IHSM La Mayora
Instituto de Hortofruticultura Subtropical y Mediterránea

UNIVERSIDAD DE MÁLAGA

The Fra Allergens and their Role in the Control of Flavonoid Biosynthesis in Strawberry Plants

Tesis Doctoral

Ana Casañal Seoane

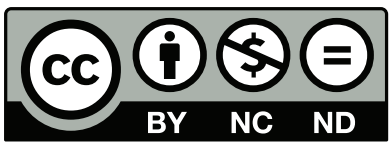
Departamento de Biología Molecular y Bioquímica
Universidad de Málaga, 2014

Publicaciones y
Divulgación Científica

UNIVERSIDAD
DE MÁLAGA

AUTOR: Ana Casañal Seoane

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga



IHSM La Mayora
Instituto de Hortofruticultura Subtropical y Mediterránea

UNIVERSIDAD DE MÁLAGA

Departamento de Biología Molecular y Bioquímica

Facultad de Ciencias

D. VICTORIANO VALPUESTA FERNÁNDEZ, Catedrático del Departamento de Biología Molecular y Bioquímica de la Universidad de Málaga.

D. JOSÉ ANTONIO MÁRQUEZ GÓMEZ, Team Leader, Head of the High-Throughput-Crystallization Facility del European Molecular Biology Laboratory (EMBL-Grenoble)

INFORMAN: que Dña. Ana Casañal Seoane ha realizado bajo nuestra dirección el trabajo experimental conducente a la elaboración de la presente Memoria de Tesis Doctoral.

Y para que así conste, y tenga los efectos que correspondan, en cumplimiento de la legislación vigente, extiendo el presente informe en Málaga, a 14 de mayo de 2014.

Fdo: Victoriano Valpuesta Fernández

Fdo: José Antonio Márquez Gómez

The present work was supported by the Ministerio de Ciencia e Innovación (MICINN), grants BIO2007-67509 and BIO2010-15630, FPI fellowship BES-2008-001964 and partially by the P-CUBE project of the European Commission FP7/2007-2013, grant No. 227764.

APPOINTED BOARD OF EXAMINERS

President

Prof. Javier Márquez Gómez

Department of Molecular Biology and Biochemistry

Universidad de Málaga

Secretary

Dr. José F. Sánchez Sevilla

Instituto de Investigación y Formación Agraria y Pesquera (IFAPA)

Junta de Andalucía, Málaga

Vocals

Dr. Carlos Fernández Tornero

Centro de Investigaciones Biológicas

CSIC, Madrid

Prof. Bruno Mezzetti

Department of Agricultural, Food and Environmental Sciences

Università Politecnica delle Marche, Italy

Dr. Araceli Díaz Perales

Centro de Biotecnología y Genómica de Plantas (UPM-INIA)

Universidad Politécnica de Madrid

Substitutes

Dr. Sonia Osorio Algar

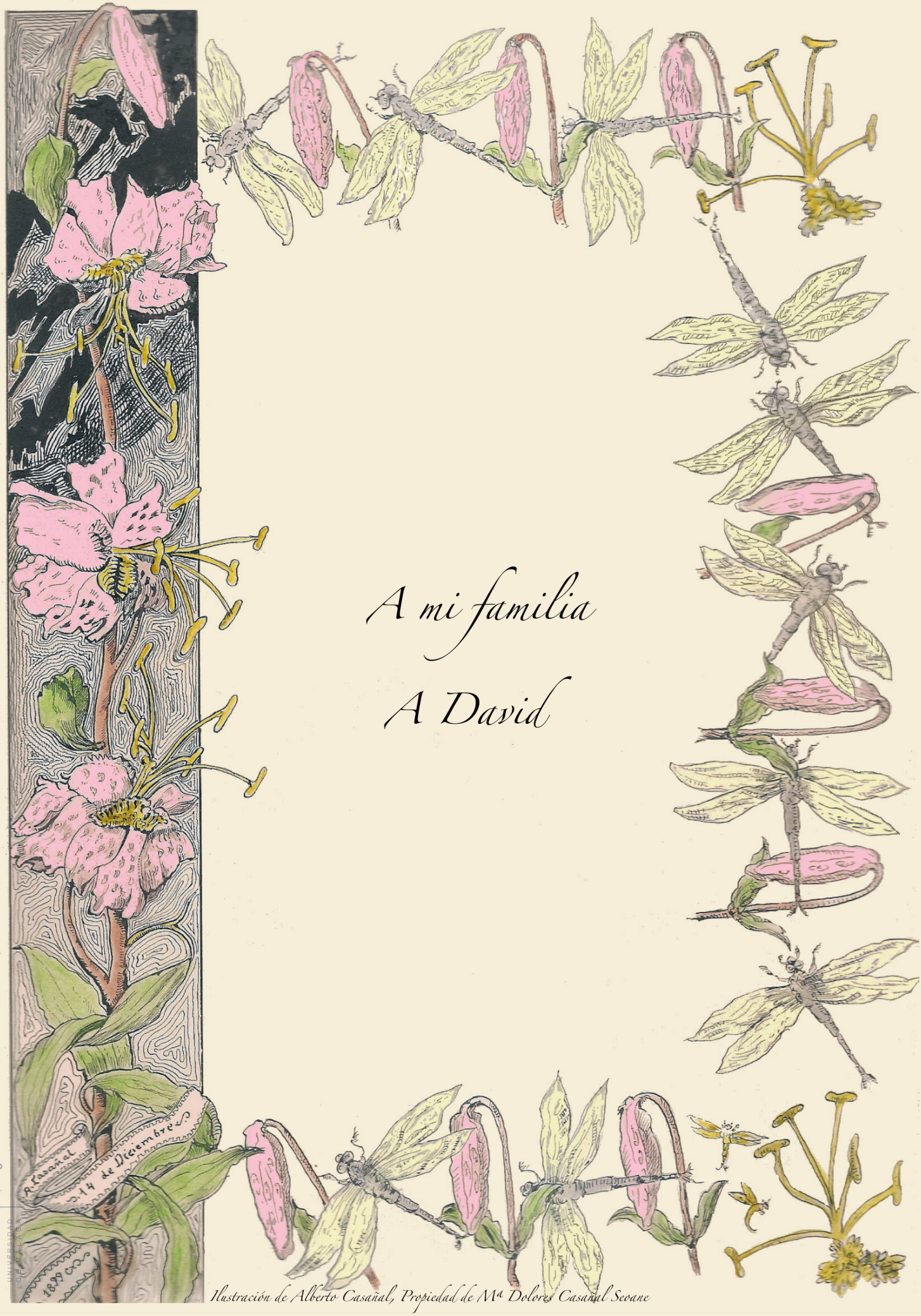
Department of Molecular Biology and Biochemistry

University of Málaga

Dr. Rui M. Tavares

Department of Biology

Universidade do Minho, Portugal



A mi familia

A David

Ilustración de Alberto Casañal, Propiedad de Mª Dolores Casañal Seoane

Casi siempre que te embarcas en un viaje largo, parece que nunca vas a llegar al final. Pero es cierto que todo llega. Ahora me encuentro en este momento, a punto de entregar mi Tesis Doctoral, que ha sido posible gracias a la ilusión y esfuerzo de todas aquellas personas que me han acompañado durante esta etapa profesional y de mi vida.

Ante todo agradezco a mis directores de tesis, Vito y Josan, sus aportaciones globales y particulares a este trabajo, que han hecho posible alcanzar esta meta.

A Vito, mi padre científico. Por confiar en mi desde el principio hasta el final, por abrirme tantas puertas y oportunidades a lo largo de estos años y preocuparse siempre de mi formación. Por la ilusión que pone en todos sus proyectos. Por sus enseñanzas para que no pierda la perspectiva científica y tenga siempre presente que lo que buscamos es contestar las preguntas biológicas que nos planteamos, y que hay que afrontarlas usando todas aquellas herramientas que sean necesarias, aprendiendo y estudiando y, sobre todo, disfrutando del camino. Por sus siempre sabias decisiones. Por transmitirme cómo ser generosa en la Ciencia, algo que creo que es fundamental para avanzar en los proyectos y en las relaciones interprofesionales. Gracias por su apoyo en los momentos difíciles, tanto profesionales como personales, por darle de nuevo sentido al trabajo que teníamos entre manos y tener alguna anécdota para sacarme una sonrisa. Porque esta última etapa, de escritura y discusión de la tesis, ha sido muy enriquecedora y amena. Simplemente, me siento afortunada de haber podido trabajar con él.

A Josan. Por introducirme en el mundo de la Biología Estructural y transmitirme sus conocimientos con ilusión. El mejor profesor que podría haber tenido en cristalografía. Es una suerte haber podido contar con su experiencia. Y el resultado ha sido que ahora voy a seguir trabajando en esta disciplina, a la que me siento totalmente "enganchada". Por enseñarme a organizar mi trabajo y a organizar mis ideas. Por su entusiasmo y positividad en todo momento. Por estar siempre dispuesto a discutir nuevas ideas y por sus decisiones siempre razonadas. El tiempo que pasé en Grenoble fue fantástico, no sólo me aportó una gran motivación profesional, sino también se convirtió en una experiencia personal de valor incalculable. Gracias por sus buenos consejos en cuanto a mi futuro científico y por su preocupación por hacer mi vida más fácil fuera de casa. Siento un gran agradecimiento hacia él.

A Miguel. Por su supervisión y ayuda en este trabajo, por aportar siempre buenas ideas y por su gran motivación y energía.

Many thanks to Krasi. For the opportunity he gave me to join his group when I was still an undergraduate student. His support and enthusiasm were essential for me to decide to go further into research.

A mi compañera de máster y tesis, Tábata. Gracias por su apoyo incondicional, por estar siempre dispuesta a ayudarme en el laboratorio, por los ratos de descanso juntas, por discutir los experimentos conmigo y por confiar siempre en mi. Ha sido una pieza esencial en esta tesis. Me ha encantado compartir con ella nuestro día a día. No es sólo mi compañera y amiga, sino mi hermana peruana.

A Natasa, por haber estado siempre ahí, por sus buenos consejos de laboratorio y también personales. Por su amistad y complicidad. Las tres hemos formado un buen equipo, y las voy a echar mucho de menos. A Juanjo, sin ti no podríamos haber sido los "cuatro fantásticos".

A mis compañeros de laboratorio, sin los que hubiese sido imposible llegar hasta el final. A Paqui y Cristina. A Cristina por enseñarme todo sobre las Fra, por haber dejado en mis manos su trabajo y por lo bien que me lo pasé trabajando con ella en el laboratorio. A Paquita, por conseguirme siempre todo lo que necesitaba para poder hacer mis experimentos, por darme serenidad cuando yo no la tenía y por sus siempre buenos consejos. A Arnaldo, por transmitirme su ilusión por la investigación, fue un compañero clave en mis inicios en el laboratorio; y por su siempre "buena honda". A Davilín, que siempre ha sido un chorro de aire fresco; gracias por su buen rollo y buenos consejos desde el día en que fue mi profesor de prácticas de laboratorio hasta hoy. Se que va a ser grande. A Carmela, por los buenos ratos que he pasado con ella, y por lo bien que trabaja en el laboratorio. A Karen, por enseñarme con emoción. A Vitor, el rey del Western. A Ali, que ha sido fundamental, sobre todo en esta recta final, gracias mil. A Jessi, por su entusiasmo y compañerismo; por aportar un halo de positividad al grupo y por ser mi química de referencia. Se que triunfará en la Ciencia. A Eli, por su gran generosidad, a la que le deseo mucho ánimo con lo que le queda de tesis. A José, mi compañero de poyata, por su ayuda en el invernadero y en el laboratorio, por los ratos interminables de café, que siempre recordaré con cariño. A Sonia, por tener siempre un buen consejo. A Aure e Itziar, por su ayuda en los comienzos de mi tesis. Y a todos los compañeros que han ido pasando por el laboratorio y han aportado su granito de arena en este proceso.

A Abeluso, por enseñarme a ver el lado positivo de las cosas, por animarme desde que le conocí, por discutir experimentos conmigo, por corregirme los abstracts, por darme ideas para la tesis y animarme con el postdoc y, sobre todo, por los buenos ratos de cous-cous, que siempre me llevaré conmigo.

A los alumnos internos, Carlos, Delphine (fantástica seguidora de las Fra), Cristóbal (mi compi en las noches del lab), Álvaro y a los "achichincles" Miguel, Sito, Pepe, Mariri y Álvaro "alas de fuego", por animar el laboratorio con su alegría, ilusión y ganas de aprender. En especial, gracias

a Carlos, que se implicó en nuestro proyecto como si fuera suyo, trabajó sin parar siempre que fue necesario y me ayudó muchísimo con la localización subcelular. Qué bien lo he pasado trabajando con él.

Gracias a todos los que han colaborado desde el IFAPA (Churriana), por su inestimable ayuda, apoyo y buena disposición. En especial a Pepe Sevilla, por haberme iniciado en el mundo del RNA-Seq, por estar siempre disponible incluso cuando no tiene tiempo, por sus sugerencias y ayuda. Gracias también a Iri, por ser siempre una persona de referencia a quien recurrir en Churriana.

A Mayte, mi compañera del alma. Trabajar con ella ha sido una de las mejores cosas que me ha pasado durante la tesis. Le agradezco enormemente sus enseñanzas sobre la fisiología de la fresa. Es la reina en el campo y en la estadística. Su contribución a mi tesis ha sido esencial. Gracias a ella todo empezó a ser realmente divertido.

A Eduardo y Araceli, por acogerme como a "una genética más". Por estar siempre dispuestos a compartir su experiencia, por sus críticas constructivas y buenas aportaciones a mi trabajo diario. Gracias a Araceli por sus clases particulares de genética y por compartir el aprendizaje de análisis de RNA-Seq conmigo.

A mis compañeros de genética. A Edgar, por sus consejos día a día en el laboratorio y por los buenos ratos bebiendo "chiqui-cervezas" en mi jardín. A Ana, por siempre estar dispuesta a ayudar y por su siempre buen humor. A Zaira, por ser nuestra mejor organizadora de eventos. A Manolo, por su transferencia de conocimiento sobre "digestiones", ¡es un crack! Cuánto nos hemos reído juntos; siempre nos quedará "el Filo". A Humbertiño, por enseñarme que se puede disfrutar mucho dando charlas, y por los ratos comiendo jamón juntos. Y a la gente del departamento, con las que he compartido tantos "lab-meetings": Carmen, Adela, Ainhoa, José, y muchos más.

A Rosa y Alberto. Porque siempre me han ayudado con sus buenos consejos, y porque son una inspiración para mi. Gracias por estar siempre ahí.

Many thanks to the people that made my work possible at the HTX Lab in Grenoble, Josan, Uli, Florine, Martin, Esther, Dominika, Daniel, Gael, Vincent and Guillaume. Thanks to Josan for giving me the opportunity to work within his group and for providing me the support I needed to advance in our project. Thanks to Uli, for his enthusiasm at the synchrotron, for teaching me the tricks of the many different programs we used together, for being always willing to help and discuss about experiments. I have learnt a lot from him, especially how fun it can be

to share a project. It was a real pleasure to have the opportunity to work with him. I thank him for his always good advices. Thanks to Florine, for her great experience in the lab, for teaching me how to purify proteins. I will never forget her favourite phrase “it depends”. Thanks as well for the good moments out of the lab. Thanks to Martin for his help with the crystallization experiments and for the great trips around Grenoble. Thanks to Gael for his advice about informatics, travelling and living in Grenoble. Thanks to Vincent for helping me with planning of the experiment and for the nice tea/coffee breaks. Thanks to Daniel, Esther and Dominika for their help in and out of the lab. Thanks to Guillaume for the good moments at the office and for taking over the ITC experiments. Thanks to all of them for making my French experience much more enjoyable.

