synthesized by reverse transcription (RT) from 2 µg of total RNA using an 18-base pair oligodT primer and a first-strand cDNA synthesis kit (Amersham Biosciences) according to manufacturer’s instructions. These cDNAs were used as templates for PCR using degenerate primers designed based on the conserved regions of [31]. The primers used were: glut-f (5’-TSNGTNGCNTAYGTNTSNTTYGG-3’) and glut-r (5’-TTCCANCCRCARTGNGTNACRAA-3’). Anchored PCR with gene-specific primers were used to analyse and identify the 5’ and 3´ ends of the glucosyltransferase. For these, 1µg of poly(A)+ RNA from stigmas was used to synthesize the 5’ and 3’ ends of the first-strand cDNA using Superscript II reverse transcriptase, using the primers 5’-CDS primer and SMARTII-A oligo for the 5’-RACE reaction and the 3’-CDS primer A for the 3’-RACE reaction supplied in the SMARTTM RACE cDNA Amplification kit (Clontech-Takara). Following dilution, the first-strand reaction product was subjected to PCR for amplification. We used the gene-specific primers UGT707B1-f1 (5’-aagacagagggatggtgtgg-3´), and UGT707B1-f2 (5’-TGGAACAGCATCTGAACGAG-3’) as forward primers, and UGT707B1-r1 (5’-CTCGTTCAGATGCTGTTCCA-3’) and UGT707B1-r2 (5’-tggattcgttctggagagct-3’) as reverse primers in combination with the universal primer mix from the SMARTTM RACE cDNA Amplification Kit (Clontech-Takara) as the reverse/forward primer with the following cycling program: one cycle at 94°C for 3 min, 10 cycles at 94°C for 20 s, 66°C-0.2°C/cycle for 20 s, and 72°C for 2 min, 30 cycles at 94°C for 20 s, 64°C for 20 s and 72°C for 2 min, and a final extension at 72°C for 5 min. The amplified PCR products were analysed by electrophoresis in 1% agarose gel. The PCR products were then cloned into pGEM-T (Promega Corporation, Madison, WI, USA). The ligated DNA was transformed into *E. coli* strain JM109. The clones (20 colonies) were picked individually and amplified in 3 ml of LB medium at 37°C overnight. The plasmid DNA from each clone was extracted using a DNA plasmid Miniprep kit (Promega Corporation, Madison, WI, USA) and then analysed by EcoRI restriction digestion. Five clones of each size were sequenced with Sp6 and T7 primers using an automated DNA sequencer (ABI PRISM 3730xl, Perkin Elmer) from Macrogen Inc. (Seoul, Korea). Computer-aided sequence similarity searches were made with the BLAST suite of programs at the National Centre for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov) Motif searches were done using PROSITE (http://expasy.hcuge.ch/sprot/prosite.html), TMPRED (http://www.isrec.isb-sib.ch/sofware/sofware.html), SignalP (http://www.cbs.dtu.dk/services/SignalP) and PSORT II (http://psort.nibb.ac.jp). Once the 5’ and 3’ sequences were determined, the full-length clone UGT707B1 was amplified from the cDNA and genomic DNA with the following primer sequences: the forward primer, 5’-ACGATGCAAATTTCTCTTGTCT-3’; and the reverse primer 5’-TCAGGAGAAACTTATACTGCCA-3’, in order to determine the presence of intron sequences.

Phylogenetic analysis

To construct the phylogenetic tree, the amino acid sequences were aligned using the BLOSUM62 matrix with the ClustalW (http://www.clustal.org) algorithm-based AlignX module from MEGA Version 4.0 (http://www.megasoftware.net/mega.html). The alignments were saved and executed by MEGA Version 4.0 to generate a Neighbour Joining Tree with bootstrapping (5000 replicates) analysis and handling gaps with pairwise deletion.

Isolation of promoter sequences

Genomic DNA was prepared from *C. sativus* leaves with a CTAB method. The *C. sativus* UGT707B1 upstream flanking sequences were isolated with the GenomeWalker Universal Kit (BD Biosciences, Palo Alto, CA, USA) using two non-overlapping gene specific primers: P-r1: 5’-AGATGAGTGACGTGGCAAAGAAGT-3’ and P-r2: 5’-AGGAGGAGAGGGATTCCATGGA-3’, based on the cDNA sequences obtained. All PCR reactions were performed using Advantage 2 Polymerase mix (BD Biosciences, Palo Alto, CA, USA). Two rounds of PCR amplification using the P-r1 primer with an adapter-specific primer and the nested PCR amplification with P-r2 plus another adapter-specific primer gave fragments around 1,400 base pairs. All PCR products were ligated to pGEM-T with the TA Cloning Kit (Promega, Madison, USA). The ligated DNA was transformed into *E. coli* strain JM109. The clones (20 colonies) were picked individually and amplified in 3 ml of LB medium at 37 °C overnight. The plasmid DNA from each clone was extracted using a DNA Plasmid Miniprep Kit (Promega, Madison, WI, USA). Plasmids were sequenced using an automated DNA sequencer (ABI PRISM 3730xl, Perkin Elmer) from Macrogen Inc. (Seoul, Korea). Promoters were analysed with the PLACE [32] and PlantCARE[33] databases.

Expression analysis

For RT-PCR, total RNA was isolated from *C. sativus* anthers, leaves, tepals, style, corms, calli and stigmas (anthesis) and from the tepals and stigmas (anthesis) of the other seven *Crocus* species and from wild-type and transgenic *A. thaliana* plants. This was done by grinding fresh tissue in liquid nitrogen to a fine powder and extracting in 1 ml of Trizol reagent (Gibco-BRL) per 100 mg of tissue fresh weight, according to the protocol of the manufacturer. The RNA was resuspended in 100 µl of RNase-free water and treated with RQ1 RNase-free DNase (Promega, Madison, WI). The DNase was heat inactivated before RT-PCR. The RNA was quantified with a spectrophotometer at OD of 260 and 280 nm and stored at -80 °C. Various initial concentrations of treated RNA, ranging over 10 fold difference, were used to demonstrate the differential accumulation of the RNA in the tissues analysed in the RT-PCR experiments. Total RNA samples were reverse transcribed with a first-strand cDNA synthesis kit (Amersham Biosciences) and random primers (Promega, Madison, WI, USA). The gene expression levels of *Crocus* samples were evaluated by PCR using 10 µM of each primer (described in the above section), 200 µM dNTPs, and 2 units of Taq polymerase (Invitrogen, Carlsbad, CA). After an initial denaturation step for 2 min, the PCR reactions were performed for 30 cycles at 94 °C for 20 s, 55 °C for 20 s, and 72 °C for 2 min. As an internal control, the mRNA level of the constitutively expressed ribosomal protein 18 was used [34]. The PhotoCaptMw program was used to quantify the intensity of the ethidium bromide stained DNA bands from the positive images of the gel.

For the analysis of expression in the *A. thaliana* samples, the primers used for UGT707B1 amplification were the same as those used above, and the *A. thaliana* actin gene was used as control as previously described [35].

Immunocytochemical analysis

Polyclonal antibodies against UGT707B1 were produced in rabbits (GenScript, Piscataway, USA), for that UGT707B1 was subcloned into a bacteria expression vector with a Trx-His tag and the purified protein used for rabbit immunization. Tissue for immunocytochemical localization was fixed in 4% p-formaldehyde in 100 mM Na-phosphate buffer at pH 7.2. After dehydration in a graded ethanol series the tissue was infiltrated and embedded in paraffin (Paraplast plus; Sigma-Aldrich). Sections 10 µm thick were obtained and mounted onto polysine slides (Sigma-Aldrich). The sections were blocked with 5% BSA (w/v) in TBS buffer [20 mM Tris-HCl (pH 8.2), 0.9% NaCl] for 30 min, incubated with the UGT707B1 antibody (1:1.000) in TBS buffer (pH 8.2), 0.1% BSA, 0.05% Tween 20 for 2 h, rinsed in the same solution, and incubated with rabbit anti-mouse IgG labelled with alkaline phosphatase for 1-2 h. Alkaline phosphatase was detected by the BCIP and NBT procedure. A purple precipitate forms at the site of localization. Controls included the use of preimmune serum and the elimination of the antibody incubation step. A bright-field microscope was used for sample visualization and photography (E600; Nikon).

Vector construction and *Arabidopsis* transformation

To produce transgenic plants in which the UGT707B1 protein was expressed under the control of the 35S promoter, the UGT707B1 ORF was cloned into the Gateway Entry vector pENTR/D TOPO (Invitrogen, Carlsbad, CA) and confirmed by sequencing. For stable transformation by *Agrobacterium tumefaciens,* the UGT707B1 clone was introduced into the Gateway plant transformation destination pB2GW7 vector [36] using Gateway LR Clonase enzyme mix with pENTR-UGT707B1 according to the manufacturer's instructions (Invitrogen, Carlsbad, CA,). The reading frame of the resulting vector, pB2GW7-UGT707B1, was confirmed by sequencing, and the vector was transformed into *A. tumefaciens* strain 53 by electroporation. A single colony containing the target construct was confirmed by PCR and used for genetic transformation of *Arabidopsis* using floral dip techniques [37]. Transgenic seedlings were selected for their resistance to the basta herbicide (CIBA). *A. thaliana* plants were grown in growth chambers under LD (16 h of light/8 h of dark).

HPLC analysis of flavonoids and anthocyanins extracted from *Crocus* tepals

For the analysis of flavonoids and anthocyanins in tepals from different *Crocus* species, 30 mg of tissue were extracted with MeOH: HCO2H: H2O (80:1:20) at 5 °C for 24 hr, and subsequently macerated with a pestle, followed by vigorous vortexing and centrifugation at 10,000g for 10 min. After centrifugation the supernatant was collected, filtered and used for HPLC analysis as previously described [28].

NMR Structural Characterization

Structure elucidation was accomplished by nuclear magnetic resonance spectroscopy (NMR). NMR spectra were recorded at 293, 303 and 313 K, using D_2O as the solvent, on a Varian SYSTEM 500 NMR spectrometer (^1H 500 MHz, ^{13}C 125 MHz) equipped with a 5 mm HCN cold probe. Chemical shifts of ^1H (δH) and ^{13}C (δC) in ppm were determined relative to an external standard of sodium [2, 2, 3, 3-$^2H_4$]-3-(trimethylsilyl)-propanoate in D_2O (δH 0.00 ppm) and 1, 4-dioxane (δC 67.40 ppm) in D_2O, respectively. One-dimensional NMR experiments (^1H, and ^{13}C) were performed using standard Varian pulse sequences. Two-dimensional [^1H, ^1H] NMR experiments (gCOSY and TOCSY) were carried out with the following parameters: a delay time of 1 s, a spectral width of 3906.2 Hz in both dimensions, 4096 complex points in t2 and 4 transients for each of 200 time increments, and linear prediction to 512. The data were zero-filled to 4096 × 4096 real points. Two-dimensional [^1H-^{13}C] NMR experiments (gHSQC and gHMBC) used the same ^{1}H spectral window, a ^{13}C spectral windows of 25133 Hz, 1 s of relaxation delay, 1024 data points, and 128 time increments, with a linear prediction to 256. The data were zero-filled to 4096 × 4096 real points. Typical numbers of transients per increment were 4 and 16, respectively.

HPLC analysis of metabolites extracted from transgenic *Arabidopsis* lines

For the analysis of the glucosides of interest in the transgenic plants, rosette leaves, cauline leaves, stems and flowers from six-week plants were cut, frozen in liquid nitrogen, and lyophilized to determine the dry weight. Flavonol analyses were performed in triplicate as previously described [38]. The dried material was incubated overnight at 4 °C in 500 µL of 80% methanol containing scopoletin 3mM as internal standard, and subsequently macerated with a pestle, followed by vigorous vortexing and centrifugation at 10,000g during 10 min. After centrifugation the supernatant was collected, filtered and used for different HPLC analysis.

The flavonoid profile analyses were performed by HPLC on a Konik HPLC system (Barcelona, Spain) equipped with a SugerlaborInertsil ODS-2.5-µm C18 column (250 × 4.6 mm) and connected on line to a photodiode array detector, with a dynamic range from ultraviolet to visible region (200-700nm). The column was developed with solvent system B: 90% acetonitrile and 0.05% TFA and A: 10% acetonitrile and 0.05%TFA, under the following conditions: 0-5 min isocratic at 5% B, 5-40 min linear gradient from 5 to 25% B, 40-48 min linear gradient up to 100% B, 48-60 min isocratic at 100% of B, 66-72 min isocratic at 100% methanol and running of 10 min at initial conditions. Using a flow rate of 0.2 mL/min. Compounds were first characterized by their elution time and their UV absorption spectrum recorded with a photodiode array detector (Konik, Spain). The quantitative evaluation was performed on 340 nm absorbance profiles for flavonol glycosides. Flavonols were quantified by measuring the peak area using as reference the peak area of the internal standard.

High pressure liquid chromatography-mass spectrometry (LC/ESI/MS/MS) analyses

The HPLC-ESI-MS/MS system adopted in this work consisted of a Surveyor HPLC system (Thermo Mod. Finnigan) plus detector and linear ion trap mass spectrometer detector LXQ (Thermo Mod. Finnigan™ LXQ™). This mass spectrometer was equipped with an electrospray ionization (ESI) source. IT source parameters were as follows: positive mode, ESI source voltage of 4.5 kV, nebulization with nitrogen at 30 psi, dry gas flow of 6.0 l/min at temperature of 300 °C and scan range, 150-1000 m/z. Metabolites were identified by their retention times, mass spectra, and product ion spectra in comparison with the data determined for authentic reference materials.

Heterologous expression in *E. coli*

The full-length open reading frame of UGT707B1 cDNA was amplified by PCR using Pfu polymerase (Promega, Madison, USA). The oligonucleotide sequences for UGT707B1 cloning were as follows: the forward primer 5’-GACGGATCCCAGATGCAAATTTCTCTTGTC-3’, and the reverse primer 5’-ATGCGGCCGCTTTAGCAGCACGAATCCTTAG-3’. Using these primers, the generated product has a site BamHI at the 5’ end and a NotI site at the 3’ end (underlined in the corresponding primers). The PCR product was cloned directionally (SmaI-NotI) into bacterial GST expression vector pGEX-5X-3 (GE Healthcare, Chalfont St Giles, UK) to create in-frame fusions at the 5’ terminus with the GST coding sequence. The construct was sequenced to confirm that the gene was in the correct reading frame. After transformation into *E. coli* Rossetta (LysS) cells, colonies were selected on LB containing 100 µg/ml ampicillin (AMP) and 60 µg/ml chloramphenicol (CM) plates. Individual colonies were grown overnight in 5 ml of LB-AMP-CM medium at 25 °C, and 2.5 ml of the culture was used to inoculate 500 ml of LB-AMP-CM fresh medium. Cells were grown at 25 °C until an A600 of 0.6 was reached, after which the culture was induced with 0.5 mM IPTG and allowed to grow for 16 h at 25 °C. The cells were harvested by centrifugation at 5,000g for 10 min and resuspended in 20 ml PBS. Resuspended cells were sonicated with a microtip probe in ice until the viscosity disappeared. After sonication, the samples were centrifuged at 10,000g for 25 min. The supernatant and pellet were tested by PAGE (polyacrylamide gel electrophoresis)/SDS for solubility of the fusion protein by coomasie stain. The soluble proteins were applied to a glutathione sepharose column for purification following manufacturer’s instructions (GE Healthcare, Chalfont St Giles, UK). Protein concentration was determined by Bradford using serum albumin as standard.

Enzyme assays and analysis of reaction products

The affinity-purified enzyme was used to determine substrate specificity and enzymatic parameters. In a final assay volume of 100 µl, various buffers and conditions were tested: buffer i) 50 mM Tris-HCl pH 8.0; ii) 50 mM K2HPO4 pH 7.5; iii) 50 mM Citrate Buffer pH 5.5; all containing 14 mM 2-mercaptoethanol, 2.5 mM UDP-glucose/UDP-galactose, the recombinant enzyme (7.0 μg) and the corresponding substrates: 100 μM quercetin, 100 μM kaempferol, 100 μM kaempferol 3-O-glucoside, 100 μM dihidrokaempferol, 100 μM rhamnetin, 100 μM isorhamnetin, 100 μM lutein, 100 μM apigenin, 100 μM myricetin, 100 μM cyanidin, 100 μM petunidin, 100 μM delphinidin, 100 μM caffeic acid, 100 μM coumaric acid, 100 μM genistein , 100 μM esculetin, 100 μM daidzein, 1 mM trans-cinnamic acid, 1 mM sinapic acid, 1 mM indole acetic acid, and 1 mM abscisic acid. The glucosyltranferase activity assays were carried out at 30 °C for 30-90 min. The reactions were terminated and the proteins precipitated by the addition of 20 μl of trichloroacetic acid (240 mg/ml). Subsequently, samples were centrifuged at 15,000g for 5 min to collect the supernatant, and aliquots were analysed by reverse-phase HPLC as previously described [28] using a C18 Ascentis, 25 x 4.6, particlesize 5 um column (Supelco, Sigma-Aldrich).

Whole-cell biotransformation and flavonoid glucoside purification and identification

Whole-cell biotransformations were performed as previously described [39] with minor modifications. In brief, *E. coli* Rossetta (LysS) cells harbouring the UGT707B1 in a pGEX-5X-3 vector were grown at 28 °C, 150 rpm, in 200 ml M9 minimal medium containing 0.4% glycerol and 100 µg/ml AMP and 60 µg/ml CM to A_{600} 0.75 and then induced with 0.1 mM IPTG and incubated for 16 h. Following the induction period, each substrate, kaempferol, quercetin, apigenin, naringenin, caffeic acid and luteolin was added to the bacterial cultures with a final concentration of 50 µM. Additional glycerol was added to the cultures, with a final concentration of 0.15% (v/v), at 4, 7, 25, and 31 h. *E. coli* cultures harbouring the empty vector were used as control. The *E. coli* cultures were collected and centrifuged 36 h after addition of substrates. The supernatant was extracted twice with ethyl acetate, and collected ethyl acetate phase was evaporated in a rotavapour. Residues were dissolved in 500 µl methanol and subjected to HPLC analysis.

Data collection and statistical analysis

Data presented are the means ±SE of at least three independent experiments, with at least 10 plant samples per transgenic line or wild-type. The experimental data were analysed by one-way ANOVA (Tukey HSD Test); the comparisons between the mean values of each transgenic line and wild-type were evaluated by the least-significant difference test at $P < 0.01$.

Results

Cloning and analysis of UGT707B1

To identify glycosyltransferases from *C. sativus* stigmas, a homology-based strategy was used, taking advantage of specific glycosyltranferase motifs located in the C-terminus region [40]. A cDNA population was prepared by reverse transcription of poly (A)+ from total RNA isolated from *C. sativus* stigmas. DNA fragments were amplified by degenerate primers and the products obtained were cloned and analysed. Sequencing of one PCR product revealed homology to glycosyltransferases. The sequence information from this clone allowed the design of PCR specific primers to obtain the full-length transcripts. We performed 5' and 3' RACE using poly(A)+ from *C. sativus* stigmas as a template. The gene obtained (1576 bp, GenBank accession number HE793682) was intronless, containing a putative open reading frame of 1413 bp encoding 471 amino acid residues with a calculated molecular mass of 52 kDa. The amino acid sequence was named according to the UGT Nomenclature Committee as UGT707B1 [41].

Because *C. sativus* is a triploid, we employed *in silico* screening of a large stigma cDNA EST database (http://www.saffrongenes.org/) as an effective method for identification of potential UGT707B1 alleles. However, no EST clones were identified.

The carboxyl terminal of the protein contained the plant secondary product glycosyltransferase (PSPG) box signature motif. Analysis of UGT707B1 sequence for N-terminal targeting signal or C- terminal membrane anchor signal using SignalP and TargetP web-based programmes predicted a signal peptide localized between amino acids 31 and 32. Prediction of sub-cellular location with PSORT yielded a chloroplast location. By contrast, the TargetP v1.1 programme predicts the presence of a signal peptide, although not a chloroplast transit peptide, and was predicted to be in the secretion pathway. Furthermore, due to the presence of a signal peptide, UGT707B1 was also analysed for the presence of glycosylated residues using the NetNGlyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/). This program predicted the presence of two putative glycosylation sites in two asparagines: N34 (NISI) and N192 (NYTW). Finally, the location of UGT707B1 in flower tissues was investigated using polyclonal antibodies (Fig. 1). The protein seems to be localized preferentially in the nuclei, but also was present in the cytosol of tepals, stigmas and style cells (Fig. 1).

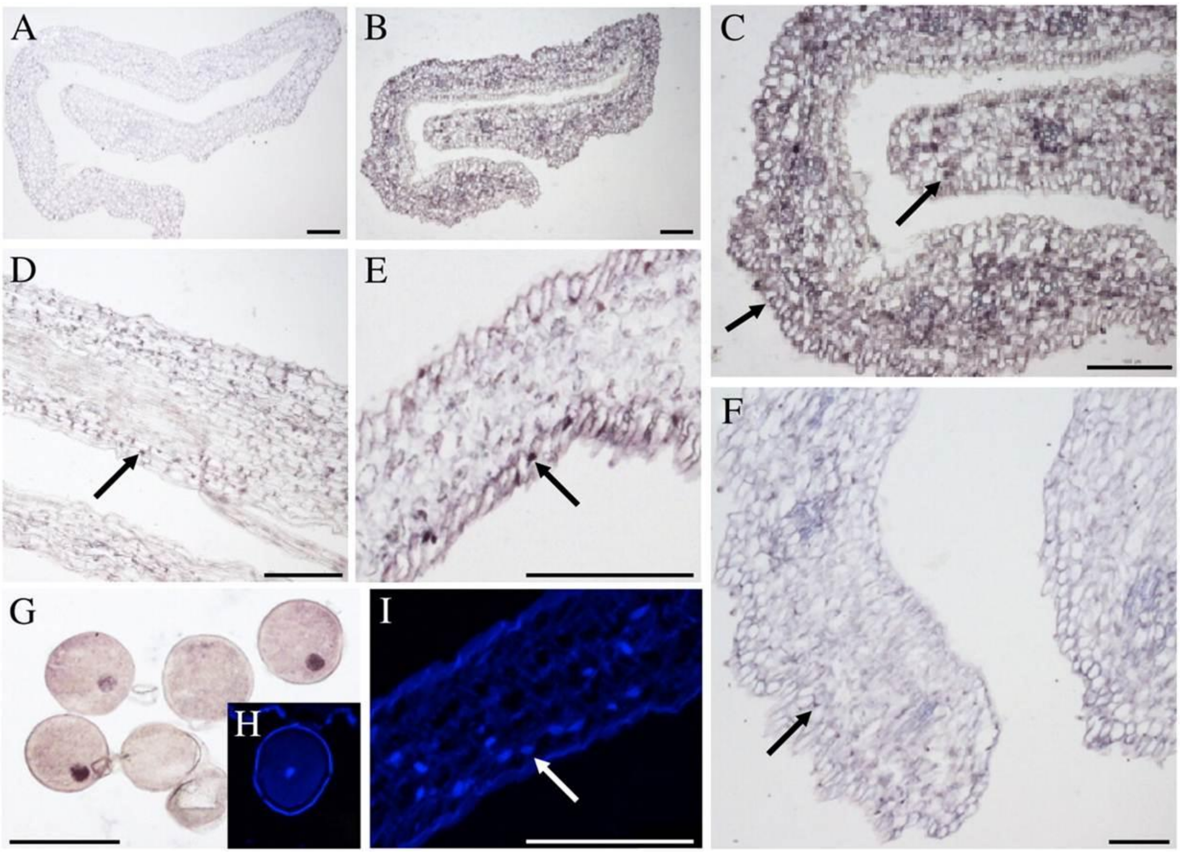


Figure 1. Immunolocalization of UGT707B1 protein in floral organs of *C. sativus* at anthesis. Paraffin-embedded sections were incubated with UGT707B1 antiserum (or pre-immune serum) and anti-rabbit IgGlabelled with alkaline phosphatase. Purple staining indicates the location of UGT707B1 in the cytoplasm and the nucleus (arrows). **A, B** Transverse sections of style incubated with pre-immune serum A and UGT707B1 antiserum B. **C** Close-up of the style section in B. **D-G** Longitudinal section of style D, transverse sections of petal E, stigma F and pollen grains G treated with UGT707B1 antiserum. **H, I** Pollen grains H and petal I sections stained with DAPI (2 µL/mL) to detect nuclei. Bars= 100 µm.

For comparative modelling, UGT707B1 was aligned with MtUGT71G1 and UGT72B1, whose crystal structures have been solved [42-43], using the Pyre2 server (http://www.sbg.bio.ic.ac.uk/phyre2/). UGT707B1 displayed 39% and 38% overall identity with MtUGT71G1 and UGT72B1, respectively (Fig. 2). UGT707B1 had a typical PSPG-box sequence including highly conserved key residues for substrate recognition and the catalysis of UGTs: a histidine residue, His-15 (His-22 in UGT71G1 and His-20 in UGT72B1) that is highly conserved among PSPGs is proposed to act as a key catalytic residue that activates the hydroxy group of the glucosyl acceptor molecule to facilitate glucosydic linkage formation. A well conserved aspartic acid residue Asp-116 (Asp-121 in UGT71G1 and Asp-119 in UGT72B1) is hydrogen-bonded with the His residue and is proposed to assist in its general acid/base role during catalysis. The residues Glu-374, Glu-390 and Gln-391 that were proposed in previous crystallographic and biochemical studies as important for the catalytic activity [44-46] are also conserved in UGT707B1 (Glu-365, Glu-381 and Gln-382).

```
UGT71G1         MSMSDINKNSELIFIPAPGIGHLASALEFAKLL--TNHDKNLYITVFCIKFPGMPFADS- 57
UGT707B1        MQIS-------LVFFPAWSAGHLTSMLEFSKLLLTTNADVNISITFLLIKLPYRTFSSAS 53
UGT72B1         MEES---KTPHVAIIPSPGMGHLIPLVEFAKRL--------VHLHGLTVTFVIAGEGPPS 49
                *. *       : ::*: . *** . :**:* *        : :  : :.:     . .

UGT71G1         --YIKSVLASQPQIQLIDLPEVEPPPQELLKSPEFYILTFLESLIPHVKATIKTILS--N 113
UGT707B1        LASMESLSSSGLQVHFHQLPEVDLP--ENSDGPEDTASTYFQLYTPHVRAFLSSHP---N 108
UGT72B1         KAQRTVLDSLPSSISSVFLPPVDLTDLSSSTRIESRISLTVTRSNPELRKVFDSFVEGGR 109
                      : :   .:    ** *: .  .     *      .    *.::  :.:     .

UGT71G1         KVVGLVLDFFCVSMIDVGNEFGIPSYLFLTSNVGFLSLMLSLK--NRQIEEVFDDSDRDH 171
UGT707B1        PVSAFLIDFFATSLIYVATEFSVPTFVYFTSTALMLGLNLHLPFLEKKIGVEFGQVEG-- 166
UGT72B1         LPTALVVDLFGTDAFDVAVEFHVPPYIFYPTTANVLSFFLHLPKLDETVSCEFRELTEP- 168
                   .:::*:* .. : *. ** :*.::: .:.. .*.: * *   :. :   * :

UGT71G1         QLLNIPGISNQVPSNVLPDACFNKDGGYIAYYKLAERFRDTKGIIVNTFSDLEQSSIDAL 231
UGT707B1        -EVEIPGVVSVPPGSMPTPLMDKKSRNYTWFVYHGRQFREAKGIVVNSVAELEPGVLSAM 225
UGT72B1         --LMLPGCVPVAGKDFLDPAQDRKDDAYKWLLHNTKRYKEAEGILVNTFFELEPNAIKAL 226
                  : :**       ..      .*.  *       .::::::**:**:. :** . :.*:

UGT71G1         YD----HDEKIPPIYAVGPLLDLKGQPNPKLDQAQHDLILKWLDEQPDKSVVFLCFGSMG 287
CsGLN21         AEGRFVEGGIMPTVYLVGPILSLADKGGG--SSSRNDECLVWLDEQPKGSVLFLCFGSMG 283
UGT72B1         QE----PGLDKPPVYPVGPLVNIGKQEAK---QTEESECLKWLDNQPLGSVLYVSFGSGG 279
                 :     .   *.:* ***::.:  :      .:...  * ***:**  **:::.*** *

UGT71G1         VSFGPSQIREIALGLKHSGVRFLWSN-----------------SAEKKVFPEGFLEWMEL 330
UGT707B1        -WFGVHQVREMATGLEQSGHRFLWSLRSMPAGDN----HQPTDANLDEVLPEGFLERTKD 338
UGT72B1         -TLTCEQLNELALGLADSEQRFLWVIRSPSGIANSSYFDSHSQTDPLTFLPPGFLERTKK 338
                  :   *:.*:* ** .*  ****                   :    .:* ****  :

UGT71G1         EGKGMICG-WAPQVEVLAHKAIGGFVSHCGWNSILESMWFGVPILTWPIYAEQQLNAFRL 389
UGT707B1        --RGMVWPSWVPQVEVLSHASVGGFVTHCGWNSVLESLWFGVPMIAWPQYAEQHLNEVEL 395
UGT72B1         --RGFVIPFWAPQAQVLAHPSTGGFLTHCGWNSTLESVVSGIPLIAWPLYAEQKMNAVLL 396
                  :*::   *.**.:**:* : ***::****** ** :  *:*: :** ****::* . *

UGT71G1         VKEWGVGLGLRVDYRKGSDVVAAEEIEKGLKDLMDKDS---IVHKKVQEMKEMSRNAVVD 446
UGT707B1        VRDMGVAVGMEVD-RKCGNFVTAAELERGVRCLMGESEESRRVRAKVADMKVAIRNALKE 454
UGT72B1         SEDIRAALRPRAG---DDGLVRREEVARVVKGLMEGEE-GKGVRNKMKELKEAACRVLKD 452
                 .:  ..:  ...    ...*   *: : :: **  ..    *: *: ::*    ..: :

UGT71G1         GGSSLISVGKLIDDITGSN--------- 465
UGT707B1        GGSSYTNLKKLAKDSCC----------- 471
UGT72B1         DGTSTKALSLVALKWKAHKKELEQNGNH 480
                .*:*   :  :  .
```

Figure 2. The isolated glycosyltransferase from saffron stigmas belongs to a new glycosyltransferase family., Amino acid sequence alignment of UGT707B1 against UGT71G1 and UGT72B1. The alignment was performed guided by conservation of secondary structure, predicted for UGT707B1 and observed from the solved crystal structures of UGT71G1 and UGT72B1. α-helices are highlighted in blue and β-strands in pink. The PSPG motif that interacts with the sugar donor, UDP-glucose, is underlined.

Phylogenetic characterization of UGT707B1

The genomic UGT707B1 gene had no introns, as the other related sequences present in the orthologous group OG7 [47]. This OG7 group contains multiple UGT families, with different functions, including abscisic acid (ABA) glucosyltransferase [48] and monolignol 4-O-glucosyltransferase [49]. As shown in a phylogenetic tree (Fig. 3) UGT707B1 showed the highest aminoacid sequence identity to *Shorgum bicolour* (53%), XP002460739; *Brachypodium distachyon* (53%) XP_003583056; to *Oryza sativa* UGT707A3 (52%) and 51% to *O. sativa* NP_001059726.1, *Zea mays* NP00114865, and *Phyllostachys edulis* FP092236.1. Interestingly, all the UGT707B1 related sequences shown had conserved the glycosylation site NYTW (data not shown) while the TargetP v1.1 programme predicted the presence of signal peptides in all these sequences. In this monocotyledoneus group only UGT707A3 has been characterized in vitro and acts as a flavonoid-3-O-glucosyltransferase [50]. When UGT707B1 is compared with the most related sequences from dicotyledoneus plants, the highest identity is found with *Vitis vinifera* (43%) CAN77507.1, and to *Populus trichocarpa* (41%) XP002305022, which are present in the UGT88 group (Fig. 3). UGT88 is considered to be a functional group represented by flavonoid 7-O-glycosyltransferases (the 7-position of isoflavones and the corresponding position of other flavonoids) [46, 51-52]. Among the functionally characterized UGTs, UGT707B1 shares the highest, 40% identity with UGT88A4 (*Maclura pomifera*), which did not exhibit any activity towards flavonoids and isoflavonoids, but catalyzed glucosylation of coumarin susbtrates, although the in vivo substrate is unknown [53]. The same identity is shown with UGT88A8 and UGT88A9 (*Hieracium pilosella*), which showed activity with flavones, flavonols, caffeic acid, esculetin, catechol, resorcinol, and hydroquinone [54].