A los españoles y latinos de Grenoble que me ayudaron en el laboratorio del EMBL y también hicieron mi estancia mucho más divertida: Cristina, Itxaso, Ester, Silvia, Juan, Bruno, Enrico, Andrés…y los que me dejo.

A la gente del Departamento de Química Física de la Universidad de Granada, que hicieron posible que terminase los experimentos de calorimetría que inicié en Grenoble. Gracias a Irene Luque por acogerme en su laboratorio, por su ayuda, discusiones y buenos puntos de vista. A Fran, “mi compi”, por hacer más llevadero el trabajo de laboratorio, ayudarme sin descanso con el ITC y por los buenos momentos que pasamos juntos tomando tapitas. Al resto de los compañeros con los que compartí mi experiencia granadina: Adela, Sara, Ana, Fadia, Fátima, Javi, Bertrand y David.

Many thanks to the people I have met in my short-term stay in Freising. Thanks to Willi Schwab and all the members of his group that gave me the support I needed and made me feel like at home.

A Martina, mi orientadora. Por apoyarme y darme buenos consejos constantemente. Por hacerme poner los pies en la tierra. Por transmitirme su positividad y templanza. Por su guía.

A Toñi, por confiar en mi cuando estuve en Bioazul, por transmitirme su fortaleza y energía, por ser siempre tan positiva y hacer que creyera en mi. Ante todo, por ser mi amiga. Es un modelo a seguir.

A Manuel y Mariquilla. A Manuel, gracias por su ayuda, su sabiduría, su compañía y las sesiones nocturnas de fotografía; y por los buenos ratos de té y fruta que hemos disfrutado juntos. Y María, mi amiga de toda la vida, con la que voy de la mano en esto. Ha sido mi serenidad y energía, sobre todo en esta última etapa. Le agradezco mucho a mi Mariquilla.

A Helena, para la que no tengo suficientes palabras para agradecer todo su apoyo siempre incondicional. Es uno de mis grandes pilares, y una pieza esencial para mi en la vida. A Gaby, también un aliado importante, gracias por su amistad.

A mi otra Elena, mi compañera y amiga de la carrera. Gracias por su ayuda y ánimos en todos los momentos difíciles. Gracias por estar ahí, por tener una visión positiva de las cosas y por su inagotable energía. Gracias por todos los buenos momentos que hemos pasado juntas desde hace ya tantos años, y por los que nos quedan aún por pasar.

A Jason, por su determinación y sus innumerables correcciones del inglés.

Gracias a mi familia, a quien va dedicado este trabajo. Sin ellos verdaderamente esta tesis no hubiese sido posible.

A Jose y a Inma, por adoptarme en la familia como a una hija más. Su apoyo ha sido vital en estos últimos años. Y gracias a Raquel también, siempre dispuesta cuando la necesitas.

A mi yeya, tan moderna y eternamente con un buen consejo que dar. Gracias por su cariño. Y al yeyo por su alegría permanente.

A mis abuelos, Eduardo y Lolita. Los llevo en el corazón. De ellos he heredado la constancia, el tesón, la ilusión y la alegría. Sin ellos no hubiese llegado hasta aquí. Ojalá estén orgullosos de mi.

A mis padres y mi hermana María. Tengo mucho que agradecerles. A Luis por confiar siempre en mi, por su apoyo absoluto, por sus valores familiares, por haberme ayudado a llegar hasta donde me propusiese, y por escucharme. A Lola, mi madre, a quien quiero muchísimo. Somos iguales. Porque con su cariño, cuidado y energía siempre ha sido todo más fácil. No puedo expresarle con palabras todo mi agradecimiento. Soy quien soy gracias a ella. Y a mi Mariri, por sus buenos consejos, su alegría y hacerme sentir que siempre estará ahí cuando la necesite.

A David, mi compañero en la vida. Por ser la persona más optimista que conozco y encontrar siempre el lado positivo a las cosas. Le admiro y le quiero. Gracias por hacer mi vida más fácil y mejor. Por quererme tal y como soy. Por estar siempre a mi lado. Por enseñarme a disfrutar de la vida. Por apoyarme en las horas de trabajo interminables. Por cuidarme. Por acompañarme sin dudarlo en la nueva etapa que se nos presenta. Por ser como es. Porque somos un equipo. Sin él, nada sería igual.

Index

General Introduction

Strawberry (*Fragaria xananassa*) is one of the most consumed berry crops in the world (Hirakawa *et al.*, 2014), not only for its delicate aroma and flavour but also for its high nutritional value (Tulipani *et al.*, 2008; Giampieri *et al.*, 2012). The production of strawberries in 2012 exceeded the 4 million metric tones, with global distributions of 42.8%, 29.2%, 18.4%, 8.8% and 0.8% in America, Europe, Asia, Africa and Oceania, respectively (FAOSTAT, 2012; http://faostat3.fao.org). Spain is a leading country in strawberry production, exportation and breeding programs, which generate important incomes both in fresh fruit sales and royalties. In 2012, Spain was the first exporter and fourth producer country of strawberries in the world (FAOSTAT, 2012; http://faostat3.fao.org), and concentrates 95% of its total production in Southern Spain (Huelva).

The strawberry belongs to the family Rosaceae in the genus *Fragaria*. The most common native species is *F. vesca*, which is diploid and proposed as a model for the genus (Shulaev *et al.*, 2011). The most important cultivated strawberry is an octoploid plant and derives from the cross of two octoploid strawberry species, *F. chiloensis* and *F. virginiana* (Rousseau-Gueutin *et al.*, 2009). The strawberry fruit results from the development of the flower receptacle that is composed of an internal pith, a fleshy cortex, an epidermis, and a ring of vascular bundles with branches leading to the achenes, the true fruits. Each achene contains a single seed and a hard pericarp and each receptacle may contain a few hundreds of achenes (Perkins-Veazie, 1995). In the cultivated strawberry *F. xananassa*, fruit development is divided into different stages: flower/anthesis, green, white, turning, and red (Perkins-Veazie, 1995; Fait *et al.*, 2008). The fruit reaches the white stage approximately 21 days after anthesis and it is ripen in 30 to 40 days, this depending on the strawberry cultivar and the growing conditions (Figure 1) (Perkins-Veazie, 1995).

Fruits can be classified as climacteric and non-climacteric based on their respiration patterns and ethylene production during fruit ripening (Grierson, 2013). Classically, the definition of a climacteric fruit has three components: (1) An increase in ethylene production; (2) An associated increase in the rates of respiration and (3) Phenotypic and genetic changes in the fruit that lead them to be identified as ripe (Iannetta *et al.*, 2006). Unlike other Rosaceae family crops, such as apple and peach, the strawberry has been traditionally considered to be non-climacteric, but recent studies in transgenic lines insensitive to ethylene has revealed not only that full ripening of the achenes, the true fruits, requires the action of this hormone, but also ripening of the receptacle partially resembles the ripening of climacteric fruits (Merchante *et al.*, 2013).

For all these reasons, strawberries are among the most studied berries from the agronomic, genomic and physiological points of view.

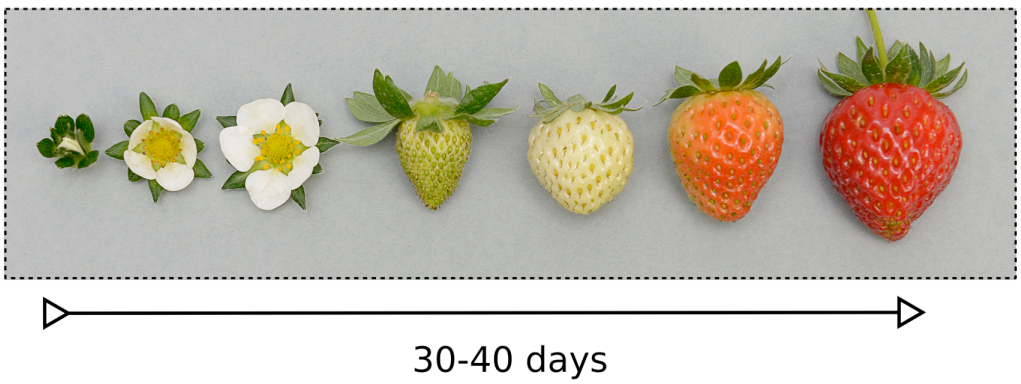


Figure 1. Strawberry development and ripening. Plants of the strawberry cultivar Camarosa grown in Southern Spain. The developmental stages shown here are, from left to right, close flower, open flower, green, white, turning and red.

Phenolic Composition of Strawberry Fruits

Bioactive compounds are defined as essential and non-essential compounds that occur in nature, are part of the food chain, and can be shown to have an effect on human health (Biesalski *et al.*, 2009). Strawberries are known to have an exceptional composition in secondary metabolites, many of them considered as bioactive compounds, thus being important the consumption of this fruit for human health (Fait *et al.*, 2008; Diamanti *et al.*, 2012; Giampieri *et al.*, 2013). In particular, strawberries are rich in minerals, vitamin C, folate and phenolic compounds, which are known by their high antioxidant capacity (Scalzo *et al.*, 2005; Tulipani *et al.*, 2008; Giampieri *et al.*, 2013). Phenolic compounds are also responsible for the colour and flavour of fruits, and provide protection against pathogens and adverse environmental conditions such as exposure to UV light (Aaby *et al.*, 2005, 2007).

Phenolic compounds originate from phenylalanine via the phenylpropanoid and the flavonoid pathways (Tohge *et al.*, 2013) and their biosynthesis occurs in a two-phase fashion during strawberry fruit development (Halbwirth *et al.*, 2006; Aaby *et al.*, 2007; Fait *et al.*, 2008). During early stages, flavonoids, mainly flavan-3-ols and proanthocyanidins, accumulate to high levels and provide an astringent flavour to immature fruit. Later in development, when fruit starts to ripen, other flavonoids, such as anthocyanins and cinnamic and coumaric acid derivatives, and flavonols accumulate to high levels and contribute to fruit and flower coloration (Halbwirth *et al.*, 2006; Fait *et al.*, 2008). Within the extensive variety of phenolics in

strawberry, the major class is represented by flavonoids (mainly anthocyanins, with flavonols, and flavanols giving a minor contribution), followed by hydrolysable tannins (ellagitannins (ETs) and gallotannins) as the second most abundant class, and phenolic acids (hydroxybenzoic acids and hydroxycinnamic acids) together with condensed tannins (proanthocyanidins) being the minor constituents (Giampieri *et al.*, 2012). Anthocyanins, quantitatively the most important type of polyphenols in ripe strawberry, are mainly represented by pelargonidin- and cyanidin- glucoside derivatives, and are known as the main pigments conferring strawberries their characteristic colour (Tulipani *et al.*, 2008; Muñoz *et al.*, 2011). To that end, the phenolic composition of strawberries not only affects significantly the quality of the fruits, contributing to their sensorial-organoleptic attributes but also to their nutritional value. Therefore, growth and ripening of strawberry fruits is an important field of research, which includes the synthesis of anthocyanins and flavour compounds during the late stages of ripening.

PR-10 Proteins

Pathogenesis-related proteins (PR) were first classified on the basis of their characteristic of being induced in plants in response to pathogens. Later, the term was extended to related proteins induced in association with pathogen resistance responses that do not necessarily imply a functional role in defense (van Loon *et al.*, 2006). More recently, not only pathogen attack conditions were considered under this definition, but also related situations mimicking pathogen stimuli, such as responses to abiotic stress, application of chemicals or wound responses (Sels *et al.*, 2008; Lebel *et al.*, 2010). The PR proteins, which are widely distributed throughout the plant kingdom, currently comprise 17 families, and are classified as PR-1-PR-17 based on their primary structure and their common biochemical and biological properties (van Loon *et al.*, 2006; Sels *et al.*, 2008; Fernandes *et al.*, 2013). Despite numerous studies, the function of most of the PR representatives remains elusive. In particular, the role of the PR-10 proteins is still not well understood. Most of PRs are extracellular, but some others are found in the cytoplasm, mainly in the vacuole. PR-10 proteins seem to be generally cytosolic and are therefore classified as intracellular PR (IPR) proteins (Lebel *et al.*, 2010; Fernandes *et al.*, 2013).

PR-10 proteins include a closely related group of major tree pollen allergens and food allergens that belong to the Bet v 1 superfamily (Markovic-Housley *et al.*, 2003). The structure of Bet v 1 proteins share a common fold characterized by the presence of three α-helices and one antiparallel β-sheet that enclose a central cavity, which is often involved in the binding of hydrophobic ligands (Mogensen *et al.*, 2002; Markovic-Housley *et al.*, 2003; Casañal *et al.*, 2013b; Seutter von Loetzen *et al*, 2013). In strawberry three new members of the Bet v 1 family,

the FaFra1, FaFra2 and FaFra3 proteins, have been identified (Hjerno *et al.*, 2006; Muñoz *et al.*, 2010). Emanuelsson and co-workers provided the first evidence of the relationship between the FaFra proteins and the flavonoid biosynthetic pathway (Hjerno *et al.*, 2006). Later, it was established that the FaFra proteins play an essential role in the pigment formation of strawberry fruit (Muñoz *et al.* 2010), which is mostly determined by the glycosylated anthocyanins of the flavonoid pathway, such as cyanidin 3-O-glucoside and pelargonidin 3-O-glucoside. Effectively, transient silencing of these genes in fruits by RNAi caused a decrease in the accumulation of these metabolites, which was parallel to a reduced expression of the genes encoding for phenylalanine ammonia lyase (*PAL*) and chalcone synthase (*CHS*), key enzymes of the anthocyanin biosynthesis pathway (Muñoz *et al.*, 2010). Interestingly, in *FaFra* silenced fruits there was an adjustment of the concentrations of some other intermediates of the flavonoid pathway, since the concentration of compounds such as kaempferol 3-O-glucoside and pelargonidin 3-malonyl-glucoside were decreased, while other intermediates as catechin and derived proanthocyanidins, were accumulated (Muñoz *et al.*, 2010). These results supported a role for FaFra proteins in the regulation of the flavonoid biosynthesis in strawberry fruits. However, the molecular basis for this function still remains unknown. The homology of the FaFra to the Bet v 1 proteins, suggests that FaFra could bind ligands, an interaction that might be of relevance for their mechanisms of action (Liu & Ekramoddoullah, 2006; Radauer *et al.*, 2008). In this work, our aim was to contribute to the understanding of the function of the FaFra proteins and their involvement in the development of strawberry fruits by investigating the presence and expression patterns of the members of the *FaFra* gene family in strawberry, searching for potential natural ligands of the FaFra proteins, and performing biochemical and structural analysis of the them. Moreover, this study could help gain significant insight into the function of the PR-10 proteins, which are widespread in plants.

Aims

1. To characterize the strawberry (*Fragaria xananassa*) *Fra* multigene family, to gain insight into their role in the control of the flavonoid biosynthesis pathway.