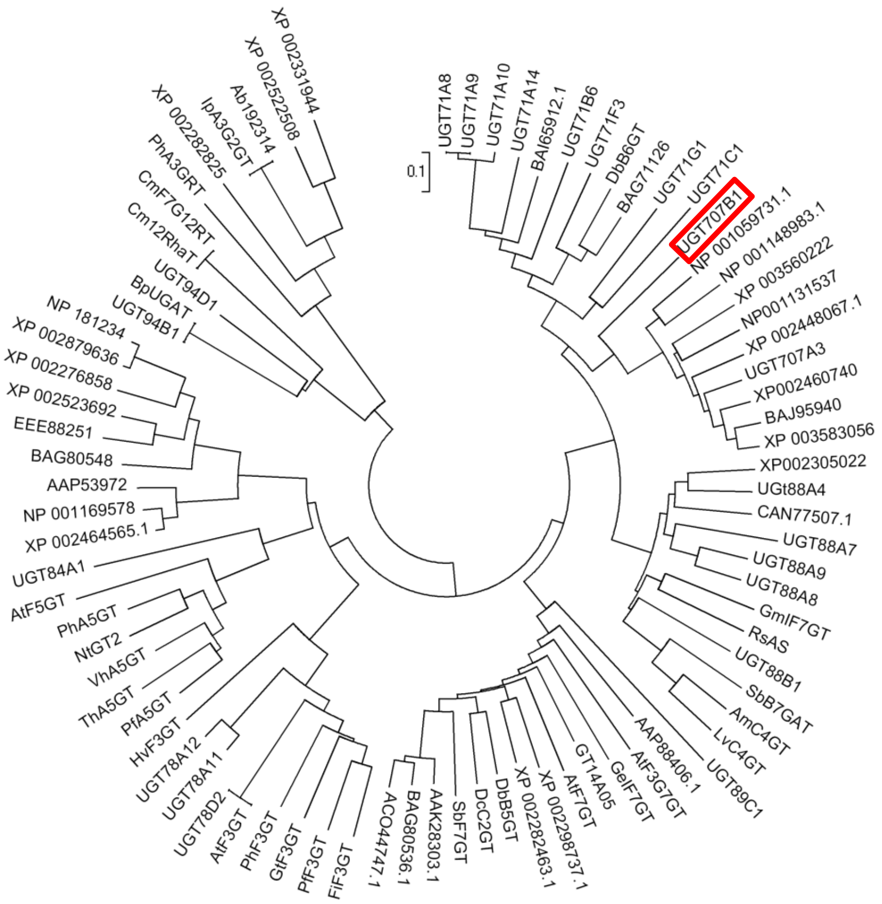


Figure 3. Non-rooted molecular phylogenetic tree of plant glycosyltransferases.

The tree was constructed as described in Material and Methods. The GenBank accession numbers not present in the tree are shown in parentheses: UGT71A8 (BAF96581), UGT71A9 (BAF96582), UGT71A10 (BAF96583), UGT71A14 (BAG80554), UGT71B6 (AEE76551), UGT71F3 (BAF75886), DbB6GT (AAL57240), UGT71G1 (AAW56092), UGT71C1 (AEC08300), UGT707B1 (HE793682), UGT707A3 (BAC83989), UGT88A4 (ABL85471), UGT88A7 (BAG31949), UGT88A9 (ACB56925), UGT88A8 (ACB56924), GmIF7GT (NP001235161), RsAS (Q4R1I9), UGT88B1 (AAR06919), SbB7GAT (Q76HR7), AmC4GT (BAE48239), LvC4GT (BAE48240), UGT89C1 (Q89C1), AtF3G7GT (NP181217), GelF7GT (BAC78438), GT14A05 (ADV71369), AtF7G7 (NP567955), DbB5GT (CAB56231), DcC2GT (BAD52006), SbF7GT (BAA83484), FIF3GT (AAD21086), PfF3GT (BAA19659), GtF3GT (BAA12737), PhF3GT (BAA89008), AtF3GT (NP180534), UGT78D2 (AED92377), UGT78A11 (CAN74919), UGT78A12 (BAI22846), HvF3GT (P14726), PfA5GT (BAA36421), BAC54093 (ThA5GT), VhA5GT (BAA36423), NtGT2 (BAB88935), PhA5GT(BAB89009), AtF5GT (AAL69494), UGT84A1 (XP002868202), UGT94B1 (Q5NTH0), BpUGAT (Q5NTH0), UGT94D1 (BAF99027), CmF7G12RT (AFB73772), CmF7G12RT (Q8GVE3), PhA3GRT (CAA50376), IpA3G2GT (BAD95582).

Expression analysis during stigma development and in other tissues

As a step toward functional analysis, we examined the expression pattern of UGT707B1 in *C. sativus* flowers using a reverse transcription (RT)-PCR procedure. These analyses revealed that UGT707B1 was expressed in all the flower organs examined (Fig. 4A). In stigmas, UGT707B1 was detected in all the developmental stages and reached the highest level at the time of anthesis. The expression pattern of UGT707B1 was also examined in vegetative organs, leaves and corms, and expression was detected in all these tissues except in the corm (Fig. 4B). The expression of UGT707B1 was also analysed along the style and stigma tissue, with higher expression levels in the style (Fig. 4C), which has been characterized by a higher content in flavonoids [55].

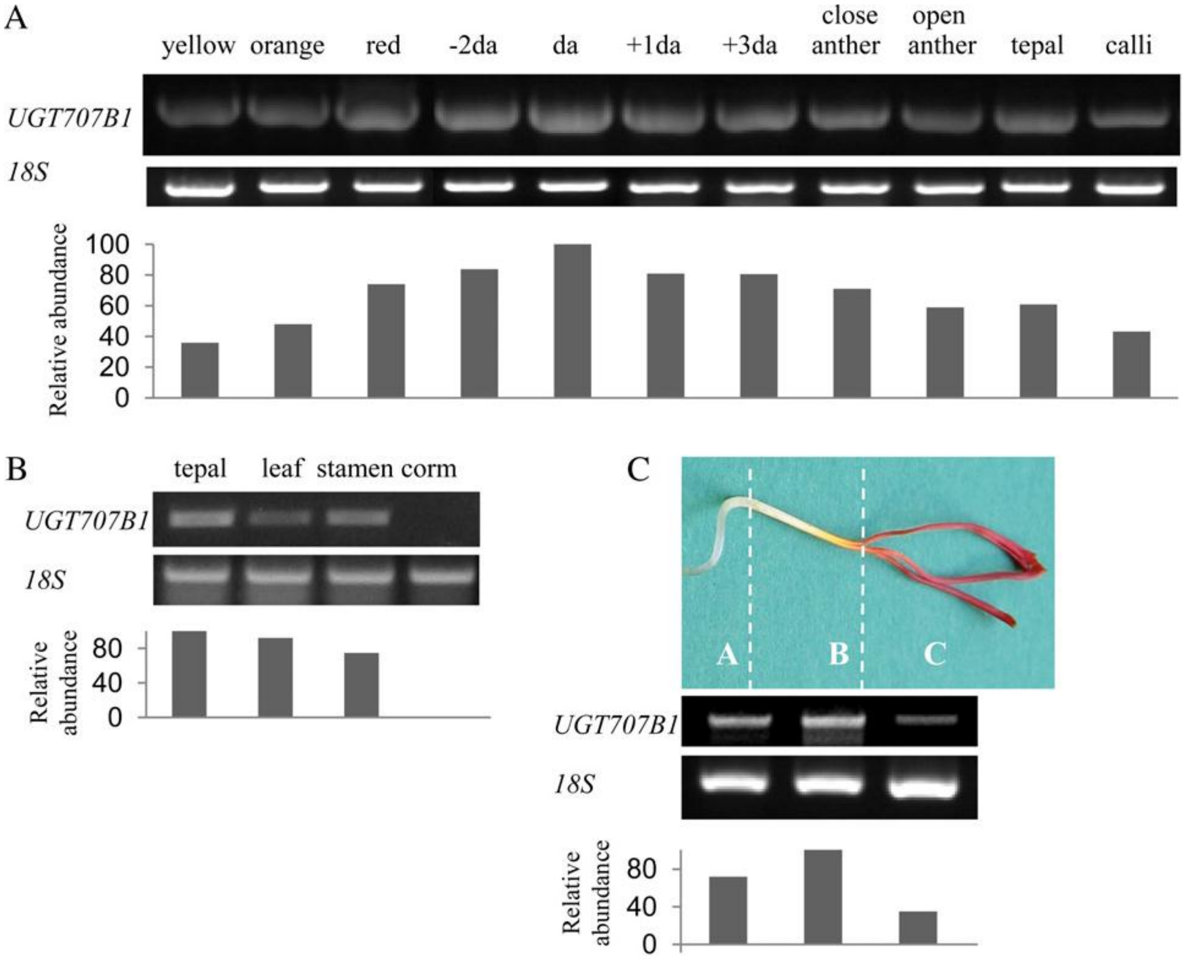


Figure 4. Expression analysis of UGT707B1 in *Crocus sativus* tissues determined by RT-PCR. **A**, The level of UGT707B1 expression was analysed in the stigma tissue of *C. sativus* in different developmental stages: yellow, orange, red, two days before anthesis (-2da), anthesis (da), one day after anthesis (+1da), and three days after anthesis (+3da). Expression levels for UGT707B1 were also analysed in closed and open anthers, tepals and calli. **B**, Levels of UGT707B1 expression in tepals,

leaves, stamens and corm. **C**, Expression levels in style and stigma tissues: A, white style; B, yellow style and C, stigma.

The expression of UGT707B1 in the tepals of different *Crocus* species was associated with the presence of kaempferol-3-O-β-D-glucopyranosyl-(1-2)-β-D-glucopyranoside.

Although *C. sativus* is mainly appreciated for its long red stigmas, other crocuses have been exploited by plant breeders and gardeners for their ornamental value. The expression of UGT707B1 was analysed in several *Crocus* species which highly differ in stigma and tepal colouration, due to differences in flavonoid and anthocyanin content [27]. Flavonols, although possessing a weak yellow colour, as pure substances, are potential co-pigments, and their sugar units are important for their interactions with anthocyanins [56]. UGT707B1 was detected in the stigmas at anthesis in all the crocuses analysed (Fig. 5). However, when the expression was analysed in tepals, the transcripts of UGT707B1 were undetectable in *C. speciosus*, *C. pulchellus* (Fig. 5) and *C. chrysanthus* (data not shown). In order to determine the possible correlation between UGT707B1 expression levels and the presence of specific glucosides, flavonols and anthocyanins were extracted from the tepals of *C. sativus*, *C. carwrigthianus*cv “albus”, *C. ochroleucus*, *C. speciosus, C. pulchellus* and *C. chrysanthus*.

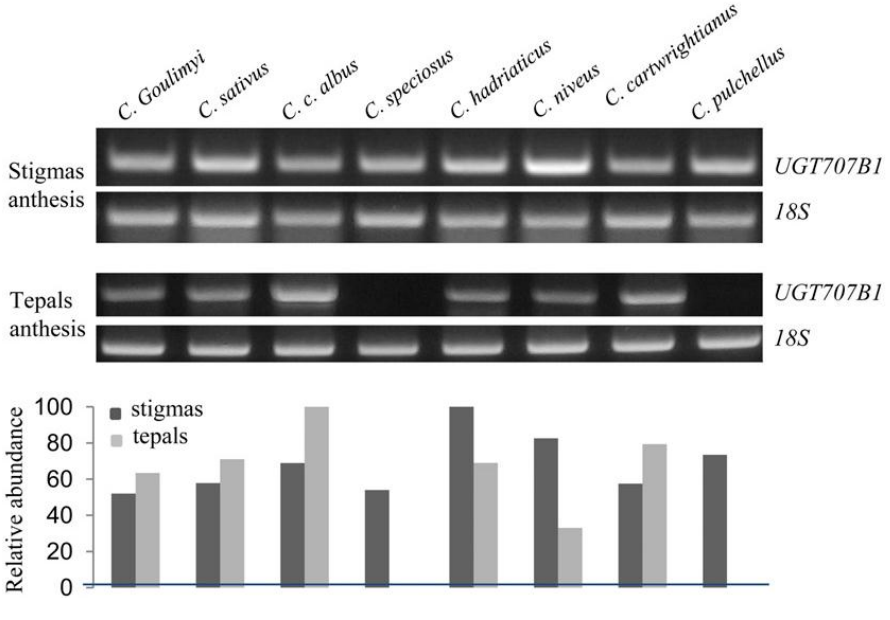


Figure 5. Expression analysis of UGT707B1 in the stigmas and tepals of different *Crocus* species. The levels of the constitutively expressed *18S*coding gene were assayed as controls. The PCR products were separated by 1% (w/v) agarose gel electrophoresis and visualized by ethidium bromide staining.

HPLC-ESI-MS/MS analyses were performed with each sample (Fig. 6). First of all, the samples were compared and searched for both different and identical peaks in terms of retention time, UV spectra and mass data. In *C. sativus*, *C. carwrigthianus* “albus” and *C. ochroleucus* samples, several major peaks were identified as common (Fig. 6A). Using the MS and UV data, these peaks were identified as kaempferol-3-O-β-D-glucopyranosyl-(1-2)-β-D-glucopyranoside (kaempferol 3-O-β-sophoroside) and quercetin-3-O-β-D-glucopyranosyl-(1-2)-β-D-glucopyranoside (quercetin 3-O-β-sophoroside) [27], whereas in *C. speciosus*, *C. pulchellus* and *C. chrysanthus* the identified major peaks were different (Fig. 6B) and corresponded to quercetin 3-O-α-(2-O-β-glucosyl)-rhamnoside-7-O-β-glucoside (compound 1) and kaempferol 3-O-α-(2-O-β-glucosyl)-rhamnoside-7-O-β-glucoside (compound 2). The presence of 3-O-β-sophorosides in *C. sativus*, *C. carwrigthianus* “albus” and *C. ochroleucus*, along with their absence from the other analysed *Crocus* species, in which UGT707B1 transcripts were not detected suggested a role of UGT707B1 in the formation of these compounds, with a putative flavonol 3-O-glucoside: 2''-O-glucosyltransferase activity.

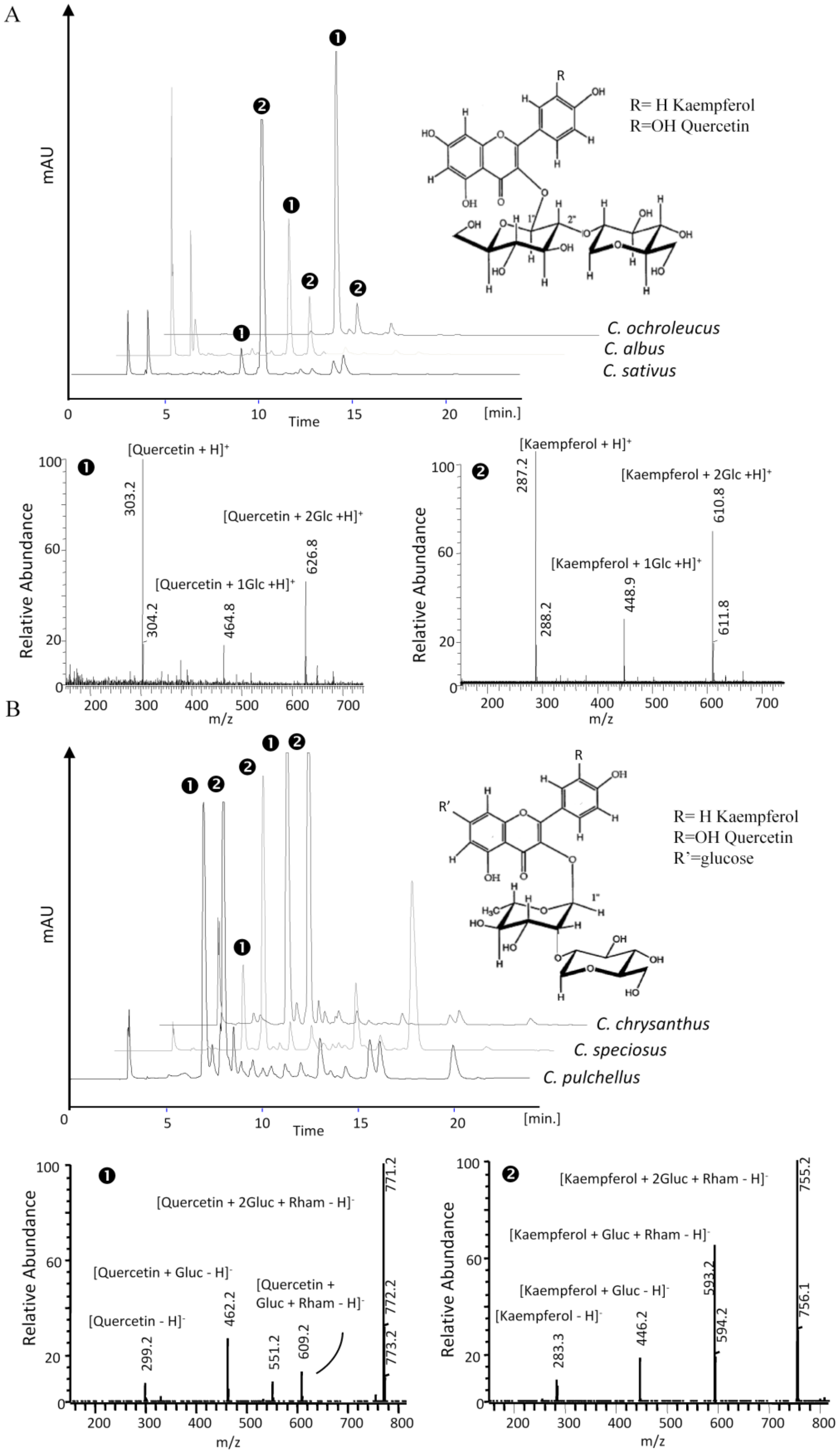


Figure 6. HPLC/DAD/MS analyses of the extracts from tepals of different *Crocus* species.
A, Flavonoid glycosides accumulated in *C. sativus*, *C. carwrigthianus* “albus”, and *C. ochroleucus*. Compound 1, which has the m/z value correspond to quercetin 3-*O*-sophoroside (m/z 626.8, RT 9.14

min). Compound 2, which has a m/z value correspond to kaempferol 3-*O*-sophoroside (m/z 610.8, RT 10.24 min). **B**, Flavonoids glycosides accumulated in *C. speciosus*, *C. pulchellus* and *C. chrysanthus*. Compound 1, which has the m/z value correspond to quercetin 3-*O*-α-(2-*O*-β-glucosyl)-rhamnoside-7-*O*-β-glucoside (m/z 771.2, RT 6.65 min). Compound 2, which has m/z value correspond to kaempferol 3-*O*-α-(2-*O*-β-glucosyl)-rhamnoside-7-*O*-β-glucoside (m/z 756.2, RT 7.78 min).

NMR characterization of the putative kaempferol 3-O-β-sophoroside

In order to elucidate the complete structure of the putative kaempferol 3-O-β-sophoroside, the *C. sativus* extract was separated by HPLC and the corresponding chromatographic peak was manually collected to be further analyzed by NMR. HPLC-ESI-MS/MS data indicated the presence of the flavonoid kaempferol and two units of glucose. Then, unequivocal structural elucidation of this compound was carried out by the combined use of 1D and 2D [^{1}H, ^{1}H] (gCOSY and TOCSY) and [^{1}H-^{13}C] NMR experiments (multiplicity-edited gHSQC and gHMBC). ^{13}C and ^{1}H NMR data confirmed the kaempferol skeleton (Table 1). The ^{1}H-NMR spectrum of the compound recorded in D2O, at 293K showed broad signals in the aglycone region. In particular, the low intensity of H-C(8) as compared to other protons of aglycone was attributed to a keto-enoltautomerism. Finally, in order to obtain resolved multiplets, experiments were carried out at 313 K.

Table 1.^{1}H (500 MHz) and ^{13}C (125 MHz) NMR spectral data for compound 2 from *C. sativus* tepals extract. Chemical shift (δ, ppm) and coupling constants (J in Hz, in parenthesis).

	Atom number	^{13}C	^{1}H
Kaempferol*	2	157.18	
	3	133.51	
	4	178.20	
	4a	104.44	
	5	160.11	
	6	94.10	6.18
	7	165.52	
	8	93.23	6.40
	8a	158.16	
	1'	121.94	
	2',6'	131.31	7.90
	3',5'	115.51	6.91
	4'	158.82	
Glucose"	1"	99.78	5.10 (7.5)
moiety	2"	79.53	3.79
	3"	76.12	3.58
	4"	69.33	3.35
	5"	76.12	3.21
	6"	60.61	3.62, 3.45
Glucose "'	1"'	102.32	4.77 (7.9)
moiety	2"'	73.71	3.27
	3"'	75.	3.42
	4"'	69.73	3.31
	5"'	76.12	3.31
	6"'	60.96	3.75, 3.62

*For comparison with corresponding values for kaempferol 3-O-glycosides in DMSO at 95ºC, see [57].

The presence of two units of glucose was confirmed from the observation of twelve carbons (two of each anomeric carbon) in the 103-60 ppm range of the ^{13}C-NMR spectrum. ^{1}H NMR spectrum showed two doublets in the anomeric region, its coupling constants indicated a β-configuration for both glucoses. The TOCSY experiment led to the identification of the ^{1}H signals of the two glucose residues while a multiplicity-edited gHSQC spectrum was used to link the carbon signals to the corresponding proton resonances. Finally, the HMBC experiment showed the linkages present in the structure. Correlations observed between the Glu'''-H^{1}anomeric proton (4.75 ppm) and the Glu''-C2 carbon (79.53 ppm) and between the Glu'''-C1 anomeric carbon (102.32 ppm) and the Glu''-H2 proton (3.77 ppm) revealed the presence of a β-D-glucopyranosyl-(1->2)-β-D-glucopyranoside moiety. In addition observed correlation between Glu’’-H1 anomeric proton (5.10 ppm) and C-3 (133.51 ppm) revealed the linkage with the aglycone moiety. Therefore compound 2 was identified as kaempferol-3-O-β-D-glucopyranosyl-(1->2)-β-D-glucopyranoside. Full assignments are collected in Table 1. ^{13}C-NMR chemical shifts were in accordance with the literature data [57]

Biochemical characterization

To identify the substrate specificity of UGT707B1, the cDNA was cloned into the expression vector pGEX-5X-3. UGT707B1 eluted after glutathione-sepharose purification did not show appreciable activity over any of the assayed substrates, including kaempferol-3-glucoside as substrate. Variations in incubation buffers, time and/or temperature did not improve the results. Thus, we used other expression vectors with different tags: pET100 (Invitrogen), pDEST14 (Invitrogen) and pE-SUMOpro3 (Lifesensors), all three with His-tags and pMAL-p5X (NEB) (with maltose binding protein). In addition, we used in combination with some of these plasmids the chaperone plasmid pGro7 (Takara), in order to increase the recovery of soluble proteins. Furthermore, different several *E. coli* strains were also tested without success. Thus, whole-cell biotransformation was carried out according to Cartwright et al. [39], but again we were unable to detect any activity.

Over-expression of UGT707B1 in *A. thaliana*

Due to the lack of activity in the in vitro assays over many of the substrates tested, we decided to develop transgenic *Arabidopsis* plants expressing UGT707B1. The CaMV 35S promoter was used to drive expression of UGT707B1 in transgenic *Arabidopsis* plants. Nine independent lines over-expressing UGT707B1 were taken to homozygosity (T3). RT-PCR was carried out to confirm the steady-state levels of its transcripts in each of the lines. The UGT707B1-expressing plants displayed phenotypes that differed from those of the wild-type plants. Figure 7 and 8 show data for six lines selected as non-expressing (lines 9 and 12), and expressing (lines 1, 3, 4 and 5), compared to wild-type (Fig. 7A). The transgenic expressing lines show aberrant cotyledon and leaf development, with a peripheral zone that is bent upward compared with Col-0 (Fig. 7B). This is referred to as hyponastic growth and is the result of asymmetric growth of the adaxial and abaxial surfaces of the cotyledon [58]. Alterations in stem thickness were observed (Fig. 7C) and flowering time was also delayed in the expressing lines up to two weeks (Fig. 7D). The flowering-time was defined by the number of rosette leaves when floral buds were clearly visible in Col-0, UGT707B1-9.1, UGT707B1-12.1 and UGT707B1-12.2 lines. The presence of flower buds is highlighted in Fig.7D by grey bars, while dark bars denote the absence of flower buds at the time of the experiment. In fact, when siliques were already present in Col-0 plants, flower buds were still absent from some of the expressing UGT707B1 transgenic lines (Fig. 9). Other phenotypic characteristics were observed in the expressing lines, the trichomes of the first rosette leaves showed an altered morphology with a significant reduction in their number compared with the wild type and non-expressing lines (Fig. 8)

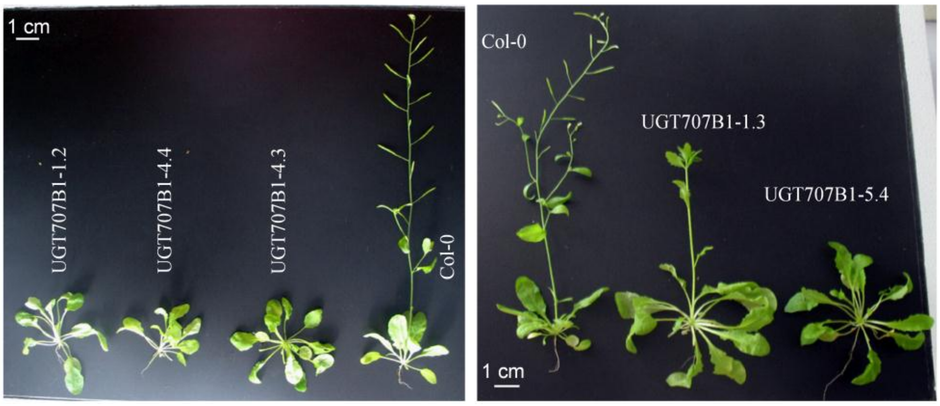


Figure 9. Late flowering transition phenotype of the UGT707B1 transgenic plants compared to wild type plants.

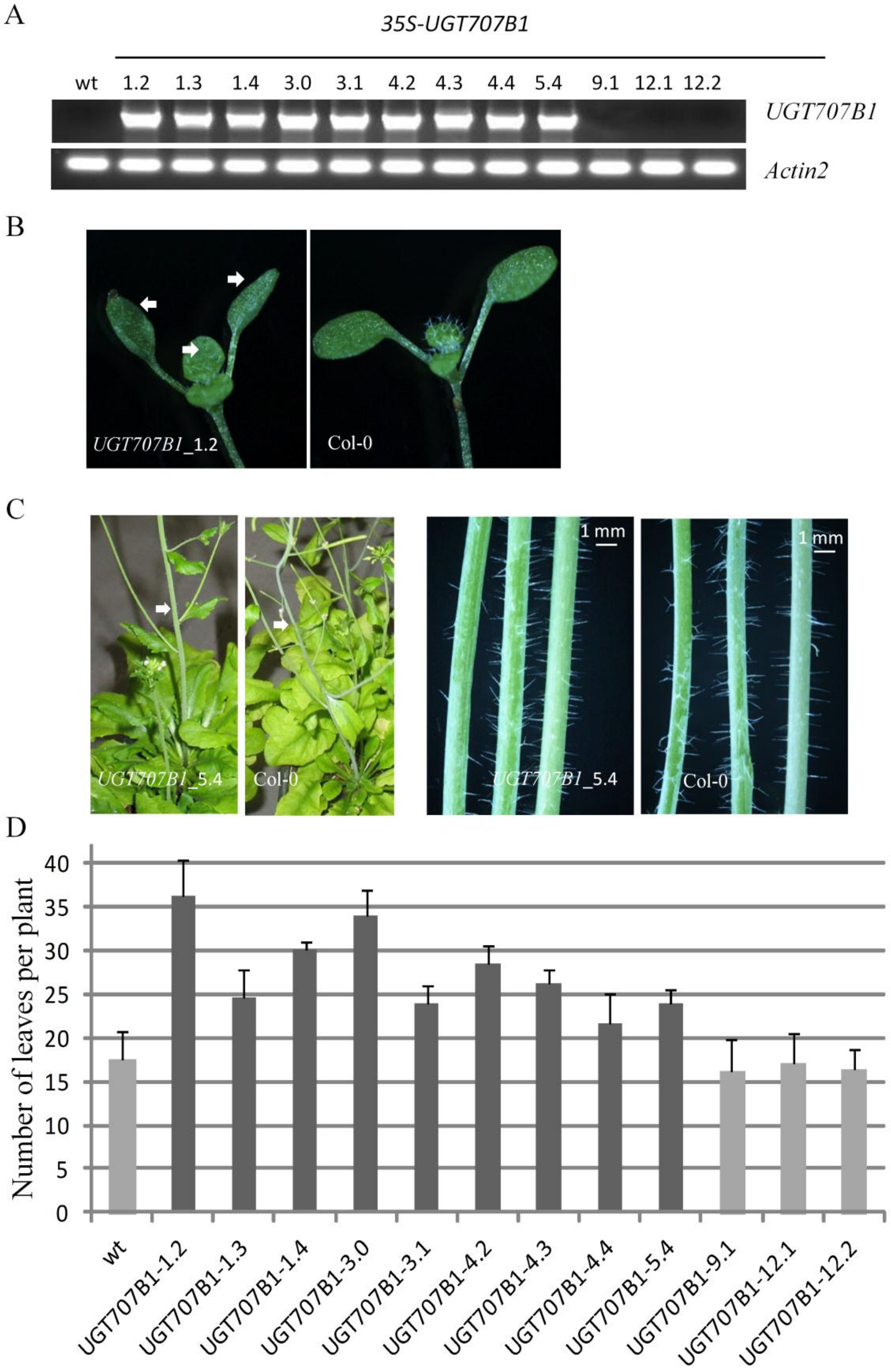


Figure 7. Phenotype of *Arabidopsis* transgenic lines expressing UGT707B1.
A, RT-PCR reactions over Col-0 wild type plants and different transgenic lines transformed with UGT707B1. **B**, Aberrant cotyledons and reduced trichome number in the transgenic lines. The periphery of cotyledons is bent upward, referred to as hyponastic growth. Also, the leaves of UGT707B1-expressing lines contain fewer trichomes (indicated by white arrows). **C**, Thicker flower shoot in UGT707B1-

expressing lines. **D**, Rosette leaf numbers at bolting of Col-0, UGT707B1-expressing plants (lines 1.2, 1.3, 3.0, 3.1, 3.2, 4.2, 4.3, 4.4 and 5.4) and UGT707B1-non-expressing plants (lines 9.1, 12.1, 12.2) grown under LD conditions. At least 20 plants were scored for each line. Error bars indicate standard deviation.

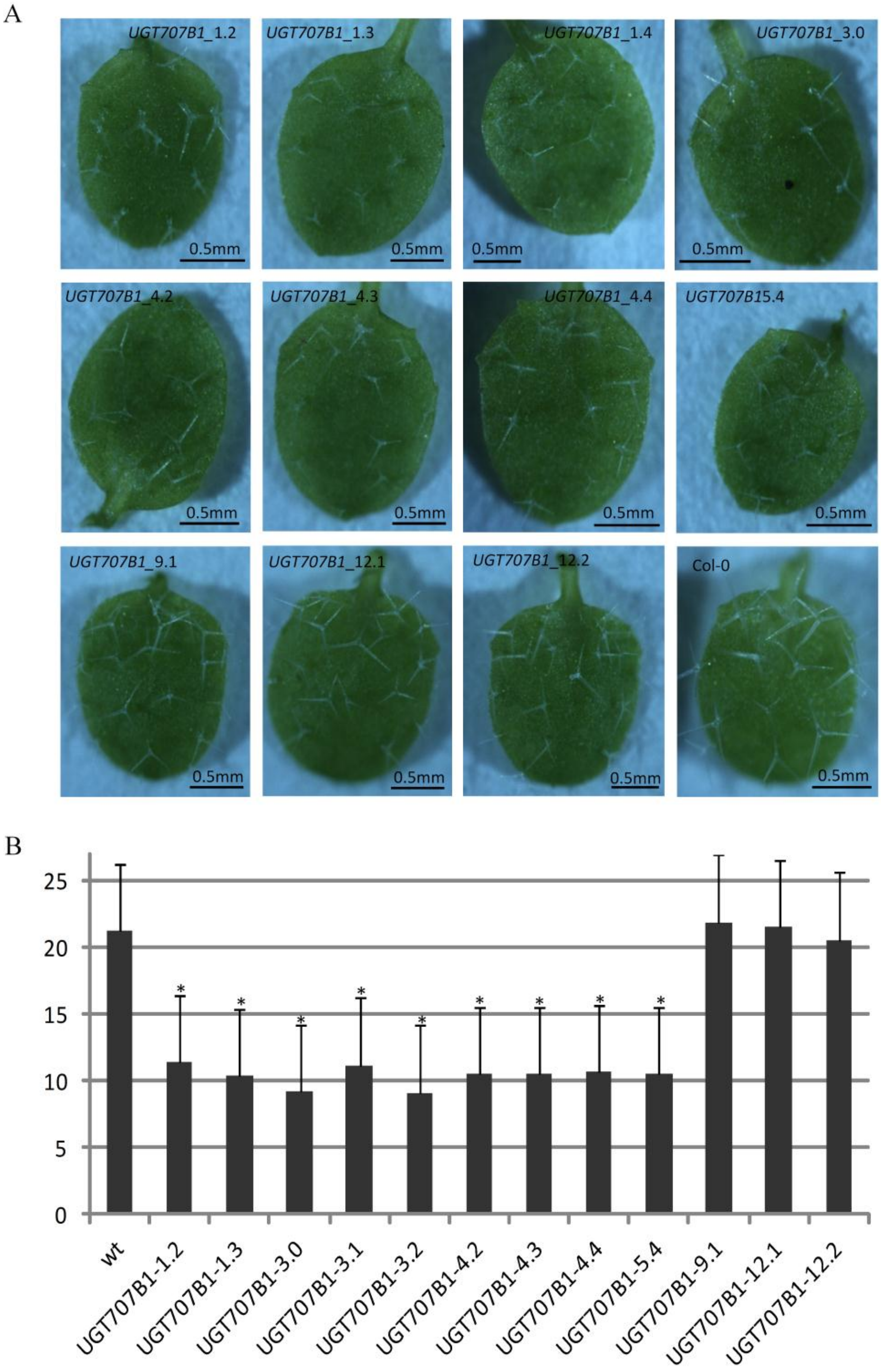


Figure 8. The UGT707B1-expressing lines showed a reduced number of trichomes. **A**, The first/second true leaves of Col-0, *UGT707B1*-expressing plants (lines 1.2, 1.3, 3.0, 3.1, 3.2, 4.2, 4.3, 4.4 and 5.4) and *UGT707B1*-non-expressing plants (lines 9.1, 12.1, 12.2). **B**, Number of trichomes in the first and second leaves for wild type, and *UGT707B1* transgenic lines. Error bars represent standard deviation, n ≥ 10. Values of mutant lines were compared with the wild type. *$p < 0.01$, calculated by Tukey HSD test.

Metabolite profiles in transgenic *Arabidopsis* plants

For the analysis of flavonol accumulation, wild-type, UGT707B1-1.2, UGT707B1-5.4, and UGT707B1-9.1 plants were grown for seven weeks, and different tissues were separately pooled and analyzed. The targeted metabolite analysis was performed by HPLC-DAD and HPLC/ESI/MS/MS. Most flavonol glycosides were identified by their UV light absorption spectra and ESI/MS/MS analysis. Recently published data were also used to interpret results [38, 59-62]. Flavonol levels were obtained by calculating the area below each HPLC peak per milligram dry weight of plant material. Figure 10 shows the elution profiles of rosette leaves, cauline leaves, stems and flowers at 254 and 340 nm. Several changes were observed between the wild type and the expressing lines. Some flavonols were less abundant than in wild-type tissues, others were elevated in the expressing lines, and new ones appear in the transgenic lines (Fig.10 and Table 2). In rosette and cauline leaves, the major flavonols, f3 (m/z 739.05, 593, 428.96, and 283.85; RT 29), f2 (m/z 593.01, 446.92, 430.95 and 284.89; RT 37.4) and f1 (m/z 577.1, 430.89, 284.88; RT 42.4), were detected in wild type and in transgenic lines (Fig. 10, Table 2). In transgenic lines an additional compound not present in wild type Col-0, fx, (m/z 755.0, 609, 431.02 and 284.85; RT 33 min) (Fig. 10 and Table 2) was detected. In flowers, changes in the relative levels of flavonols were also appreciated (Fig. 10 and Table 2). The *Arabidopsis* flowers contain the most flavonol derivatives in comparison with leaves and stems (Fig. 10). The flavonols f8 (m/z 757, 611, 449 and 303; RT 25.6), f3, f6 (m/z 611, 449 and 303; RT 33.3), f15, f26 (m/z 627.9 , 465 and 303; RT 35.4), f2, f5 (m/z 592.98, 446.94, 300.88, 178.82; RT 38.1), f14 (623, 476.94, 313.86; RT 38.1), f21, f1 and f30 (m/z 606.96, 460.95, 314.9; RT 43.1) were detected in wild type plants. In transgenic lines clear differences in the pattern of accumulation were observed. We detected the presence of f8, f3, f26, f14, f21, f1 and f30 at approximately equal levels as in the wild-type; f5 and f2 were reduced 3 to 4 times in comparison to the wild-type; while f6 and f15 flavonols were not detected. Instead, fx was again detected in this tissue, and high levels of f21 (m/z 608.99, 446.97, and 284.88; RT 39.4) and f26 were observed (Fig. 10 and Table 2). Stems showed a relatively low number of flavonol types in the wild type with the presence of the kaempferol derivatives f3, f2, and f1 at high levels (Fig. 10). In the transgenic lines expressing UGT707B1, an important reduction of these flavonols was observed, with up to three times less for f1 and f3, and more than a 30 times reduction for f2 (Table 2). Instead, f21 (m/z 609 and 286; RT 40.1 min) was detected as the second major peak in addition to f1 (m/z 578.83, 433.02 and 287.05; RT 42.08 min), other minor peaks were also detected as f3 (m/z 740.90 and 594.98; RT 29.6), f2 (m/z 595, 442.8, 287; RT 37.8 min) and fx (m/z 755.0, 609, 431.02 and 284.85; RT 33 min).

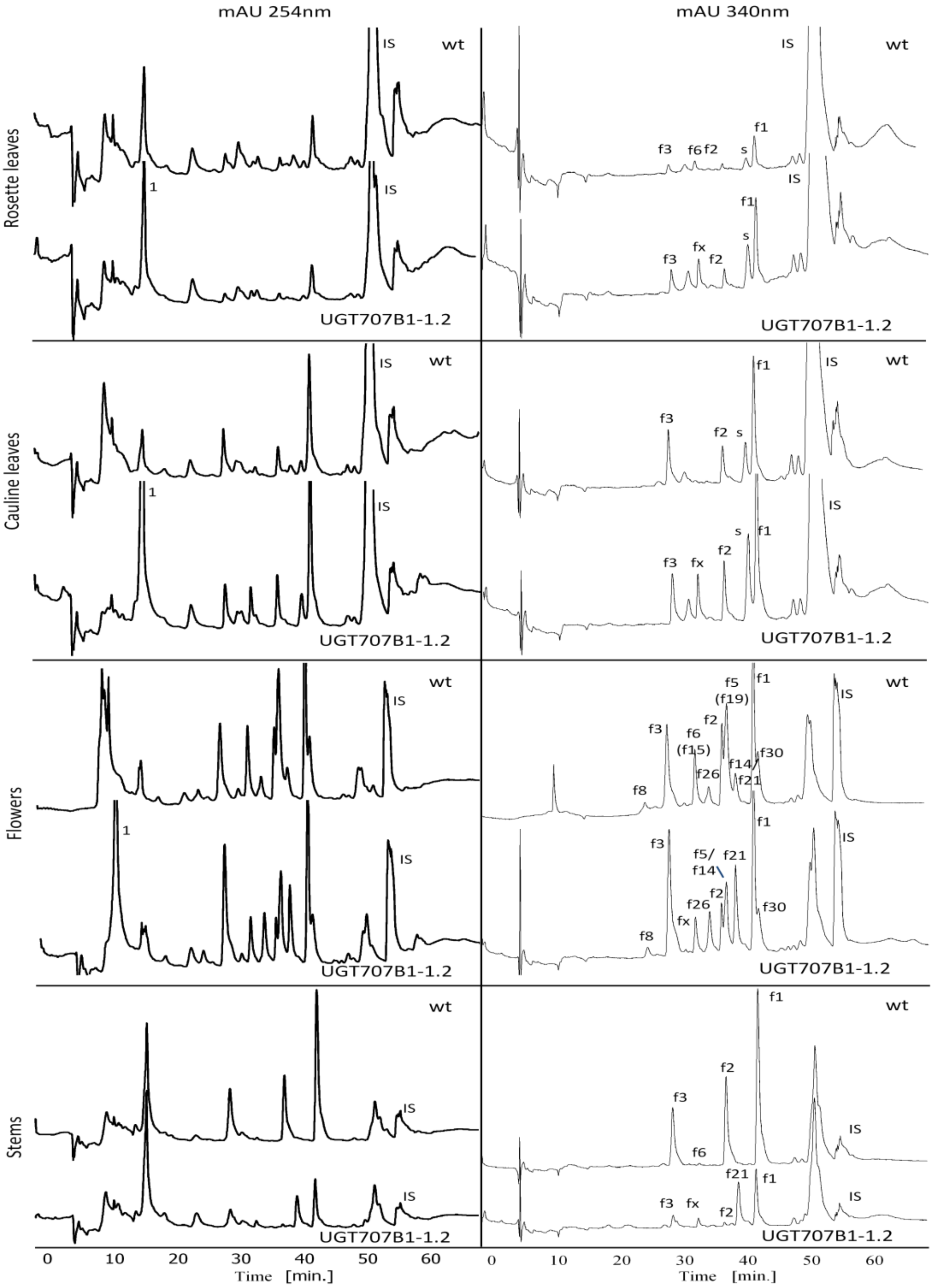


Figure 10. HPLC/DAD analyses of the extracts from rosette leaves, cauline leaves, stems and flowers of wild type and UGT707B1-expressing plants. Chromatograms of aqueous methanol extracts of wild type and UGT707B1-expressing plants. Absorbances at 254 and 340 nm were used for the detection of flavonols. Labels correspond to compounds previously identified [62] and are shown in Table 3.

Table 2. The flavonol profiles of leaves, stems and flower extracts from wild type and UGT707B1-*1.2* plants.

compound			WT			UGT707B1-1.2			UGT707B1-1.2:WT ratio
	WT	UGT707B1-1.2	Rt (min)	Uv spectra	Peak area	Rt (min)	Uv spectra	Peak area	
Rossete leaves	f3	**f3**	29	268, 348	33.08	29	268, 348	35.4	1.07
	f6	**fx**	33	280, 310	17.36	33	265, 345	32.45	
	f2	**f2**	37.4	265, 348	21.9	37.4	265, 348	17.4	0.79
	f1	**f1**	42.4	264, 342	146	42.4	264, 342	133.4	0.91
Cauline leaves	f3	**f3**	29.3	268, 348	221	29.3	268, 348	181	0.82
	-	**fx**	33.2	n.d.	9.8	33.2	265, 345	46.4	
	f2	**f2**	37.7	265, 348	173.5	37.3	265, 348	135.7	0.78
	f1	**f1**	42.4	264, 342	431.6	42.4	264, 342	536.8	1.24
stems	f3	**f3**	29.6	268, 348	135.75	29.6	268, 348	31.05	0.22
	f6	**fx**	33.6	271, 340	3.7	33.6	265, 345	20.54	
	f2	**f2**	38	265, 348	174.4	37.8	265, 348	5.43	0.03
	-	**f21**	n.d.	n.d.	n.d.	40.1	265, 348	105.6	
	f1	**f1**	42.8	264, 342	302	42.4	264, 342	116.03	0.38
flowers	f8	**f8**	25.6	260, 352	62.7	25.9	260, 352	60.9	0.97
	f3	**f3**	29	268, 348	499	29.2	268, 348	720.2	1.44
	f6+f15	**fx**	33.3	261, 353	280.8	33.3	265, 345	191.4	
	f26	**f26**	35.5	260, 352	106.13	35.4	260, 355	281.04	2.64
	f2	**f2**	37.3	265, 348	255.5	37.3	265, 348	140.12	0.54
	f5+f14	**f5+f14**	38.1	258, 353	740.7	38.1	258, 351	438.5	0.59
	f21	**f21**	39.4	265, 350	145.5	39.4	265, 348	350.81	>2.41
	f1	**f1**	42.3	264, 342	596	42.3	264, 342	563.2	0.95
	f30	f30	43.1	263, 350	259.17	43.1	265, 340	241.5	0.93

Flavonols were quantified by measuring peak area using a response value of the peak area of the internal standard. N.d., not detected.

The new peaks and increasing peaks in the transgenic UGT707B1-*1.2* line are highline with a grey background. Flavonols correspond to: f1, kaempferol-3-*O*-rhamnoside-7-*O*-rhamnoside; f2, kaempferol-3-*O*-glucoside-7-*O*-rhamnoside; f3, kaempferol-3-*O*-rhamnosyl(1-->2)-glucoside-7-*O*-rhamnoside; f5, quercetin-3-O-rhamnoside-7-*O*-rhamnoside; f6, quercetin-3-*O*-glucoside-7-O-rhamnoside; f8, quercetin-3-*O*-rhamnosyl(1-->2)-glucoside-7-*O*-rhamnoside; f14, isorhamnetin-3-*O*-glucoside-7-*O*-rhamnoside; f21, kaempferol-3-*O*-sophoroside; f26, quercetin-3-*O*-sophoroside; f30, isorhamnetin-3-O-rhamnoside-7-*O*-rhamnoside; fx, kaempferol-3-*O*-glucosyl(1-->2)-glucoside-7-*O*-rhamnoside.

The new compound that accumulates in UGT707B1-expressing lines was identified as a kaempferol with a rhamnose molecule and two glucoses. Mild acid hydrolysis of fx led to the detection of the aglycone kaempferol and kaempferol 3-sophoroside (f21). The UV study in methanol after the addition of UV shift reagents [63] showed that fx was a kaempferol derivative, with the hydroxyls at the 7 and 3-positions blocked. Furthermore, we determined the possible structure of fx based on the ions obtained through its fragmentation [64]. The ESI+ full MS resulted in the detection of a main peak of m/z 755.01, the MS2 of that peak resulted in a fragment of m/z 609.00 [755.01-146.01], suggesting the loss of a rhamnose molecule; the MS3 of the 609 peak generated a fragment of m/z 284.85 [609-(2x162)]. Taking into consideration the presence of 3 and 7- substitutions on the *Arabidospsis* flavonols, this successive pattern of fragmentation of fx suggests a 7-substitution in the kaempferol molecule with a rhamnose molecule and a 3-substitution with a sophoroside molecule. Although this compound has not been detected in *Arabidopsis* Col-0 wild-type plants [62, 65], its presence has been reported in rosette leaves of the *Arabidopsis* ecotype C-24 [38].

Flavonol aglycones were not detected in any of the plant lines analyzed. Table 2 shows the spectral data of each of the identified flavonol species and the induction/repression factor in the expressing lines compared with the wild type. In general, flavonols containing 3-O-sophoroside were increased several fold. The clearest alteration was found for fx (kaempferol 3-O-sphoroside-7-O-rhamnoside), newly detected in all the tissues analysed in the transgenic lines which expressed UGT707B1. This new presence can be partially explained by the reduction observed in f2 (kaempferol 3-O-glucoside-7-O-rhamnoside). The second differentially detected compounds were f21 and f26. F21 (kaempferol-3-O-sophoroside), which has been only detected in wild-type flowers [62, 65], was detected in UGT707B1 expressing-lines in flowers up to 2.2 fold and newly detected in the stems. In the flowers of UGT707B1 expressing-lines, f26 (quercetin 3-O-sophoroside) was increased up to 2.3 fold. Thus, the increased levels in 3-O-sophoroside compounds may be responsible for the phenotype observed in UGT707B1-expressing lines (Fig.11)

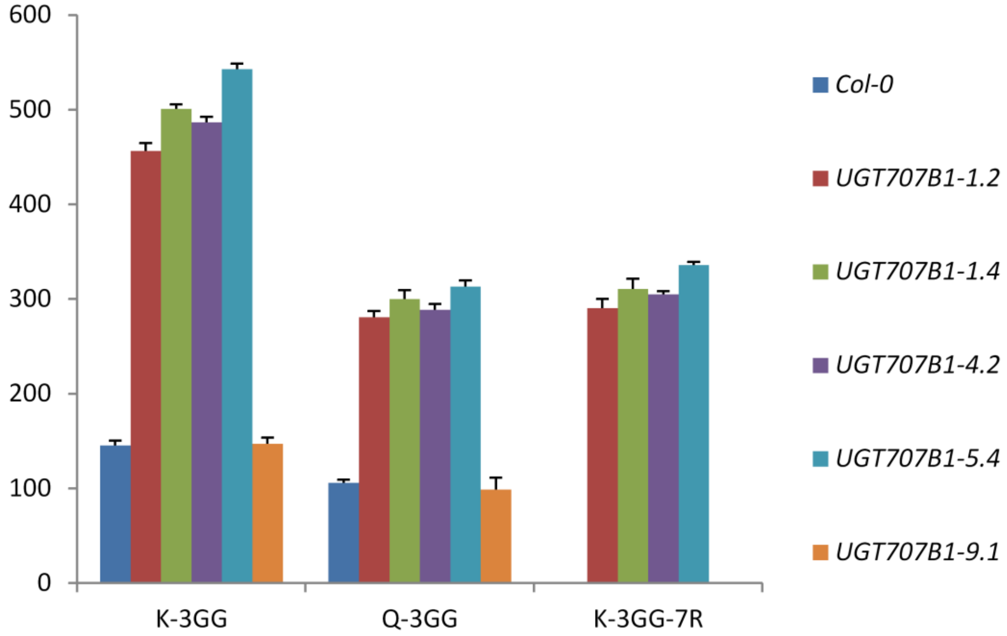


Figure 11. The Abundance of K-3GG (f21), Q-3GG (f26) and K-3GG-7R (fx) correlates with the phenotype observed in UGT707B1-expressing plants. K-3GG, Q-3GG and K-3GG-7R fulfill the predicted requirements for the flavonol species that cause the observed phenotypes. These compounds are found in small amounts or are absent in the wild type Col-0 and in the transgenic non-expressing line *UGT707B1-9.1* But are present in *UGT707B1-1.2*, *UGT707B1-1.4*, *UGT707B1-4.2* and *UGT707B1-5.4*. Mean ± se amounts of shoot flavonols are shown (n = 3). The area under the HPLC peak was used as the measurement for the amount of the flavonols, and corrected with the area of the internal standard.

Promoter analysis of UGT707B1

Variation in the flavonol pattern in different organs, and thus the formation and accumulation of specific flavonol compounds, could be caused by differences in the set of genes coding for those enzymes responsible for flavonol modification. In *Arabidopsis*, there is clear evidence that R2R3-MYB transcription factors are able to activate glycosyltransferase expression [62, 65-66], controlling flavonol accumulation in an organ- and development-dependent manner. Hence, we isolated and analysed the promoter region of UGT707B1 in order to identify elements that could be recognized by R2R3-MYB transcription factors. The promoter sequence showed multiple light-regulatory units (LRUs) (Table 3) as have been shown in the promoter sequences of genes of the phenylpropanoid biosynthesis [67].

Table 3. Putative cis-elements identified in the UGT707B1 promoter

Category	Cis-acting Element	Sequence	position
Light	ACE	ACGTGGA	-778, -630
	ABRE	GACACGTGGC	-729
	ATCT-motif	AATCTAATCC	-672
	G-Box	CACGTG	-727
	G-box	ACACGTGGC	-409, -729
	GATA-motif	GATAGGA	-468
	GTGGC-motif	CATCGTGTGGC	-967, -1199
	I-box	GATAAGGTC	-191, -970
	Chs-Unit 1ml	ACCTAACCTCC	-11
Salicylicacid response	SARE	TTCGACCTCCTT	-237
MYB-flavonol activation	MYBPZM	CCWACC	-967, -1110
MYB binding site involved in drought-inducibility	MBS	CAACTG	-302, -467
ABA response	CE3	GACGCGTGTC	-729
MYB binding site	MRE	AACCTT	-630, -1011
Core promoter element	TATA-box	ATATAT	-109
Common cis-acting element in promoter and enhancer regions	CAAT-box	CAAT	-97, -248, -662, -918, -1092

In *Arabidopsis,* these LRUs contain two distinct types of cis-regulatory elements, namely the ACGT-containing element (ACE; core sequence, CACGT) and the MYB-recognition element (MRE; core sequence, ACCTT). It has been demonstrated that MRE acts via binding MYB12, an R2R3-MYB type transcription factor [68]. These two elements are present in the promoter region of UGT707B1. In addition, a putative MYBPZM binding element was present, which is present in the A1 gene, required for 3-deoxy flavonoid and phlobaphene biosynthesis and recognised by Myb homolog P gene from maize [69]. There are no data about cis-elements recognized by other transcription factors known to modulate the expression of glycosyltransferases in different tissues [65, 70], therefore we do not know if such elements could be present in the promoter of UGT707B1. Nevertheless, in *A. thaliana* are not known the glucosyltransferase enzymes responsible for the formation of K-3GG and Q-3GG are unknown, as are the MYB transcription factors, which differentially influence the spatial accumulation of specific these flavonol derivatives.

Discussion

Molecular- and metabolite-analysis approaches were used to identify UGT707B1-metabolite correlations in order to determine the role of UGT707B1 in flavonol glucosylation in *C. sativus*. Since coupling the analysis of metabolites with co-expression has been used as an efficient method for identifiying glycosyltransferase functions [15,16, 62, 66, 71], we have used this strategy on several *Crocus* species which differ in the expression of a glucosyltransferase gene, UGT707B1, highly expressed in saffron flowers. Furthermore, these *Crocus* species differed in the activity of 3-O-glucoside-2"-O-glycosyltransferases in tepals. The non-expressing UGT707B1 group was characterized by high activity of a flavonol 3-O-glucoside-2"-O-rhamnosyltransferase, while the expressing-UGT707B1 group had an active flavonol 3-O-glucoside-2"-O-glucosyltransferase. In addition, the use of transgenic *Arabidopsis* lines expressing UGT707B1, effectively support the role of UGT707B1 in saffron as a flavonol 3-O-glucoside-2”-O-glucosyltransferase involved in the formation of flavonol 3-O-sophorosides.

The UDP-glucosyltransferase, UGT707B1, was identified in saffron stigmas based on the conservation of amino acid domains of plant UDP-glycosyltransferases associated with secondary metabolism [40, 71]. In the C-terminal domain, UGT707B1 contains a motif that shows similarity with the 44 amino acid-long PSPG sequence found in most known UDP-glycosyltransferases. The motif comprises amino acid residues playing a key role in binding the nucleotide-diphosphate sugar [72]. Particularly, tryptophan (W), glutamic acid (E) and glutamine (Q) residues found in the PSPG sequence of UGT707B1 (at positions 22, 43 and 44 of the motif, respectively), have been shown to be conserved in several plant UDP-glucosyltransferases, preferring UDP-glucose as the sugar donor substrate [73]. In fact, glucose is the unique glycoside present in saffron stigmas [55], where UGT707B1 was expressed throughout all the developmental stages, although its expression was developmentally regulated, with higher levels at the time of anthesis characterized by the highest accumulation of kaempferol sophorosides[28]. In addition, UGT707B1 transcripts were detected in other flower parts such as the style, rich in flavonols[55]; stamens and tepals, and were also present in leaves but absent from the corm.

The phylogenetic relationships did not allow the determination of the possible activity of UGT707B1, while only one of the most related sequences, UGT707A3 (with a 52% identity), showed a 3-O-glucosyltransferase activity toward kaempferol and quercetin[50]. Moreover, the biochemical basis of the enzymes from the most related family, the UGT71, is indecisive due to the significant structural diversity of their substrates. For example, *Medicago truncatula* UGT71G1 and *A. thaliana* UGT71C1 have glucosylation activity toward triterpene and coumarin in vitro, respectively [74-75]. *A. thaliana* UGT71B proteins are able to glucosylate a spectrum of different compounds such as abscisic acid and hydroxybenzoic acid [43, 48, 76] whereas UGT71A9 from *Sesamun indicum* catalyzed the glucosylation at the 2-hydroxyl group of (+)-sesaminol [77]. Since sequence homology is far from being a definitive argument to describe the precise enzymatic activity of glycosyltransferase enzymes [78] the functional characterization of these enzymes is still necessary. Furthermore, correlations between transcripts and metabolites could facilitate an efficient narrowing-down of candidate products of glycosyltransferase activities [62, 79-80].

Expression analysis of UGT707B1 in the stigmas and tepals of other *Crocus* species that differ in flavonol and anthocyanin accumulation allowed the determination of the possible products of UGT707B1 activity. The transcripts were detected in all the stigmas analysed, but were absent from the tepals of several *Crocus* species. Consistent with this, a comparison of the glycosylated metabolites from the tepals of these non-expressing species with those from the UGT707B1-expressing species showed the absence of kaempferol and quercetin 3-O-glucosyl-2"-O-glucosides, suggesting that these compounds could be the result of UGT707B1 activity. Furthermore, the amino acid sequence identity (25-29%) of UGT707B1 with UGTs catalyzing glycosylation at the sugar moiety of a glycoside [47, 81-82], is not far from that observed among these UGTs.

We were unable to determine the in vitro activity of UGT707B1 using a recombinant enzyme produced in *E. coli* and several substrates, including kaempferol and kaempferol-3-O-glucoside. Several reasons can explain the lack of activity of the recombinant enzyme; UGTs involved in secondary metabolism pathways are known to be unstable enzymes that occur only in minute amounts in plants [14], thus the first reason can be that UGT707B1 is highly labile, with loss of enzyme activity upon purification. However, when the activity was assayed in culture cells without purification no products were detected. The second reason could be that the produced recombinant enzyme in *E. coli* is inactive. UGT707B1 is predicted to contain a signal peptide and to be glycosylated, which suggests that it may play an important role in establishing or maintaining the activity. And the third but not last reason is that perhaps the substrates adequate for UGT707B1 activity were not present in the reaction. It cannot be excluded that UGT707B1 might not directly use UDP-glucose as a donor.

We therefore decided to functionally characterize UGT707B1 in vivo. The transgenic *Arabidopsis* lines expressing UGT707B1 showed characteristic phenotypes and were affected in the flavonol glucosylation profile. The observed phenotype of UGT707B1-expressing *Arabidopsis* lines closely resembles the phenotype previously described in the *Arabidopsis* rol-mutants, affected in the rhamnose synthase RHM1 [83], which accumulate K-3G and kaempferol-3-O-(6″-acetyl-glucoside) and show hyponastic cotyledon growth and aberrant trichomes on the first rosette leaves, most probably caused by kaempferol-induced modulation of auxin transport [84]. In addition to these phenotypes, the UGT707B1-expressing *Arabidopsis* lines showed a reduced number in trichomes, a delay in flowering-time and thicker stems. In *Arabidopsis*, flavonoid biosynthesis and modification have been shown to be involved in trichome formation [85-86], the transition to flowering [86] and flowering time [87]. The UGT707B1-expressing transgenic plants displayed a strong increase in 3-diglucosylated flavonols, with variations in the different tissues analysed. Compared with the wild type, there is a reduction in total amounts of flavonols in the shoots (f3, f2 and f1), along with the presence of new flavonols not detected in the wild type f21 (K-3GG) and fx (K-3GG-7R), where accumulations in shoots were correlated with the development of the observed phenotype. These compounds were completely absence in lines developing a wild-type shoot phenotype. In flowers, the pattern of flavonol accumulation is more complex [16, 62]. In the flowers of UGT707B1-expressing lines f2, f5 levels were reduced. By contrast, the levels of f26 (Q-3GG) and f21 (K-3GG) were increased, and fx (K-3GG-7R) appears as a new flavonol in this tissue. The observed reduction in f2 (K-3G7R) levels in the tissues of the transgenic lines could be explained if this compound is converted by UGT707B1 into fx (K-3GG7R). In addition, the reduction of kaempferol and quercetin 3-O-rhamnosides observed in flowers and stems could be due to the reduced availability of quercetin and kaempferol for their conversion by rhamnosyl-transferases, due to the increased transformation of those compounds by UGT707B1 in 3-O-diglucosides.

The K-3GG-7R, Q-3GG and K-3GG flavonols have been previously identified in the rosette leaves from *A. thaliana* C24 ecotype [38], which is not the only ecotype that has shown substantial qualitative and quantitative differences in flavonoid composition in comparison with Col-0 [66]. Interestingly, *A. thaliana* C24 is a vernalization-responsive, moderately late flowering ecotype in comparison with the early flowering ecotype Col-0 [88]. The UGT707B1-expressing lines were characterized as well by a delay in flowering transition, suggesting a possible implication of these glucosylated flavonols in the control of the flowering process. Recently, delayed flowering was reported in flavonol synthase (FLS) silenced tobacco lines, which could be due to a decrease in level of indole acetic acid (IAA) in the apical region of their shoots [89], while the glycosyltransferase UGT87A2 activity, in a way still unknown, regulates flowering time via the flowering repressor FLC [87]. Genetic and pharmacologic evidence clearly demonstrate a role for flavonols as negative regulators of auxin transport. The flavonols quercetin and kaempferol have been identified as the most active flavonoids acting as regulators for the transport of auxins[6, 90-91] and can displace synthetic auxin transport inhibitors in vitro [90] and in vivo [84]. Furthermore, flavonol biosynthesis was recently shown to be induced by auxin through a transport inhibitor response 1 (TIR1) auxin receptor-dependent pathway [92].

Plant glycosyltransferase enzymes are assumed to be cytosolic [78]. However, recently, the UGT87A2 enzyme, controlling flowering time in *Arabidopsis,* was located in both cytoplasm and nucleus [87]. In addition, the results presented in this work suggest UGT707B1 as being both a cytoplasmic and nuclear protein alike, although the TARGETP program, which has high sensitivity and specificity for plant sequences, showed the presence of a signal peptide in the UGT707B1 sequence which was predicted to be in the secretory pathway. The detection in the nuclei was unexpected due to the absence of a nuclear location signal (NLS). Nevertheless, in other well known plant proteins that are localized in the nuclei, the NLS domain is not necessary for nuclear localization [93]. Typically, small molecules (~50 kDa) diffuse freely in and out of the nucleus through the nuclear pores, and the calculated 48 kDa of the mature UGT707B1 is smaller than the reported permeability of the nuclear pore complex by passive diffusion [94]. Although the prevailing model suggests that flavonoids are synthesized in a complex of metabolic enzymes located on the cytoplasmic face of the endoplasmic reticulum [95], flavonols accumulate in the nucleus and cytoplasm of different plant species [96-98] as the flavonoid biosynthetic enzymes CHS, CHI [99] and FLS [84], which suggests that the flavonol biosynthetic machinery is also active in the nucleus where flavonols might modulate gene expression [86, 100-102]. Interestingly, the UGT87A2 enzyme has been found in the cytosol and nucleus [87], a

Medicago truncatula (iso)flavonoid β-glucosidase, which digests pre-stored (iso)flavonoid glucoconjugates, was found to localize in the nucleus of the cells under stress conditions [103] and an isoflavone malonyltransferase involved in the formation of (iso)flavonoid conjugates localize in the nucleus as well [104]. Thus, the observed nuclear location of UGT707B1 suggests that in addition to flavonoid biosynthetic enzymes, flavonoid modifying enzymes are also present in the nuclei and can modulate the activity of flavonoids in this compartment, controlling different developmental processes in the plant.