2. To apply biophysical techniques to identify potential natural ligands of the FaFra family members.

3. To perform a structural analysis of FaFra proteins in order to understand their molecular function.

Chapter 1

Characterization of the Strawberry (*Fragaria xananassa*) *Fra* Multigene Family

INTRODUCTION

The study of secondary metabolic pathways is the subject of intense research, as their intermediate metabolites have important medical and biotechnological applications. Plants are potential sources of natural bioactive compounds, and an impressive and growing number of them have been identified that have important health benefits. These compounds can act as antioxidants, enzyme inhibitors and activators, suppressors of receptor activities, and inducers and inhibitors of gene expression, among other actions (Kris-Etherton *et al.*, 2004). Flavonoids and phenolic acids are among the most important secondary metabolites in plants. The study of dietary flavonoids is currently of great interest due to their medicinal, pharmaceutical and nutritional properties. This is the reason why they have been sometimes termed as "nutraceuticals" (Hichri *et al.*, 2011). These metabolites present potential antioxidative and anticarcinogenic activities (Diamanti *et al.*, 2012; Ghasemzadeh & Ghasemzadeh, 2011). Antioxidants are specific compounds that protect cells against the damaging effects of free radicals. In the case of flavonoids, they exert their antioxidant activity by scavenging radical species, such as reactive oxygen species (ROS), suppressing their formation, or up-regulating antioxidant defense (Cotelle, 2001). In strawberries, which are rich in flavonoids and present important antioxidant capacity (Giampieri *et al.*, 2013), polyphenols have been proved to protect gastric mucosa against the damages caused by agents such as ethanol (Alvarez-Suarez *et al.*, 2011) and, further consumption of strawberries, has been correlated with beneficial effects on the plasma antioxidant status and erythrocyte resistance to oxidative damage (Tulipani *et al.*, 2011a). The anticancer properties of flavonoids are also associated with their role as antioxidants and a significant number of studies support their involvement in the prevention and inhibition of cancer growth (Ghasemzadeh & Ghasemzadeh, 2011). For example, isolated flavonoids from strawberry, including kaempferol, quercetin, anthocyanins, coumaric acid and ellagic acid, were shown to inhibit the growth of human oral (KB, CAL-27), colon (HT-29, HCT-116), and prostate (LNCaP, DU-145) tumor cell lines (Zhang *et al.*, 2008).

In addition to their health benefits flavonoids exhibit a diverse spectrum of biological functions and play an important role in the interaction between plants and their environment. Flavonoids protect plants against ultraviolet light damage by absorption of UV radiation (Aaby *et al.*, 2005, 2007a, 2007b), serving as UV screen. Thus, one of the most general responses of plant seedlings to UV light is the transcriptional activation of flavonoid biosynthetic genes (Koes *et al.*, 1994, Harborne & Williams, 2000). Quercetin glycoside, a flavonoid involved in UV-protection, has been reported to accumulate in apple fruit, being this effect related to its

protective role against the UV-induced damage to photosystem II in apple skin (Solovchenko *et al.*, 2003). Other class of flavonoids, such as anthocyanins, which are pigments absorbing visible light and responsible for the colour of pollen, flowers and fruits, are well known by their implication in the attraction of pollinators and seed dispersal (Lepiniec *et al.*, 2006). In addition, flavonoids are also important in fertility and reproduction as, for instance, they control pollen fertility and auxin transport (Thompson *et al.*, 2010). Certain flavonoids participate in the interaction between plants and other organisms such as symbiotic bacteria and parasites. This is the case of *p*-coumaric acid and naringenin, metabolites exudated from roots, that play a role in stimulation and inhibition of quorum-sensing, respectively (Schaefer *et al.*, 2008; Vikram *et al.*, 2010). Flavonoids also contribute to resistance against pathogens, ranging from bacteria, to fungi and insects, being quercetin one of the most widely studied flavonols with such function (Hassan & Mathesius, 2012). Further, flavonoids like proanthocyanidins, which are responsible for the bitter taste in plants, protect against feeding by herbivores (Harborne & Williams, 2000; Aron & Kennedy, 2008).

The flavonoid pathway is one of the best-studied biosynthetic pathways in plants and most of the enzymatic steps have been established (Hassan & Mathesius, 2012). To date, more than 10,000 flavonoid compounds have been identified and new structures are still being reported (Hassan & Mathesius, 2012). This specific pathway starts with the condensation of one molecule of *p*-coumaroil-coenzyme A (CoA) and three molecules of malonyl-CoA yielding naringenin chalcone, catalysed by the action of the enzyme chalcone synthase (CHS). These precursors originate from the shikimate pathway and the pyruvate metabolism, respectively (Figure 1). *p*-coumaroil-CoA is synthesised from the amino acid phenylalanine by three successive enzymatic steps, that initiate the general phenylpropanoid pathway. Malonyl-CoA is produced by carboxylation of acetyl-CoA, a key metabolite in the intermediary metabolism. The naringenin chalcone is subsequently isomerised by the enzyme chalcone flavanone isomerase (CHI) to produce naringenin, which is the first compound of the pathway that presents the basic general structure of this family of metabolites. From this central intermediate, built up of two aromatic rings, named A and B, and an heterocycle unit (C) of 3 carbon atoms, the pathway diverges into several side branches, each giving as a result the different class of flavonoids, as depicted in Figure 1 (flavonols, flavan-3-ols, anthocyanidins, anthocyanins, proanthocyanidins). Their classification is based upon the oxidation level of the central C heterocycle, the presence of hydroxyl and methyl substitutions on the A and B rings, and other modifications such as glycosylation or polymerization.

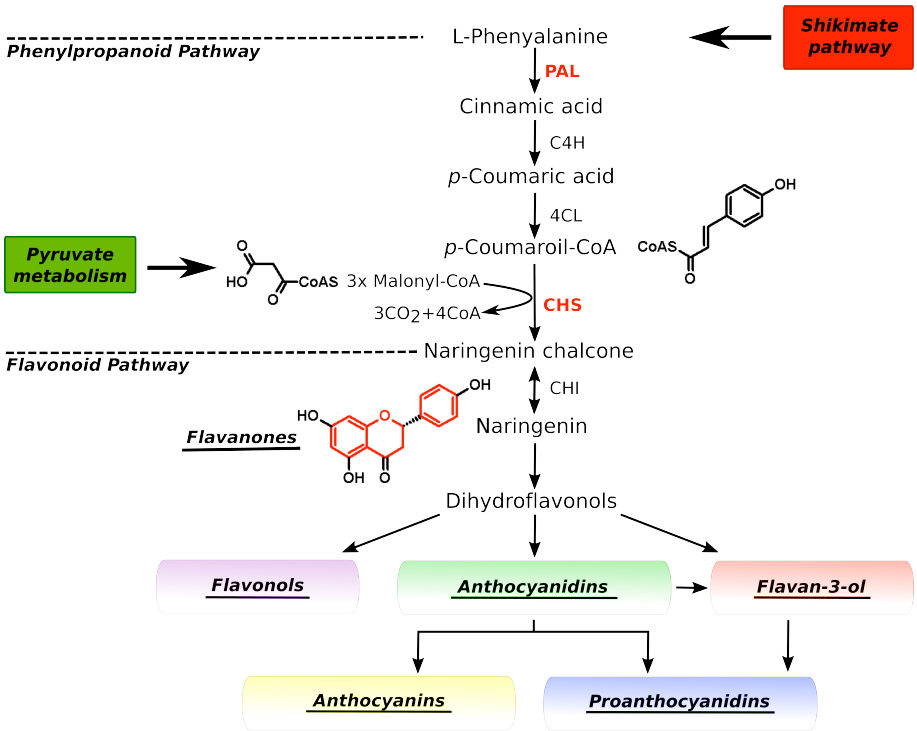


Figure 1. The flavonoid biosynthesis pathway. A simplified representation of the flavonoid biosynthesis pathway is shown. Major families of flavonoid compounds are highlighted. The three rings basic structure of flavonoids is represented by the naringenin molecule. Suppression of *FaFra* genes affects the expression of phenylalanine ammonia lyase (*PAL*) and chalcone synthase (*CHS*) (Muñoz *et al.*, 2010). C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarol-coenzyme A ligase; CHI, chalcone isomerase.

Although flavonoid biosynthesis is well established, little is yet known about the proteins transporting anthocyanins, glycosylated quercitins, and kaempferols. Flavonoids are synthesized in the cytosol and are mainly transported to the vacuole for storage. They are also found in cell walls, the nucleus, chloroplasts and the endoplasmic reticulum (ER), depending on the plant species, the tissue, or the stage of development (Zhao & Dixon, 2009; Thompson *et al.*, 2010; Hichri *et al.*, 2011). Given the different functions that flavonoids play in plants, additional research is still necessary to know better how their trafficking takes place within the cell.

Regulation of flavonoid biosynthesis is under a complex network of diverse elements. Plants produce flavonoids in a coordinated manner as a result of genetic, environmental, and developmental regulatory factors. In general, environmental factors mostly affect the quantitative and qualitative composition of flavonoids, whereas genetic background of

species/variety determines which metabolites occur at the different tissues of the plant (Carbone *et al.*, 2009; Muñoz *et al.*, 2011). MYB/bHLH-WD40 transcription factors synchronize the expression of the structural genes of the pathway (Vogt *et al.*, 2010; Schaart *et al.*, 2012; Jaakola, 2013). Their activity is directly related to the developmental control of flavonoid metabolism, and it is affected by the action of plant hormones during development (Giovanonni, 2004; Jia *et al.*, 2011; Zifkin *et al.*, 2012; Seymour *et al.*, 2013). Under this scenario, many key players in this highly integrated regulatory network remain to be identified. Furthermore, crosstalk between plant hormones, development, environmental actors and transcription factors requires deeper investigation to gain insight into the mechanisms underlying flavonoid biosynthesis control (Jaakola, 2013).

Strawberry (*Fragaria xananassa*) is one of the most important edible fruits worldwide due to its remarkable organoleptic properties and nutritional quality, which are correlated with a high content of vitamin C, folate, and a large quantity and diversity of flavonoids (Tulipani *et al.*, 2011b; Giampieri *et al.*, 2011; Csukasi *et al.*, 2012). Strawberry belongs to the Rosaceae family in the *Fragaria* genus and it is considered a model for non-climacteric fruits, as it does not exhibit a sharp peak in respiration rate and ethylene does not appear to be critical for the ripening process (Seymour *et al.*, 2013), although a small production of ethylene (Iannetta *et al.*, 2006) and differential expression of genes related to ethylene synthesis, perception and signalling during fruit ripening have been reported (Trainotti *et al.*, 2005; Merchante *et al.*, 2013). The strawberry fruit initiates from a single inflorescence and it is composed of a receptacle (false fruit), which results from the engrossment of the floral receptacle, and the achenes (true fruits) that are connected through vascular bundles to the fleshy part (receptacle) of the fruit (Perkins-Veazie, 1995) (Figure 2). These two closely connected organs exhibit coordinated developmental programmes but their functions differ significantly in terms of cell ontogeny and function (Fait *et al.*, 2008; Csukasi *et al.*, 2012). Indeed, a metabolic profiling of primary and secondary metabolites performed in both the strawberry receptacle and the achene across several developmental stages displayed a high level of metabolic synchrony and specialization during development (Fait *et al.*, 2008). Anthocyanins are quantitatively the most important type of flavonoids in strawberry, being pelargonidin- and cyanidin- glucosides the dominant anthocyanin compounds in fruits (Kosar *et al.*, 2004; Halbwirth *et al.*, 2006; Fait *et al.*, 2008). These major metabolites are responsible for the bright red colour of the fruits, which is considered one of the most relevant quality characteristics of cultivated strawberry from the commercial point of view. Strawberries also contain smaller amounts of other phenolic compounds such as flavonols and flavanols. The content and composition of flavonols has been

studied (Giampieri *et al.*, 2012), being these compounds identified mainly as derivatives of quercetin and kaempferol, with quercetin derivatives being the most abundant ones. On the other hand, flavanols are the only class of flavonoids that do not occur naturally as glycosides. They are found in strawberries in monomeric (flavan-3-ols) and polymeric forms (proanthocyanidins or condensed tannins). Although not extensively represented, proanthocyanidins are commonly found in strawberry receptacles and achenes (Fait *et al.*, 2008). In this regard, it has been established that flavonoids not only occur in an organ specific manner but also follow a developmental pattern (Halbwirth *et al.*, 2006). While unripe strawberry fruits produce large amounts of flavan-3-ols and proanthocyanidins, during the late stages of development the content of anthocyanins and flavonols increase continuously, until they reach their maximum levels in ripe fruits. This differential accumulation of metabolites during fruit ripening happens in concert with the expression of the genes encoding for key enzymes of the pathway (Halbwirth *et al.*, 2006). The highly coordinated developmental regulation of the whole pathway is considered a general strategy to fulfil different functions in the plant, and to ensure maximum plant fitness. Thus, whereas flavan-3-ol mediate fruit protection at early growth stages, anthocyanins promote fruit consumption and seed dispersal at full fruit ripening (Halbwirth *et al.*, 2006; Carbone *et al.*, 2009).

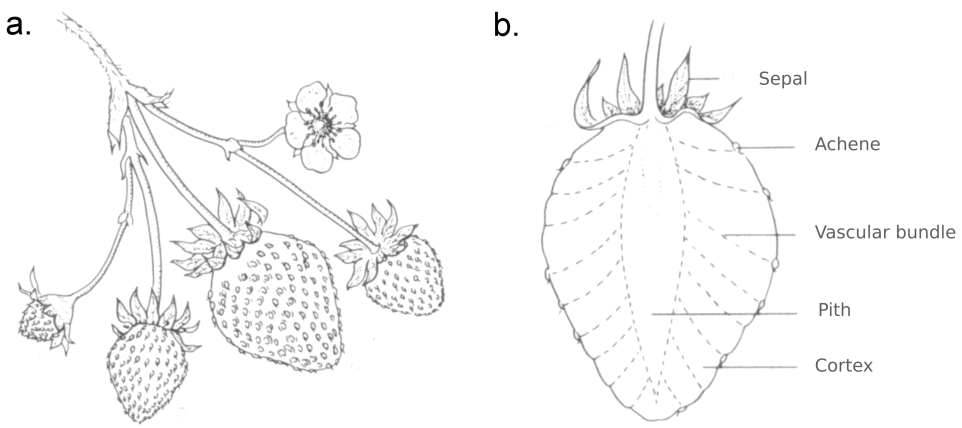


Figure 2. Strawberry fruit morphology. The strawberry plant is a nonwoody perennial made up of a crown, leaves, runners, and a root system. **a.** Flowers are born in clusters (inflorescences) from the crown and yield strawberry fruits. **b.** Strawberry is a false fruit. The achenes are the true fruits and are connected by vascular bundles to the floral receptacle. The receptacle (fleshy part) is composed of an internal pith, a cortex layer and an epidermal layer (Illustrations by Linda Chandler).

The strawberry Fra proteins belong to the ubiquitous family of plant pathogenesis- related- 10 proteins (PR-10), which have been implicated in the response of plants to pathogenic infections

and abiotic stress (Markovic-Housley *et al.*, 2003). However, despite numerous studies, the physiological functions of PR-10 proteins remain unclear (Fernandes *et al.*, 2013). Our group identified two new members of this family, the FaFra2 and the FaFra3 allergens of strawberry (Muñoz *et al.*, 2010). The other member of this family in strawberry, FaFra1, had been previously described (Karlsson *et al.*, 2004). These three strawberry FaFra proteins are differentially expressed during fruit development (Muñoz *et al.*, 2010), each presenting a different expression pattern. Whereas the *FaFra2* expression increases through the maturation of strawberry fruits, *FaFra1* and *FaFra3* present their maximum transcript level at the green fruit stage. In addition to their properties as food allergens (Karlsson *et al.*, 2004), it has been reported that FaFra proteins play an important role in the control of flavonoid biosynthesis and are thus required for the development of colour during fruit ripening (Hjerno *et al.*, 2006; Muñoz *et al.*, 2010). The first study that reported a relationship between FaFra and the flavonoid biosynthesis was based in the observation that, in white varieties of strawberry, fruits were free of FaFra1 protein, and the content of some enzymes of the pathway, such as chalcone synthase (CHS), flavanone 3- hydroxylase (F3H) and dihydroflavonol reductase (DFR), were significantly lower in these varieties (Hjerno *et al.*, 2006). More recently, suppression of the expression of *FaFra* genes in strawberry fruits led to a decrease in the accumulation of the two main anthocyanins responsible for the red colour of fruits, and selective changes in other intermediates of the pathway (Muñoz *et al.*, 2010). Furthermore, down-regulation of genes encoding for key enzymes of the flavonoid pathway (*PAL*, *CHS*) was also shown (Muñoz *et al.*, 2010) (Figure 1). The molecular mechanism through which the FaFra proteins participate in the control of flavonoid biosynthesis is still unknown. However, the fact that these proteins are predicted to have cavities for the binding of small ligands (Seutter von Loetzen *et al.*, 2012; Casañal *et al.*, 2013b) suggests that the FaFra proteins might exert their function by binding metabolic intermediates of the flavonoid pathway.

Recently, the genome of the model species *Fragaria vesca* has been released (Shulaev *et al.*, 2011). This facilitates the identification and analysis of *Fra* genes in strawberry, whose joint study might shed light on the role of the encoded proteins. By using the available databases and high-throughput mRNA sequencing (RNA-Seq) we have characterized the main sequences of the Fra family of *Fragaria xananassa* and monitored the expression of the candidate genes in various tissues and developmental stages of strawberry. To further investigate Fra biological roles, here we also present the RNA-Seq analysis of down-regulated *FaFra* genes in strawberry fruits. Our results further support the role of the FaFra proteins in the control of the secondary metabolism in strawberry plants.

MATERIALS AND METHODS

Plant Material

Strawberry plants (*Fragaria xananassa* cv. Camarosa) were grown under field conditions in the strawberry-producing area of Huelva, Spain. Fruits were harvested at four ripening stages: green (G), white (W), turning (T) and red (R) (see results, Figure 13a). The vegetative tissues collected for analysis were the leaves (L) and roots (Rt) of adult plants. All tissues were flash frozen immediately after collection in liquid nitrogen and stored at -80°C until use. For quantitative PCR (qPCR), Western-blot and RNAseq analysis, achenes and receptacles were isolated and ground separately in liquid nitrogen.

For transient down-regulation assays, strawberry plants of the same cultivar (*Fragaria xananassa* cv. Camarosa) were grown in a greenhouse of Instituto Andaluz de Investigación y Formación Agraria y Pesquera (IFAPA Churriana, Málaga, Spain). Fruits were flash frozen immediately in liquid nitrogen after collection and stored at -80°C until use. For qPCR and immunoblotting, achenes were carefully separated from receptacles by using the tip of a scalpel. Receptacles were ground in liquid nitrogen and stored at -80°C until further analysis.

Nicotiana benthamiana and *Arabidopsis thaliana* (Columbia ecotype) plants were grown in soil using a mixture of organic substrate and vermiculite (4:1 v/v) at 22°C in long day conditions (16 h light/ 8 h dark photoperiod). For *in vitro* assays, surface-sterilized and cold-stratified *Arabidopsis* seeds were sown onto Murashige and Skookg (MS) agar-solidified medium (MS salts, 10 g/l sucrose, 0.25 g/l MES pH 5.7 and 6 g/l phytoagar [Roko, http://www.rokoagar.com]) in Conviron growth chambers (http://www.conviron.com) under long day conditions at 22°C.

RNA Extraction

Always, a minimum of three independent total RNA samples were performed from pools of fruits and vegetative tissues. Total RNA was extracted from strawberry tissues according to the method previously described (Manning, 1991, Osorio *et al.*, 2008). Briefly, 2 g of macerated strawberry tissue were added to 50 ml Falcon tubes containing 10 ml of extraction buffer (0.2M Tris-HCl pH 7.6, 10mM EDTA, 100mM NaCl, 5% SDS, 2% β-mercaptoethanol, DEPC water) and then vortexed. The mixture was shaken with an equal volume of phenol/ chloroform/ isoamilyc alcohol (25:24:1), incubated 5-15 min at room temperature. The upper aqueous phase was recovered after a centrifugation step of 20 min at 17000 rpm and 4°C. Nucleic acids were

precipitated directly from the aqueous phenol extract by a two-step addition of 2-BE. To separate RNA from DNA, total nucleic acids were dissolved in water and adjusted to 3M LiCl by adding 12M LiCl. After overnight incubation at 4°C, precipitated RNA was recovered by centrifugation at 13000 rpm, 20 min and 4°C. The pellets were successively washed with 70% and 100% ethanol, air dried and resuspended in 50 ml of DEPC water. Total RNA was quantified using micro-spectrophotometry (NanoDrop Technologies). Contaminating DNA was removed with Turbo DNAfree (Ambion) according to the manufacturer's instructions. Removal of DNA from the RNA samples was confirmed by performing PCR on 100 ng of total RNA using the *FaRib413* (Casado-Díaz *et al.*, 2006) and glyceraldehyde-3-phosphate dehygrogenase (*GAPDH*) primer sets (Salvatierra *et al.*, 2010; Merchante *et al.*, 2013).

Gene Expression Analysis: Quantitative Real-Time PCR (qPCR)

First-strand cDNA was generated from 1 mg of total RNA using the iScript cDNA synthesis kit (Bio-Rad) according to the manufacturer's instructions. The synthesized first-strand cDNA was used for quantitative PCR (qPCR) analysis, using *FaFra* specific primers (Muñoz *et al.*, 2010). For qPCR the reaction mixture consisted of primer mix (10 µmol each) and SsoFast Eva Green Supermix (Bio-Rad) according to the manufacturer's instructions. The reactions were performed using a MyiQ real time cycler (Bio-Rad). A relative quantification qPCR method was used to compare expression of different strawberry tissues (Pfaffl *et al.*, 2001). Relative quantification describes the change in expression of the target gene in a test sample relative to a calibrator sample. Each data point is the mean value from three experimental replicate determinations. Gene-specific primers were chosen so that the PCR products were 100–250 bp. The absence of primer dimers was confirmed by melting curve examination. *FaRib413* and *GAPDH* were used as housekeeping genes to normalize the data (Casado-Diaz *et al.*, 2006; Salvatierra *et al.*, 2010; Merchante *et al.*, 2013).

Protein Extraction SDS-PAGE Analysis

For total protein extraction 0.1 g of plant tissues were ground in liquid N_2 and homogenized in 200 ml of sample buffer containing 125 mM Tris-HCl pH 6.8, 4% (w/v) SDS, 20% (v/v) glycerol, and 2% (v/v) 2-mercaptoethanol, 0.01% (w/v) bromophenol blue) (Laemmli, 1970). The extracts were then centrifuged at 14000 rpm for 20 min at 4°C. The resulting supernatants were collected and the soluble protein concentration was measured by the modified Lowry method (Hartree, 1972) in the case of strawberry extracts. Equal amounts of total protein were loaded onto a 1 mm thick 15% (w/v) polyacrylamide gel and run for 90 min at 120 V. For

visualization of total proteins, the gels were stained in Coomassie Blue (0.25% [w/v] in 50% [v/v] methanol and 10% [v/v] acetic acid) and distained in 30% (v/v) methanol and 10% (v/v) acetic acid. The extracted proteins were subjected to Western blotting.

Western Blot Analysis

For western blotting, proteins were transferred from the SDS-PAGE gels to a Trans-Blot® nitrocellulose membrane using a Trans-Blot® Turbo™ Transfer System (Bio-Rad). After the transfer, the membrane was incubated with 5% blocking solution (5% fatty acid free powder milk in TTBS buffer containing 20mM Tris-HCl pH 7.6, 140 mM NaCl, 0.05% v/v Tween20) for 2 h at room temperature, following an overnight incubation with the primary antibody at 4°C. The anti-FaFra2A primary polyclonal antibody was used at a 1:500 dilution in 1% blocking solution (1% fatty acid free powder milk in TTBS buffer containing 20mM Tris-HCl pH 7.6, 140 mM NaCl, 0.05% v/v Tween20). Chemiluminescent immunodetection was performed using ECL Western Blotting Detection Reagents (GE Healthcare) and anti-rabbit peroxidase-conjugated secondary antibodies. To generate the anti-FaFra2A polyclonal antibody, the full length FaFra2A purified protein (Casañal *et al.*, 2013a) was used to immunize rabbits. Affinity purified fractions of the collected immune serum were used for Western Blot experiments. Anti-rabbit secondary antibodies were purchased from Santa Cruz Biotechnology (www.scbt.com).

Transient Down-Regulation of FaFra Genes in Strawberry Fruits

RNAi-mediated down-regulation of the *FaFra* genes in the receptacle of strawberry fruits (*Fragaria xananassa* cv. Camarosa) was performed as described by Muñoz *et al.* (2010), with some minor modifications. The *Agrobacterium tumefaciens* strain AGL0 harbouring the pBI–*FaCHS*i, or pBI–Intron, or pBI–*Fra1ei* plasmids (Muñoz *et al.*, 2010), were grown at 28°C in LB medium with appropriate antibiotics according to Hoffmann *et al.* (2006). When the culture reached an OD^{600} of about 0.8, *Agrobacterium* cells were harvested and re-suspended in a modified MMA medium (MS salts, 10 mM MES pH 5.6, 20 g/l sucrose). The *Agrobacterium* suspension was inoculate in half fruits by using a sterile 1ml hypodermic syringe. Fruits were harvested after 7 days, flash frozen in liquid nitrogen, and stored at -80°C until use.

Library Construction and RNA-Seq

Total RNA was extracted as described above. RNA quality and quantity were determined based on absorbance ratios at 260 nm/280 nm and 260 nm/230 nm using a Nanodrop 2000

spectrophotometer (Thermo Scientific). RNA integrity was confirmed by the appearance of ribosomal RNA bands and lack of degradation products after separation in agarose gel electrophoresis and RedSafe™ staining (Intron Biotechnology). The integrity of RNA samples was further verified using a 2100 Bioanalyzer instrument (Agilent Technologies) and RIN values ranged between 7.5 and 8.0 for all the samples. The libraries consisted of a) separate samples of achenes and receptacles at four ripening stages (green, white, turning and red) as well as leaves and roots (3 biological replicates) of strawberry plants and b) control (3 biological replicates) and RNAi (5 biological replicates) receptacles agroinfiltrated for down-regulation of *FaFra*. Samples were pooled and 10 mg of total RNA from each pool was used for library construction and RNA-Seq analysis. Addition of adapters, size selection, PCR amplification and RNA-Seq (Illumina HiSeq 2000) were performed at a) Centro Nacional de Análisis de Genómica (CNAG) (Barcelona, Spain) and b) Beijing Genome Institute (BGI) (Hong Kong). Briefly, for each sample, one paired-end library with approximately 300 bp insert size was prepared. Libraries were sequenced using 2 x 100 bp reads. More than 30 and 20 million reads, for CNAG and BGI respectively, were generated for each sample.

De- novo Assembly of Fragaria xananassa RNA-Seq Reads

Raw RNA-Seq reads were first filtered to remove the adapter sequences and low quality sequences, with an assurance that >80% of all bases passing filter had a quality value (Q-value) of at least 30. The Q-value represents the sequencing quality of related nucleotides. Clean reads were used in *de novo* assembly and read-mapping to the transcriptome to obtain the transcripts expressed in *Fragaria xananassa*. RNA-Seq data was *de novo* assembled using the assembly program Trinity (Grabherr *et al.*, 2011). The transcript contigs most similar to the *Fragaria vesca* candidate genes were identified by BLAST search.

RNA extraction, RNA-Seq generation data, and transcriptome assembly in the strawberry fruit developmental assay were performed as part of a running project in our laboratory (Sánchez-Sevilla *et al.*, unpublished).

Mapping RNA-Seq Reads to the Reference Genome, Generation of Read Counts and Differential Expression Analysis

Same as for *de novo* assembly, RNA-Seq reads were first filtered to remove the adapter sequences and low quality sequences, with an assurance that >80% of all bases passing filter had a quality value (Q-value) of at least 30. RNA-seq reads were aligned to the *Fragaria vesca*

reference genome (v1.1) and CDS (v1.0) (Shulaev *et al.*, 2011) using the program TopHat v2.0.6 (Kim *et al.*, 2013). Default parameters of TopHat were used, allowing 40 multiple alignments per read and a maximum of 2 mismatches when mapping reads to the reference.

The aligned read files were processed by Cufflinks v2.0.2 (Trapnell *et al.*, 2012). Reads were assembled into transcripts, their abundance estimated, and tests for differential expression and regulation between the samples were performed. The normalized RNA-Seq fragment counts were used to measure the relative abundances of transcripts expressed as **F**ragments **P**er **K**ilobase of exon per **M**illion fragments mapped (FPKM).

Once all short read sequences were assembled with Cufflinks, the output files were processed with Cuffcompare using a reference annotation file, downloaded from Genome Database for Rosaceae (GDR) database (*Fragaria vesca* Whole Genome v1.1 Assembly & Annotation. http://www.rosaceae.org/). This classified each transcript as known or novel. The file produced by Cuffcompare was passed to Cuffdiff along with the original alignment files produced by TopHat to identify differentially expressed transcripts between the pools. The Cuffdiff algorithm was used to re-estimate the abundance of transcripts and simultaneously to test for differential expression between the pools using a rigorous statistical analysis (Trapnell *et al.*, 2012).

Phylogenetic and Sequence Analyses

Multiple sequence analysis was performed with ClustalW (McWilliam *et al.*, 2013) at the default settings. The neighbor– joining method was used to generate the tree using MEGA5 (Tamura *et al.*, 2011). The percentages of replicate trees in which the associated proteins clustered together in the bootstrap test (1000 trials) are shown next to the branches. The trees are drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method, and the scale is the number of amino acid differences per site. The evolutionary analyses were conducted in MEGA5 (Tamura *et al.*, 2011).

Generation of Overexpressing Transgenic Lines

Isolation of *FaFra1E*, *FaFra2A* and *FaFra3* ORFs (EMBL accession numbers CAJ85645, GQ148818 and GQ148819) from strawberry (*Fragaria xananassa* cv. Camarosa) has been described before (Muñoz *et al.*, 2010). The *FaFra1-GFP*, *FaFra2-GFP* and *FaFra3-GFP* constructs were obtained by Gateway® cloning (Invitrogen, http://www.invitrogen.com).

FaFra1E, *FaFra2A* and *FaFra3* full-length sequences were amplified by PCR using gene-specific primers and the plasmids *Fra1-pGEMT, Fra2-pGEMT* and *Fra3-pGEMT* as a templates (Muñoz *et al.*, 2010). Primers used for amplification were as follows: FwF1-pENTRY CACCATGGGTGTTTACACTTATG/ RvF1-pENTRY GTTGTATTCGCTGGGGTGGTC; FwF2-pENTRY CACCATGGGTGTGTTCACTTATG/ RvF2-pENTRY ACAGTATTCATTAGGATTGGC; FwF3-pENTRY CACCATGGGTGTGTTCACATAC/ RvF3-pENTRY GTTGTATTCCTCAGGATGGG. The forward primers were designed to include the CACC sequence required at the 5' end of the *FaFra* ORFs to ensure directional sub-cloning into the pENTR™/D-TOPO® vector. To enable fusion of the *FaFra* sequences in frame with C-terminal tags, the reverse PCR primers were design to remove the native stop codon in the *FaFra* genes. For optimal expression of the FaFra-GFP heterologous proteins, the resulting PCR products were transferred into pENTR™/D-TOPO® vector and subsequently into pMDC43 and pMDC85 Gateway Destination vectors (Curtis & Grossniklaus, 2003) by TOPO and LR clonase reactions according to the manufacturer's instructions (TOPO® Cloning Kit, Gateway® LR Clonase® II enzyme mix, Invitrogen). DNA sequencing confirmed that the recombinant vectors encoded the expected sequences. The resulting plasmids were introduced into *Agrobacterium tumefaciens* strain GV3101, and then into *A. thaliana* by the floral dip method (Clough & Bent, 1998), and agroinfiltrated into *N. benthamiana* leaves for transient expression assays (Voinnet *et al.*, 2003).