Flavonols are involved in a vast array of biological functions in plants, and the presence of different flavonol-glycosylation profiles in plant tissues suggests specific functions of these compounds. As example, the expression of UGT707B1 in the floral organs of *C. sativus* is responsible for the presence or absence of a specific set of flavonols. In tepals, flavonoids could directly contribute to colour formation [105] in order to attract pollinators [106]. In the stigma tissue UGT707B1 reaches its highest expression levels at the time of anthesis coincident with the highest accumulation of kaempferol sophorosides[28]. The accumulation of flavonols at the time of anthesis have been also observed in other plant species [107] and could be associated with floral morphogenesis and pollination [108]. Nevertheless, one of the most important roles of flavonols is to influence the transport of the plant hormone auxin. In this study, the phenotype displayed by UGT707B1 transgenic *Arabidopsis* plants suggests an implication of these flavonols in auxin transport. Auxin gradients have been proposed to determine flower morphology: high auxin levels in the apical region promote differentiation and proliferation of the style and stigma, intermediate levels specify the ovary, and low levels in the basal region specify the gynophore [109]. In *Arabidopsis*, alteration of these gradients by inhibition of polar auxin transport leads to an elongation of the style and stigma in the apical part whereas the ovary size concomitantly decreases [110]. The resulting plants showed a phenotype that could resemble the morphology of *C. sativus* flowers. The female organs of *C. sativus* consist of a trilocular ovary (0.5-0.7 cm long), a very long style (10-12 cm), and 3 red stylar branches (3-4 cm long) which are folded to give the stigmas a trumpet-like structure [111]. Auxins participated in the elongation of floral tube in *Crocus* [112], and the different parts of the female organs of *C. sativus* differ in their flavonol content. No flavonoids have been detected in ovaries, and between stigma and the style, flavonoids are present at higher levels in the style than in the stigma [55]. In addition, expression analysis showed higher expression levels of UGT707B1 in the style than in the stigma and this distribution could have a direct implication in the control of the auxin gradient and thus in flower morphology.

The precise identification of flavonols in conjunction with expression analysis has therefore led to the identification of a gene function. The data in the present study indicate that UGT707B1 is an enzyme involved in the formation of flavonol sophorosides in *C. sativus*. Based on the accumulation of certain flavonols in the transgenic UGT707B1 *Arabidopsis* plants, we suggest that UGT707B1 has a flavonol-3-O-glucoside: 2''-glucosyltransferase activity.

References

1. Harborne J, Williams C: **Advances in flavonoid research since 1992**. *Phytochemistry* 2000, **55**(6):481-504.
2. Buer CS, Imin N, Djordjevic MA: **Flavonoids: new roles for old molecules**. *J Integr Plant Biol* 2010, **52**(1):98-111.
3. Anderson Ø, Markham K: **Flavonoids: chemistry, biochemistry, and applications**. Boca Raton: CRC Press,; 2006.
4. Grotewold E: **Plant metabolic diversity: a regulatory perspective**. *Trends Plant Sci* 2005, **10**(2):57-62.
5. Lepiniec L, Debeaujon I, Routaboul JM, Baudry A, Pourcel L, Nesi N, Caboche M: **Genetics and biochemistry of seed flavonoids**. *Annu Rev Plant Biol* 2006, **57**:405-430.
6. Taylor LP, Grotewold E: **Flavonoids as developmental regulators**. *Curr Opin Plant Biol* 2005, **8**(3):317-323.
7. Subramanian S, Stacey G, Yu O: **Distinct, crucial roles of flavonoids during legume nodulation**. *Trends Plant Sci* 2007, **12**(7):282-285.
8. Ross JA, Kasum CM: **Dietary flavonoids: bioavailability, metabolic effects, and safety**. *Annu Rev Nutr* 2002, **22**:19-34.
9. Bischoff SC: **Quercetin: potentials in the prevention and therapy of disease**. *Curr Opin Clin Nutr Metab Care* 2008, **11**(6):733-740.
10. Cazarolli LH, Zanatta L, Alberton EH, Figueiredo MS, Folador P, Damazio RG, Pizzolatti MG, Silva FR: **Flavonoids: prospective drug candidates**. *Mini Rev Med Chem* 2008, **8**(13):1429-1440.
11. Tanaka Y, Filippa B: **Flower color**. In: *Flowering and Its Manipulation.* Edited by Ainsworth C, vol. 20. London: Blackwell Publishers; 2006: 201–239.
12. Bowles D, Lim EK, Poppenberger B, Vaistij FE: **Glycosyltransferases of lipophilic small molecules**. *Annu Rev Plant Biol* 2006, **57**:567-597.
13. Li Y, Baldauf S, Lim EK, Bowles DJ: **Phylogenetic analysis of the UDP-glycosyltransferase multigene family of Arabidopsis thaliana**. *J Biol Chem* 2001, **276**(6):4338-4343.
14. Vogt T, Jones P: **Glycosyltransferases in plant natural product synthesis: characterization of a supergene family**. *Trends Plant Sci* 2000, **5**(9):380-386.
15. Yonekura-Sakakibara K, Tohge T, Niida R, Saito K: **Identification of a flavonol 7-O-rhamnosyltransferase gene determining flavonoid pattern in Arabidopsis by transcriptome coexpression analysis and reverse genetics**. *J Biol Chem* 2007, **282**(20):14932-14941.
16. Yonekura-Sakakibara K, Fukushima A, Nakabayashi R, Hanada K, Matsuda F, Sugawara S, Inoue E, Kuromori T, Ito T, Shinozaki K *et al*: **Two glycosyltransferases involved in anthocyanin modification delineated by transcriptome independent component analysis in Arabidopsis thaliana**. *Plant J* 2012, **69**(1):154-167.
17. Morita Y, Hoshino A, Kikuchi Y, Okuhara H, Ono E, Tanaka Y, Fukui Y, Saito N, Nitasaka E, Noguchi H *et al*: **Japanese morning glory dusky mutants displaying reddish-brown or purplish-gray flowers are deficient in a novel glycosylation enzyme for anthocyanin biosynthesis, UDP-glucose:anthocyanidin 3-O-glucoside-2''-O-glucosyltransferase, due to 4-bp insertions in the gene**. *Plant J* 2005, **42**(3):353-363.
18. Ono E, Ruike M, Iwashita T, Nomoto K, Fukui Y: **Co-pigmentation and flavonoid glycosyltransferases in blue Veronica persica flowers**. *Phytochemistry* 2010, **71**(7):726-735.
19. Sawada S, Suzuki H, Ichimaida F, Yamaguchi MA, Iwashita T, Fukui Y, Hemmi H, Nishino T, Nakayama T: **UDP-glucuronic acid:anthocyanin glucuronosyltransferase from red daisy (Bellis perennis) flowers. Enzymology and phylogenetics of a novel glucuronosyltransferase involved in flower pigment biosynthesis**. *J Biol Chem* 2005, **280**(2):899-906.
20. Frydman A, Weisshaus O, Bar-Peled M, Huhman DV, Sumner LW, Marin FR, Lewinsohn E, Fluhr R, Gressel J, Eyal Y: **Citrus fruit bitter flavors: isolation and functional characterization of the gene Cm1,2RhaT encoding a 1,2 rhamnosyltransferase, a key enzyme in the biosynthesis of the bitter flavonoids of citrus**. *Plant J* 2004, **40**(1):88-100.
21. Masada S, Terasaka K, Oguchi Y, Okazaki S, Mizushima T, Mizukami H: **Functional and structural characterization of a flavonoid glucoside 1,6-glucosyltransferase from *Catharanthus roseus***. *Plant Cell Physiol* 2009, **50**(8):1401-1415.
22. Montefiori M, Espley RV, Stevenson D, Cooney J, Datson PM, Saiz A, Atkinson RG, Hellens RP, Allan AC: **Identification and characterisation of F3GT1 and F3GGT1, two glycosyltransferases responsible for anthocyanin biosynthesis in red-fleshed kiwifruit (*Actinidia chinensis*)**. *Plant J* 2011, **65**(1):106-118.

23. Calderon-Montano JM, Burgos-Moron E, Perez-Guerrero C, Lopez-Lazaro M: **A review on the dietary flavonoid kaempferol**. *Mini Rev Med Chem* 2011, **11**(4):298-344.
24. Straubinger M, Jezussek M, Waibel R, Winterhalter P: **Two kaempferol sophorosides from Crocus sativus**. *Nat Prod Lett* 1997, **10**:4.
25. Tarantilis PA, Polissiou M: **Isolation and identification of the aroma constituents of saffron (*Crocus sativa*)**. *Journal of Agriculture and Food Chemistry* 1997, **45**:4.
26. Hosseinzadeh H, Younesi HM: **Antinociceptive and anti-inflammatory effects of *Crocus sativus* L. stigma and petal extracts in mice**. *BMC Pharmacology* 2002, **2**(7).
27. Nørbæk R, Brandt K, Nielsen J, Orgaard M, Jacobsen N: **Flower pigment composition of Crocus species and cultivars used for a chemotaxonomic investigation**. *Biochemical Systematics and Ecology* 2002, **30**:28.
28. Moraga AR, Mozos AT, Ahrazem O, Gomez-Gomez L: **Cloning and characterization of a glucosyltransferase from *Crocus sativus* stigmas involved in flavonoid glucosylation**. *BMC Plant Biol* 2009, **9**:109.
29. Termentzi A, Kokkalou E: **LC-DAD-MS (ESI+) analysis and antioxidant capacity of crocus sativus petal extracts**. *Planta Med* 2008, **74**(5):573-581.
30. Rubio A, Rambla JL, Santaella M, Gomez MD, Orzaez D, Granell A, Gomez-Gomez L: **Cytosolic and plastoglobule-targeted carotenoid dioxygenases from *Crocus sativus* are both involved in beta-ionone release**. *J Biol Chem* 2008, **283**(36):24816-24825.
31. Ross J, Li Y, Lim E, Bowles DJ: **Higher plant glycosyltransferases**. *Genome Biol* 2001, **2**(2):REVIEWS3004.
32. Higo K, Ugawa Y, Iwamoto M, Korenaga T: **Plant cis-acting regulatory DNA elements (PLACE) database: 1999**. *Nucleic Acids Res* 1999, **27**(1):297-300.
33. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S: **PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences**. *Nucleic Acids Res* 2002, **30**(1):325-327.
34. Moraga AR, Nohales PF, Perez JA, Gomez-Gomez L: **Glucosylation of the saffron apocarotenoid crocetin by a glucosyltransferase isolated from *Crocus sativus* stigmas**. *Planta* 2004, **219**(6):955-966.
35. Gomez-Gomez L, Felix G, Boller T: **A single locus determines sensitivity to bacterial flagellin in Arabidopsis thaliana**. *Plant J* 1999, **18**(3):277-284.
36. Karimi M, Inze D, Depicker A: **GATEWAY vectors for Agrobacterium-mediated plant transformation**. *Trends Plant Sci* 2002, **7**(5):193-195.
37. Clough SJ, Bent AF: **Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana***. *Plant J* 1998, **16**(6):735-743.
38. Stobiecki M, Skirycz A, Kerhoas L, Kachlicki P, Muth D, Einhorn J, Mueller-Roeber B: **Profiling of phenolic glycosidic conjugates in leaves of *Arabidopsis thaliana* using LC/MS**. *Metabolomics* 2006, **2**(4):197-219.
39. Cartwright AM, Lim EK, Kleanthous C, Bowles DJ: **A kinetic analysis of regiospecific glucosylation by two glycosyltransferases of *Arabidopsis thaliana*: domain swapping to introduce new activities**. *J Biol Chem* 2008, **283**(23):15724-15731.
40. Hughes J, Hughes MA: **Multiple secondary plant product UDP-glucose glucosyltransferase genes expressed in cassava (Manihot esculenta Crantz) cotyledons**. *DNA Seq* 1994, **5**(1):41-49.
41. Mackenzie PI, Owens IS, Burchell B, Bock KW, Bairoch A, Belanger A, Fournel-Gigleux S, Green M, Hum DW, Iyanagi T *et al*: **The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence**. *Pharmacogenetics* 1997, **7**(4):255-269.
42. He XZ, Wang X, Dixon RA: **Mutational analysis of the Medicago glycosyltransferase UGT71G1 reveals residues that control regioselectivity for (iso)flavonoid glycosylation**. *J Biol Chem* 2006, **281**(45):34441-34447.
43. Brazier-Hicks M, Offen WA, Gershater MC, Revett TJ, Lim EK, Bowles DJ, Davies GJ, Edwards R: **Characterization and engineering of the bifunctional N- and O-glucosyltransferase involved in xenobiotic metabolism in plants**. *Proc Natl Acad Sci U S A* 2007, **104**(51):20238-20243.
44. Shao H, He X, Achnine L, Blount JW, Dixon RA, Wang X: **Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from Medicago truncatula**. *Plant Cell* 2005, **17**(11):3141-3154.

45. Offen W, Martinez-Fleites C, Yang M, Kiat-Lim E, Davis BG, Tarling CA, Ford CM, Bowles DJ, Davies GJ: **Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification**. *EMBO J* 2006, **25**(6):1396-1405.
46. Noguchi A, Saito A, Homma Y, Nakao M, Sasaki N, Nishino T, Takahashi S, Nakayama T: **A UDP-glucose:isoflavone 7-O-glucosyltransferase from the roots of soybean (glycine max) seedlings. Purification, gene cloning, phylogenetics, and an implication for an alternative strategy of enzyme catalysis**. *J Biol Chem* 2007, **282**(32):23581-23590.
47. Yonekura-Sakakibara K, Hanada K: **An evolutionary view of functional diversity in family 1 glycosyltransferases**. *Plant J* 2011, **66**(1):182-193.
48. Priest DM, Ambrose SJ, Vaistij FE, Elias L, Higgins GS, Ross AR, Abrams SR, Bowles DJ: **Use of the glucosyltransferase UGT71B6 to disturb abscisic acid homeostasis in Arabidopsis thaliana**. *Plant J* 2006, **46**(3):492-502.
49. Lanot A, Hodge D, Jackson RG, George GL, Elias L, Lim EK, Vaistij FE, Bowles DJ: **The glucosyltransferase UGT72E2 is responsible for monolignol 4-O-glucoside production in Arabidopsis thaliana**. *Plant J* 2006, **48**(2):286-295.
50. Ko JH, Kim BG, Kim JH, Kim H, Lim CE, Lim J, Lee C, Lim Y, Ahn JH: **Four glucosyltransferases from rice: cDNA cloning, expression, and characterization**. *J Plant Physiol* 2008, **165**(4):435-444.
51. Hirotani M, Kuroda R, Suzuki H, Yoshikawa T: **Cloning and expression of UDP-glucose: flavonoid 7-O-glucosyltransferase from hairy root cultures of Scutellaria baicalensis**. *Planta* 2000, **210**(6):1006-1013.
52. Ono E, Nakayama T: **Molecular breeding of novel yellow flowers by engineering the aurone biosynthetic pathway**. *Transgenic Plant Journal* 2007, **1**:14.
53. Tian L, Blount JW, Dixon RA: **Phenylpropanoid glycosyltransferases from osage orange (Maclura pomifera) fruit**. *FEBS Lett* 2006, **580**(30):6915-6920.
54. Witte S, Moco S, Vervoort J, Matern U, Martens S: **Recombinant expression and functional characterisation of regiospecific flavonoid glucosyltransferases from Hieracium pilosella L**. *Planta* 2009, **229**(5):1135-1146.
55. Rubio-Moraga A, Trapero A, Ahrazem O, Gomez-Gomez L: **Crocins transport in Crocus sativus: the long road from a senescent stigma to a newborn corm**. *Phytochemistry* 2010, **71**(13):1506-1513.
56. Meng H, Duan C, Xiao F, Yang S, Zha Y, Wen G: **Molecular cloning and expression analysis of sucrose synthase gene from Dendrobium officinale**. *Zhongguo Zhong Yao Za Zhi* 2011, **36**(7):833-837.
57. Markham KR, Ternai B, Stanley R, Geiger H, Mabry TJ: **Carbon-13 NMR studies of flavonoids-III**. *Tetrahedron* 1978, **34**:9.
58. Kang BG: **Epinasty**. In: *Physiology of Movements.* Edited by Haupt W, Feinleib ME. Berlin: Springer-Verlag; 1979: 647-667.
59. Kerhoas L, Aouak D, Cingoz A, Routaboul JM, Lepiniec L, Einhorn J, Birlirakis N: **Structural characterization of the major flavonoid glycosides from Arabidopsis thaliana seeds**. *J Agric Food Chem* 2006, **54**(18):6603-6612.
60. Le Gall G, Metzdorff SB, Pedersen J, Bennett RN, Colquhoun IJ: **Metabolite profiling of Arabidopsis thaliana (L.) plants transformed with an antisense chalcone synthase gene**. *Metabolism* 2006, **1**:17.
61. Veit M, Pauli GF: **Major flavonoids from Arabidopsis thaliana leaves**. *J Nat Prod* 1999, **62**(9):1301-1303.
62. Yonekura-Sakakibara K, Tohge T, Matsuda F, Nakabayashi R, Takayama H, Niida R, Watanabe-Takahashi A, Inoue E, Saito K: **Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in Arabidopsis**. *Plant Cell* 2008, **20**(8):2160-2176.
63. Mabry TJ, Markham KR, Thomas MB: **The systematic identification of flavonoids** New York Springer-Verlag 1970.
64. Ferreres F, Llorach R, Gil-Izquierdo A: **Characterization of the interglycosidic linkage in di-, tri-, tetra- and pentaglycosylated flavonoids and differentiation of positional isomers by liquid chromatography/electrospray ionization tandem mass spectrometry**. *JOURNAL OF MASS SPECTROMETRY* 2004, **39**:10.
65. Stracke R, Jahns O, Keck M, Tohge T, Niehaus K, Fernie AR, Weisshaar B: **Analysis of production of flavonol glycosides-dependent flavonol glycoside accumulation in *Arabidopsis thaliana* plants reveals MYB11-, MYB12- and MYB111-independent flavonol glycoside accumulation**. *New Phytol* 2010, **188**(4):985-1000.

66. Tohge T, Nishiyama Y, Hirai MY, Yano M, Nakajima J, Awazuhara M, Inoue E, Takahashi H, Goodenowe DB, Kitayama M *et al*: **Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor**. *Plant J* 2005, **42**(2):218-235.
67. Hartmann U, Sagasser M, Mehrtens F, Stracke R, Weisshaar B: **Differential combinatorial interactions of cis-acting elements recognized by R2R3-MYB, BZIP, and BHLH factors control light-responsive and tissue-specific activation of phenylpropanoid biosynthesis genes**. *Plant Mol Biol* 2005, **57**(2):155-171.
68. Mehrtens F, Kranz H, Bednarek P, Weisshaar B: **The Arabidopsis transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis**. *Plant Physiol* 2005, **138**(2):1083-1096.
69. Grotewold E, Drummond BJ, Bowen B, Peterson T: **The myb-homologous P gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset**. *Cell* 1994, **76**(3):543-553.
70. Stracke R, Ishihara H, Huep G, Barsch A, Mehrtens F, Niehaus K, Weisshaar B: **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):660-677.
71. Gachon CM, Langlois-Meurinne M, Henry Y, Saindrenan P: **Transcriptional co-regulation of secondary metabolism enzymes in Arabidopsis: functional and evolutionary implications**. *Plant Mol Biol* 2005, **58**(2):229-245.
72. Osmani SA, Bak S, Moller BL: **Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling**. *Phytochemistry* 2009, **70**(3):325-347.
73. Shao H, He X, Achnine L, Blount J, Dixon R, Wang X: **Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula***. *Plant Cell* 2005, **17**:3141-3154.
74. Lim EK, Baldauf S, Li Y, Elias L, Worrall D, Spencer SP, Jackson RG, Taguchi G, Ross J, Bowles DJ: **Evolution of substrate recognition across a multigene family of glycosyltransferases in Arabidopsis**. *Glycobiology* 2003, **13**(3):139-145.
75. Achnine L, Huhman DV, Farag MA, Sumner LW, Blount JW, Dixon RA: **Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume Medicago truncatula**. *Plant J* 2005, **41**(6):875-887.
76. Lim EK, Doucet CJ, Li Y, Elias L, Worrall D, Spencer SP, Ross J, Bowles DJ: **The activity of Arabidopsis glycosyltransferases toward salicylic acid, 4-hydroxybenzoic acid, and other benzoates**. *J Biol Chem* 2002, **277**(1):586-592.
77. Noguchi A, Fukui Y, Iuchi-Okada A, Kakutani S, Satake H, Iwashita T, Nakao M, Umezawa T, Ono E: **Sequential glucosylation of a furofuran lignan, (+)-sesaminol, by *Sesamum indicum* UGT71A9 and UGT94D1 glucosyltransferases**. *Plant J* 2008, **54**(3):415-427.
78. Caputi L, Malnoy M, Goremykin V, Nikiforova S, Martens S: **A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land**. *Plant J* 2011.
79. Kovinich N, Saleem A, Arnason JT, Miki B: **Combined analysis of transcriptome and metabolite data reveals extensive differences between black and brown nearly-isogenic soybean (*Glycine max*) seed coats enabling the identification of pigment isogenes**. *BMC Genomics* 2011, **12**:381.
80. von Saint Paul V, Zhang W, Kanawati B, Geist B, Faus-Kessler T, Schmitt-Kopplin P, Schaffner AR: **The Arabidopsis glucosyltransferase UGT76B1 conjugates isoleucic acid and modulates plant defense and senescence**. *Plant Cell* 2011, **23**(11):4124-4145.
81. Richman A, Swanson A, Humphrey T, Chapman R, McGarvey B, Pocs R, Brandle J: **Functional genomics uncovers three glucosyltransferases involved in the synthesis of the major sweet glucosides of Stevia rebaudiana**. *Plant J* 2005, **41**(1):56-67.
82. Shibuya M, Nishimura K, Yasuyama N, Ebizuka Y: **Identification and characterization of glycosyltransferases involved in the biosynthesis of soyasaponin I in Glycine max**. *FEBS Lett* 2010, **584**(11):2258-2264.
83. Ringli C, Bigler L, Kuhn BM, Leiber RM, Diet A, Santelia D, Frey B, Pollmann S, Klein M: **The modified flavonol glycosylation profile in the Arabidopsis rol1 mutants results in alterations in plant growth and cell shape formation**. *Plant Cell* 2008, **20**(6):1470-1481.
84. Kuhn BM, Geisler M, Bigler L, Ringli C: **Flavonols accumulate asymmetrically and affect auxin transport in Arabidopsis**. *Plant Physiol* 2011, **156**(2):585-595.

85. Walker AR, Davison PA, Bolognesi-Winfield AC, James CM, Srinivasan N, Blundell TL, Esch JJ, Marks MD, Gray JC: **The transparent testa glabra1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in Arabidopsis, encodes a WD40 repeat protein**. *Plant Cell* 1999, **11**(7):1337-1350.
86. Gou JY, Felippes FF, Liu CJ, Weigel D, Wang JW: **Negative regulation of anthocyanin biosynthesis in Arabidopsis by a miR156-targeted SPL transcription factor**. *Plant Cell* 2011, **23**(4):1512-1522.
87. Wang B, Jin SH, Hu HQ, Sun YG, Wang YW, Han P, Hou BK: **UGT87A2, an Arabidopsis glycosyltransferase, regulates flowering time via flowering locus c**. *New Phytol* 2012.
88. Stangeland B, Rosenhave EM, Winge P, Berg A, Amundsen SS, Karabeg M, Mandal A, Bones AM, Grini PE, Aalen RB: **AtMBD8 is involved in control of flowering time in the C24 ecotype of *Arabidopsis thaliana***. *Physiol Plant* 2009, **136**(1):110-126.
89. Mahajan M, Ahuja PS, Yadav SK: **Post-transcriptional silencing of flavonol synthase mRNA in tobacco leads to fruits with arrested seed set**. *PLoS One* 2011, **6**(12):e28315.
90. Jacobs M, Rubery PH: **Naturally occurring auxin transport regulators**. *Science* 1988, **241**(4863):346-349.
91. Brown DE, Rashotte AM, Murphy AS, Normanly J, Tague BW, Peer WA, Taiz L, Muday GK: **Flavonoids act as negative regulators of auxin transport in vivo in arabidopsis**. *Plant Physiol* 2001, **126**(2):524-535.
92. Lewis DR, Ramirez MV, Miller ND, Vallabhaneni P, Ray WK, Helm RF, Winkel BS, Muday GK: **Auxin and ethylene induce flavonol accumulation through distinct transcriptional networks**. *Plant Physiol* 2011, **156**(1):144-164.
93. Fleck B, Harberd NP: **Evidence that the Arabidopsis nuclear gibberellin signalling protein GAI is not destabilised by gibberellin**. *Plant J* 2002, **32**(6):935-947.
94. Keminer O, Peters R: **Permeability of single nuclear pores**. *Biophys J* 1999, **77**(1):217-228.
95. Burbulis IE, Winkel-Shirley B: **Interactions among enzymes of the Arabidopsis flavonoid biosynthetic pathway**. *Proc Natl Acad Sci U S A* 1999, **96**(22):12929-12934.
96. Buer CS, Muday GK: **The transparent testa4 mutation prevents flavonoid synthesis and alters auxin transport and the response of Arabidopsis roots to gravity and light**. *Plant Cell* 2004, **16**(5):1191-1205.
97. Feucht W, Dithmar H, Polster J: **Nuclei of tea flowers as targets for flavanols**. *Plant Biol (Stuttg)* 2004, **6**(6):696-701.
98. Feucht W, Treutter D, Polster J: **Flavanol binding of nuclei from tree species**. *Plant Cell Rep* 2004, **22**(6):430-436.
99. Saslowsky DE, Warek U, Winkel BS: **Nuclear localization of flavonoid enzymes in Arabidopsis**. *J Biol Chem* 2005, **280**(25):23735-23740.
100. Naoumkina MA, He X, Dixon RA: **Elicitor-induced transcription factors for metabolic reprogramming of secondary metabolism in *Medicago truncatula***. *BMC Plant Biol* 2008, **8**:132.
101. Gilbert ER, Liu D: **Flavonoids influence epigenetic-modifying enzyme activity: structure - function relationships and the therapeutic potential for cancer**. *Curr Med Chem* 2010, **17**(17):1756-1768.
102. Yin R, Messner B, Faus-Kessler T, Hoffmann T, Schwab W, Hajirezaei MR, von Saint Paul V, Heller W, Schaffner AR: **Feedback inhibition of the general phenylpropanoid and flavonol biosynthetic pathways upon a compromised flavonol-3-O-glycosylation**. *J Exp Bot* 2012.
103. Naoumkina M, Farag MA, Sumner LW, Tang Y, Liu CJ, Dixon RA: **Different mechanisms for phytoalexin induction by pathogen and wound signals in *Medicago truncatula***. *Proc Natl Acad Sci U S A* 2007, **104**(46):17909-17915.
104. Yu XH, Chen MH, Liu CJ: **Nucleocytoplasmic-localized acyltransferases catalyze the malonylation of 7-O-glycosidic (iso)flavones in *Medicago truncatula***. *Plant J* 2008, **55**(3):382-396.
105. Nishihara M, Nakatsuka T: **Genetic engineering of flavonoid pigments to modify flower color in floricultural plants**. *Biotechnology Letters* 2011, **33**(3):9.
106. Mol J, Cornish E, Mason J, Koes R: **Novel coloured flowers**. *Curr Opin Biotechnol* 1999, **10**(2):198-201.
107. Beliaeva RG, Evdokimova LI: **Variability of flavonol contents during floral morphogenesis in *Papaver somniferum* L**. *Ontogenez* 2004, **35**(1):16-22.
108. Mo Y, Nagel C, Taylor LP: **Biochemical complementation of chalcone synthase mutants defines a role for flavonols in functional pollen**. *Proc Natl Acad Sci U S A* 1992, **89**(15):7213-7217.

109. Staldal V, Sundberg E: **The role of auxin in style development and apical-basal patterning of the Arabidopsis thaliana gynoecium**. *Plant Signal Behav* 2009, **4**(2):83-85.
110. Okada K, Ueda J, Komaki MK, Bell CJ, Shimura Y: **Requirement of the Auxin Polar Transport System in Early Stages of Arabidopsis Floral Bud Formation**. *Plant Cell* 1991, **3**(7):677-684.
111. Grilli M, Canini A: **Ultrastructure of chromoplast and other plastids in *Crocus sativus* L. (Iridaceae)**. *Plant Biosystems* 2004, **138**(1):10.
112. Stark D: **anatomical and physiological studies of floral tube elongation of *Crocus vernus* (Iridaceae)**. *American Journal of Botany* 1982, **69**(9):7.

Chapter 5

“*In silico* identification of new β-glycosidases and glycosyltransferases present in different stages of stigmas of *Crocus sativus*”

Abstract

Plants produce a high diversity of secondary metabolites that play important biological roles in their environments. However, the biosynthetic pathways of several plant secondary metabolites and their regulation are still poorly understood. Saffron, the desiccated stigma of C*rocus sativus,* which is rich in different forms of apocarotenoids and flavonoids, is a good candidate for expanding knowledge and understanding about how these compounds are generated. The high diversity of metabolites present in this plant tissue is due to secondary modifications which are produced by different classes of enzymes including glycosyltransferases (GTs) and β-glycosidases (GHs) that play an important role in these processes.

Expressed sequences tag (ESTs) collections have been used to identify GTs and GHs involved in the modification of apocarotenoids and flavonoids present in the stigma. ESTs were obtained by sequencing clones from six libraries created from six different developmental stages of stigma. The ESTs corresponding to GTs and GHs were assembled in order to determine the number of different GTs and GHs, to increase sequence length and to obtain information on the expression profile of the genes coding for these enzymes. A total of 63 different sequences were obtained and identified as putative GTs and GHs proteins; 43 and 20 sequences encoded GT-proteins and GH-proteins, respectively. In addition, we tentatively described their putative functions based on similarity to other described plant proteins and we also look into their corresponding level of expression in each stigma developmental stage.

Keywords: apocarotenoids, β-glycosidases, *Crocus sativus,* ESTs, flavonoids, glycosyltransferases, stigmas

Introduction

Plants produce thousands of different compounds as a reaction to environmental adaptation and also by the interaction with other organisms. All these compounds belong to the secondary metabolism of plants. Based on their biosynthetic origins, secondary metabolites are divided into three main groups: terpenoids, alkaloids and phenylpropanoids. Terpenoids also called "isoprenoids" are built from five-carbon isoprene units (IPP). Alkaloids are a group of nitrogen-containing bases, most of which are drugs, and phenylpropanoids are a large family of plant compounds including anthocyanin, proanthocyanidin and phlobaphene pigments, as well as flavonol, flavone and isoflavone. The high number and diversity of secondary metabolites observed in nature is mainly due to secondary modifications on the basic skeleton of these compounds. One of these secondary modifications is glycosylation. Glycosylation is the transfer of a single or multiple sugar units from activated sugars to an aglycon to form glycosidic bonds. This reaction is catalyzed by glycosyltransferases (GT) [1], which are active over a wide range of substrates. These enzymes are very attractive for many reasons such as, their potential role in developmental and metabolic homeostasis, and their function in detoxification processes.

Other important modification in the secondary metabolites is catalysed by glycosyl hydrolases (GH), which are involved in the hydrolysis of the glycosidic bond between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety. The β-glycosidases are one of the most important families inside of this huge group. These enzymes play important roles in numerous biological process, such as biosynthesis of phytohormones [2], lignification [3], cell wall degradation [4], and aroma generation [5], etc.