Transient Expression Assays in Nicotiana benthamiana

Transient expression assays in *N. benthamiana* leaves were performed using the method described by Voinnet (Voinnet *et al.*, 2003). Briefly, *A. tumefaciens* carrying the construct of interest were grown at 28°C in Luria-Bertani (LB) medium supplemented with 50 μg/ml kanamycin, 50 μg/ml rifampicin and 5 μg/ml tetracycline to stationary phase. The cells were harvested by centrifugation at 4000 rpm for 15 min at 25°C and re-suspended in 10mM $MgCl_2$, 10 mM MES pH 5.7 and 1mM acetosyringone. After a dark incubation of 2h, the abaxial surface of leaves of 2–3-week-old *N. benthamiana* plants were co-infiltrated by using a needleless syringe with two strains of *Agrobacterium* carrying transgenes for the 35S:Fra-GFP proteins and the p19 silencing suppressor (Voinnet *et al.*, 2003) at an OD_{600} of 1.0. GFP expression was observed under UV light and samples (two leaves per plant from three plants) were taken at 4 days post-infiltration. All the assays were performed at least twice.

Arabidopsis thaliana Transformation by the Floral Dipping Method

Transgenic *A. thaliana* plants were obtained using the floral dip method (Clough & Bent, 1998; Zhang, *et al.*, 2006). Briefly, sterilized and cold-stratified *A. thaliana* seeds were sown on MS medium for 1-2 weeks in short day conditions (8h light/ 16h dark, 22°C), and then transfer to soil (4 seedlings/pot). To induce flowering, plants were grown in a growth chamber under long day conditions (16h light/8 h dark, 22°C) for 3-4 weeks. Single *A. tumefaciens* colonies carrying the construct of interest were inoculated into 5 ml liquid LB medium containing 50 μg/ml rifampicin and 5 μg/ml tetracycline, the appropriate antibiotic for binary vector selection (50 μg/ml kanamycin), and grown at 28°C for 2 days. This feeder culture was used to inoculate 200 ml of liquid LB medium supplemented with 50 μg/ml kanamycin. The cells were grown to stationary phase, collected by centrifugation at 4000 rpm for 15 min at 25°C, and gently re-suspended in 200 ml of freshly made 5% (w/v) sucrose solution containing Silwet L-77 0.05% (v/v). Plants were inverted and aerial parts dipped in the *Agrobacterium* cell suspension for 10 s. Dipped plants were laid down, covered with a plastic film and maintained in darkness overnight. The cover was removed the next day and the plants were allowed to grow normally in the growth chamber for 1 month. When siliques were brown and dry, seeds were collected. *In vitro* screening of primary transformants was accomplished using MS selection plates containing cefotaxime and kanamycin. Observation of GFP signals was carried out on non-selecting plates with segregating T2 seedlings. Between three and six independent transformants per construct were scored, and those with the best expression levels were chosen for imaging and microscopy. The presence of the FaFra-GFP fusion proteins was verified by Western blot.

Imaging and Microscopy

A Nikon AZ-100 multizoom microscope coupled to a Nikon Digital Sight DS-5Mc camera (http://www.nikon.com) was used for the light microscopy observation and documentation of *N. benthamiana* leaves and *A. thaliana* seedlings. The confocal microscopy was performed on a Leica SP5II confocal microscope (http://www.leica.com), equipped with argon and krypton lasers. For excitation of GFP, the 488-nm line of the argon laser was used and emission was detected with the setting of the acousto optical beam splitter set to 500 to 560 nm. *Arabidopsis* and *Nicotiana* samples were examined using x40 oil immersion lenses. For double labelling using propidium iodide and GFP, propidium iodide was visualized at wavelengths 605–665 nm and GFP at wavelengths 521–561 nm. For confocal analysis, seedlings of transgenic *Arabidopsis* were grown on vertically oriented half-strength MS Petri dishes and analysed at 3–5 days old. For transiently transformed leaves of *N. benthamiana,* pieces of ~1 cm^2 were cut out and mounted

between slide and cover slip in distilled water. For labelling with propidium iodide, *Nicotiana* leaves were stained using 10 μg/ml propidium iodide (Sigma) for 5 min. Images were processed using Leica LAS AF software.

RESULTS

Identification of Fragaria xananassa Fra Transcripts

In order to determine *Fragaria xananassa Fra* transcripts, we performed Illumina RNA sequencing (RNA-Seq). Recent advances in high throughput mRNA sequencing not only provide an effective way to obtain large amounts of transcription data from many organisms and tissue types, but also enable abundance estimation and differential gene expression measurements (Grabherr *et al.*, 2010; Trapnell *et al.*, 2010). RNA was extracted from bulked pools of strawberry fruits at four ripening stages (green, white, turning, red) and vegetative tissues (leaves and roots), and were used as biological replicates for RNA-Seq. Cultivated strawberry (*Fragaria xananassa*) is an octoploid (2n= 8x= 56) species; its complex genome is composed of three main diploid genomes (Rousseau-Gueutin *et al.* 2009). The *Fragaria xananassa* genome has a high level of conservation with the model species *Fragaria vesca* (2n= 2x= 14) (Bombarely *et al.*, 2010). Although the *F. vesca* diploid genome sequence is available (Shulaev *et al.* 2011), and it can be used as a reference genome within the *Fragaria* genus, the fact that the current *F. vesca* sequenced genome and the gene model are still incomplete, makes its use for transcript reconstruction unsuitable in many cases, as the prediction of transcripts might result inaccurate. For this reason, we have performed a *de novo* assembly of the RNA-Seq reads to identify all expressed *Fragaria xananassa Fra* transcripts. This method does not rely on aligning reads to a reference genome but uses the reads to assemble transcripts directly. In this case, Trinity (Grabherr *et al.*, 2010) has been used as the assembly-first (*de novo*) method to determine the *FaFra* transcripts of interest. We first identified 28 contigs, 1230 to 212 bp long, with homology to the *FaFra1E*, *FaFra2* and *FaFra3* genes (Muñoz *et al.*, 2010). Some of them contained full length coding sequence (CDS), 5' UTR and 3' UTR regions. However, some others were incomplete, lacked UTR regions and in some cases, presented only a partial CDS. To reduce the number of contigs and obtain a realistic representation of the *FaFra* transcripts, we took into account only those transcripts with abundances higher than 1.0 **F**ragments **P**er **K**ilobase of exon per **M**illion fragments (FPKM), and lower levels were considered as not expressed. Furthermore, we also selected transcripts containing full CDS, with the only exception of those either showing a considerable number of reads, or with high similarity to

F.vesca sequences, not previously represented by other identified sequences. Taking into consideration these conditions, we ended up with 15 *FaFra* sequences (Table 1), 12 sequences with a complete CDS and 3 additional incomplete ones (*FaFra6*, *FaFra7a* and *FaFra7b*). Following the nomenclature proposed by Musidlowska-Persson and co-workers (Musidlowska-Persson *et al.*, 2007), small letters denote different DNA sequences, while capital letters are used for amino acid sequences. Additionally, serial numbers after the letter identify different DNA sequences that have the same deduced amino acid sequence.

Table 1. *FaFra* transcripts identified by RNA-Seq and the Trinity method.

***Fanannassa* sequence**	**Length (bp)**	**Protein**	**Length (aa)**	***Fvesca* gene**	***Fvesca* assigned name**
FaFra1c1	848	FaFra1C	159	gene07080-v1.0-hybrid	*FvFra1*
FaFra1c2	879				
FaFra2b	766	FaFra2B	160	gene07086-v1.0-hybrid	*FvFra2b*
FaFra4a1	958	FaFra4A	159	gene07085-v1.0-hybrid	*FvFra4*
FaFra4a2	763				
FaFra4b1	921	FaFra4B	159		
FaFra4b2	726				
FaFra5	777	FaFra5	160	gene07084-v1.0-hybrid	*FvFra5*
FaFra6	238	FaFra6	79	gene05185-v1.0-hybrid	*FvFra6*
FaFra7a	437	FaFra7A	129	gene32299-v1.0-hybrid	*FvFra7*
FaFra7b	661	FaFra7B	124		
FaFra8	1230	FaFra8	160	gene11094-v1.0-hybrid	*FvFra8*
FaFra9a1	861	FaFra9A	163	gene07077-v1.0-hybrid	*FvFra9*
FaFra9a2	1136				
FaFra10	1081	FaFra10	160	gene07076-v1.0-hybrid	*FvFra10*

Gene number according to the *Fragaria vesca* genome draft (www.Strawberrygenome.org).

Alignments, including UTRs and newly identified *FaFra* CDS, are represented in Figures 4-7, and the percentages of nucleotide identity are given in Figure 3. All sequences are very similar and show at least 55% of nucleotide identity among them. Lower percentages represented in Figure 3, are due to partial CDS coverage of *FaFra6*, *FaFra7a* and *FaFra7b* transcripts. The proposed *FaFra* CDS range from 477 to 489 bp long, and code for small proteins of 159-163 amino acids (Table 1), which is in accordance with the previous reports on *FaFra* genes and *pr-10* genes (Musidlowska-Persson *et al.*, 2007; Muñoz *et al.*, 2010, Fernandes *et al.*, 2013). To date, three different *Fra* genes have been described in *Fragaria xananassa: FaFra1, FaFra2 and FaFra3* (Musidlowska-Persson *et al.*, 2007; Muñoz *et al.*, 2010). From the 15 *FaFra* sequences identified here, 2 of them, *FaFra1c1* and *FaFra1c2*, showed high sequence identity with the

already published *FaFra1c* (>99% CDS identity, Figure 3), encoding protein FaFra1C (Musidlowska-Persson *et al.*, 2007). The corresponding *F. vesca* predicted gene in the strawberry genome sequence was gene07080 (hereafter named *FvFra1*, Table 1).

CDS	*FaFra1e*	*FaFra1c*	*FaFra1c1*	*FaFra1c2*	*FvFra1*
FaFra1e		97,93	98,14	98,14	98,76
FaFra1c			99,79	99,79	98,76
FaFra1c1				100,00	98,96
FaFra1c2					98,96
FvFra1					

CDS	*FaFra2a*	*FaFra2b*	*FvFra2a*	*FvFra2b*
FaFra2a		90,68	97,93	96,07
FaFra2b			89,44	91,10
FvFra2a				95,24
FvFra2b				

CDS	*FaFra3*	*FvFra3*	*FaFra4a1*	*FaFra4a2*	*FaFra4b1*	*FaFra4b2*	*FvFra4*	*FaFra5*	*FvFra5*
FaFra3		99,79	89,58	89,58	87,08	87,08	89,58	85,3	85,09
FvFra3			89,79	89,79	87,29	87,29	89,79	85,51	85,3
FaFra4a1				100,00	92,71	92,71	99,58	89,65	86,54
FaFra4a2					92,71	92,71	99,58	89,65	86,54
FaFra4b1						100	92,71	82,4	82,61
FaFra4b2							92,71	82,4	82,61
FvFra4								89,23	86,13
FaFra5									96,27
FvFra5									

CDS	*FaFra6*	*FvFra6*	*FaFra7a*	*FaFra7b*	*FvFra7*	*FaFra8*	*FvFra8*	*FaFra9a1*	*FaFra9a2*	*FvFra9*	*FaFra10*	*FvFra10*
FaFra6		49,16	45,90	47,48	37,89	29,4	28,99	30,08	30,08	30,08	31,26	31,47
FvFra6			59,01	58,18	76,60	58,39	57,97	57,72	57,72	57,52	57,97	57,76
FaFra7a				94,62	79,50	48,65	48,03	46,34	46,34	46,54	50,52	49,9
FaFra7b					75,57	47,2	47	45,33	45,33	45,12	48,86	48,24
FvFra7						62,11	61,28	57,93	57,93	58,13	61,28	61,08
FaFra8							96,89	65,24	65,24	65,45	65,42	65,22
FvFra8								64,23	64,23	64,43	63,98	63,77
FaFra9a1									100	99,8	61,38	61,99
FaFra9a2										99,8	61,38	61,99
FvFra9											61,18	61,79
FaFra10												98,55
FvFra10												

Figure 3. Sequence similarity between the CDS of *Fra* from *Fragaria xananassa* and *F. vesca*. Upper block shows percentages of nucleotide similarity. Highest values are highlighted in red. The values were obtained from sequence comparison in CLC Main Workbench 7.

Figure 4. Multiple DNA sequence alignment of the strawberry allergen *Fra1*. *FaFra1e* and *FaFra1c* correspond to previously described *FaFra1* genes. *FaFra1c1* and *FaFra1c2* correspond to the newly identified transcripts through RNA-Seq. *FvFra1* represents the *Fra1* homologous in *F. vesca*. UTRs and CDS are shown for newly identified transcripts and *FvFra1*. First nucleotide in CDS is flagged as number 1. Letters and numbers in labels designate different nucleotide sequences; same letters indicate same deduced amino acid sequence. Colour code for nucleotides: Adenine (red), Guanine (yellow), Thymine (green) and Cytosine (blue). Matching nucleotides are indicated as dots. Alignment was performed using ClustalW (EBI server; McWilliam et al., 2013), and the figure was generated by CLC Main Workbench 7.

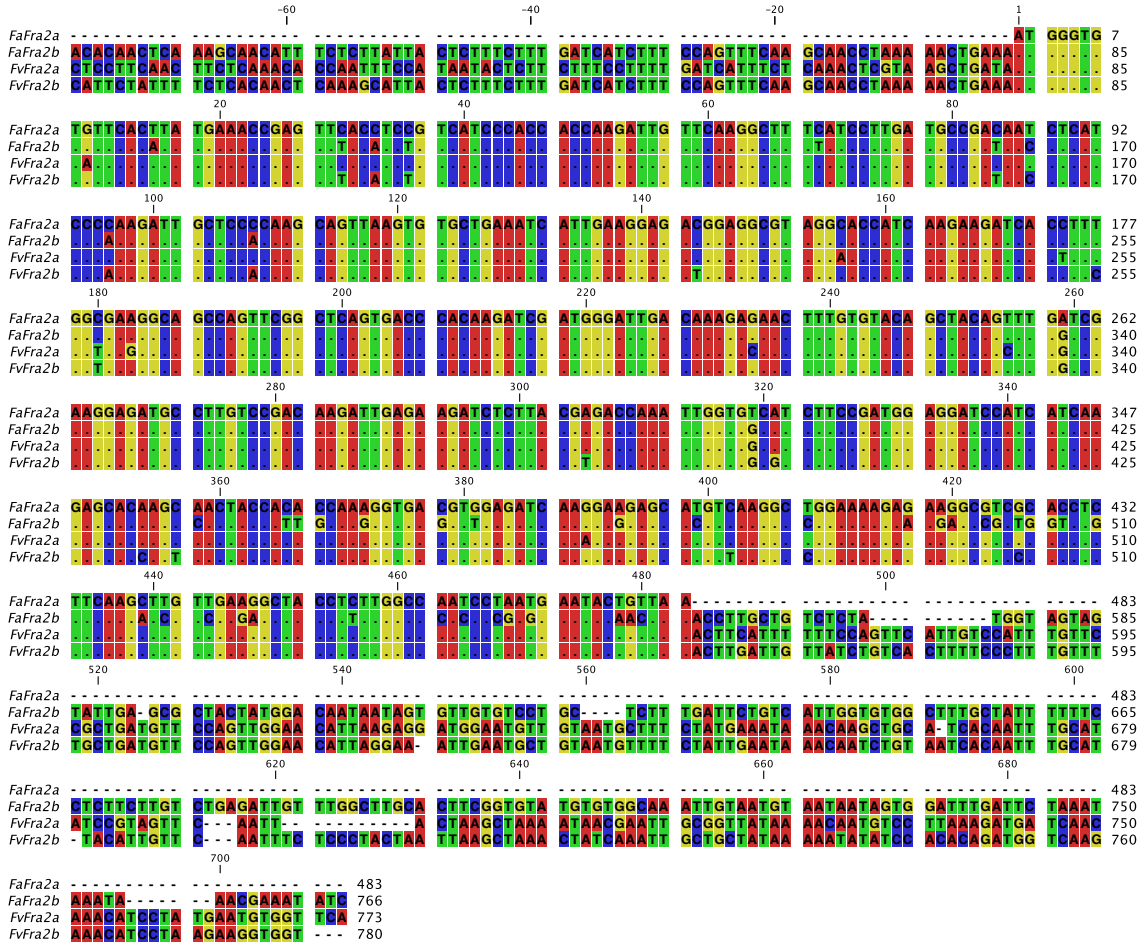


Figure 5. Multiple DNA sequence alignment of the strawberry allergen *Fra2*. Sequence *FaFra2a* correspond to the previously described *Fra2* gene in *Fragaria xananassa*. *FaFra2b* correspond to the newly identified transcript by RNA-Seq. *FvFra2a* and *FvFra2b* represent the *Fra2* homologous in *F. vesca*. UTRs and CDS regions are shown for the identified transcript and *F. vesca* sequences. First nucleotide in CDS is flagged as number 1. Letters and numbers in labels designate different nucleotide sequences; same letters indicate same deduced amino acid sequence. Colour code for nucleotides: Adenine (red), Guanine (yellow), Thymine (green) and Cytosine (blue). Matching nucleotides are indicated as dots. The sequence alignment was performed using ClustalW (EBI server; McWilliam et al., 2013), and the figure was generated by CLC Main Workbench 7.