The two major secondary metabolites found in saffron are apocarotenoids and flavonoids [6]; [7]. The main apocarotenoids are cis- and trans-crocins, picrocrocin (β-D-glucopyranoside of hydroxyl-β-cyclocitral), and its degradation product, the odor-active safranal. The principal flavonoids are different glycosylated forms of kaempferol. In spite of that the main pathways for the biosynthesis of carotenoids and flavonoids in plants have been elucidated [8], much less is known about the secondary modifications that suffer these metabolites. Therefore, the analysis of the transcriptome of *C. sativus* stigmas can help to gain insights on these modification reactions.

Nowadays, there are several strategies to identify and to analyze genes that encode interesting enzymes, such as GTs and GHs. One of these strategies is the analysis of the expressed sequence tags (ESTs) through bioinformatics tools. The use of these ESTs allows the elucidation, in several cases, of the nucleotide sequence of plant genes. ESTs are determined by single-pass sequencing of the 5’ or 3’ ends of cDNA clones. The cDNA libraries are generated from tissue, developmental stage or specific mRNA samples. Therefore, ESTs are a powerful strategy for gene identification, gene expression studies and identification of important genetic variants such as single nucleotide polymorphisms (SNP). ESTs have been used to identify GTs involved in the glycosylation of secondary metabolites in *Stevia rebaudina* leafs [9], in the glucosylation of flavonoids in *Maclura pomifera* fruits [10], to identify and to characterize GTs with *in vitro* activity against isoflavones, flavones and flavonols in *Medicago truncculata* [11], to identify a UDP-Glc: cinnamate glucosyltransferase involved in the biosynthesis of cinnamic acid, a precursor of some metabolites responsible for flavor in strawberries [12], to a GT involved in the biosynthesis of phloridzin, the most important phenolic compounds in apple [13], and 6 GTs involved in glycyrrhizin biosynthesis in *Glycyrrhiza uralensis*, one of the most popular medicinal plants in the world which is used in food and tobacco flavoring [14]. Recently, 454 pyrosequencing technology have been used as first step for isolation and characterization of GTs and GHs in *Consolida orientalis* from a ESTs library, which are involved in the benzoxazinoid biosynthesis, one of the multiple metabolites secondary involved in the defense against herbivores and pathogens attack [15].

An important database for *C. sativus* has been designed recently to manage and explore ESTs from saffron stigma cDNA library [16]. The collection contains more than 6600 ESTs which have been grouped into 1893 clusters. The collection is accessible through the Saffron Genes database at http://www.saffrongenes.org. To get more insights about the expressed genes in saffron stigma, the group of Green Biotechnology laboratory (ENEA, Italy) has produced new collections of ESTs from six developmental stages of *C. sativus* stigma: three immature stages and three mature stages. The immature stages correspond to yellow, orange and red stigma while the mature stages correspond to two days before anthesis, anthesis and two days after anthesis (Fig. 1).

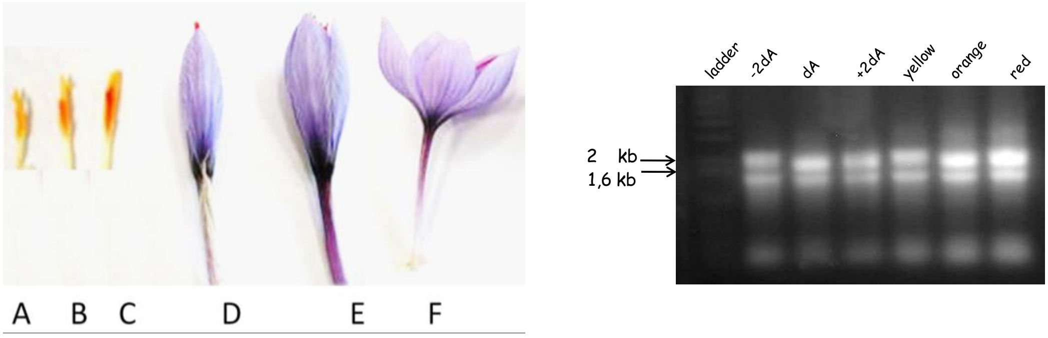


Figure 1. **Developmental stages of *C. sativus* from which ESTs libraries were constructed and total RNA extracted from each developmental stage**. A, yellow stigma; B, orange stigma; C, red stigma; D, two days before anthesis; E, anthesis; F, two days after anthesis stigma.

The ESTs abundance in each contig can be indicative of the relative mRNA abundance in the stigma tissue. Such information could be useful to determine the expression levels of the transcripts in each developmental stage of the stigma in *C. sativus*, and also the relationship between the expression of each transcript and the level of glycosylated metabolites in stigma tissues. In this context, the principal aim of this work is to search *in silico* for new glycosyltransferases (GTs) and β-glycosidases (GHs) from *C. sativus* stigma using the above mentioned ESTs collections and to determine their expression profiles during the different stages of development.

Material and Methods

ESTs collection of *C. sativus* stigmas

ESTs collections were constructed from cDNA libraries obtained from different developmental stages of saffron stigma. The developmental stages of stigmas were defined as follow: yellow, closed bud inside the perianth tubes (around 0.3 cm in length); orange, closed bud inside the perianth tubes (around 0.4 cm in length); red, closed bud inside the perianth tubes (0.8 cm in length); -2da, two days before anthesis, dark red stigma in closed bud outside the perianth tubes (3 cm in length); da, day of anthesis, dark red stigma (3 cm in length); +2da, two days after anthesis, dark red stigma [17].

Total RNA was extracted as previously described [6] and each cDNA library was built as reported by D'Agostino 2007 [16]. Briefly, cDNA was prepared using the SMART PCR cDNA synthesis kit. The library was then directly sequenced using 454 pyrosequencing FLX system (454 Life Sciences, Roche Company). The reads were assembled using Newbler assembler software (454 Life Sciences, Roche Company), which generated a tentative consensus sequence (TCs) and single sequence (singletons). Then, in order to assign a preliminary function, the TCs and singletons were compared using BLASTX to the UniProtKB/Swiss-Prot database.

***In silico* search and identification of new glycosyltransferase and β-glycosidase sequences**

In order to identify putative glycosyltransferases present in the different ESTs libraries, we blasted against all the libraries sequences of saffron glycosyltransferases from published works CsGLN2 (Accession Number AY262037), CsGLN3 (Accession Number AY290820) and CsGT45 (Accession Number FJ194947) [18]; [7] and sequences isolated from our laboratory using degenerate primers (CsGLN10, CsGLN21, CsGLN122, CsGLN19, CsGLN1DPA, CsGLN70 and CsGLN27). On the other hand, partial sequences obtained in our laboratory using degenerated primers were used to identify β-glycosidases. The sequences used for the queries are available in annex 1.

An automatic list of putative contigs and singletons was obtained. The contig was defined as unigene or gene index sequence which are composed of the consensus of an alignment of two or more EST sequences; and singleton as sequence which has been determined not to overlap sufficiently with any other sequence in the input data set. The selected sequences were subject to a second assembling procedure using CondeCore Aligner software [19] and the generated consensus sequences were compared using BLASTX to dbNCBI. A manual checking was done to distinguish between GT- and GH-sequences.

To determine the relative abundance of transcripts, the number of ESTs in each contig was used as register of expression level in each developmental stage. The plots were drawn using Excel Office 2010.

Results

In silico identification of glycosyltransferases and β-glycosidases

ESTs collections from ENEA Green Biotechnology laboratory are currently made by 65000 ESTs. Blasting the sequences mentioned in materials and methods against the ESTs libraries, we obtained 3897 ESTs encoding putative glycosyltransferases and β-glycosidases, representing 6% of total ESTs. All GT and GH ESTs were subjected to a clustering/assembling procedure. This procedure generated 45 contigs or tentative consensus sequences, formed by 3791 ESTs, and non-grouped ESTs yielding 106 unique (non-overlapping) sequences or singletons (Fig 2).

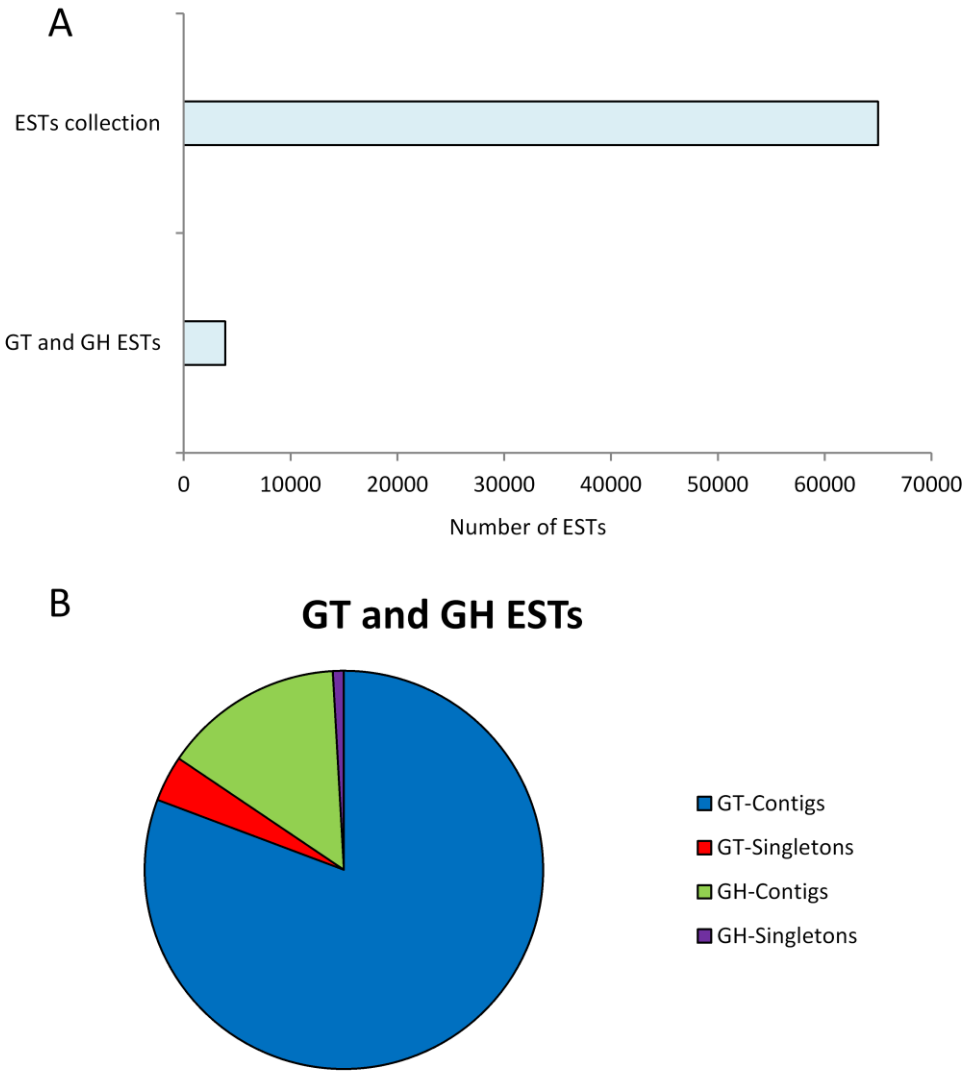


Figure 2. A. Representation of GT and GH ESTs versus overall ESTs collections. The GT and GH ESTs represented 6% of total ESTs **B. Distribution of GT and GH ESTs**: blue color (81%) shows ESTs grouped in contigs that encoded proteins similar to glycosyltransferase; red color (3%) represents ESTs non grouped (singleton) that encoded proteins similar to glycosyltransferase; green color (14%) shows

ESTs grouped in contigs that encoded proteins similar to β-glycosidases; purple (0.3%) represents ESTs non grouped (singleton) that encoded proteins similar to β-glycosidases.

A second assembling procedure was done to check the real number of different sequences obtained. For this purpose, the GT and GH sequences from ESTs collection were assembled using CondeCore Aligner software, a total of 40 contigs were obtained by the second assembling, 15 of them were new contigs corresponding to 37 sequences which have been considered as single sequences (singletons) after the first assembling. Other 13 contigs correspond to 67 sequences from which 21 contigs and 46 singletons were obtained by the first assembling and the remaining 12 contigs coincide with 12 contigs from the first assembling. The number of singletons was reduced to 23 sequences instead of the 106 unassembled sequences found after the first assembling and the lengths of these sequences have increased considerably (Fig. 2B).

BlastX approach was used in order to maximize the output and to identify genes with different nucleotides but similar amino acidic sequence. Each contig and singleton was compared to protein database from the National Center for Biotechnology Information (NCBI). Then, a manual check was performed in order to distinguish between putative β-glycosidases and glycosyltransferases.

Table 1 shows a total of 43 sequences of putative GTs including contigs and singletons. Among them, 24 sequences seemed to be located on the positive strand while 19 sequences were located on the minus one. The lengths of ESTs were ranged between 40 and 515 aminoacids while the sequence coverage of these ESTs matching to homologues proteins were between 28% and 89%. Table 1 shows also the number of ESTs obtained for each sequence.

From the 25 GT-contigs built, the contig3 was the highest in number of ESTs while the contigs 20, 22, 23, 24 and 28 were formed by only 2 ESTs each one. After checking the GT-sequences, 17 contigs that encoded partial glycosyltransferase were identified as new possible GTs from stigma (Table 1). The contig29 encoded a protein similar to UDP-glycosyltransferase of *Linum usitatissimum* with 48% of identity [20]. The contig31 encoded protein sharing 98% of identity with crocetin-glycosyltransferase 2 from *C. sativus*; this enzyme is involved in the synthesis of crocin, one of the apocarotenoids responsible for the color and bitter taste of saffron [18]. The contig36 encoded a fragment similar to putative glucosyltransferase which is involved in the biosynthesis of zeatin (the most active and ubiquitous cytokinin) in seed from *Phaseo luslunatus* [21].

When the GT-contigs were checked, 8 contigs were marked out because they have been already isolated in our laboratory using degenerate primers. These contigs were contig 19, 12, 11, 10, 8, 5, 3 and 1 (Table 1). The contig1 is grouped by 14 sequences encoding CsGLN27 (UGT85V1). UGT85V1 is present in group G of plant secondary glycosyltransferases, where *A. thaliana* GTs glycosylated terpenoids [22], trans-zeatin [23] and cyanogenic compounds [24]; [25]. Their biochemical characterization is in progress. The contig3 is grouped by 11 sequences encoding CsGLN122 (UGT85U1) and CsGLN10 (UGT85U2) glycosyltransferase in *C. sativus* stigma. Both enzymes shared 98% of identity each other, which probably could be allelic sequences. UGT85U1 and UGT85U2 are present in group G, same as UGT85V1, which explain the broad spectrum activity described in this group [22]; [23]; [24]; [25]. The biochemical characterization of UGT85U1 and UGT85U2 is in progress. The contig5 is clustered by 9 unigenes that encodes already a described crocetin-glucosyltransferase 2 in *C. sativus*. The CsGLN2 (Accession Number AY262037) [18] is involved in the biosynthesis and accumulation of apocarotenoids. The CsGLN2 was cloned using degenerate primers and the recombinant protein showed activity over crocetin β-D-glucosyl and crocetin β-D-gentibiosyl *in vitro*. The expression of CsGLN2 was mainly expressed in fully developmental stigma of saffron. CsGLN2 showed glucosylation activity against crocetin, crocetin β-D-glucosyl ester and crocetin β-D-gentibiosyl ester, and is considered as one of the glucosyltransferases responsible for the glucosylation of these apocarotenoids in the stigma of *C. sativus* [18].

The contig8 was identified as CsGLN19 (UGT703C1). The complete DNA sequence has been isolated in our laboratory, but its activity remains unknown. Contig10 was made by 5 unigenes that encodes a glycosyltransferase isolated in our laboratory, GLN5 (UGT703B1), which enzymatic activity is in progress. The contig11 encodes an already described glycosyltransferase CsGT45 (UGT75P1) (Accession Number FJ194947) from *C. sativus* stigma [7]. *In vitro* analyses showed that CsGT45 involved in flavonoid glucosylation, with different specificities. As described in chapter 1, CsGT45 was expressed in stigmas and stamen and found to be more active in anthesis. The contig12 is grouped by 4 sequences encoding CsGLN70 (UGT75P2), enzyme also isolated in our laboratory. UGT75P2 share 98% of identity with UGT75P1, thus this enzyme could also be involved in flavonoid glucosylation. Even so the UGT75P2 biochemical characterization is in progress. The contig19 encoded a 1DPA-like protein (UGT86B2). UGT86B2 belongs to group K of plant secondary glycosyltransferases, which is characterized by a low number of genes in all sequenced plant genomes: 2 in *A. thaliana*, *Vitis vinifera*, *Populus trichocarpa* and *Glycine max*, and only one in *Cucumis sativus, Zea mays, O. sativa* and *S. bicolor* [22], and none of the putative enzymes from this group within those identified in the other plants has been characterized functionally.

Regarding the singletons, we found 18 unassembled sequences that encoded partial glycosyltransferases (Table 1). The singleton 9 encoded a protein sharing 61% of identity with flavonoid glucosyltransferase from *Allium cepa*; this protein is involved in the glycosylation of flavonoids mono and di-glucosides in the epidemical layer of yellow onion [26]. The singleton 4 encoded proteins sharing 67% with crocetin-glycosyltransferase 2 from *C. sativus*; this enzyme is involved in the synthesis of crocin, one of the apocarotenoids responsible for the color and bitter taste of saffron [18]. The singleton 8 encoded a fragment similar to UDP-glucosyltransferase of *Gardenia jasminoides* with 67% of identity; this protein is involved in the glycosylation of geniposide, a major iridoid compound in *G. jasminoides* [27]. The singleton 17 is a fragment of 410 bp of UGT707B1, an enzyme characterized in chapter 4, which is involved in the formation of flavonol sophorosides.

Table 1. Identification of contigs and singletons corresponding to glycosyltranferases.

Assembled	GenBank protein annotation	GenBank protein ID	Identity (%)	Length AA	Sequence coverage (%)	ESTs	Strand
Contig1	UDP-glycosyltransferase 85A1 isoform 1 *Vitis vinifera*	XP_002285771.1	50	435	66	375	+
Contig3	UDP-glycosyltransferase 85A1 isoform 1 *Vitis vinifera*	XP_002285771.1	83	476	75	1574	-
Contig5	Crocetin glucosyltransferase 2 *Crocus sativus*	Q6X1C0	88	395	83	973	-
Contig8	UDP-glucosyltransferase putative *Ricinus communis*	XP_002518718	41	465	80	39	-
Contig10	UDP-glucosyltransferase putative *Ricinus communis*	XP_002518724.1	49	91	84	36	-
Contig11	Flavonoid glucosyltransferase *Crocus sativus*	ACM66950	69	481	80	3	-
Contig12	UDP-glucuronosyltransferase putative *Ricinus communis*	XP_002530117	63	427	54	25	-
Contig13	UDP-glycosyltransferase 85A1 *Vitis vinifera*.	XP_002268487	62	76	67	4	+
Contig14	UDP-glycosyltransferase 85A1-like *Vitis vinifera*	XP_002276823.1	47	515	76	83	+
Contig15	UDP-glycosyltransferase 87A2-like *Vitis vinifera*	XP_002274420.2	79	91	70	3	-
Contig19	UDP-glycosyltransferase 86A1-like isoform 1 *Brachypodium distachyon*	XP_003573976.1	65	268	81	33	+
Contig22	UDP-glucosyltransferase putative *Ricinus communis*	XP_002532392.1	42	173	86	2	+
Contig23	UDP-glycosyltransferase 73B3-like *Glycine max*	XP_003520068.1	33	52	71	2	+
Contig24	UDP-glucosyltransferasefamily 1 protein *Medicago truncatula*	XP_003625715.1	51	40	80	2	+
Contig28	Flavonoid glucosyltransferase *Crocus sativus*	ACM66950	67	48	28	2	-
Contig29	UDP-glycosyltransferase 1 *Linum usitatissimum*	AFJ52994.1	48	68	32	15	+
Contig30	UDP-glycosyltransferase 85A2-like *Vitis vinifera*	XP_002285780	60	137	89	11	+
Contig31	Crocetin glucosyltransferase 3 *Crocus sativus*	Q6WFW1.1	98	155	29	38	+
Contig33	Anthocyanidin 3-O-glucosyltransferase *Zea mays*	NP_001147674	51	104	81	6	+
Contig34	Putative glucosyltransferase *Phaseo luslunatus*	AAM09517	61	133	75	15	-
Contig36	UDP-glycosyltransferase 85A2-like *Vitis vinifera*	XP_002285780	58	82	54	4	+
Contig37	UDP-glycosyltransferase 91C1-like *Brachypodium distachyon*	XP_003560664.1	55	154	77	6	+
Contig38	UDP-glycosyltransferase 85A1 *Vitis vinifera*	XP_002268487.2	58	79	54	15	+

Contig39	UDP-glycosyltransferase 1 *Linum usitatissimum*	AFJ52946.1	51	125	70	12	-
Contig40	Glycosyltransferase UGT88A8 *Hieracium pilosella*	ACB56924	71	64	92	6	+
Singleton1	Glycosyltransferase GT14A05 *Pueraria montana*	ADV71369.1	49	90	25	1	-
Singleton2	UDP-glucuronosyltransferase putative *Ricinus communis*	XP_002513937.1	73	59	93	1	+
Singleton3	Flavonol 3-o-glucosyltransferase partial *Arachishy pogaea*	AAL01623.1	73	43	45	1	+
Singleton4	Crocetin glucosyltransferase 3 *Crocus sativus*	Q6WFW1.1	67	82	99	1	-
Singleton5	Anthocyanin 7-O-glucosyltransferase *Agapanthus africanus*	BAM29304	68	43	46	1	+
Singleton6	Isoflavonoid glucosyltransferase *Glycyrrhiza echinata*	BAC78438.1	40	140	99	1	-
Singleton7	Anthocyanin 3'-O-beta-glucosyltransferase-like *Glycine max*	XP_003536714.1	49	92	76	1	-
Singleton8	UDP-glucoseglucosyltransferase *Gardenia jasminoides*	BAK55744	67	33	56	1	-
Singleton9	Flavonoid glucosyl-transferase *Allium cepa*	AAP88406	66	94	61	1	-
Singleton10	Anthocyanin 3'-O-beta-glucosyltransferase *Medicago truncatula*	XP_003637924	46	103	56	1	-
Singleton11	UDP-glycosyltransferase 85A3-like *Glycine max*	XP_003517688.1	70	85	58	1	+
Singleton12	UDP-glycosyltransferase 85A1 isoform 1 *Vitis vinifera*	XP_002285767	63	120	91	1	+
Singleton13	UDP-glycosyltransferase *Avena strigosa*	ACD03249	73	65	68	1	+
Singleton14	Indole-3-acetate beta-glucosyltransferase *Zea mays*	DAA57209	49	118	89	1	+
Singleton15	UDP-glucoseglucosyltransferase *Gardenia jasminoides*	BAK55747	68	23	95	1	-
Singleton16	UDP-glycosyltransferase 87A2-like *Vitis vinifera*	XP_002278049	52	97	73	1	+
Singleton17	Glucosyltransferase *Crocus sativus*	CCG85331.1	98	116	99	1	+
Singleton18	UDP-glucosyltransferase 74 F2 *Vitis vinifiera*	XP_002278455	60	46	63	1	-

+ translated in 5’3’ sense; - translated in 3’5’ sense.

We finally selected the longest contigs in order to compare them to each other by making a global alignment of their corresponding proteins using ClustalW (Fig. 3).

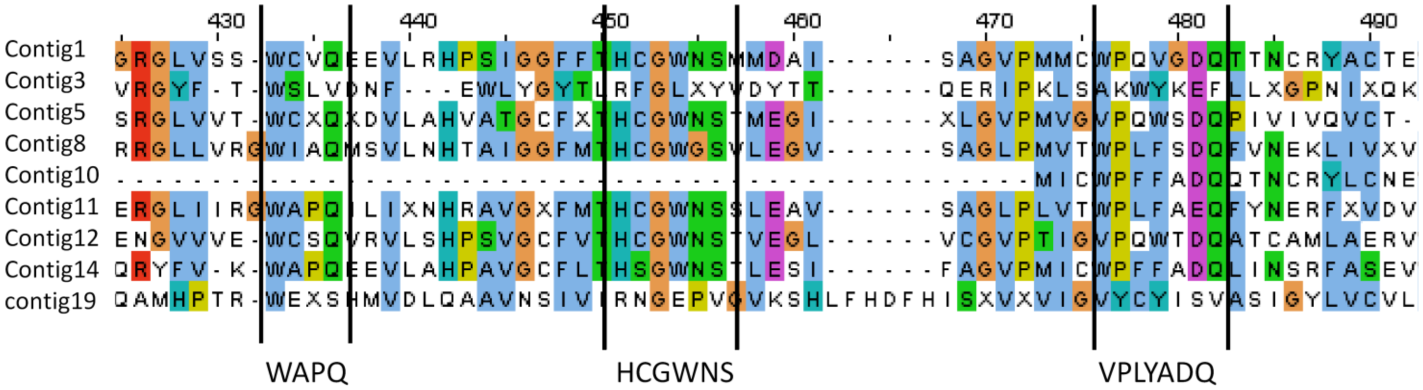


Figure 3. ClustalW alignments of deduced GT proteins. PSPG motif is formed by WAPQ, HCGWNS, and WPLYADQ residues.

All GT enzymes have a typical PSPG box involved in the catalytic activity found in glycosyltransferases. The PSPG box is located in the C-terminal and included conserved key residues that activate the hydroxy group of the glycosyl acceptor molecule to facilitate glycosydic linkage formation. Taking into consideration the high level of similarity with other GTs and the presence of the PSPG box, we can affirm that contig1, 3, 5, 8, 10, 11, 12, 14 and 19 encode active GTs in *C. sativus* stigma.

Regarding β-glycosidases, as far as we know no glycosidase has been described for *C. sativus* only small fragments have been isolated in our laboratory, using degenerated primers based on conserved motifs from orthologues genes present in other plants, such as *Hordeum vulgare* and *Orzya sativa*. Table 2 shows 20 sequences that encoded putative GHs, which 14 of them being located on the positive strand while 6 sequences were located on the minus one. Table 2 described also the number of ESTs obtained for each contig and singleton. 15 contigs encoding putative GHs were found. Contig2 was the highest contig in number with 205 ESTs, following by contig4 with 116 and contig6 with 103 ESTs, while the contig21, 25, 26 and 27 had only 2 ESTs each one. The lengths of ESTs were ranged between 40 and 510 amino acids while the sequence coverage of these ESTs matching to homologues proteins were between 58% and 90%.

After the second assembling procedure, most partial isolated GH-sequences were grouped into only an additional contig: contig4. This contig encodes a protein similar to β-glucosidase 24-like from *Brachypodium distachyon* (Accession Number XP_003563902.1) with 69% of identity. This enzyme belongs to family GH1, and share also some similarities to a β-glucosidase involve in the hydrolysis of hydroxynitrile glucosides in aerial parts of *Lotus japonicas* [28]. The contig7 encodes a protein which is similar to β-glucosidase 10 from *Brachypodium distachyon* (Accession Number XP_003566492). The contig2 encodes a protein sharing 71% of identity with a hypothetical protein from *Sorghum bicolor* (SORBIDRAFT_06g022510). This enzyme belongs to the GH1 family. The contig35 encodes a β-glucosidase similar to that described for *Arabidopsis thaliana* (Accession Number BAD94684) with 75% of identity and contig32 which encoded a protein similar to β-mannosidase 4 isolated from *Oncidium* Gower Ramsey sharing 65% of identity with β-glucosidase involved in lignification by hydrolysing monolignol glucosides in *Arabidopsis thaliana* [29]. The encoded protein is also similar to β-mannosidase (GH1 family) involved in the carbohydrate modulation during the flowering in *Oncidium orchid* [30]. The others GH-contigs: contig9, 16, 18, 21, 25, 26 and 27, are formed by small fragments of putative β-glucosidase sequences (Table 2). Only 5 sequences were found to be singletons encoding small fragments similar to β-glucosidases, e.g singleton22, encodes a β-glucosidase similar to that described for *Medicago trunculata* with 76% of identity or singleton23 which encoded a protein similar to β-glucosidase involved in lignification of *Pinus contorta* [31](Table 2).

Table 2: Identification of different assembled unigenes corresponding to glucosidases

Assembled	GenBank protein annotation	GenBank protein ID	Identity (%)	Length AA	Sequence coverage (%)	ESTs	Strand
Contig2	hypothetical protein SORBIDRAFT_06g022510 *Sorghum bicolor*	EES12507	71	510	70	205	+
Contig4	β-glucosidase 24-like *Brachypodium distachyon*	XP_003563902	69	466	78	116	+
Contig6	β-glucosidase 14-like *Brachypodium distachyon*	XP_003581423	64	416	67	103	-
Contig7	β-glucosidase 10-like isoform *2 Brachypodium distachyon*	XP_003566492	70	246	63	84	-
Contig9	β-glucosidase 4-like *Brachypodium distachyon*	XP_003564837	69	128	96	5	+
Contig16	β-mannosidase 1 *Oncidium* Gower Ramsey	ABC55718.1	79	160	99	3	-
Contig17	Os04g0474500 *Oryza sativa* Japonica Group	NP_001053067.1	58	89	62	3	+
Contig18	β-glucosidase 22-like *Brachypodium distachyon*	XP_003566327	68	553	81	50	+
Contig20	Os01g0897600 *Oryza sativa* Japonica Group	NP_001045089.2	61	75	34	2	+
Contig21	β -glucosidase 41-like *Vitis vinifera*	XP_002280323.2	69	20	63	2	+
Contig25	β-mannosidase 1 *Oncidium* Gower Ramsey	ABC55718.1	87	90	34	2	-
Contig26	non-cyanogenic β-glucosidase precursor partial *Oryza sativa*	ABR26210	74	48	42	2	-
Contig27	β-glucosidase 15 *Oryza sativa*	Q7XPY5	75	40	32	2	+
Contig32	β-mannosidase 4 *Oncidium* Gower Ramsey	ABC55715	65	93	55	9	+
Contig35	β-glucosidase like protein *Arabidopsis thaliana*	BAD94684	75	130	69	6	+
Singleton19	β-mannosidase 1 *Oncidium* Gower Ramsey	ABC55718.1	90	102	55	1	+
Singleton20	β-mannosidase 3 *Oncidium* Gower Ramsey	ABC55716	83	41	99	1	+
Singleton21	β-glucosidase 11-like isoform 1 *Vitis vinifera*	XP_002268147.1	65	86	41	1	+
Singleton22	β-glucosidase G4 *Medicago truncatula*	AES69718	76	64	28	1	+
Singleton23	β-glucosidase *Pinus contorta*	AAC69619.1	47	90	81	1	-

+ translate in 5’3’ sense; - translate in 3’5’ sense.

The longest putative GH-proteins were aligned using ClustalW. Based on the conserved motif found in *Zea mays* glucosidases, we searched for some motifs responsible for the enzymatic activity. The conserved motifs AYRFSISWSR, TFNEP, EPY were found in six contigs, these motifs are located in the N-terminal region and are characteristic of GH family 1 (Fig. 4). Therefore, the putative GH-proteins from contig 2, 4, 6, 7, 16 and 18, belong to GH family 1 and they seem to be active enzymes.

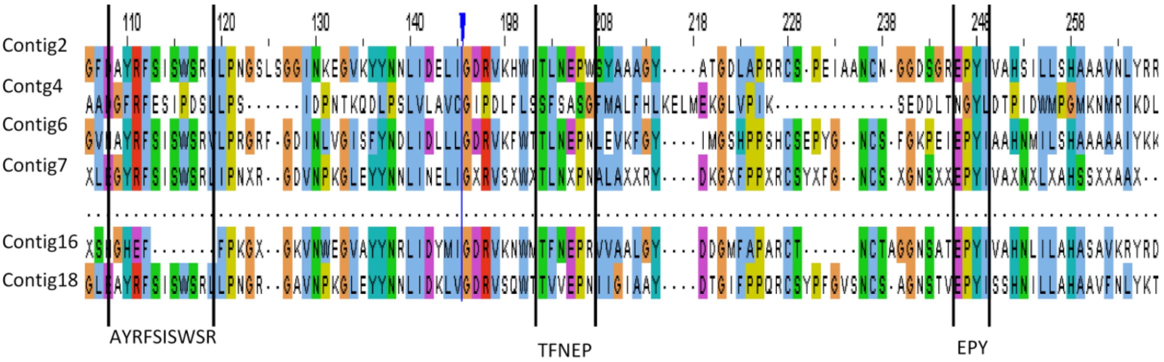


Figure 4. Amino acids sequence alignments from putative GH-protein. Conserved motifs involved in the enzymatic activity are highlighted.