Surprisingly, no transcripts were found for the already described *FaFra2* (hereafter named *FaFra2a*) and *FaFra3* sequences (Muñoz *et al.*, 2010); their counterparts in *F. vesca* would be the predicted gene07065 (*FvFra2a*) and gene07082 (*FvFra3*), respectively (Table 1, Figure 5, Figure 6). *FaFra2b* is closely related to gene07086 (*FvFra2b*) and shows more than 90% homology to *FaFra2a*. *FaFra4a1-a2*, *FaFra4b1-b2* are very similar to gene07085 (*FvFra4*), showing identities higher than 99% and 92%, respectively; *FaFra5* presents high similarity to

gene07084 (*FvFra5*), higher than 96%. *FaFra6*, *FaFra7a* and *FaFra7b* are partial transcripts, but match *FaFra1* and *FaFra3* with up to 79% and 85% nucleotide identity, when the covered CDS are compared, respectively. Furthermore, they show high identity to gene05185 and gene32299, respectively (in both sequences over 97% on the covered sequence). *FaFra8*, *FaFra9a1-a2* and *FaFra10* represent the most divergent transcripts in comparison to the published *FaFra1*, *FaFra2a* and *FaFra3* sequences, but still show at least 60% of homology to them. Their *F. vesca* counterparts are gene11094, gene07077, and gene07076, with 96%, 99% and 98% identities, respectively.

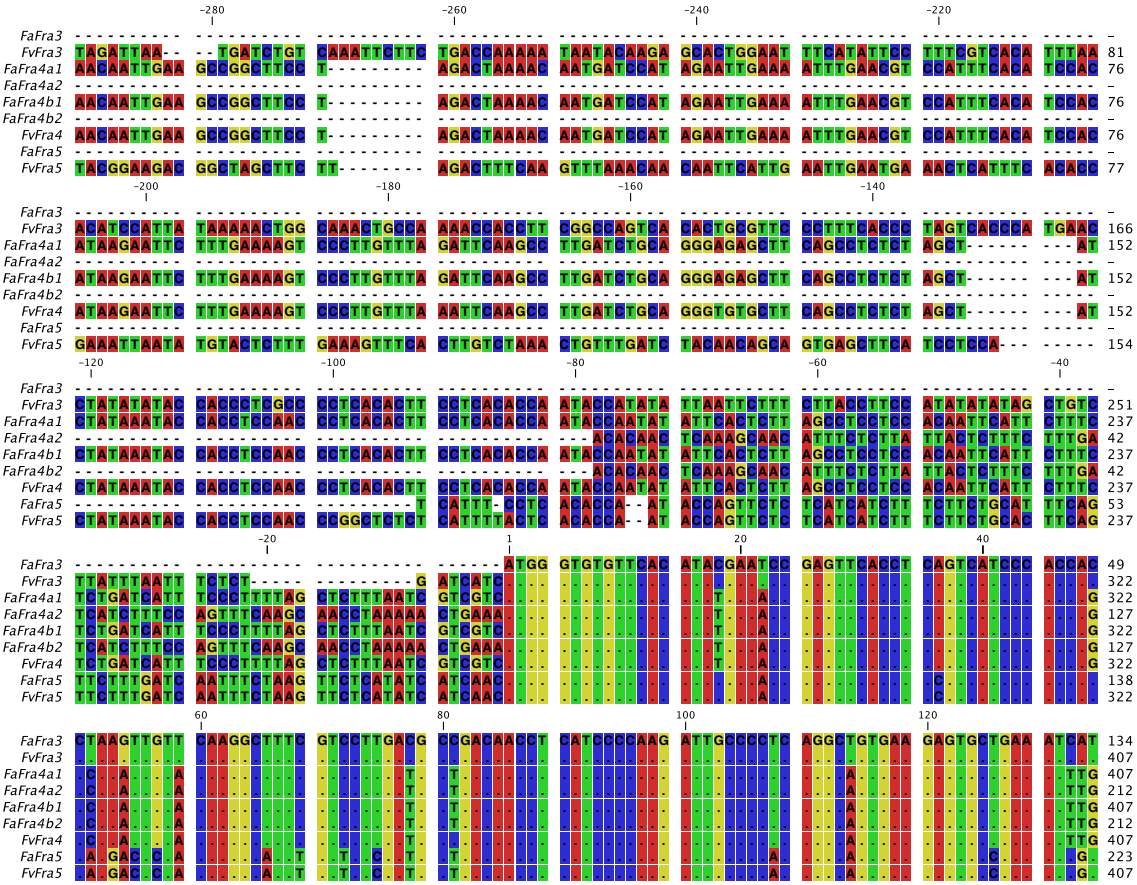


Figure 6. Multiple DNA sequence alignment of the strawberry allergens related to *Fra3*. *FaFra3* corresponds to the previously described *Fra3* gene in *Fragaria xananassa*. *FaFra4a1, FaFra4a2, FaFra4b1, FaFra4b2 and FaFra5* correspond to new identified *Fra* transcripts by RNA-Seq. *F.vesca* homologous are also illustrated. UTRs and CDS regions are shown for newly identified transcripts and *F. vesca* sequences. First nucleotide in CDS is flagged as number 1.

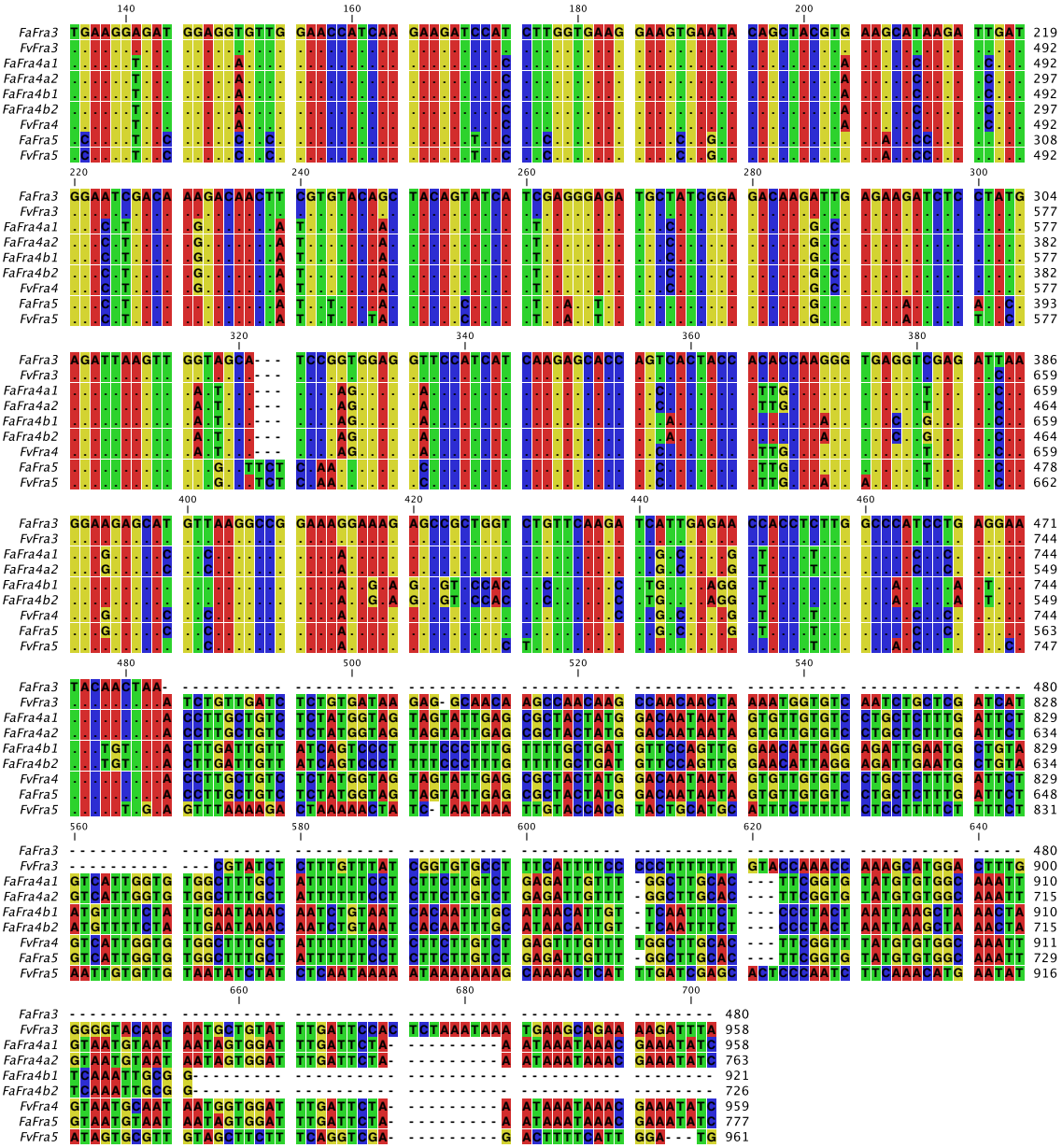


Figure 6. *(Continued.)* Letters and numbers in labels designate different nucleotide sequences; same letters indicate same deduced amino acid sequence. Colour code for nucleotides: Adenine (red), Guanine (yellow), Thymine (green) and Cytosine (blue). Matching nucleotides are indicated as dots. The sequence alignment was performed using ClustalW (EBI server; McWilliam et al., 2013), and the figure was generated by CLC Main Workbench 7.

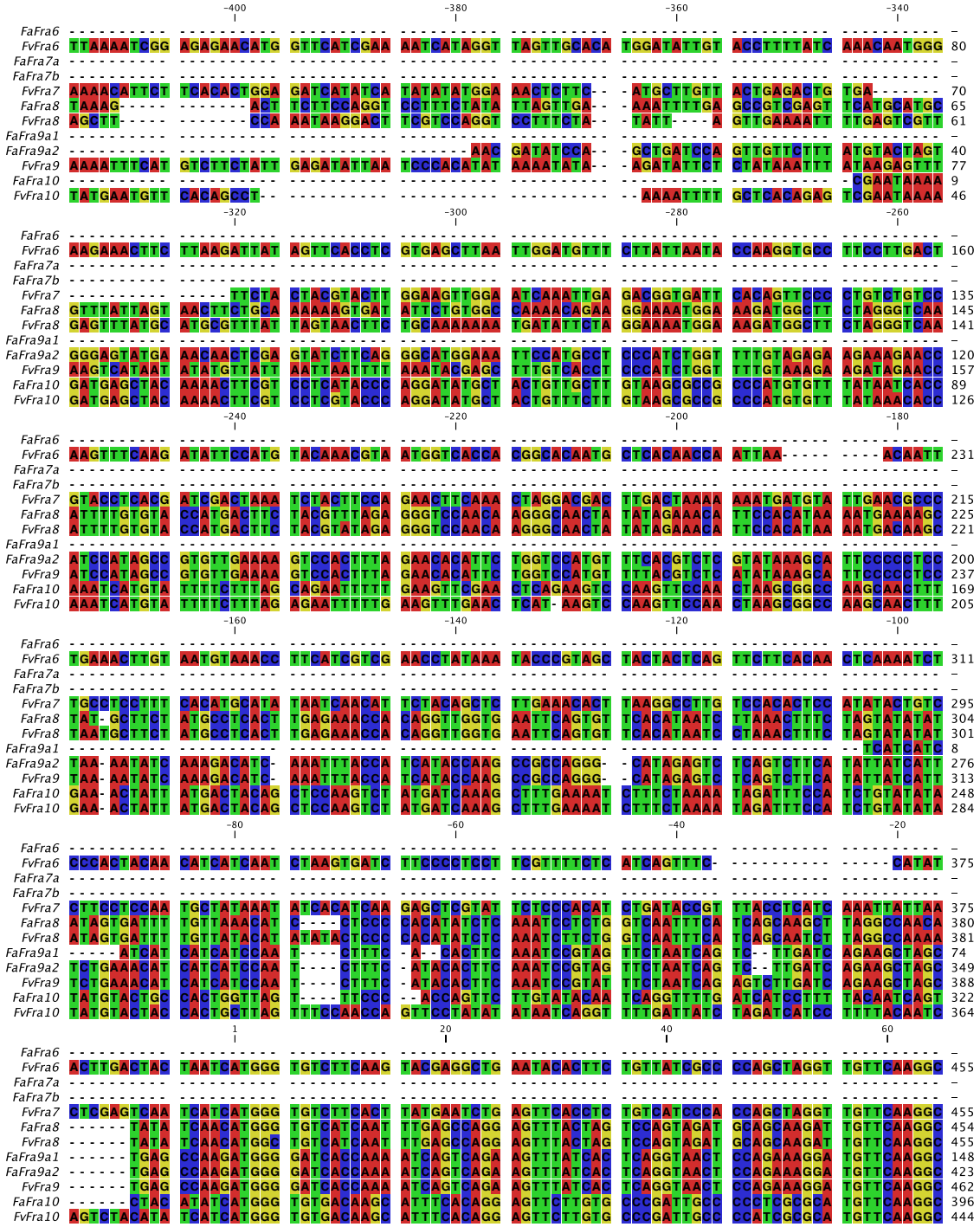


Figure 7. Multiple DNA sequence alignment of the most divergent *Fra* strawberry allergens. *FaFra6*, *FaFra7a*, *FaFra7b*, *FaFra8*, *FaFra9* and *FaFra10* correspond to newly identified *Fra* transcripts in RNA-Seq. *F. vesca* homologous are also illustrated. UTRs and CDS regions are shown for identified transcripts and *F. vesca* sequences.

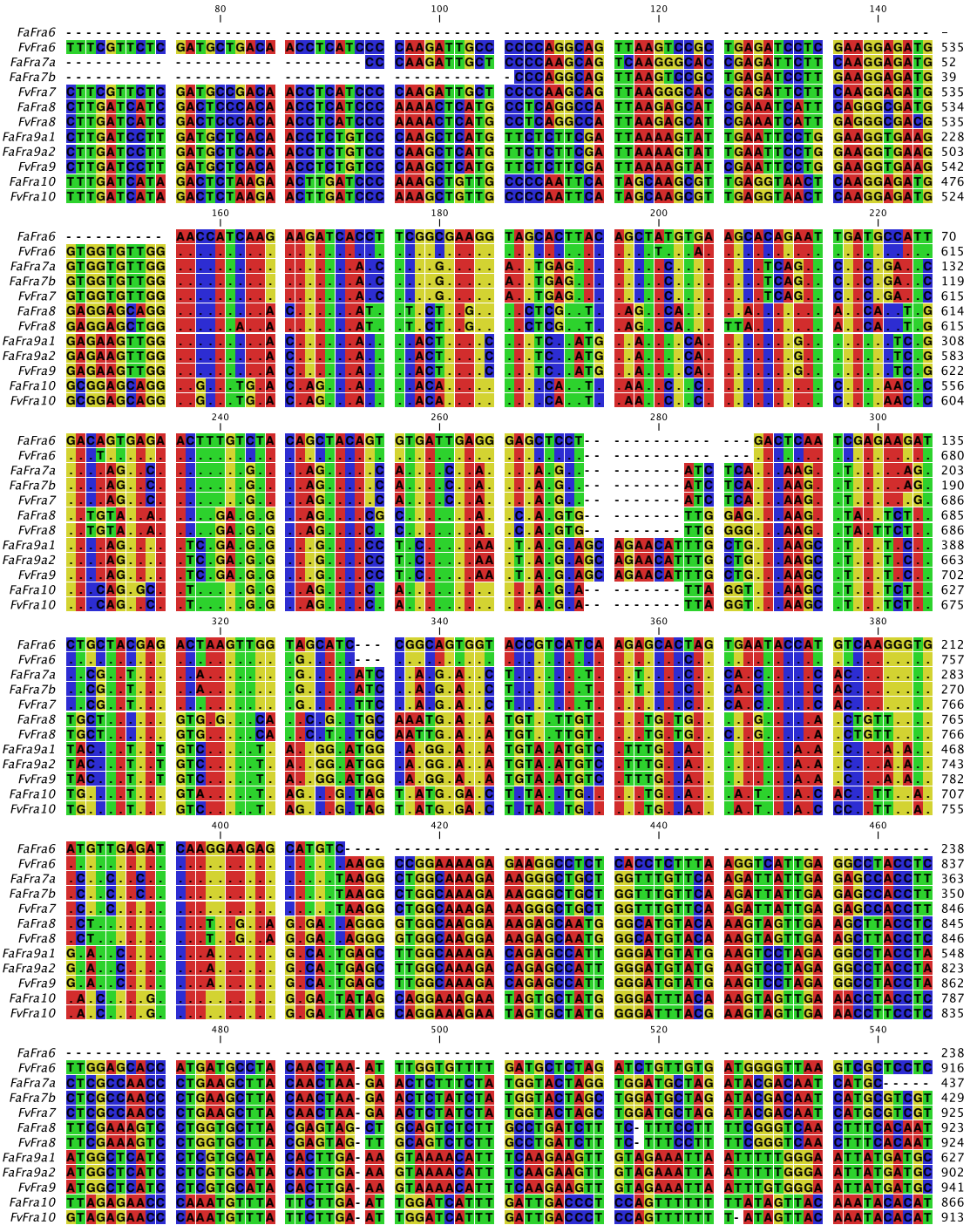


Figure 7. *(Continued.)* First nucleotide in CDS is flagged as number 1. Letters and numbers in labels designate different nucleotide sequences; same letters indicate same deduced amino acid sequence. Colour code for nucleotides: Adenine (red), Guanine (yellow), Thymine (green) and Cytosine (blue). Matching nucleotides are indicated as dots. The sequence alignment was performed using ClustalW (EBI server; McWilliam et al., 2013), and the figure was generated by CLC Main Workbench 7.

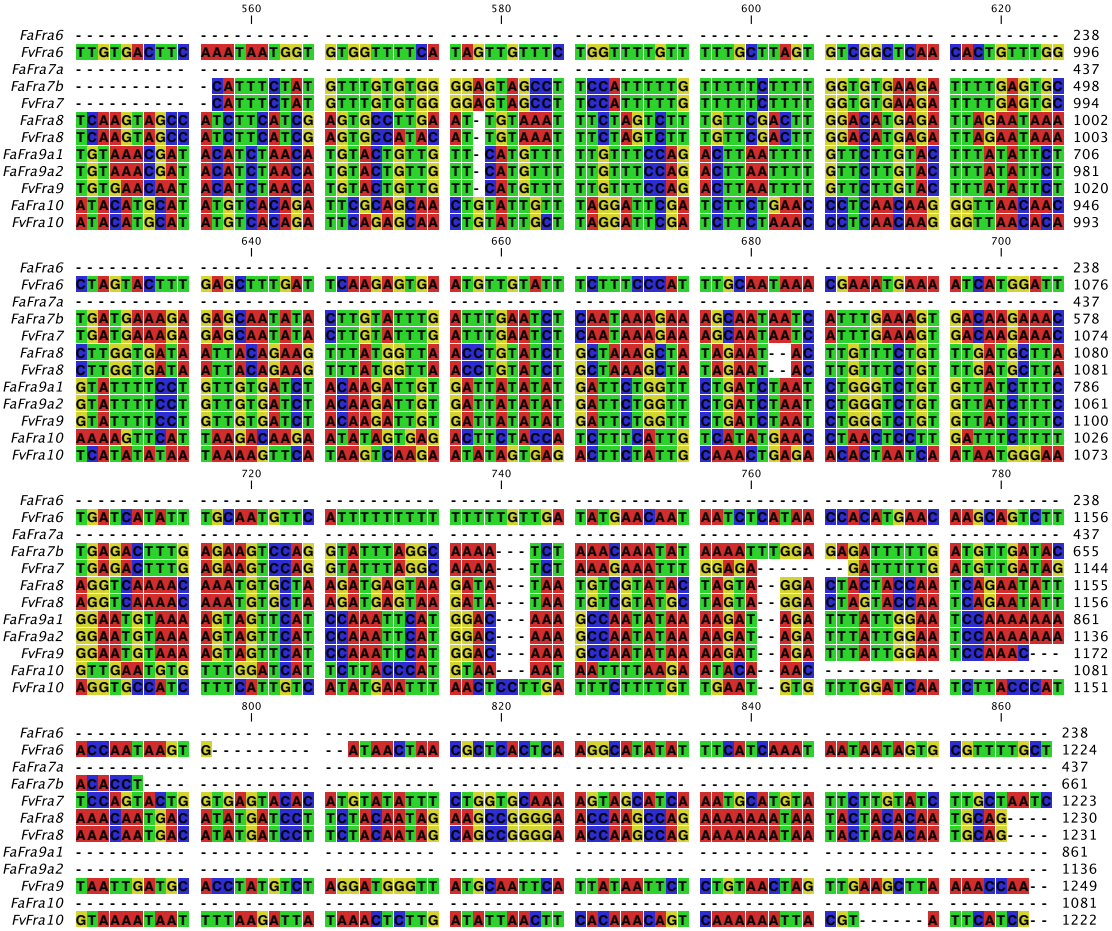


Figure 7. *(Continued.)*

Characterization of Deduced FaFra Proteins and Phylogenetic Analysis

Putative protein sequences were deduced from nucleotide sequences, analysed and compared. All proteins are 159- 163 aa long (Table 1, Figure 8) and have a calculated molecular mass ranging from 17.5 to 18.6 KDa, consistent with the described FaFra1E and FaFra3 proteins (Seutter von Loetzen *et al.*, 2012; Casañal *et al.*, 2013a). Furthermore, the four amino acids invariable across the PR-10 family, Gly-47, Gly-52, Asp-77 and Leu-152, are also conserved among these sequences (Liu *et al.*, 2006) (Figure 8). All proteins were modelled using Swiss-Model automated mode to predict their structure. All of them are composed of three α-helices and a seven-stranded antiparallel β-sheet embracing a central hydrophobic cavity, as expected for the PR-10 members and this allergen subclass (Seutter von Loetzen *et al.*, 2012; Casañal *et al.*, 2013b, Fernandes *et al.*, 2013). The phylogenetic tree shows the relationship between these proteins and indicates the presence of two major clusters (Figure 9). The first cluster includes all proteins

related to FaFra1, FaFra2 and FaFra3 and the closely related homologs FaFra4, FaFra5, FaFra6, and FaFra7, as well as their corresponding *F. vesca* representatives. The second group, consisting of FaFra8, FaFra9, FaFra10, and their homologs in *F.vesca*, present the most divergent amino acid sequences to the previously characterized FaFra1, FaFra2 and FaFra3 proteins, and therefore come out in the phylogenetic tree as the most external group of sequences.

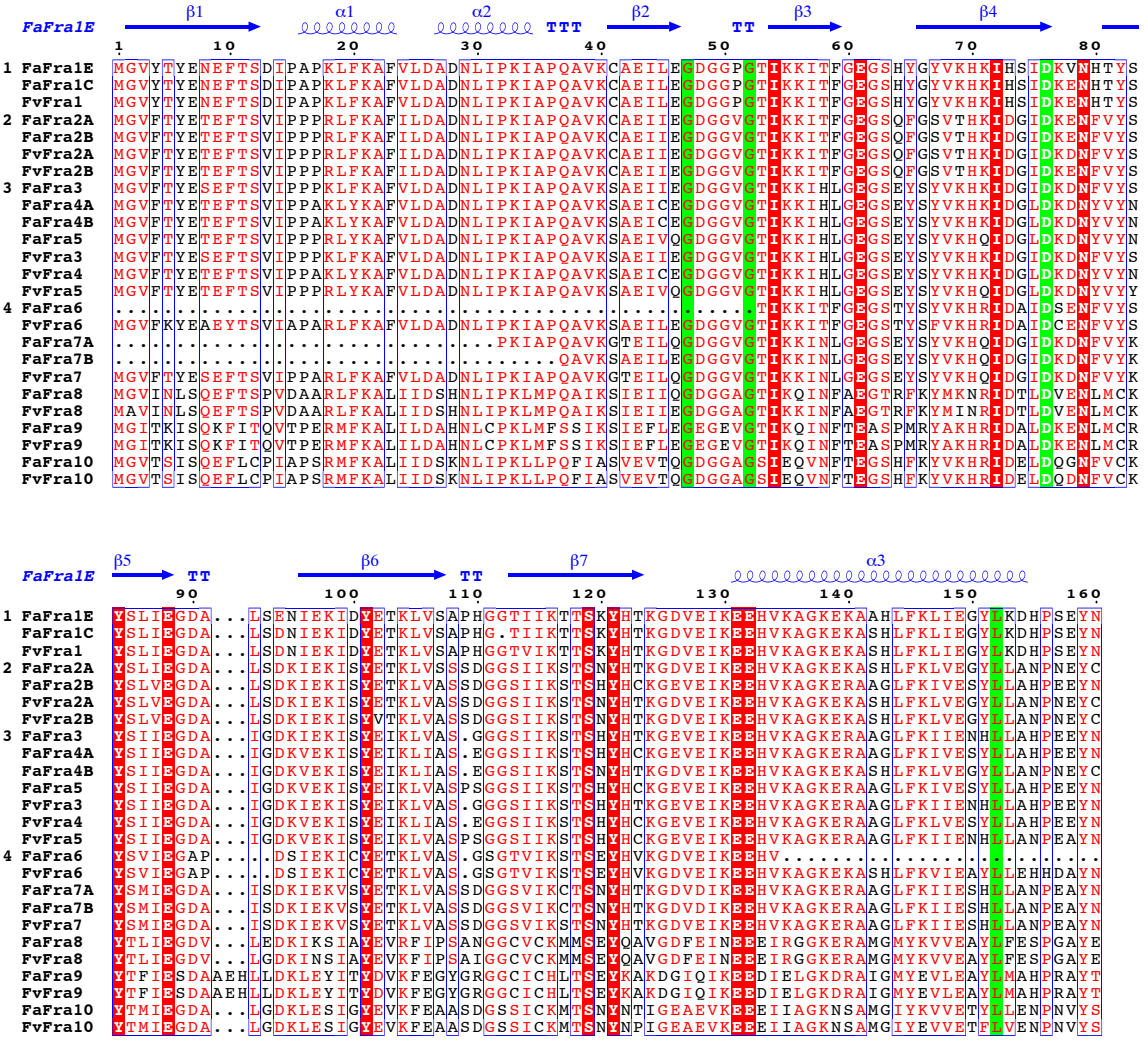


Figure 8. Multiple amino acid sequence alignment of the Fra proteins from strawberry. FaFra1E, FaFra3 correspond to already published sequences of FaFra (Seutter von Loetzen et al., 2012; Casañal et al., 2013a; Casañal et al., 2013b). The rest of the sequences correspond to deduced amino acid sequences of either *FaFra* identified transcripts or their *F.vesca* homologs. Amino acid sequences have been classified in four main groups. Groups 1, 2 and 3 include sequences closely related to the Fra1, Fra2 and Fra3 proteins.

(**Figure 8** ***Legend Continued***) The fourth group includes the incomplete amino acid sequences of FaFra6, FaFra7A and FaFra7B that were obtained from partially identified transcripts by RNA-Seq, and the most divergent FaFra proteins FaFra8, FaFra9 and FaFra10. The four residues that are strictly conserved in PR-10 proteins are highlighted in green. The secondary structure elements correspond to FaFra1E (PDB 4C9C). The sequence alignment was performed using ClustalW (EBI server; McWilliam et al., 2013), and the figure was generated by ESPript (Gouet et al., 2003).

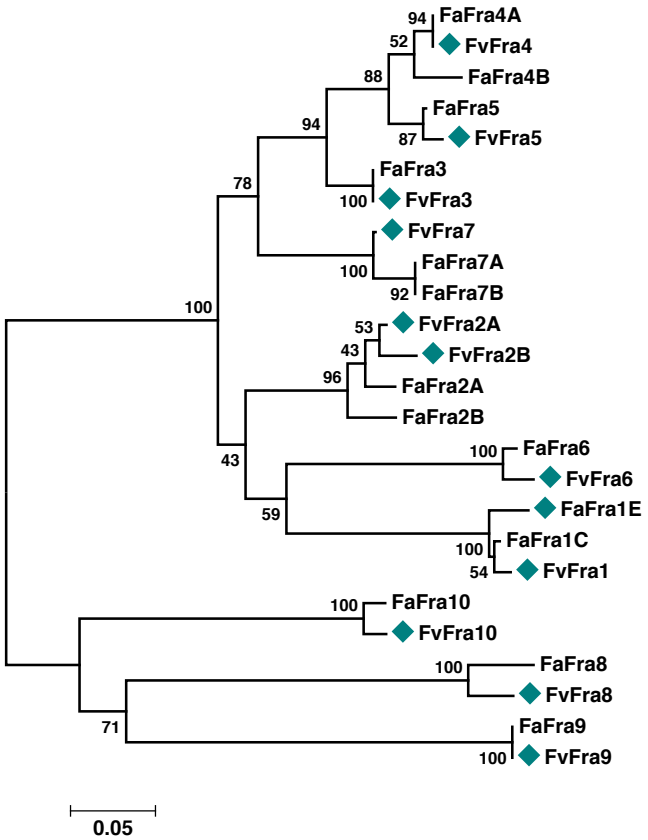


Figure 9. Phylogenetic tree of the Fra proteins of *F. xananassa* (Fa) and *F. vesca* (Fv). Two major clusters are depicted, the first including proteins related to FaFra1, FaFra2 and FaFra3 and, the second represented by the most divergent Fra proteins (Fra8-10). FvFra proteins are highlighted with diamond-shaped symbols (turquoise). Bootstrap values (%) for 1000 replicates are indicated at the nodes. MEGA5 (Tamura et al., 2011) was used to perform the phylogenetic analysis and generate the figure.

Expression of Fragaria xananassa Fra Transcripts

High throughput mRNA sequencing offers the possibility not only to identify new genes and transcripts but also to determine transcript expression levels. Trinity software was used for transcript abundance estimations of the *FaFra* genes. All methods for RNA-Seq analysis rely on

the number of reads produced from an RNA transcript, which is a function of that transcript's abundance. This is the case for Trinity as well; to calculate abundance levels, first the original reads are aligned to the Trinity transcripts and then transcript quantification is computed (Haas *et al.*, 2013). Visualization of read alignments and Trinity transcripts was performed using the Integrated Genomics Viewer (IGV) software available at http://www.broadinstitute.org/igv/.

The expression level of the 15 *FaFra* genes, as Fragments Per Kb per Million reads (FPKM) values, is shown in Table 2 and represented in Figures 10 and 11. Abundance (FPKM values) in achenes and receptacles, at green (G), white (W), turning (T) and red (R) stages, as well as that in leaves and roots is shown for each gene.