The GH1 family contains the highest number of GHs enzymes, between 40-50 enzymes per genome. Although the experimental characterization of the enzymes is very limited, several important roles have been proposed for these enzymes: (1) in numerous biological processes: growth and development via release of phytohormones (auxins, gibberellins, cytokinins) from their inactive glycoconjugated forms, host-parasite interactions, lignification, cell wall degradation in the endosperm during germination, circadian rhythm of leaf movements, in floral development and pigmentation, and (2) in biotechnological applications: food detoxification, biomass conversion, and flavour enhancement in beverages. Indeed the intensive research carried out over the past decades has demonstrated that, in a great number of fruit and other plant tissues, important flavour compounds accumulate as non-volatile and flavourless glycoconjugates, which make up a reserve of aromas to be exploited.

Expression profiles of GTs and GHs in the different developmental stage of stigmas.

The frequency of EST for individual genes in diverse cDNA libraries can be used to estimate the expression patterns of individual genes. Using this approach, we were able to associate some contigs and singletons with their expression levels, which was calculated as number of mapped reads for each stage of *C. sativus* stigma: yellow, orange, red, 2 days before anthesis (-2da), anthesis day (da) and 2 day after anthesis day (+2da). This approach allowed us to follow the kinetics expression during the developmental stage of these putative β-glycosidases and glycosyltransferases genes and also to score the most active genes in each developmental stage. In this context, there are of particular interest those transcripts that possess an evident time-specific expression. The patterns of ESTs expression are shown in Figures 6 and 7.

Regarding the singletons, their expression could not be estimated because they are represented by a unique read from one cDNA library, for example singleton1 which was detected as a unique read from cDNA library from red stage or singleton11 in yellow stage, as well as singleton20 found in two days after anthesis cDNA library or singleton23 in orange stage.

On the other hand, there are some contigs that we were unable to determine their time-profile because they were detected only as unique read in one cDNA library, this is the case of contig9 found in cDNA library from red stage and contig20 detected in cDNA library from yellow stage. These data suggest that these enzymes are only active in that specific developmental stage (Fig. 5).

Figure 5 showed that GT contigs and singletons were more abundant in the mature stages than in the immature stages, thus 18% of GT-sequences were expressed in the yellow stage, 11% of GT-sequences were expressed in the orange stage and 15% in the red stage, while 16%, 15% and 25% of GT-sequences were expressed in anthesis, 2 day after anthesis and 2 day before anthesis, respectively. However, the opposite was found for the GH contigs and singletons that were found to be more abundant in the immature stages than in the mature stages, 10% of GH-sequences are expressed in 2 day after anthesis, 15% in anthesis and 14% in 2 day before anthesis while 29%, 15% and 17% of GH-sequences are expressed in yellow stage, orange and red stage respectively.

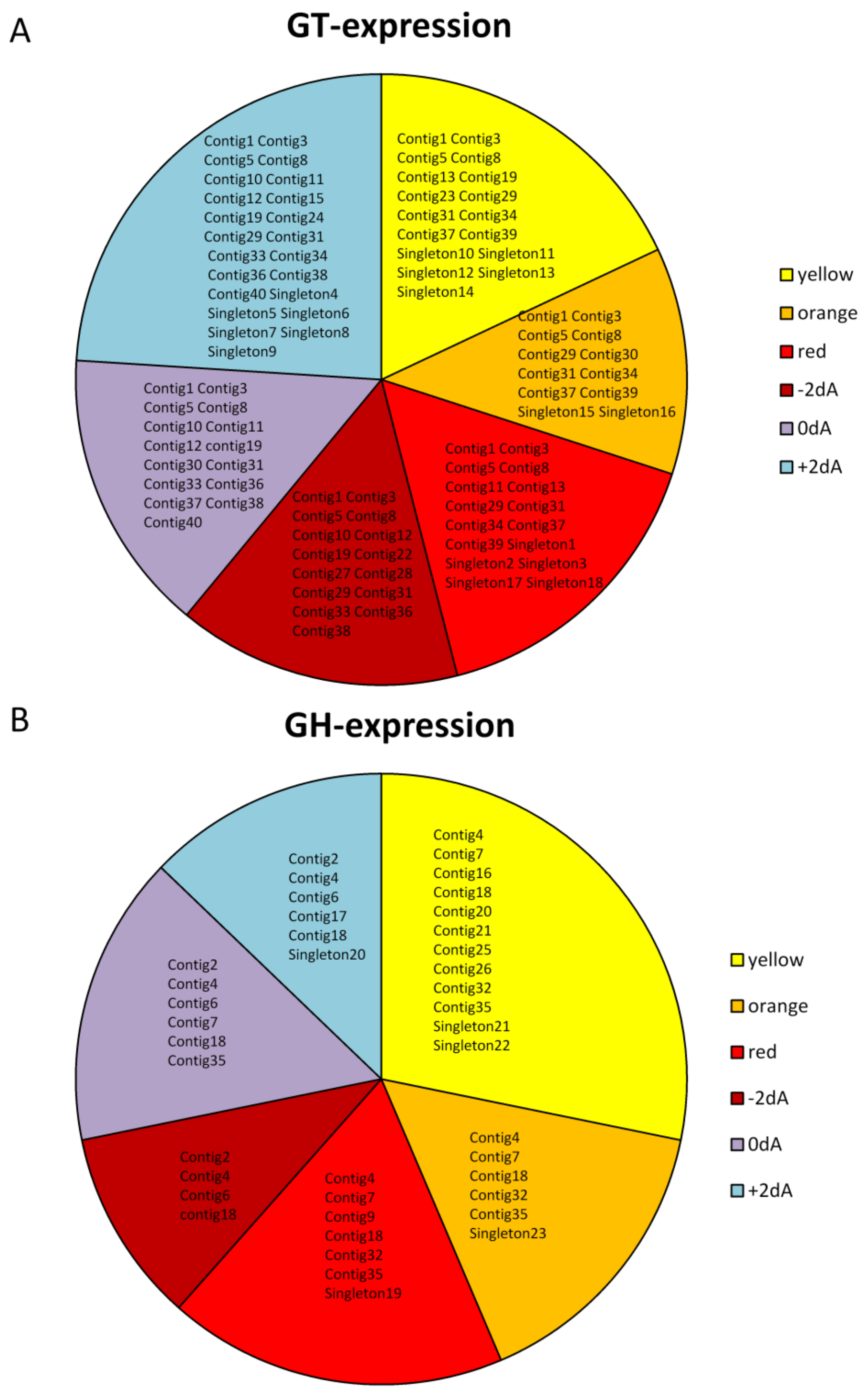


Figure 5. Distribution of the expression profile of contigs and singletons in each developmental stage of the stigma. **A**. **GT contigs and singletons**: yellow color corresponds to contigs and singletons expressed in yellow stage; orange color corresponds to contigs and singletons expressed in orange stage; red corresponds to contigs and singletons expressed in red stage; dark red corresponds to contigs and singletons expressed in 2 day before anthesis (-2da); purple color corresponds to contigs and singletons expressed in anthesis (0da); blue correspond to contigs and singletons expressed in 2 day after anthesis (+2da). **B. GH contigs and singletons**: yellow color corresponds to contigs and singletons expressed in yellow stage; orange color corresponds to contigs and singletons expressed in orange stage; red color

corresponds to contigs and singletons expressed in red stage; dark red color corresponds to contigs and singletons expressed in 2 day before anthesis (-2da); purple color corresponds to contigs and singletons expressed in anthesis (da); and blue color corresponds to contigs and singletons expressed in 2 day after anthesis (+2da).

Some GT contigs were expressed throughout all the developmental stages such contig1 (Fig. 6A). By contrast, contig3 (Fig. 6A) had a time-specific expression, where the maximum number of mapped reads was observed 2 days before anthesis day (-2da). The high expression level of this glycosyltransferase in this stage, matches with the high accumulation of flavonoids in the last developmental stages of the stigma [7] and suggest that contig3 might encode a protein which acts over flavonoids present in the last developmental stages. Additionally, there were sequences that were detected only in the immature stages as contig37 and other sequences were present in the mature stages as contig30.

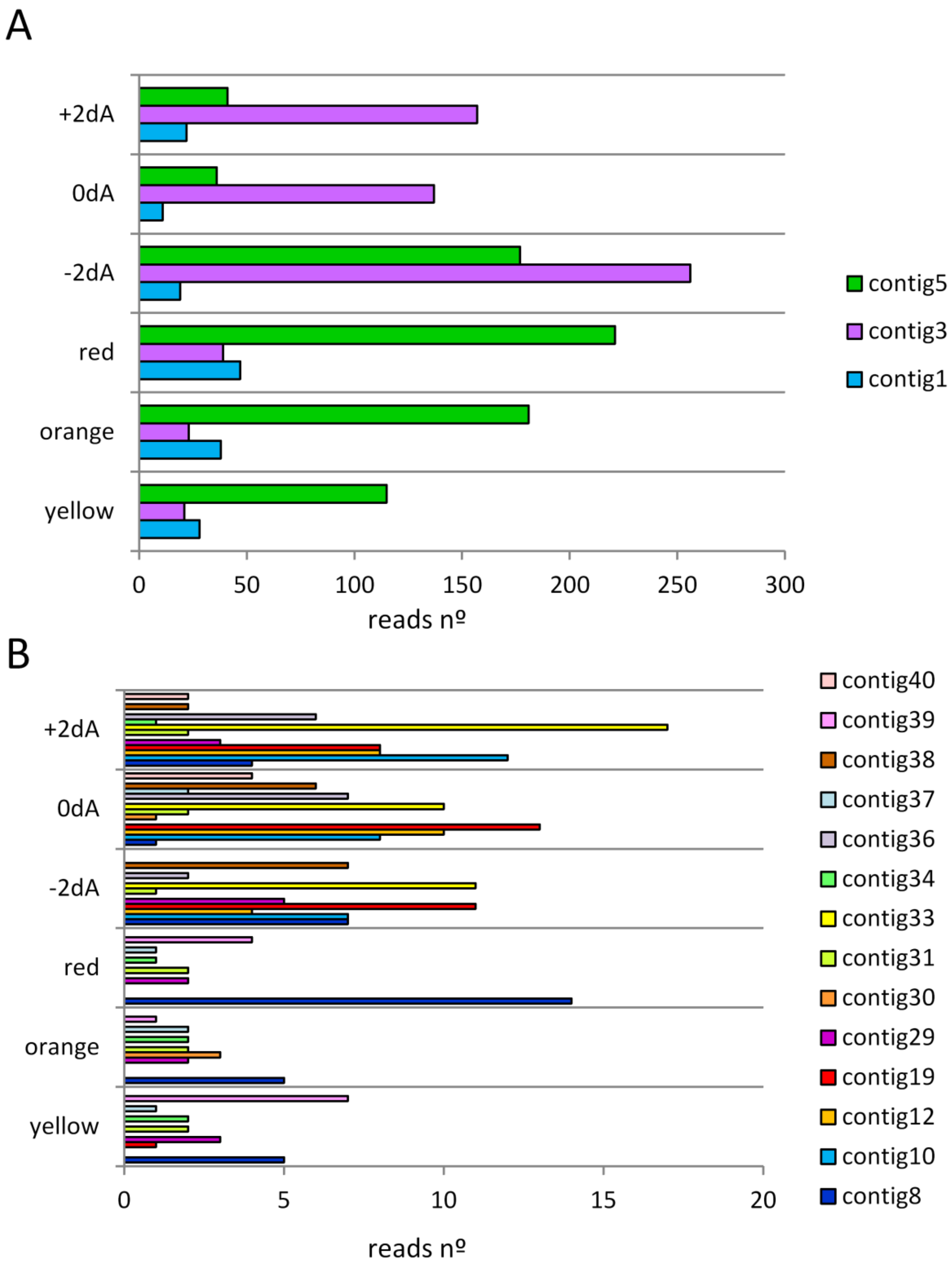


Figure 6. Expression level of different GT-contigs in different developmental stages of *C. sativus* stigmas: yellow stigma, orange stigma, red stigma, 2 days preanthesis, day of anthesis, 2 days postanthesis. **A**. Expression level of main contigs. **B**. Expression level of minor contigs.

Regarding GHs, contig2 encodes a protein very similar to a β-mannosidase which belongs to GH family 1 involved in the carbohydrate modulation during the flowering in *Oncidium orchid*, and has a maximum expression level in two days after anthesis (+2da). Hence, this protein could be involved in flower opening in *C. sativus* (Fig. 7).

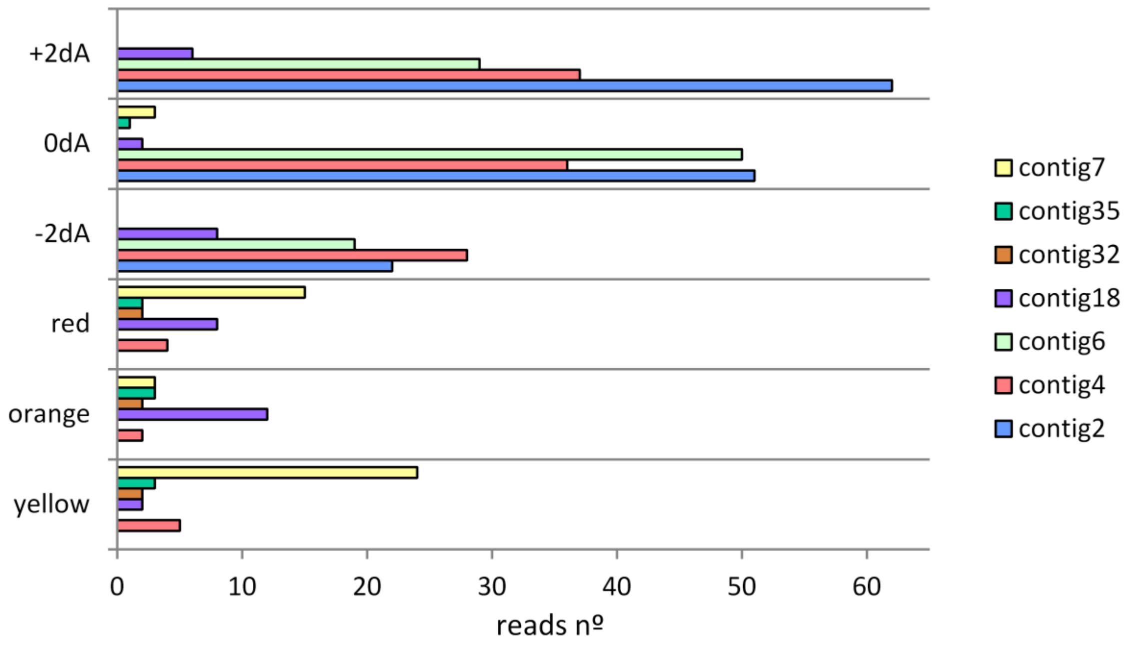


Figure 7. **Levels of expression of different GH-contig in different development stage of *C. sativus* stigmas:** yellow stigma, orange stigma, red stigma, 2 days before anthesis, day of anthesis, 2 days after anthesis.

Contig7 showed a sinusoidal pattern with a maximum expression level in the yellow stage, the expression level decreased in the orange stage and increased again in the red stage, disappeared in -2dA and appeared finally at 0dA. Contig6 showed a maximum expression level at anthesis and was not expressed in the immature stage. However contig18 was expressed throughout all the developmental stages. Contig35 and contig32 showed low expression levels in the early developmental stages when the flower remains closed.

Discussion

Glycosyltransferases and β-glycosidases are involved in the secondary metabolism of plant cells. Although many of these enzymes have been isolated from several plant species and assayed *in vitro*, in many cases their roles in the secondary metabolism of these plants have not been elucidated. This work focuses on the *in silico* identification of β-glycosidases and glycosyltransferases genes in saffron and the determination of their expression profile during the developmental stages of the stigma.

To reach our objective we used ESTs collections generated from six different developmental stages from *C. sativus* stigma, 3897 ESTs were found and 3791 ESTs were assembled in 45 contigs or tentative consensus sequences and 106 unique (non-overlapping) sequences. After a second assembling procedure of the selected ESTs, we obtained a total of 40 contigs and 23 singletons as showed in Table 1 and 2. Using the databases of NCBI, a tentative function has been assigned to most contigs and singletons on the basis of sequence similarity, to known proteins or genes with either a demonstrated function or a putative function based on sequence similarly to other genes. After a manual checking, 25 contigs and 18 singletons encoded putative glycosyltransferases and 15 contigs and 5 singletons encoded putative β-glycosidases in *C. sativus* stigma.

Other additional information that could be drawn from the ESTs collection is the determination of the expression levels. This approach could estimate in which developmental stage genes are expressed, and it could be useful to identify putative substrates depending on the specific metabolite profile.

Although the number of ESTs sequences is lower than others obtained in other species (Table 3) for example California poplar (*Populus trichocarpa*) with 100000 ESTs isolated from 17 different tissues (leaves, roots, flowers, etc.) and from which 840 GTs and 600 GHs were identified [32], corresponding to 2.7% of total ESTs, the percentages of encoding GT and GH ESTs are higher in *C. sativus* which is represented by 6% of total 65000 ESTs. In other plant species, the number of sequences is closer to that obtained in saffron, this is the case of *Medicago truncatula*, where 200000 ESTs were obtained and from which 150 GTs have been identified [11].

Table 3: Number of ESTs in some plant species and the number of GTs and GHs identified proteins in each species.

Plant	ESTs	GT proteins	GH proteins	Reference
Zea mays	2019137	125	1163	NCBI Protein database
Arabidopsis thaliana	1529700	488	2115	NCBI Protein database
Glycine max	1461722	199	568	NCBI Protein database
Triticum aestivum	1286372	13	98	NCBI Protein database
Oryza sativa	1253557	174	2094	NCBI Protein database
Hordeum vulgare	501838	31	425	NCBI Protein database
Medicago truncatula	269501	107	636	NCBI Protein database
Solanum tuberosum	249761	115	129	NCBI Protein database
Sorghum bicolor	209835	87	480	NCBI Protein database
Populus trichocarpa	89943	60	588	NCBI Protein database
Crocus sativus	65000	43	20	Data from this work
Ricinus communis	62592	48	486	NCBI Protein database
Fragaria x ananassa	10855	23	79	NCBI Protein database
Cucumis sativus	8152	156	174	NCBI Protein database

The high number of GT and GH sequences obtained in *C. sativus* stigmas could be related to the more active secondary metabolism in saffron compared to *P. trichocarpa* or *M. truncatula*. Saffron is constituted by a complex group of metabolites; most of them are found in glycosylated form. Different glycosidic structures for any given secondary metabolite suggest the existence of different families of GHs and GTs that could hydrolyse or add a sugar to these compounds [33]. The triploid nature of *C. sativus* might be other probably reason for the highest number of GTs and GHs identified in comparison to other plants.

We were able to classify some GTs on the basis of their homology with known genes. Since sequence homology is far from being a definitive argument to describe the precise enzymatic activity of glycosyltransferase enzymes [34] the functional characterization of these enzymes is still necessary. Furthermore, correlations between transcripts and metabolites could facilitate an efficient narrowing-down of candidate products of glycosyltransferase activities [35]; [36]; [37]. Nevertheless a tentative function was proposed for the identified sequences. The contig34, 37, 40 and singleton 7, 10 and 15 could act as rhamnosyltransferases because they share around 40% identity with a rhamnosyltransferase from citrus involved in the glycosylation of some flavonoids responsible of bitter flavor [38]. The singleton13, 16, 19, 28 and 27 could acts as flavonoid glucosyltransferases, due to their similarities with reported flavonol glucosiltransferases, for example flavonol 3-O-glucosyltransferase expressed in strawberry fruit with an *in vitro* activity over quercetin and kaempferol [39]. The contig1, 14, 29, 31 and singleton 2 and 12 are GTs could be active over zeatin, with 35% to 50% of identity with AtZOG2 from *Arabidopsis thaliana* [40]. The contig12 shares 42% of identity with a GT which acts over indole-3-acetic acid [41]. Using the possible substrates as key for enzymatic classification, all identified GT-sequences seem to belong to family 1. The GT1 family comprise a wide variety of enzymes catalyzing sugar addition to diverse compounds such as hormones, including indole-3-acetic acid, zeatin, salicylic acid, monolognols and secondary metabolites as flavonols and anthocyanins [42].

The analysis of the conserved amino acids encoded GTs of family 1 supported the presence of 16 distinct groups designated A to P [22]. Thus, contig19 belongs to group K or contig3 belongs to group G, which contains different enzymes that use terpenoids, benzoates and flavonoids as substrates. It was not possible to classify each GT-sequence into a group due to the lack of full length sequences.

The GHs are involved in diverse functions as releasing of phytohormones and volatile compounds, lignification, cell wall degradation, floral development and pigmentation, among others. The most interesting GHs in saffron might be those able to hydrolase the glycosidic bonds of compounds involved in flavor, modifying the organoleptic character of the spice. In this work, GHs involved in resealing the volatiles were not found. One of the possible reasons could be attribute to low expression levels of these enzymes. However the principal activities of the putatives GHs identified in saffron stigma could be reported using sequence homology, thus the contig2, 9, 16, 18, 32 and 35 share similarities with Os4BGlu4 from *Orzya sativa* [43] and could be implicated in lignification and cell wall degradation.

Nowadays, the identification of enzymes by sequence homology is not sufficient; these studies must be complemented by biochemical characterization describing their enzymatic activity. Although the principal aim of using ESTs collection in this study was not to obtain the full-enzyme sequences, we have been close to determining some full-sequences as in contig1 or contig2; both of them lacking only 21 and 30 bp respectively. All reported GT and GH sequences could be a start point to determine the full-length enzyme using different molecular strategies.

In summary, the analysis of ESTs is a powerful tool to study the genome function, especially in organisms where other genomics technologies are difficult or intractable, and in organisms of relatively minor collective importance.

Reference

1. Kumar R, Sangwan RS, Mishra S, Sabir F, Sangwan NS: **'In silico'motif diversity analysis of the glycon preferentiality of plant secondary metabolic glycosyltransferases**. *Plant Omics* 2012, **5**(3):200.
2. Duroux L, Delmotte F, Lancelin J, Keravis G, Jay-Allemand C: **Insight into naphthoquinone metabolism: beta-glucosidase-catalysed hydrolysis of hydrojuglone beta-D-glucopyranoside**. *Biochemical Journal* 1998, **333**(Pt 2):275.
3. Leah R, Kigel J, Svendsen I, Mundy J: **Biochemical and molecular characterization of a barley seed b-glucosidase**. *J Biol Chem* 1995, **26**:15789-15797.
4. Leah R, Kigel J, Svedsen I, Mundy J: **Biochemical and molecular characterization of a barley seed beta-glucosidase**. *J Biol Chem* 1995, **270**:15789 - 15797.
5. Winterhalter P, Skouroumounis G: **Glycoconjugated aroma compounds: occurrence, role and biotechnological transformation**. *Biotechnology of aroma compounds* 1997:73-105.
6. Moraga ÁR, Rambla JL, Ahrazem O, Granell A, Gómez-Gómez L: **Metabolite and target transcript analyses during *Crocus sativus* stigma development**. *Phytochemistry* 2009, **70**(8):1009-1016.
7. Moraga AR, Mozos AT, Ahrazem O, Gomez-Gomez L: **Cloning and characterization of a glucosyltransferase from *Crocus sativus* stigmas involved in flavonoid glucosylation**. *BMC Plant Biol* 2009, **9**:109.
8. Forkmann G, Martens S: **Metabolic engineering and applications of flavonoids**. *Current Opinion in Biotechnology* 2001, **12**(2):155-160.
9. Richman A, Swanson A, Humphrey T, Chapman R, McGarvey B, Pocs R, Brandle J: **Functional genomics uncovers three glucosyltransferases involved in the synthesis of the major sweet glucosides of *Stevia rebaudiana***. *The Plant Journal* 2005, **41**(1):56-67.
10. Tian L, Blount JW, Dixon RA: **Phenylpropanoid glycosyltransferases from osage orange (Maclura pomifera) fruit**. *FEBS letters* 2006, **580**(30):6915-6920.
11. Modolo LV, Blount JW, Achnine L, Naoumkina MA, Wang X, Dixon RA: **A functional genomics approach to (iso) flavonoid glycosylation in the model legume *Medicago truncatula***. *Plant molecular biology* 2007, **64**(5):499-518.
12. Lunkenbein S, Bellido M, Aharoni A, Salentijn EM, Kaldenhoff R, Coiner HA, Muñoz-Blanco J, Schwab W: **Cinnamate metabolism in ripening fruit. Characterization of a UDP-glucose: cinnamate glucosyltransferase from strawberry**. *Plant physiology* 2006, **140**(3):1047-1058.
13. Gosch C, Halbwirth H, Schneider B, Hölscher D, Stich K: **Cloning and heterologous expression of glycosyltransferases from *Malus x domestica* and *Pyrus communis*, which convert phloretin to phloretin 2′-O-glucoside (phloridzin)**. *Plant Science* 2010, **178**(3):299-306.
14. Li Y, Luo H-M, Sun C, Song J-Y, Sun Y-Z, Wu Q, Wang N, Yao H, Steinmetz A, Chen S-L: **EST analysis reveals putative genes involved in glycyrrhizin biosynthesis**. *BMC genomics* 2010, **11**(1):268.
15. Dick R, Rattei T, Haslbeck M, Schwab W, Gierl A, Frey M: **Comparative Analysis of Benzoxazinoid Biosynthesis in Monocots and Dicots: Independent Recruitment of Stabilization and Activation Functions**. *The Plant Cell Online* 2012, **24**(3):915-928.
16. D'Agostino N, Pizzichini D, Chiusano ML, Giuliano G: **An EST database from saffron stigmas**. *BMC plant biology* 2007, **7**(1):53.
17. Rubio A, Rambla JL, Santaella M, Gomez MD, Orzaez D, Granell A, Gomez-Gomez L: **Cytosolic and plastoglobule-targeted carotenoid dioxygenases from *Crocus sativus* are both involved in beta-ionone release**. *J Biol Chem* 2008, **283**(36):24816-24825.
18. Moraga AR, Nohales PF, Perez JA, Gomez-Gomez L: **Glucosylation of the saffron apocarotenoid crocetin by a glucosyltransferase isolated from Crocus sativus stigmas**. *Planta* 2004, **219**(6):955-966.
19. **http://www.codoncode.com**
20. Barvkar VT, Pardeshi VC, Kale SM, Kadoo NY, Gupta VS: **Phylogenomic analysis of UDP glycosyltransferase 1 multigene family in Linum usitatissimum identified genes with varied expression patterns**. *BMC genomics* 2012, **13**(1):175.
21. Martin RC, Mok MC, Mok DW: **Isolation of a cytokinin gene, ZOG1, encoding zeatin O-glucosyltransferase from Phaseolus lunatus**. *Proceedings of the National Academy of Sciences* 1999, **96**(1):284-289.

22. Caputi L, Malnoy M, Goremykin V, Nikiforova S, Martens S: **A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land**. *The Plant Journal* 2012, **69**(6):1030-1042.
23. Hou B, Lim E-K, Higgins GS, Bowles DJ: **N-Glucosylation of Cytokinins by Glycosyltransferases of *Arabidopsis thaliana***. *Journal of Biological Chemistry* 2004, **279**(46):47822-47832.
24. Jones PR, Møller BL, Høj PB: **The UDP-glucose: p-hydroxymandelonitrile-O-glucosyltransferase that catalyzes the last step in synthesis of the cyanogenic glucoside dhurrin in sorghum bicolor isolation, cloning, heterologous expression, and substrate specificity**. *Journal of Biological Chemistry* 1999, **274**(50):35483-35491.
25. Hansen KS, Kristensen C, Tattersall DB, Jones PR, Olsen CE, Bak S, Møller BL: **The in vitro substrate regiospecificity of recombinant UGT85B1, the cyanohydrin glucosyltransferase from *Sorghum bicolor***. *Phytochemistry* 2003, **64**(1):143-151.
26. Kramer CM, Prata RT, Willits MG, De Luca V, Steffens JC, Graser G: **Cloning and regiospecificity studies of two flavonoid glucosyltransferases from *Allium cepa***. *Phytochemistry* 2003, **64**(6):1069-1076.
27. Nagatoshi M, Terasaka K, Nagatsu A, Mizukami H: **Iridoid-specific Glucosyltransferase from *Gardenia jasminoides***. *Journal of Biological Chemistry* 2011, **286**(37):32866-32874.
28. Morant AV, Bjarnholt N, Kragh ME, Kjærgaard CH, Jørgensen K, Paquette SM, Piotrowski M, Imberty A, Olsen CE, Møller BL: **The β-glucosidases responsible for bioactivation of hydroxynitrile glucosides in *Lotus japonicus***. *Plant physiology* 2008, **147**(3):1072-1091.
29. Escamilla-Trevino L, Chen W, Card M, Shih M-C, Cheng C, Poulton J: **Arabidopsis thailiana beta-glucosidases BGLU45 and BGLU46 hydrolyse monolignol glucosides**. *Phytochemistry* 2006, **67**:1651 - 1660.
30. Wang C-Y, Chiou C-Y, Wang H-L, Krishnamurthy R, Venkatagiri S, Tan J, Yeh K-W: **Carbohydrate mobilization and gene regulatory profile in the pseudobulb of *Oncidium orchid* during the flowering process**. *Planta* 2008, **227**(5):1063-1077.
31. Dharmawardhana DP, Ellis BE, Carlson JE: **cDNA cloning and heterologous expression of coniferin β-glucosidase**. *Plant Molecular Biology* 1999, **40**(2):365-372.
32. Geisler-Lee J, Geisler M, Coutinho PM, Segerman B, Nishikubo N, Takahashi J, Aspeborg H, Djerbi S, Master E, Andersson-Gunnerås S *et al*: **Poplar Carbohydrate-Active Enzymes. Gene Identification and Expression Analyses**. *Plant physiology* 2006, **140**(3):946-962.
33. Coutinho PM, Stam M, Blanc E, Henrissat B: **Why are there so many carbohydrate-active enzyme-related genes in plants?** *Trends in plant science* 2003, **8**(12):563-565.
34. Caputi L, Malnoy M, Goremykin V, Nikiforova S, Martens S: **A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land**. *Plant J* 2011.
35. Yonekura-Sakakibara K, Tohge T, Matsuda F, Nakabayashi R, Takayama H, Niida R, Watanabe-Takahashi A, Inoue E, Saito K: **Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene–metabolite correlations in Arabidopsis**. *The Plant Cell Online* 2008, **20**(8):2160-2176.
36. Kovinich N, Saleem A, Arnason JT, Miki B: **Functional characterization of a UDP-glucose: flavonoid 3- *O*-glucosyltransferase from the seed coat of black soybean (*Glycine max*(L.) Merr.)**. *Phytochemistry* 2010, **71**(11):1253-1263.
37. von Saint Paul V, Zhang W, Kanawati B, Geist B, Faus-Keßler T, Schmitt-Kopplin P, Schäffner AR: **The Arabidopsis glucosyltransferase UGT76B1 conjugates isoleucic acid and modulates plant defense and senescence**. *The Plant Cell Online* 2011, **23**(11):4124-4145.
38. Frydman A, Weisshaus O, Bar-Peled M, Huhman DV, Sumner LW, Marin FR, Lewinsohn E, Fluhr R, Gressel J, Eyal Y: **Citrus fruit bitter flavors: isolation and functional characterization of the gene Cm1, 2RhaT encoding a 1, 2 rhamnosyltransferase, a key enzyme in the biosynthesis of the bitter flavonoids of citrus**. *The Plant Journal* 2004, **40**(1):88-100.
39. Griesser M, Vitzthum F, Fink B, Bellido ML, Raasch C, Munoz-Blanco J, Schwab W: **Multi-substrate flavonol O-glucosyltransferases from strawberry (*Fragaria× ananassa*) achene and receptacle**. *Journal of experimental botany* 2008, **59**(10):2611-2625.
40. Woo H-H, Jeong BR, Hirsch AM, Hawes MC: **Characterization of Arabidopsis AtUGT85A and AtGUS gene families and their expression in rapidly dividing tissues**. *Genomics* 2007, **90**(1):143-153.

41. Eudes A, Bozzo GG, Waller JC, Naponelli V, Lim E-K, Bowles DJ, Gregory JF, Hanson AD: **Metabolism of the Folate Precursor p-Aminobenzoate in Plants: glucose ester formation and vacuolar storage**. *Journal of Biological Chemistry* 2008, **283**(22):15451-15459.
42. Ross J, Li Y, Lim E-K, Bowles DJ: **Higher plant glycosyltransferases**. *Genome Biol* 2001, **2**(2):3004.3001-3004.3006.
43. Opassiri R, Pomthong B, Onkoksoong T, Akiyama T, Esen A, Ketudat Cairns J: **Analysis of rice glycosyl hydrolase family 1 and expression of Os4bglu12 beta-glucosidase**. *BMC plant biology* 2006, **6**(1):33.

Conclusion

Conclusion

1- We have determined the role of CsGT45 in the transfer of glucose on 7-OH of flavonoids, together with its implication in the generation of *C. sativus* flavonoids in the stigma tissue. The characterization of this glucosyltransferases involved in the formation of these compounds in *C. sativus* will help to understand the biosynthesis and regulation of these glucosides and their implications in the nutraceutical properties of saffron.

2- We have analysed CCD4 gene structure in relation to CCD4 genes from related *Crocus* species and also from other plant species, revealing a history of intron loss and gain during the evolution of these genes. In addition, we report the sequences of the putative promoter regions of CsCCD4a and CsCCD4b genes, and a functional evaluation of a 1,400-bp genomic DNA fragment from the CsCCD4a promoter. Gene structures analyses suggest that the CCD4 family is highly dynamic, with relative recent changes found in the *Crocus* sequences from series Crocus. The data presented provide an overview for the CCD4 plant family, which can facilitate their functional studies.

3- The remobilization of substances from unfertilised flower parts could be considered as part of this life strategy in *C. sativus*. Similar pattern of relative apocarotenoid composition and accumulation in the ovary and in the corm, although it was in this latter organ where a significant accumulation was observed. In contrast to leaves they are a sink rather than a source of photosynthates, a feature of stigma senescence that supports their role in the remobilization of cellular constituents from the senescing organs to other structures that continue to develop.

4- UGT707B1 is an enzyme involved in the formation of flavonol sophorosides in *C. sativus*. Based on the accumulation of certain flavonols in the transgenic UGT707B1 Arabidopsis plants, we demonstrate that UGT707B1 has a flavonol-3-O-glucoside: 2''-glucosyltransferase activity.

5- Through the using of bioinformatics tools and ESTs collection of partial sequences from six developmental stages of *C. sativus*, 43 encoded-sequences glycosyltransferases and 20 encoded-sequences β-glucosidases have been identified. The putative function of each one have been reported on base homology sequences other species and their expression pattern have been reported, thus the majorly of GT-sequences are more active in maturated developmental stages of stigma while GH-sequences are more active in immaturated. Although for describing the precise enzymatic activity of GTs and GHs enzymes, the functional characterization of these enzymes is still necessary.

Annex

Annex 1

GT-sequence and partial GH-sequence used for identifying of new sequence in ESTs collection.

GLT2_ AN_AY262037

GGACAGCCAGCAGAAATCACAATGGACGAGAAAGATGTTGAACGGCAACAAATGCCACATCCTCCTCCTC
CCCTGTCCAGCACAAGGCCATATCAACCCCATTCTCCAATTCGGCAAGCGACTGGCCTCTCACAACCTCC
TCACCACACTCGTCAACACCCGGTTCCTCTCCAACTCGACCAAATCCGAACCCGGTCCGGTCAACATCCA
GTGCATATCCGACGGATTCGATCCCGGTGGGATGAATGCAGCACCTAGTCGTCGGGCCTATTTCGACCGA
CCTCAGAGTCGGTCGGGTCAGAAACATGTTGGCCTCATCGAGTCACTCCGGTCCAGAGGCCGACCCGGCG
CATGTTTTGGTCTACGACCCGTTCCCCTCTGGGCAATGAACGTAGCCGAGCGGTCCGGGCTGCGGTCGGT
CGCATTCTTTACTCAGCCGTGTGCAGTGGACACAATCTACCGCCACGTGTGGGAGGGACGAATTAAGGTT
CCGGTTGCCGAGCCGGTCCGGTTGCCTGGATTGCCCCCACTCGAGCCTAGTGATCTGCCGTGCGTTCGTA
ACGGATTCGGACGAGTCGTTAACCCGGATCTCTTGCCACTTCGAGTGAATCAGCACAAAAATTTGGACAA
AGCTGACATGATGGGGAGGAACTCCATATACGAGCTAGAGGCGGATTTGTTGGATGGGTCCCGATTACCA
TTACCGGTGAAGTCGATTGGGCCGACCGTGCCGTCTACTTATCTCGACAACCGAATCCCGTCGGATTCAC
ACTATGGATTCAATCTATACACCCCTGACACTACTCCCTACTTGGACTGGCTCGACTCCAAGGCCCCGAA
CTCTGTCATCTATGTCTCATTTGGTAGTTTGTCCAGTCTGAGCCCTGATCAAACCAATGAGATTGCGTCG
GGTCTGATCGCCACCAACAAAAGCTTTATTTGGGTGGTACGCACTTCGGAGTTAGCCAAACTCCCTGCCA
ACTTCACCCAAGAGAATGCGAGCCGTGGGCTGGTTGTGACCTGGTGCGACCAACTCGATCTCCTCGCACA
CGTTGCGACCGGTTGCTTCGTGACTCACTGCGGATGGAACTCGACTATGGAAGGTGTCGCACTTGGTGTG
CCGATGGTGGGAGTGCCGCAGTGGTCGGACCAGCCTATGAATGCCAAGTATGTAGAGGATGTGTGGAAGG
TCGGCGTGAGGGCGAAGACTTATGGGAAGGATTTTGTGAGAGGGGAGGAGTTCAAAAGATGTGTGGAGGA
GGTGATGGATGGAGAGAGGAGTGGAAAGATTAGAGAGAATGCCGCAAGGTGGTGTAAGTTGGCCAAGGAC
TCTGTTAGTGAGGGTGGGAGTTCTGACAAGTGTATCAAGGAGTTCATACATCAGTGTTGCAATGACTCCA
AAATTTCCTTAGTTTAATCACCCTGTTAATAAAGTTTTGTTTCATGTGTCAAACCTGAGTCAATCTATTA
AGAGAAAAATTGTGCATAACAGAAGA

GLT3_ AN_AY290820

CTCATTGGCATGGCGAAGGAACACATAGTATTGTTTCCATTCATGTCACAAGGCCATATCATCCCTTTCC
TTTCCTTAGCCAAACTCATTTCGGAACGCCACCCGACTTACACCATCACCCTCCTCAACACCCCACTCAA
CATCCTTAACCTCCAATCCACACTCCCTCCAAACTCCAACATCCATCTCAAGTCCCTCCCCTACCGAAGT
TCCGACTTCGGCCTCCCGCCCGATCGCGAAAACACCGACTCCTTACCCTTTCCCTTGGTACTAAGTTTCT
ACCAATCCGGCGAATCCCTCGCAACCCACTTCACCCACTTCGTCTCCGACCTGACCCGTCAAAACCACGA
TACCCCCCCTCTTCTAATTGTAGCCGATGTTTTCTTCGGATGGACGGCCGAGATCGCAAAGAGGTTGAAC
ACCCATGTGAGCTTCTCTACTTGCGGTGCCTATGGCACGGCGGCCTACTTTTCCGTATGGCTCCACTTGC
CGCACGCGGAAACCGACTTACCGGACTTCACAGCCCCGGGGTTCCCGGAAACTTTCAAGCTGCAACGGAA
CCAACTCTCTACTTACTTGAAGAAGGCCGACGGTAGTGATCGGTGGTCGAAGTTTTTCCAACGGCAGATA
TCGCTGTCCTTAACTTCGGACGCGATGATATGCAACACCGTTGAGGAAATGGAGGCGGAGGGGCTGAGGC
TGCTTAGAAAGAACACCGGGCTAAGAGTCTGGTCCATCGGGCCCTTACTTCCTTCGCTTCCTCCGAATTC
GTCACTTGGCCGAAGCGGAAGGAAGTCCGGAATGGAAGTGAGTTACATTATGAAGTGGCTCGATTCGCAC
CCACCGGGTTCTGTGGTTTACGTGTCTTTCGGGTCAATACACGATACGGCTGCCCAGATGACGAGTCTCG
CCGTCGGGCTTGCGGTCGAGCTGGCCACCCGTTCGTGTGGGCACTCAGGCCGCCGCTTTGGTGGGAACCG
AAATCGGAATTCAAACCCGAATGGCGTTCCGGACGAGTTCGAGGCGAGAATGCGGGGGAGCGGGCGGGGG
ATTCTGATACACGGGTGGGCCCCGCAGCTGGAGATTCTTGAGCATGAGTCCACGGGGGCATTCGTGAGTC
ACTGCGGGTGGAACTCGACCCTAGAGAGCTTGAGCAGAGGAGTGTGCATGATAGGGTGGCCGCTGGCCGC
CGAACAATTCTACAACTCAAAGATGGTGGAGGAGGATTGGGAGTGGGGTGGAACTTGCGAGGGGAGCGGC
GGAGGAGTGAGGAGCGAGGAGGTGGAGAGACTGGTTAGGTTGGTGACGGAGGACGAGAAAGGAAGCGATG
AGGAGAATGAGCAATATGATGAGATGATAGGAGGCTATGAAGAGAAGGGAGGTGAAGGTTCTCTATCAGG
TCAGTTGATTAAGTTTATTGGAATGGAATCCCAATAAGGAAGCAATGCATCCCAAGGGTAATGGTTTTGC
A

CsGT45_ AN_FJ194947
ATAGCAAGCAGTCCAGAATCACTCAACTACAATCTCTTTTCTTGCTCTGGCTTCTGCTTCTGCTTCGCAA
GTAAGACAAGACGACACAGATGGAGCAACATCAGCCTAACATTCTCGTGGTAACCTATCCAGCACAGGGC
CACATCAACCCTGCCCTCCACCTGGCCAAGCACCTCGCCGCCGACACGAAAGGCCTCCTCATCACCTTCT
CCACCGCCATCTCCGCCCATCGCAAGATGTTCCCCGAGTCAACCGAACCCGACCAAGAAGTGGAAGACGG
ACCCATTACCTACATACCCTTCTCCGACGGATACGACGACGGCTTCCAGAGAGACAAGCACGACGGCAAA
CAATTCAGATCCCGGTTCAAAACCGTAGGATCCAACACCCTCTCCGCTATAATCCAAAACCTCGAGCATC
GCGGCCGGAAGGTCAGCTGCGTCATCTATACCTTCTTCGTCTCTTGGGCCGCTGACGTGGCACGACAGCA
TGCCATTCCTTCGGTCCAGTACTGGATCCAGCCGGCCACCGTCTTTGCCATCTACTACCATTACTTCCAT
GGCTACGAGTCCGTCGTGGCTGCCCACTCTCACGACCCGTCCTATCCGATTAATTTACCAGGCCTGTCAC
CCGTGCAAGTGCGTGACCTCCCTTCTTTCCTGACGATTAAGCCTGACGACCCATACGCCGTAGTCCTCTC
TATGATTAGAGACTCGTTCGAAGGTCTGGACCGGGAAGAGACGAAGACCAAGGTTTTGGTGAACACATTC
GGCCAGCTGGAGGCCGATGCGATTTTAGCTGTCGATAAGATGGATATCATCCCGGTGGGTCCGATACTCC
CGTGCAAGGGCGGGGTTTCCAGGGGCGATCTCTTGAAGGAGGACGAGAAGGGATACATGGAGTGGCTGGA
CTCGAAGCCCGAGAATTCGGTAGTCTACGTGTCATTGGAAAGCTTGGCAGTGCTGAAGAAGCAGCAGAAG
TTTTTGATTCTCAAGGGGCTCAAGGATAGCGGGAGGCCGTATCTTTGGGTGGTGAGAAGGGACTCCGGAT
TGGAAGGAGTCGAGTTGGGGGACTGGGATGGGGATGGGGATGGGGATAATGGAATGGTGGTGGGATGGTG
CTCTCAGGTGAGTGTCTTATCGCATCCTTCGGTCGGATGTTTTGTGACACACTGCGGGTGGAATTCGACG
ATGGAGAGCTTGGCTTCCGGAGTGCCGACAATTGGAGTGCCTCAGTGGTCAGACCAGCCGACCTGTGCAG
CGTTGGCAGAGAAGGATTGGGGGATTGGAGTGAGGAGTGAGGTTAACGGAGATGGGATATTGGAAGGTGG
GGAGCTGAAGAGGTGTTTGGACTTGGTATTAGGAGATGGGGAGAGAGGTGTTGAGATAAGGAGAAAAGTT
GAGTTTTGGAAGGATAAAGCTACGGAAGCAATCAGTTTTGGAGGGTCTTCTGATAAAAATTTGCGGACGT
TCGTGGATCAGATAACTGGGTACGATCCAAATGTTTCTTCAGTTGGGTTTTGGAAGGATAAAGCTATTCA
TAGACAGCATTGCGAGGAGTCTGAGTCTGCTTCACCCTCTTAAAGGAGGGTGA

CsGLN21_AN_ CCG85331.1
AGAAGACGACGATGCAAATTTCTCTTGTCTTCTTTCCGGCATGGAGCGCCGGCCACCTGACCTCCATGCT
AGAGTTCTCGAAGCTTCTACTGACCACTAACGCTGACGTCAACATTTCCATCACATTCCTCCTCATCAAA
CTCCCCTATAGAACCTTCTCCTCCGCCTCCCTCGCCTCCATGGAATCCCTCTCCTCCTCCGGCCTACAAG
TCCACTTCCACCAGCTCCCGGAAGTCGACCTACCCGAAAACTCCGACGGCCCCGAGGACACCGCTTCCAC
CTACTTCCAGCTCTACACGCCCCACGTCAGAGCCTTCCTGTCTTCCCACCCGAACCCCGTCTCCGCTTTC
CTCATCGACTTCTTTGCCACGTCACTCATCTACGTCGCCACCGAGTTCTCCGTACCCACCTTCGTCTACT
TCACTTCCACCGCCCTCATGCTCGGTCTCAATCTCCACCTCCCTTTTCTCGAAAAGAAGATCGGCGTCGA
GTTCGGCCAGGTCGAAGGGGAAGTGGAGATCCCCGGGGTCGTGTCGGTGCCGCCGGGATCGATGCCGACC
CCTTTGATGGACAAGAAGAGCAGGAATTACACGTGGTTCGTTTACCACGGGAGGCAGTTTCGGGAGGCCA
AGGGCATTGTTGTGAATAGTGTTGCCGAGTTGGAACCGGGGGTTTTGTCGGCGATGGCGGAGGGGAGGTT
CGTCGAGGGTGGCATTATGCCGACTGTCTATCTGGTGGGCCCGATTCTTTCGCTTGCTGACAAGGGTGGT
GGTAGTTCTAGCCGGAATGATGAATGTTTGGTGTGGTTGGATGAGCAGCCCAAAGGCTCTGTTTTGTTCC
TTTGCTTTGGAAGTATGGGATGGTTTGGAGTTCACCAGGTGAGGGAAATGGCGACAGGGCTAGAACAGAG
CGGACACCGCTTCCTCTGGAGCCTCCGCAGCATGCCAGCTGGCGACAACCACCAACCTACCGACGCGAAT
CTCGACGAGGTACTCCCCGAAGGCTTCCTGGAGAGAACCAAAGACAGAGGGATGGTGTGGCCTTCGTGGG
TCCCACAGGTGGAGGTACTGTCCCATGCTTCGGTGGGGGGGTTCGTGACGCACTGCGGGTGGATTCGTTC
TGGAGAGCTGTGGTTTGGGGTGCCGATGACCGCGTGGCCCCAGTACGTGGAACAGCATCTGAACGAGGTC
GAGCTGGTGAGGGATATGGGGGTGGCGGTGGGGATGGAAGTGGACCGTAAGTGCGGGAACTTCGTTTCGG
CGGCGGAACTGGAGAGGGGAGTGAGGTGTCTGATGGGGGAAAGTGAGGAAAGTAGGAGGGTGAGGGCAAA
GGTGGCAGATATGAAGGTGGCAATCAGGAATGCCCTGAAGGAAGGAGGGTCCTCCTATACTAATCTGAAG
AAACTGGCTAAGGATTCGTGCTGCTAAATGTGGACATGTGATGGTCTTAGTATGTGTGGCAGTATAAGTT
TCTCCTGAGTCTGGTGTGATAAATTGATACTTTATTTGCACATCTGTGATCAGACGGGTGCAATATTTCT
AAATAATCAATTGATCCTGGAGAATTTCTAGTAAAA

CsGLN5
ACTAGTGATTACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGATTAGAATTGAACATTTTGATATTA
TATCTTGTACTCAATCCGATAAGAACCGACCCACGAACCCGAATTGCCATTCCTAGACGTTTCGGTCATT
TTCGCTCTGGAGAATGAGAACTTGCTCTGGCAAAATGGATTCTCGATTCGCTCTGTCGGATTTTCTGTAT
ATATATAGAGCCAGTGGCATTCAAGTTCCCACAACTCCAAAAAAAGAAAGACCACCTTCATTCAATATAG
ATTGACAAGACATTCTTGTAATTCTCTTGACCAGCCATGGGCTCAGAAGATCGCAGTCTCAGCATTCTCT
TCTTCCCCTTCATGGCCCAAGGCCACATGCTCCCGATGCTCGACATGGCCAAGCTCTTCGCCCTTTACGG
CGTCAAGTCCATCGTCGTCACCACTCCGGCCAACGTCCCCATCGTCAACTCCGTCATCGACCAACCCGAC
GTCTCGACTCTCCATCCGATCCAACTCCGTCTCATCCCGTTCCCTTCGGACACCGGCCTTCCGGAAGGTT
GCGAGAACGTCTCCTCCATCCCTCCACGGGACATGCCGACCGTCCACGTCACCTTCTTCTCCGCCACGGC
GAAACTCCGCGAGCCGTTCGGGAAGGTTCTGGAAGACCTCCGCCCGGACTGCATCGTGACCGACATGTTC
TTCCCCTGGACGTACGATGTCGCCGCCGAGCTCGGGATACCGAGGATCGTATTCCATGGCACCAATTTCT
TCTCCCTCTGCGTCACTGATAGTTTGGAGCGGTACAAGCCGGTCGAGAACCTTCGGTCGGATGCCGAGTC
GGTGATCATCCCCGGCCTTCCCCATCGGATCGAAGTGCTCAGGTCTCAGATACCTGAATATGAAAAATCG
AAGGCGGATTTTGTTAGGGAGGTGACAGAGTCCGAATCCAAGAGCTACGGTGCGGTGGTCAACAGCTTCT
TCGAGTTGGAGCCCGACTATGCTCGTCACTACAGGGAGGTAGTCGGGAGAAGAGCATGGCATATTGGCCC
CCTTGCTCTAGCCAACGATTCCACCACCGACAAATCTTCGCGCGGATACAAGACCGCGATCGACCGCAAC
GATTGCCTGAAATGGCTCGACAGCAAGAGCCTCCGCTCGGTCGTGTACGTTTGTTTCGGGAGCATGAGCG
ACTTCAGCGACGCTCAGCTCCGCGAGATGGCTTCGGGCCTCGAGGCCTCCAACCATCCGTTCATCTGGGT
GGTGAGGAAGAGCGGGAAGGAATGGCTGCCCGAAGGGTTCGAAGAAAGAGTGCAGGAAAGAGGTCTGATC
ATAAGAGGATGGGCCCCGCAGATACTGATCCTGAACCACCGGGCGGTCGGTGGGTTCATGACGCACTGCG
GGTGGAACTCGAGCCTCGAGGCGGTGAGTGCCGGCCTGCCGTTGGTGACGTGGCCTCTATTTGCCGAGCA
GTTCTACAACGAGAGGTTTATCGTCGACGTGCTGAGGATTGGAGTGTCGGTCGGAGCGAAGCGTCACGGG
ATGAGCGCGGAGGAGAGGGAAGTGGTGGAAGCGGAAATGGTAAAGAAGGCAGTGGATGGAGTTATGGGTG
ATGGAGAGGAGGCAGAGGGAAGGAGGAGAAGGGCAAGGGAGCTTGGTGAGAAAGCGAGGAAGGCAGTAGA
AAAGGGGGGATCGTCGTATGAGGACATGAGGAATCTGGTTGAAGAATTGAAGAGTGATTCAAAGCTAACA
GTCGGCTGCTGAACAAATCTCATGGATGCAATTTTAGAATTTCTCCTTTTCTGCTTTCAATTGATTAGAA
TGTGTCTAAAGTGTAATACTGGTTGTGTTTATTTACTTTATCTAGTCAATTTTTGCTG

CsGLN10
CACTAGTGATTCTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACATGGGGATTGTG
AGTCTTACAAGTAGAAGAACACATCAAAGAGCTAGCAATTAGGATGGAAAAACCCCATGCCATCTGCATG
CCATTCCCAGCACAAGGCCACATAAACGCGATGCTCAAAGTGGCAAAGCTCCTCCACTTCAGAGGCTTCC
ACGTCACCTTCGTCCACACCGAGTACAACTCGGCCCGAATCGGATGCTCCAACGAGATGCCGCCTCTCCT
CGCGGCCGACGGATTCCGATTCGAATCGATCCCCGACAGCCTCCTGCCTTCCATCGACCCGAACACCAAA
CAGGACCTGCCGTCGCTTGTCCTCGCGGTCTGTAACGAGTTCCTGTCTCCGTTCCGTGACCTGATGAAGA
GGCTCAACGACCCGTCGTCTGGGGTCCCACCGGTGACCTGTGTGGTGTCCGACTGCATCACCGGATTCAC
GCTGGAGATCGCTAAGGAGTTCGGCATCCCCGATCTCTTCCTATCCTCCTTCAGCGCGTCCGGCTTCATG
GCCCTCTTCCATCTTAAAGAGTTGATGGAGAAAGGACTCGTGCCGATAAAAAGTGAAGATGACCTTACCA
ATGGATATCTGGACACCCCCATCGACTGGATGCCCGGAATGAAGAACATGCGAATCAAAGACCTACCCAG
TTTCGTACGCTCAACAGATCCTAACGACATCTTCTTCAACTTCTCCAAAGACGAAATACAATGCGCACTC
CGAGCCGACGCGATCATTATAAACACGTTCGACGACTTGGAAGAGCCCATTCTCGACGCCATGTCATCCT
CTCTCCTCCCTCCCATCTACTGCTTGGGCCCACTACACAAGCTATGCTCGCAGTACTGTAACAACCAACT
CGAGTCCATCAACGGTAACATGTGGGAGGAGCACACCAACTGTCTGGCCTGGCTCGACCAAAAATCGCCC
GGGTCCGTAATTTACGTGAATTTCGGGAGCATAACGGTAATGACGAGCCAACATCTGGTAGAATTTGCGT
GGGGGCTGGTGAAGAGCGGCTACGAATTCCTGTGGATCATCAGACCCGATGTCGTAATCGGGGAATCGTC
GGTTTTGCCGAAGGAGCTCTTGGACGAGACGATGAAAAGAGGTCTGGTGACGAGCTGGTGCCCTCAGGTG
GACGTTCTGGAACACGCGTCGATCGGAGGATTCTTGACGCACTGTGGGTGGAACTCGACGATGGAGGCTT
TGAGTGCCGGAGTTCCGATGATATGTTGGCCTTTTTTTGCGGATCAGCAGACGAATTGTCGGTACGTGTG
CGCGAATTGGGGGGTGGGGACGGAGATCGATGTCAGAGTGAAGAGAGAGGAGGTGGGGGAGAGGATCAGA
GAGATGATGGGAGGTGGAAAGAAGGGGGAGGAGATGAGGAAGATGGCGTCCAAGTGGAAGGAGAGTGCGT
CCAGAACGATTAGTCCTGGCGGGAAGTCTTTGGAGAACTTTGAGAAAGTTGTCAAACTTCTTATGAAGGA
GAAAGATGGTTAATTAGGGTATATATATATGATCTCATTGTGAATCAATAATACTTCGTAATATATATGT
ATATATACGGAGTCCTTGGACAAATCCAATGTCCTTCCTATATATTATTATTATCTGGAATAATGTCATA

ATAAAATTCGGTGGATCTCCCGCTGAAGGATTCTGCCGACGTATCATTACTCGGTGTCAATGTAACAGAT
CGCTTGCAGTAAAAAAAAAAAAAAAAAAAAAA

CsGLN122
TGTTATCAACGCAGAGTACGCGGGGATTGTGAGTCTTACAAGTAGAAGAACACATCAAAGAGCTAGCAAT
TAGGATGGAAAAACCCCATGCCATCTGCATGCCATTCCCAGCACAAGGCCACATAAACGCGATGCTCAAA
GTAGCAAAACTCCTCCACTTCAGAGGCTTCCACGTCACCTTCGTCCACACCGAGTACAACTCGGCCCGAA
TCGGAAGCTCCAACGAGATGTCGCCTCTCCTCGCGGCCGACGGATTCCGGTTCGAGTCGATCCCCGACAG
CCTCCTGCCTTCCATCAACCCGACCACCAAACAGGACCTGCCGTCGCTTGTCCTCGCGGTCTGTAACGAG
TTCCTGTCTCCGTTCCGTGACCTGATGAAGAGGCTCAACGACCCGTCTGGGGTACCACCGGTGACCTGTG
TGTCCGACTGCATCACCGGATTCACGCTGGAGATCGCTAAGGAGTTCGGCATCCCCGATCTCTTCCTATC
CTCCTTCAGCGCGTCCGGCTTCATGGCCCTCTTCCATCTTAAAGAGTTGATGGAGAAAGGACTCGTGCCG
ATAAAAAGTGAAGATGACCTTACCAATGGATATCTGGACACCCCCATCGACTGGATGCCCGGAATGAAGA
ACATGCGAATCAAAGACCTTCCCAGTTTCGTACGCTCAACAGATCCTAACGACATCTTCTTCAACTTCTC
CAAAGACGAAATACAAAGCGCATTCCGAGCCGACGCGATCATTATAAACACGTTCGACGACTTGGAAGAG
CCCATTCTGGACGCCATGTCATCCTCTCTCCTCCCTCCGATCTACTGCTTGGGCCCATTAACACAAGCTA
TGCTCGCAGTAGATGACAACCAACTCGAGTCCATCAACGGTAACATGTGGGAGGAGCACACCAACTGTCT
GGCCTGGCTCGACCAAAAATCGCCCGGGTCCGTAATTTACGTGAATTTCGGGAGCATAACGGTAATGACG
AGCCAACATCTGGTAGAATTTGCGTGGGGGCTGGTGAAGAGCGGCTACGAATTCCTGTGGATCATCAGAC
CCGATGTCGTAATCGGGGAATCGTCGGTTTTGCCGAAGGAGCTCTTGGACGAGACGATGAAAAGAGGTCT
GGTGACGAGCTGGTGCCCTCAGGTGGACGTTCTGGAACACGCGTCGATCGGAGGATTCTTGACGCACTGT
GGGTGGAACTCGACGATGGAGGCTTTGAGTGCCGGAGTTCCGATGATATGTTGGCCTTTTTTTGCGGATC
AGCAGACGAATTGTCGGTACGTGTGCGCGAATTGGGGGGTGGGGACGGAGATCGATGTCAGAGTGAAGAG
AGAGGAGGTGGGGGAGAGGATCAGAGAGATGATGGGAGGTGGAAAGAAGGGGGAGGAGATGAGGAAGATG
GCGTCCAAGTGGAAGGAGAGTGCGTCCAGAACGATTAGTCCTGGCGGGAAGTCTTTGGAGAACTTTGAGA
AAGTTGTCAAACTTCTTATGAAGGAGAAAGATGGTTAATTAGGGTATATATATATGATCTCATTGTGAAT
CAATAATACTTCGAATATATATGTATATATACGGAGTCCTTGGACAAATCCAATGTCCTTCCTATATATT
ATTATTATCTGGAATAATGTCATAA

CsGLN19
CGGGCCACATGATTCCCCTGACGAACGTGGCTCGAATCTTTGCCAGCCGTCCCAACGTTAAGGCCACCAT
CATCACCGTCCCCTCGCACGTCAACCATTTCAAACCCGCACTCGACATCTGTCCCTACCCCATCCACCTC
CTCACAATCAAATTCCAGTTGCCCGCCCACGATAAGCTCGTCTCCTCCGATCTCATACCGGATAGGCTCT
TCTCTGACGCCTACTTCGAAGCCATAGACTCCCTGGAGGGTCCCTTTGGCCGCCTGCTTGATGAGCACTG
TCCCGATTGTGTCGTCTCTAGCGGGAACTTCCCGTGGTCCGCCGCGCTAGCAAATGCCCGCTGCATCCCC
CGTTTGGTCTTTTACGGGAGCGCTGCCATCACCTCCATTGTCATGCGAACGATTAGTGGCCTCAAGCTCA
CTGAGGGCCGAAGCGACCACGAGCCCTTCATTGTGCCGGGGCTGCCTCATCCCTTGGAGTTCACCAAGGC
ACAAATCATGAGCTCCGCGTCGTATCCAGGTGACACACTCGCAAGGATGGTTAAAGCCCATGCCCAAAGC
TATGGAACAGTTGGGAACAGTTCCTACGAGCTGGAACCAGGCTATGCAAGCAAACTCGACAAATCATATT
TCATCGGGCCCGTCGCACTTTGCACCGGAGACTCAACCGGCGAGGGAGGAAACCGCGACTCCGACGAATA
TCTGAATTGGCTCGACAACCAGAAGAAGTGCTCGGTCCTTTATGTGTGTTTCGGTAGCATTGCTCAGTTC
AGCAATGCACAACTCTTCGAGATCGCAACAGGGCTCGAAGCCTCTGGCCATCCGTTCCTTTGGGTGGTAC
GGGACACAGGGAACACGGCGGAGTGGCTTCCGGAAGGTTTTGTGGAAAGAGTAGGGAGGAGAGGATTGCT
CGTTAGAGGATGGATCGCGCAGATGTCGGTGCTCAACCATACTGCAATCGGAGGGTTCATGACACATTGC
GGATGGGGATCGGTATTGGAGGGAGTGAGTGCCGGATTGCCGATGGTGACTTGGCCGCTTTTTTCGGATC
AGTTCGTGAACGAGAAGCTGATTGTGGACGTGCTGAGAATCGGGGTGCCCGTTGGATCGAAGGTTAACAG
CTCGCGGGAGAAGGAGAGGACATTTGTGGCCGGTGAAATGGTGAGAAAGGCCGTTGGAGAAGTGTTGAGT
GAGGATGGTGAGGAAGCCGAGGAGAGGAGGAAGAGAGCGAGGAAGGCGGCAGAGATGTCAAGGAAAGCAG
TTGAAGAAGGGGGTTCTTCTTATAATGATGTGAGCAAGTTAATACATGAACTAAGAGAGATCAAGAAGAG
TAAAGCAACCAGGTAGAGCTAAGATGCCCATGACAGATTTCAGTATAGTTTGCCGAGTGTTGGAGATTTC
TCCAAATAAAGAGGAACGGAACCTGGGGATATAGGTTTCCCCAACCCAAGTCTTCTTAATGTCAGTTTGT
TGTTGGGTCTCAGCTCAAGTGCTTTAAGGAAATCCATCTGGAATTGGTTAGAAAAAAATGCCCAGGTAGC
CATTTTCAAAACGCCACCCAACCGGGCGGATAAAAATAAAAAAAAAAAAACAAAAAACAAAAGACTCT

CsGLN1DPA

GTGGTATCAACGCAGAGTACCTGGGATTGCGGATAGCCTAGATCCTCTTCTTCCTTTCTCTTCTCTCTCT
CTCTCTCTCTCTCTCTCTCTCTCCAATGGCAAACTCCACCAACAAGCCCCATGCCCTCCTCATCACCTAC
CCTCTCCAAGGCCACATCATCCCATCCGTCCACCTGGCAACAAAGCTCGCCTCCCGCGGCATCGTCGTCA
CCTTCGTCAACACCGAATCCGTCCACCGACAAACTTCCCTAGCAAACAACGGCCGCTCCGACATCTTCTC
TTCCGCCGGAACCGAAGGCCTCGTCGACATCCGTTACCGAGTCGTAAGTGACGGCATGCCGGTCGAGTTC
GACCGCTCTCTCGGCCACGACGAGTTCATGGAGTCCCTACTGCACAATTTTTCCGATCACGTCGAGGCGC
TAGCGCGGAAATTGATGCACGACGATCCCCCGATCACGTGCATAGTTGCCGATACGTTCTTCGTGTGGCC
GTCGGCGTTGTCGAAGAATCTCGGACTTGTTTACGTGTCGTTCTGGACGGAGCCGGCGCTCGTGTTCACG
CTCTACTATCATCTTGATCTCCTGAGGATAAATGGCCATTTTGACTCTCCTAAAGGATCTCGAAAGGACA
CAATCACATACATACCGGGGGTCTCCGCAATCGAGCCTTCCGAACTCATGTCGTACCTCCAAGAATCCGA
CACGACTTCCGTGGTCCACCGAATCATCTTCAAAGCATTCGACGATGCGAAGGGCGCATACTACATACTA
TGCAACACCGTCGAAGAACTCGAGCCCACCACGATCTCGGCCCTACAGAAGATCAAGCCATTCTACGCAA
TCGGTCCGATCTTCCCCAAACGGCTTCACCAAAAGCAGCGTCGCGACGAGTCTATGGGACGAGTCGGACT
GCGCCCAATGGCTCGACTCGAACGCGCGGGATACGTCTTGTACTTGTCTTTTGGGAGTTACGCGCATGTC
AGCAAGAAAGAGCTCGAGGAGATCTCATGCGGTGTTCGGGACAGCAATGTGAGCTTCTTGTGGGTCTTAC
GTCCCGATATCGTTAGCTCAGACGAACCGCACCCGCTTCCGGAAGGTTTCATGGAGGAGATTAGGGATCG
GGGGAAAGTGGTGCAATGGTGCAGCCAGATAGACGTTTTATCACACGGGTCGATAGGAGGGTTGTTGACT
CACTGCGGTTGGATCACTAGTGAAGATGCTATTAGTGCCGGGGTGCCTATGATGTGCTGGCCACAGGTTG
GAGATCAGACGACAAATTGCCGGTACGCTTGCACCGAGTGGGGAATAGGGATGGAGATTGATGCAGATGT
GAGGAGGGAGGAGGTAGGGAAGGTGATAAGAGAGTTGATGGGGGGAGAGAAGGGGAAGGAGATGAAGAGG
AGGGCAGTGGAGTGGAAAGAAAGTGCATTCAGAGCAACAGAACCAGGTGGGAAGTCATCCCTACATATTG
AAAGGCTTCTTAAGGAGGTTATGCTTAAATGAAGCTATCCATGTATGTGATCTTCATCCAAGTTCAACTT
GTATATTTTTGTTATCC

CsGLN70

AAGTTCGCAAGATCTCAATTCCTTCTTCTTCCCAAAAAGCCCTCACTGATGGAGCAACATTTCCTGCTGG
TAACCTATCCGGCACAAGGCCACATCAACCCGGCCCTCCACCTGGCAAAGCGCCTCGCCGAAACCGCCAC
CGCCCGGATCACGTTCTGCACCACGGTCGCAGCCCAACGGAAGATGTTCCCCGAGATCGCCGCCAAACCT
GGCGAGGAAATCCAAGACGGGCCCATCACCTACATCCCCTTACCCGACGGATACGACGACGACGGCGACA
ACAACGGCTCCAAAAGTGGGACCCACGACCTCAACCTCTACTTCTCCCGGTTCAAGACCACGGGCTCCAA
AACCCTCTCCTCCGTGATCGAAACCCTCCGACACCGCGGCCGCCGCGTCAGCTGCATCATATACACCTTG
CTCCCCTGGGTCGCTGACGTGGCACGCCACCACGGGATCCCCTCGGCCCTGTACTGGATCCAGCCGGCCA
CCGTGCTCGCCATTTACTACCACTACTTCCACGGCTACAGCTCCCTCGCGCCGTCCGCTCAGGGCATCGT
AAATTTACCGGGTCTGCCCCCGTTGCGAACCCGCGACCTCCCGTCCTTCCTAACAGTCGAGTCCACGGAC
GACCCCTACGCCGGGATACTCGACACGCTCGGAGAGACGCTCCACTCGCTGGACCGCGAGGCGGAGCGGT
CGAACTCGAAGACGAAGGTCCTAGTGAACACGTTCGACGAGCTGGAAGCGGACGCGCTGTCGGCCGTTGA
CAAGTTGGATATGATTCCGGTGGGTCCCGTACTCCCATGCTCCTCTTCCAGAGGTGACCTCATCAAGCAG
GACGGGGAGGGGTATATGGAGTGGCTGGACTCTAAGGAGGGCAATTCGGTAATCTACGTGTCGTTCGGAA
GCGTGGCAGTGCTGGGAAAGGAGCAGATGGAGGAGATTCTGAAGGGGGTGAGGGGTAGCGGGAGGCCGTA
TTTATGGGTGGTGAGGAGGGAGTGCGGACTGGAAGGAGTCGAGTTGGGTGACGGAGAGAATGGGGTGGTG
GTGGAGTGGTGCTCTCAGGTGAGGGTTTTATCGCATCCTTCGGTGGGATGTTTCGTGACGCACTGCGGGT
GGAACTCGACGGTGGAGGGGTTGGTTTGCGGAGTGCCGACGATTGGGGTGCCGCAGTGGACGGACCAGGC
GACGTGTGCGATGCTGGCGGAGAGGGTTTGGGGGACTGGAGTGAGGGGGGAGGTGAATGGAGATGGGGTA
TTGGAAGGGGAGGAGCTGAAGAGGTGTTTGGATTTGGTTTTGGGAGAGGGGGAGAGGAGTCTTGAGATGA
GGAGGAAAGTGGGGTTTTGGAAGGAGAGAGCTATGGCGGCTGTAAGCTTTGGAGGGTCTTCTGATAGAAA
TTTGAGGGCGTTCATGGATCAGATAGATAACTGTGTAGAATGAAAAATGAAATTAATGTTTTCTAGTTGC
TTACTAAAGTTCATTGTTTGGTAGTGTGAGGCAATGCCATCTATGTACCATGTCATCTATGTACCATGTC
ATCTTTTGCTATTTGGTGGATGTATCAATCCATGGCAAGTCACATGATTAATAAAACTCTTTTTCTCTTT
TTTTCAAAAAAAAAAAAAAAAAAA

CsGLN27

ATCGCAGAGTACATGGGGATAACAAGAACGAGGAACTGAGAGAGAGAGAGAGAGAGAGAGATGGAGTCCA
TAGTAGTGGGAGAGAGGCCCCATGCTGTTTGCATACCAAGCCCATTCCAAGGACACATAAACGCCATGCT

GAAAATGGCGAAGATCCTCTACTCCAAAGGTTTCCGCATCACTTTTGTCCACAATGAGTCCAACTACAAC
CGCATCGTGAGCTCCGGCAGCACCGAGTCCTCACTCGGTAGCCTCGATAACTTCCGGTTTGAGACCGTCC
ATGGTGAAGTTAATTACTCTTCTGCCACCATAACTGAAGACCTCCGGGATCCCATCCGCAGCCTATTGAG
CAGGCTCAACGATCCCCCGGTGACCTGCATAGTCACAGACAGCTTCATGAGCTTCAATCTTGGATTTGCG
GCCGAGATGGGGGTTCCGAGCGTATCCTTCTGCTCGGTGAGCGCTTGTAGCTTCATGGGCTTCCTCCATT
ACAAGGAGCTCCTTGAAAGGGGCATTGTTCCATTTAGGAGTGAAAGCGATCTTCGCAGCGAATACCTCGA
CACGGTAATAGAGTGGATACCAGGGATGACGAAGAACACGCGACTGCGGGATCTTCCAAGTTTCATCCGG
ACAGCAGATCCTAACCTGCAACCATGGTTCCACTTCTGTAAAAATGAATCGCAAGCCGCGTTTCAAGCCA
CCGCAATCGTTATGCACACTTTCGATGAGTTGGAAGGCGCGGTTCTTGATGCACTGTCTTCCATCCTCCC
AGCTCCAGTGTACTCCATAGGCCCTCTGAACCTTCCCTGTTCGCGAATTACCGACGACAAAATCAACCAC
ATTGGGAATAGCCTGTGGAAAGAGGACTCCAGTTTCATGGACTGGTTGGAGGGAAGAAAGGCGGGATCGG
TGGTTTACGTGAACTTTGGGAGTTTGACGAGTATGACGAGCGAACAATCGTTAGAAGTTGCGTGGGGGAT
AGCCGGTAGTGGCCATGACTTCCTGTGGGTGATCAGACCGAACATGGTAGAGGGCGGCGCTTCGGGTATT
CTGCCGTCGGAGTTCTTATCCGAGACCAAAGGAAGGGGTCTGGTGAGTAGCTGGTGCGTACAGGAGGAGG
TTCTCAGACACCCTTCGATCGGGGGGTTCTTTACTCACTGCGGATGGAATTCGATGATGGATGCTATTAG
TGCCGGGGTGCCTATGATGTGCTGGCCACAGGTTGGAGATCAGACGACAAATTGCCGGTACGCTTGCACC
GAGTGGGGAATAGGGATGGAGATTGATGCAGATGTGAGGAGGGAGGAGGTAGAGAAGGTGATAAGAGAGT
TGATGGGGGGAGAGAAGGGGGCACGATCCGCCTAAGA

Partial sequence GH1
CNANNGNAGTCATCGGGCCAGGTCAGATCCTGTGGTGTTCTTTTCGGGGTTCACTCACTATACTCGCCGT
CTTGGTATCCTCTCCACCTCTGCGTGGTGTCTGATTCTCCCCGTGACGACGGCACTGCTAACTTCATCCG
GAGCAGCTTCCCGACCGGATTCATCTTCGGGACAGCGTCCTCGGCGTATCAGTACGAAGGAGCTGCAAAA
GAAGGTGGCAGGGGACCTAGCCTTTGGGATACTTTCACCCACAAGCACGCAGACCAAGTAGCCGATGGAA
GCAACGGAGATGTAGCAATAGACTCCTATCACCGATACAAGGGATATGTGAAAATCATGAAAAAGATGGG
ATTTGACGCCTACAGGTTCTCCATTTCTAATCACTATTGAATTCACGGCCGCCTGCAGGTCGACCATATG
GGAAAGCTCCCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAAT
CATGGTCATAGTGGTTTTCCTGACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGC
TTGGCGTAATCATGGTCATACTTGTTTCCTGANCCAAAATNCTGGGGGTTTTTTTGGCCAATNTTTTCGG
GGGNAAACACACANGGTGAAATGTGTANACGAACTCACTATAGGGCGGAATTGGGCCCGAACGTTCGCAA
GCTCCCGGCCGCCATGGCGGCCGCGGGAATTCGATTCTATACGACTCACTATAGGGCAAGCAGTGGTATC
ACGCAGAGTACATGGGGAAAGGTCAGATCTCTCATGAAATTAATGGGGTTGCTACATCTCTCCCGTCTTT
CTATTACTCTCATCTCCGTGGNGTCTGCTTCGTCTCGTGACGACGCACTGCTACATCAGCCGAGCAGCTT
CCGACCGGGATTCATCTTCNGACAGCGTCCTCCGCCGTAT

Partial sequence GH2
CGATTGGCCCGTCCATCTGGGATACTTTCACCCACAAGCACGCAGACCAAGTAGCCGATGGAAGCAACGG
AGATGTAGCAATAGACTCCTATCTCCGATACAAGGGAGATGTGAAAATCATGAAGGAGATGGGATTTGAC
GCCTACAGGTTCTCCATTTCTTGGAGCAGGATCCTACCAAATGGAAGCCTAAGCGGAGGAATTAACAAGG
AAGGTGTCAAATACTACAATAATTTGATTGATGAACTTATATCCAAAGGCTTGACACCCTTTGTGACAAT
CTTCCATTGGGACACTCCGCAAGGCCTGGAAGACAAATATGGAGGTTTCTTATCCAACCGAATAGCAGAA
GATTTTCGCGATTATGCCGGAGTTTGCTTCCGGGAGTTTGGTGACCGAGTGAAGCACTGGATCACCTTGA
ACGAACCGTGGAGTTATGCTGCGGCCGGTTACGCTACGGGAGATCTAGCACCACGCAGGTGCTCCCCTGA
GATAGCAGCGAATTGCAACGGTGGGGACTCGGGACGCGAACCTTACATTGTGGCACACAGTATATTGTTA
TCCCATGCTGCAGCTGTCAATCTGTACAGACGCAAATATCAGGCGTCTCAGAAGGGAAAAATAGGCGTCA
CACTGATCACAAATTGGATGGTACCCTACTCTAACTCCAAGAGGGGTATGGATGCTGTCGAAAGAGGCAT
AGATTTTATGTTTGGATGGTTTATGGATCCTCTAACTCAAGGGGACTATCCACTTAGCATGAAGGCACTT
GTCCGAGACCGATTGCCAAAATTCACAGAGGAGCAGTCCAAAATGGTTAAGGGATCCTTTGATTTCCTTG
GAGTGAACTATTATTCCACAAACTACGCTAGAAGTATGCCAATAAACGTTATAGCCAAGGATCTCTACAG
TACAGACAGTTGTGCTAACAGAACGCAGGTACGCAATGGAATTCCAATTGGTCCAAGGGCAGCTTCAAAC
TGGCTCTACGTCTACCCAAAAGGAATACTTGATCTCTTGCTCTACACCAAGGAAAAGTACAACAACCCAC
TCATCTACATCACAGAAAATGGTGTCGATGACGCCAACTGTGAAACCGTGATCACCTCGTATTATTCCCT
ATAGGATGATGCAAGTATACACTCTTACCAACAACACCTGTTTATCTTCGAAGAACTATTAGAGAGGAGT
TGACGTGAGGGGATACTTTGCATG

Partial sequence GH3
GGGANCCTTTGATTTCCTTGGAGTGAACTATTATTCCACAAACTACGCTAGAAGTCTGCCAATAAACGTT
ATAGCCAAGGATCTCTACAGTACAGACAGTTGTGCTAACAGAACGCAGGTACGCAATGGAATTCCAATTG
GTCCAAGGGCAGCTTCAAACTGGCTCTACGTCTACCCAAAAGGAATACTTGATCTCTTGCTCTACACCAA
GGAAAAGTACAACAACCCACTCATCTACATCACAGAAAATGGTGTCGATGAGGCCAACAATGAAACCTTA
TCACTCGAAGAAGCCCTAAAGGATGATGCAAGAATACACTTTTACCAGCAACACCTGTTTTATCTTCAAA
GAGCTATTAGAGAAGGAGTTGACGTGAGGGGATACTTTGCATGGTCACTTCTGGACAACTTTGAATGGTT
AAATGGTTATACTGTCAGATTCGGCCTTAACTATGTCGATTACAAGAATGATCTGAAGAGATACCCCAAG
AGTTCTGCTATTTGGTTTAAGAAATTTCTGAAGAAATAGAATTGTAAATCATGTATCTCATTATAGTGAT
AACAATAAAATAAAGTGGGGTTGAATGCAATGCAACAAAGCAAAAAAAAAAAAAAAAAAAAAACGGGGGGA
AA

Partial sequence GH4
CGACGTTGTAAAACGACGGCCAGTGAATTGTAATACGACTCACTATAGGGCGAATTGGGCCCGACGTCGC
ATGCTCCCGGCCGCCATGGCGGCCGCGGGAATTCGATTCTAATACGACTCACTATAGGGCAAGCAGTGGT
ATCAACGCAGAGTACATGGGGAAGGTCAGATCCTCTCCATGAAATTAATGGGGTTCGCTACATCTCTCCC
CGTCTTTCTATTACTCTCCATCCTCGCCGTGGTGTCTGCTTCGTCTCGTGACGACGGCACTGCTAACATC
AGCCGGAGCAGCTTCCCGACCGGATTCATCTTCGGGACAGCGTCCTCGGCGTATCAGTACGAAGGAGCTG
CAAAAGAAGGTGGCAGGGGACCGAGCCTTTGGGATACGTTCACCCACAAGCACGCAGACCAAGTAGCCGA
TGGAAGCAACGGAGATGTAGCAATAGACTCCTATCACCGATACAAGGGAGATGTGAAAATCATGAAAGAG
ATGGGATTTGACGCCTACAGGTTCTCCATTTCTAATCACTAGT

Partial sequence GH5
ACTAGTGATTTTAATGGGGTTCGCTACATCTCTCCCCGTCTCTCTATTACTCTCCATCCTCGCCGTGGTG
TCTGCTTCGTCTCGTGACGACGGCGCTGCTAACATCAGCCGGAGCAGCTTCCCGACCGGATTCATCTTCG
GGACAGCGTCCTCGGCGTATCAGTACGAAGGAGCTGCAAAAGAAGGTGGCAAGGGACTGAGCCTTTGGGA
TACGTTCGTTGAGAAGCACGCAGACCAAGTAGCGGATGGAAGCAATGGAAATGTAGCAATAGACTCGTAT
CACCGATACAAGGGAGATGTGAAAATCATGAAGGAGATGGGATTTGACGCCTACAGGTTCTCCATTTCTT
GGAGCAGGATCCTACCAAATGGAAGCCTAAGCGGAGGAATTAACAAGGAAGGTGTCAAATACTACAATAA
TTTGATTGATGAACTTATATCCAAAGGCTTGACACCCTTTGTGACAATCTTCCATTGGGACACTCCGCAA
GGCCTGGAAGACAAATATGGCGGTTTCTTATCCAACCGAATAGCAGAAGATTTCCGCGATTATGCCGGAG
TTTGCTTCCGGGAGTTTGGTGACCGAGTGAAGCATTGGATCACCTTGAACGAACCGTGGAGTTACGCTGC
GGCCGGTTACGCTACGGGAGATCTAGCACCACGCAGGTGCTCCCCTGAGATAGCAGCGAATTGCAACGGT
GGGGACTCAGGACACGAACCTTACATTGTGGCACACAGTATTTTGTTATCCCATGCTGCAGCTGTCAATC
TGTACAGACGCAAATATCANGCGTCTCANAAGGGAAAAATAGGCATCACACTGATCACCAATTGGATGGT
ACCCTACTCTAACTCCAAGAGGGATATGGATGCTGTCGAAAGAGGC

Partial sequence GH6
AGTCTCCTGGCCGCCATGGTTCGCGGGATTCTTGGAAGACAAATATGGAGGTTTCTTATCCTATCTAATA
GCAGAAGATTTCCGCGATTATGCCGAAGTTTGCTTCCGGGAGTTTGGTGACCGAGTGAAGCACTGGATCA
CCTTGAGCGAACCGTGGAGTTATGCTGCGGACGGTTACGCTACGGGAGATCTAGCACCACGCAGGTGCTC
CCCTGAGATAGCAACGAATTGCAACGGTGGGGACTCGGGACGCCAACCTTACACTGCCCTTCACAGTATG
TTGTTATCCCATGCTGCAGGCTGTCAATCTGTACAGACGCAAGTATCAGGCGTCTCAGAAGGGGAAAATA
GGCATCACACTGGTCACCAATTGGATGGTACCCTATTCTGACTCCAAGAGGGATATGGATGCTGTCGAAT
GAGGCCTAGATTTTATGTTTGGATGGTTTATGGATCCTCTAACTCAAGGGGACTATCCACTTAGCATGAA
GGCACTTGTCCGAGACCGATTGCCAAAATTCACAGAGGAGCAGTCCAAAATGGTTAAGGGATCCTTTGAT
TTCCTTGGAGTGAACTATTATTCCACAAACTACGCTGGAAGTCTGCCAATAAACGTTATAGCCAAGGATC
TCTACAGTACAGACAGTTGTGCTAACAGAACGCATGTACGCAATGGAATTCCAATTGGTCCAAGGGCAGC
TTCAAACTGGCTCTACGTCTACCCAAAAGGAATACTTGATCTCTTGCTCTACACCAAGGAAAAGTACAAC
AACCCACTCATCTACATCACAGAAAATGGTGTCGATGACGCCAACTGTGAAACCTTATCACTCGAATAAT
CCCTATAGGATGATGCAAGAATACACTTTTACCAGCAACACCTGTTTATCTTCAAGAACCATTACAGAGG
AGTTGACGTGAGGGGATACTTTGCATGGTCCCC

Partial sequence GH7
GCAGCTTCAAACTGGCTCTACGTCTACCCAAAAGGAATACTTGATCTCTTGCTCTACACCAAGGAAAAGT
ATAACAGCCCACTCATCTACATCACAGAAAATGGTGTCGATGAGCCACATGAAACCTTATCACTCGAAGA
AGCACTAAAGGATGATGCAGATACACTTTACCAGCACACCTGTTTATCTTCAAGAGCTATTAATCTTCAA
GAACCATTACAGAGGAGTTGACGTGAGGGGATACTTTGCATGGTCCCC

Partial sequence GH8
GATTGGCCCGTCCATCTGGGACACATTCACCCACAAGCACGCAGACCAAGTAGCCGATGGAAGCAACGGA
GATGTAGCAATAGACTCCTATCACCGATACAAGGGAGATGTGAAAATCATGAAGGAGATGGGATTTGACG
CCTACAGGTTCTCCATTTCTTGGAGCAGGATCCTACCAAATGGAAGCCTAAGCGGAGGAATTAACAAGGA
AGGTGTCAAATACTACAATAATTTGATTGATGAACTTATATCCAAAGGCTTGACACCCTTTGTGACAATC
TTCCATTGGGACACTCCGCAAGGCCTGGAAGACAAATATGGAGGTTTCTTATCCAACCGAATAGCAGAAG
ATTTCCGCGATTATGCCGGAGTTTGCTTTCGGGAGTTTGGTGACCGCGTCAAGCACTGGTTCACCATGAA
CGAATCACTAGTG

Partial sequence GH9
CACTAGTGATTCGTTCATGGTGAACCAGTGCTTGACGCGGTCACCAAACTCCCGAAAGCAAACTCCGGCA
TAATCGCGGAAATCTTCTGCTATTCGGTTGGATAAGAAACCTCCATATTTGTCTTCCAGGCCTTGCGGAG
TGTCCCAATGGAAGATTGTCACAAAGGGTGTCAAGCCTTTGGATATAAGTTCATCAATCAAATTATTGTA
GTATTTGACACCTTCCTTGTTAATTCCTCCGCTTAGGCTTCCATTTGGTAGGATCCTGCTCCAAGAAATG
GAGAACCTGTAGGCGTCAAATCCCATCTCCTTCATGATTTTCACATCTCCCTTGTATCGGTGATAGGAGT
CTATTGCTACATCTCCGTTGCTTCCATCGGCTACTTGGTCTGCGTGCTTGTGGGTGAATGTGTCCCAGAT
GGACGGGCCAATC

Partial sequence GH10
GGACTCCGNGCCGCCNTGGTCGCGTGTGATTTCTTGGAAGACGAATATGGAGGTTTCTTATCCTTTCGAA
TAGCAGAAGATTTTCGCGATTATGCCGGAGTTTGCTTCCGGGAGTTTGGTGACCGAGTGAAGCACTGGAT
CACCTTGAACGAACCGTGGAGTTATGCTGCGGCCGGTTACGCTACGGGAGATCTAGCACCACGCAGGTGC
TCCCCTGAGATAGCAGCGAATTGTAACGGTGGGGACTCGGGACCCGAACCTTACATTGTGGCACACAGTA
TATTGTTATCCCATGCTGCAGCTGTCAATCTGTACAGACGCAAATATCAGGCGTCTCAGAAGGGAAAAAT
AGGCGTCACACTGATCACAAATTGGATGGTACCCTACTCTAACTCCAAGAGGGATATGGATGCTGTCTAA
AGAGGCATAGATTTTATGTTTGGATGGTTTATGGATCCTCTAACTCAAGGGGACTATCCACTTAGCATGA
AGGCACTTGTCCGAGACCGATTGCCAAAATTCACAGAGGAGCAGTCCAAAATGGTTAAGGGATCCTTTGA
TTTCCTTGGAGTGAACTATTATTCCACAAACTACGCTAGAAGTATGCCAATAAACGTTATAGCCAAGGAT
CTCTACAGTACAGACAGTTGTGCTAACAGAACGCATGTACGCAATGGAATTCCAATTGGTCCAAGGGCAG
CTTCAAACTGGCTCTACATCTACCCAAAAGGAATACTTGATCTCTTGCTCTACACCAAGGAAAAGTACAA
CAACCCACTCATCTACATCACAGAAAATGGTGTCGATGACGCCAACTGTGAAACCGTGATCACCTCGTAT
TATTCCCTATAGGATGATGCAAGTATACACTCTTACCAACAACACCTGTTTATCTTCGAAGAACTATTAG
AGAGGAGTTGACGTGAGGGGATACTTTGCATG

Partial sequence GH11
ACTAGTGATTTGTAAACCGTGATCACCTCCACGTAGGTAACGGATCTGTATCTTCGTGCCCTCGTTTTGT
ATCTTGTACGATTATAGGACTTTTGGACTAACGGTTATT

Partial sequence GH12
CGCGGNCACGTCAGATCCCCTGNATGTTCTTATTCGGGGTTCACTCACTCTCTCGCAATCTTGGTTTCCT
CTCCAACCTCTGCGTGAAGGTCTGATTCTCCTCATGACGACAATGCTGCTAGCTTCATCTCGAGCAGCTT
CCCGACCGGATTCATCTTCGGGACAGCGTCCTCGGCGTATCAGTACGAAGGAGCTGCAAAAGAAGGTGGC
AAGGGACCTAGCCTTTGGGATACTTTCACCCACAAGCACGCAGACCAAGTAGCCGATGGAAGCAACGGAG
ATGTAGCAATAGACTCCTATCACCGATACAAGGGATATGTGAAAATCATGAAAAAGATGGGATTTGACGC
CTACAGGTTCTCCATTTCTAATCACTATTGAATTCACGGCCGCCTGCAGGTCGACCATATGGGAGAGCTC
CCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCAT
ACTGGATTTCCTGACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAA

TCATGGTCATACTTGTTTCCTGATCTAAAAAATCGGGGGCCATTTTTGCCATATCGTTTGGCGGGGAGAA
AAAAAACAGTGAATTGGAATACGACTCAATATAGGGCGAAATGGGCCCGACGTCGCAGCTCCCGGCCGCC
ATGGCGGCCGCGGGAATTCGATTCTATACGACTCACTATAGGGCAAGCAGTGGTATCACGCAGAGTACAT
GGGGGAAGGTCAGATCTCTCATGAAATTAATGGGGTTGCTTACATCTCTCCCGTCTTTCTATTACTCTCA
TCTCCCGTGNTGTCTGCTTCGTCTCGTGACGACGCACTGCTACATCAGCCGGAGCAGCTTCCGACCGGGA
TCATCTTCTGGACAGCGGTCTCGGCGTATC

Partial sequence GH13

ACTAGTGATTCGTTCATGGTGAACCAGTGCTTAACGCGGTCACCAAACTCCCGAAAGCAAACTCCGGCAT
AATCGCGGAAATCTTCTGCTATTCGGTTGGATAAGAAACCTCCATATTTGTCTTCCAGGCCTTGCGGAGT
GTCCCAATGGAAGATTGTCACAAAGGGTGTCAAGCCTTTGGATATAAGTTCATCAATCAAATTATTGTAG
TATTTGACACCTTCCTTGTTAATTCCTCCGCTTAGGCTTCCATTTGGTAGGATCCTGCTCCAAGAAATGG
AGAACCTGTAGGCGTCAAATCCCATCTCCTTCATGATTTTCACATCTCCCTTGTATCGGTGATAGGAGTC
TATTGCTACATCTCCGTTGCTTCCATCGGCTACTTGGTCTGCGTGCTTGTGGGTAAAAGTATCCCAGATG
GACGGGCCAATCGAATTCCCGCGGCCGCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCT
ATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCG

Partial sequence GH14

AGTGATTCGTTCATGGTGAACCAGTGCTTGACACGGTCACCAAACTCCCGAAAGCAAACTCCGGCATAAT
CGCGGAAATCTTCTGCTATTCGGTTGGATAAGAAACCTCCATATTTGTCTTCCAGGCCTTGCGGAGTGTC
CCAATGGAAGATTGTCACAAAGGGTGTCAAGCCTTTGGATATAAGTTCATAAATCAAATTATTGTAGTAT
TTGACACCTTCCTTGTTAATTCCTCCGCTTAGGCTTCCATTTGGTAGGATCCTGCTCCAAGAAATGGAGA
ACCTGTAGGCGTCAAATCCCATCTCCTTCATGATTTTCACATCTCCCTTGTATCGGTGATAGGAGTCTAT
TGCTACATCTCCGTTGCTTCCATCGGCTACTTGGTCTGCGTGCTTGTGGGTAAACGTATCCCAGATGGAC
GGGCCAATCGAATTCCCGCGGCCGCCATGGCGGCCGGGAGCATGCGACG

Partial sequence GH15

CTAGTGATTCTGGAAGACAAATATGGAGGTTTCTTATCCAACCGAATAGCAGAAGATTTTCGCGATTATG
CCGGAGTTTGCTTCCGGGAGTTTGGTGACCGAGTGAAGCACTGGATCACCTTGAACGAACCGTGGAGTTA
TGCTGCGGCCGGTTACGCTACGGGAGATCTAGCACCACGCAGGTGCTCCCCTGAGATAGCAGCGAATTGC
AACGGTGGGGACTCGGGACGCGAACCTTACATTGTGGCACACAGTATATTGTTATCCCATGCTGCAGCTG
TCAATCTGTACAGACGCAAATATCAGGCGTCTCAGAAGGGAAAAATAGGCGTCACACTGATCACAAATTG
GATGGTACCCTACTCTAACTCCAAGAGGGGTATGGATGCTGTCGAAAGAGGCATAGATTTTATGTTTGGA
TGGTTTATGGATCCTCTAACTCAAGGGGACTATCCACTTAGCATGAAGGCACTTGTCCGAGACCGATTGC
CAAAATTCACAGAGGAGCAGTCCAAAATGGTTAAGGGATCCTTTGATTTCCTTGGAGTGAACTATTATTC
CACAAACTACGCTAGAAGTATGCCAATAAACGTTATAGCCAAGGATCTCTACAGTACAGACAGTTGTGCT
AACAGAACGCAGGTACGCAATGGAATTCCAATTGGTCCAAGGGCAGCTTCAAACTGGCTCTACGTCTACC
CAAAAGGAATACTTGATCTCTTGCTCTACACCAAGGAAAAGTATAACAGCCCACTCATCTACATCACAGA
AAATGGTGTCGATGAGCCACATGAAACCTTATCACTCGAAGAAGCACTAAAGGATGATGCAGATACACTT
TACCAGCACACCTGTTTATCTTCAAGAGCTATTANNGAAGGNNTGACGTGA