Overall, the most represented *FaFra* transcripts in strawberry correspond to *FaFra1c1*, *FaFra1c2*, *FaFra2b* and *FaFra4a1* genes (Figure 10a). Interestingly, the highest expression levels of *FaFra* transcripts are observed in roots (Figure 10b). *FaFra1c1* and *FaFra1c2* show expression levels in roots with values over 30000 and 17000 FPKM. Previous qPCR results also displayed that *FaFra1* was mainly expressed in this tissue (Muñoz *et al.*, 2010). *FaFra4a1*, *FaFra5* and *FaFra2b* are highly transcribed in roots as well, with values above 6000, 1800 and 1300 FPKM, respectively. Also in roots, *FaFra6*, *FaFra4b2* and *FaFra4b1* display transcripts levels ranging from 600 to 250 FPKM (Table 2, Figure 10b, Figure 11). In general, the expression of *FaFra* in roots has to be considered of relevance and, as depicted in Figure 11 with asterisks, differences are significant in relation to other tissues.

In achenes, *FaFra1c1* and *FaFra1c2* together accumulate the maximum transcript levels, both over 300 FPKM, (Table 2, Figure 10c, Figure 11), and display complementary expression patterns, since *FaFra1c1* expression increases from white to red achenes, while *FaFra1c2* exhibits opposite pattern along ripening. *FaFra2b*, *FaFra4a1*, *FaFra9a1* and *FaFra9a2*, also present considerable levels of expression in this organ. Previous to RNA-Seq, we performed qPCR experiments using the primers for the known *FaFra* genes at that time (Muñoz et al., 2010) (Figure 12). It should be indicated that we did not identify in our analysis of the strawberry fruit transcriptome any sequence identical to the previously cloned *FaFra1e, FaFra2a* and *FaFra3.* However, the primers used for qPCR analysis perfectly aligned not only to the *FaFra1e* sequence but also to all the *FaFra1c* ones. This was not the case for primers designed for *FaFra2a.* In the case of *FaFra3*, PCR simulations using the Lasergene software (DNASTAR, Inc.) showed that primers used for *FaFra3* qPCR analysis might also amplify *FaFra4a1*, *FaFra4a2* and *FaFra5* sequences.

Table 2. FPKM values of *FaFra* transcripts as determined by Trinity in the achenes and receptacles of strawberry during ripening (green (G), white (W), turning (T) and red (R)), as well as in leaves (L) and roots (Rt).

		FaFra1c1	*FaFra1c2*	*FaFra2b*	*FaFra2c*	*FaFra4a1*	*FaFra4a2*	*FaFra4b1*	*FaFra4b2*	*FaFra5*	*FaFra6*	*FaFra7a*	*FaFra7b*	*FaFra8*	*FaFra9a1*	*FaFra9a2*	*FaFra10*
Achene	G	231,1	219,8	95,8	0,0	69,2	0,0	18,1	21,1	22,1	0,4	1,5	0,5	3,2	43,5	29,7	0,1
	W	103,9	356,0	21,4	0,0	164,6	0,0	8,7	8,6	18,8	0,4	1,6	0,0	2,1	79,2	35,7	0,0
	T	188,5	285,7	34,7	0,0	102,7	0,0	7,6	10,3	30,6	0,3	2,4	0,8	1,7	24,9	17,9	1,2
	R	301,7	156,6	32,9	0,0	83,0	1,8	9,6	4,3	16,8	0,4	1,2	0,4	0,6	24,4	11,2	0,1
Receptacle	G	2540,7	131,3	72,6	0,0	35,8	4,3	20,0	14,4	23,0	4,0	1,8	0,6	1,0	148,8	26,5	0,3
	W	1139,5	57,4	27,1	66,5	3,5	7,0	17,4	0,8	5,7	0,5	1,2	0,2	0,5	48,6	8,8	0,0
	T	249,6	17,9	126,9	345,7	0,0	3,7	12,7	55,9	17,8	0,1	1,1	0,2	0,4	48,9	8,7	0,0
	R	38,0	3,2	147,5	388,4	1,8	5,6	19,8	63,8	37,0	0,0	1,6	0,4	0,1	49,4	8,9	0,0
	L	438,4	180,6	35,6	0,0	74,1	2,3	12,7	4,4	18,3	2,8	2,0	0,8	0,7	30,5	6,1	0,0
	Rt	33594,9	17496,4	1307,3	0,0	6280,2	0,0	253,9	318,9	1838,0	675,7	80,1	36,9	86,5	113,3	19,2	5,8

Accordingly, with reference to the qPCR analysis of *FaFra1* (Figure 12), the observed results could be considered as the sum of all *FaFra1e* and *FaFra1c* expressed genes (Figure 11). The total count of FPKM values of *FaFra1c1* and *FaFra1c2* (Figure 11) would give an expression profile in achenes not coincident with the qPCR profile (Figure 12). In relation to *FaFra3* qPCR results, *FaFra4a1* would be the gene that concentrates the maximum FPKM values within the group amplified by *FaFra3* set of primers, and it could be considered as the gene defining this group in achenes. In this case, qPCR profile is characterized by a decrease expression levels along ripening (Figure 12), similar to the observed pattern for *FaFra4a1* (Figure 11).

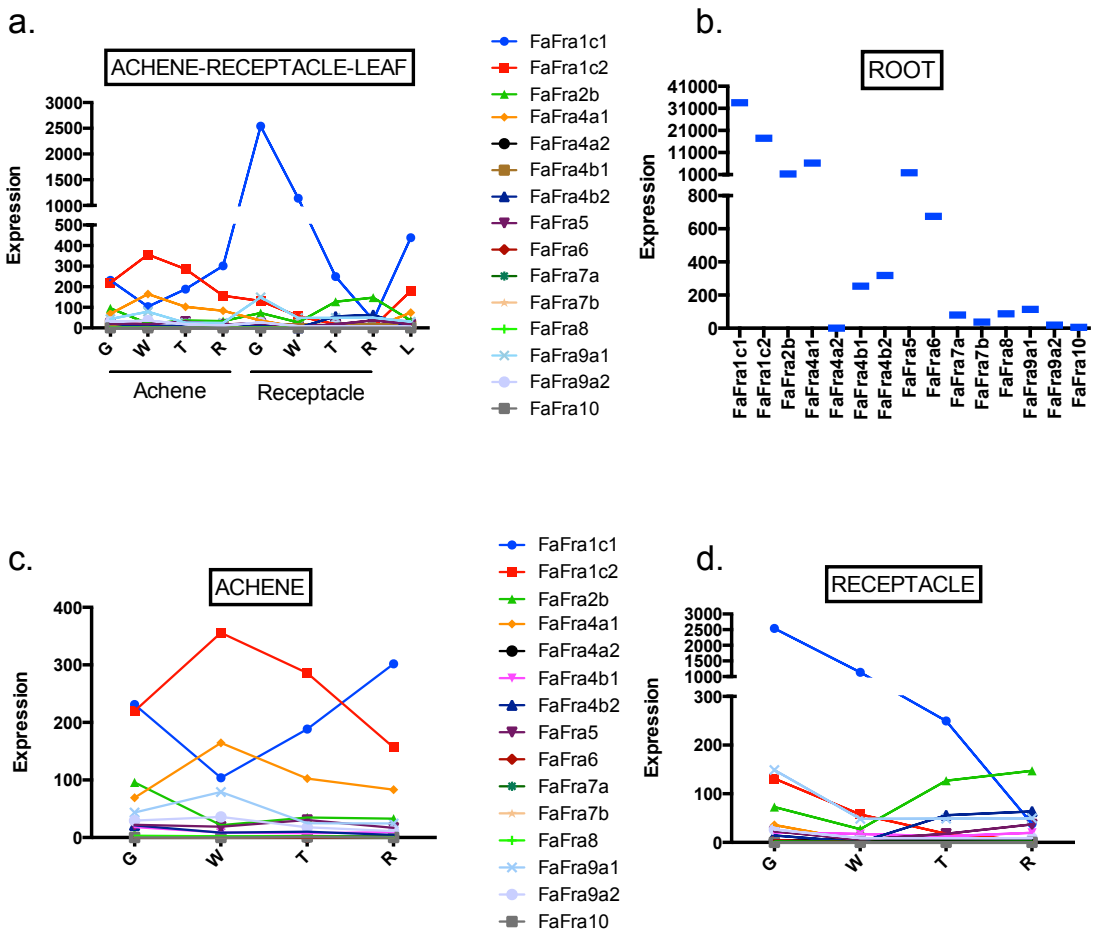


Figure 10. Expression of *Fragaria xananassa Fra* transcripts. a. Expression of *FaFra* in achenes, receptacles and leaves. **b.** Expression of *FaFra* in roots is shown separately so that expression of genes in other tissues is not misled. **c.** Detailed expression of *FaFra* in achenes. **d.** Detailed expression of *FaFra* in receptacles. Absolut expression levels are expressed as FPKM values.

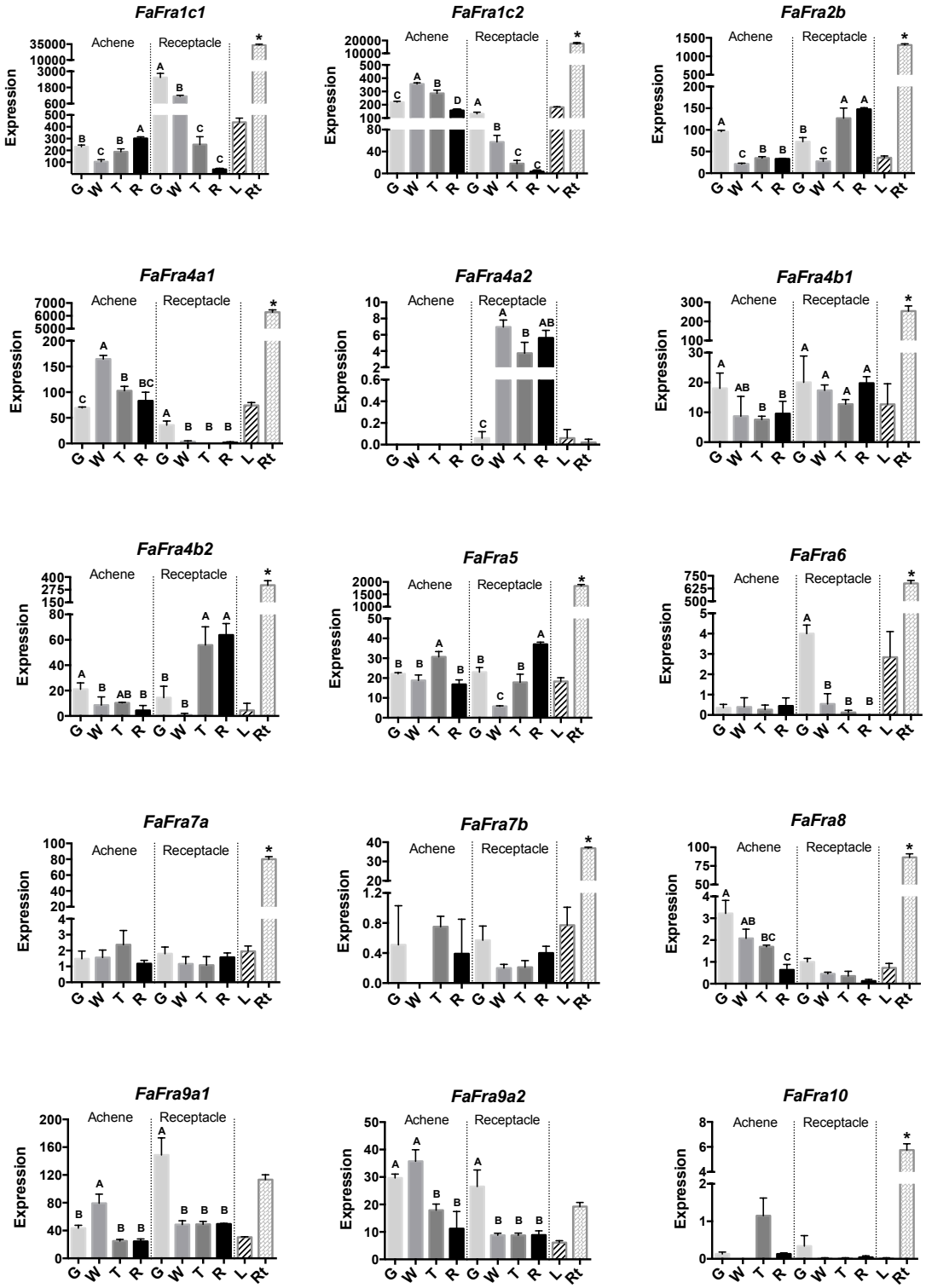


Figure 11. *(Legend appears on following page.)*

Figure 11. Expression of *Fragaria xananassa Fra* transcripts. *FaFra* genes show different expression patterns, as determined by RNA-Seq analysis, in the achenes and receptacles of strawberry fruits during ripening (green (G), white (W), turning (T) and red (R)), as well as in leaves (L) and roots (Rt). Statistical analysis was performed with Statistix 9 (Analytical Software) using analysis of variance (ANOVA). The means were compared by Tukey HSD test adjusted to a 95% significance level. To avoid underestimation of significant differences among data, statistical analysis was applied to achenes, receptacles, and all tissues separately. Capital letters represent significant differences within achenes and receptacles. Asterisks are used to illustrate significant differences observed in roots after statistical comparison of all tissues. Error bars indicate standard deviations from three biological replicates. Absolut expression levels are expressed as FPKM values.

In receptacles, the highest expression levels are found for transcript *FaFra1c1*, with values around 2500 (Table 2, Figure 10d, Figure 11). *FaFra1c2*, *FaFra2b* and *FaFra9a1* are the successive ones in expression, as they also display high levels within this tissue, and all of them with values over 100 FPKM. While *FaFra1c1* is the most expressed transcript in green receptacle, *FaFra2b* is the highest in red receptacle. Precisely, *FaFra1c1* transcript levels decrease from the green to the ripe fruit; same expression pattern is observed for *FaFra1c2*; and, in contrast, the *FaFra2b* gene is highly transcribed at the late stages of receptacle ripening. These results were validated by qPCR, which confirmed the described profiles in receptacle (Figure 12). Western analysis, using anti-FaFra2A polyclonal antibodies, also supported the fruit developmental pattern for *FaFra1* and *FaFra2* genes (Figure 13). Regarding immunoblot assays, it has to be clarified that: 1) anti-FaFra2A polyclonal antibodies were effective to detect the three recombinant proteins purified from *E. coli* (FaFra1E, FaFra2A and FaFra3 (Casañal *et al.*, 2013a), which were used as experimental controls; 2) anti-FaFra2A antibodies preferentially recognized FaFra2A, therefore recombinant proteins were loaded in different concentrations for their optimal visualization (FaFra2A 10 times less concentrated than FaFra1E and FaFra3) and 3) the differences in size observed between recombinantly expressed and strawberry extracted proteins rely on the fact that three foreign amino acids remain at the N-terminal end of purified proteins from *E. coli* purified proteins (Casañal *et al.*, 2013a). Concerning *FaFra3* qPCR results, *FaFra4a1* (Figure 11) would be the gene that best fits *FaFra3* expression profile in receptacle (Figure 12), which is defined by a reduction of transcripts levels as ripening increases.

In leaves, the maximum transcripts levels correspond to *FaFra1c1* and *FaFra1c2* genes, with *FaFra4a1* being the one that follows, and FPKM values around 430, 180 and 70 FPKM, respectively (Table 2, Figure 10a, Figure11).

Despite the wide range of expression levels and developmental patterns here reported for *FaFra* genes in strawberry, it is noteworthy to point out that *FaFra1c2* and *FaFra4a1* present very high similarity, in both transcripts level and developmental pattern in fruits (Figure 11).

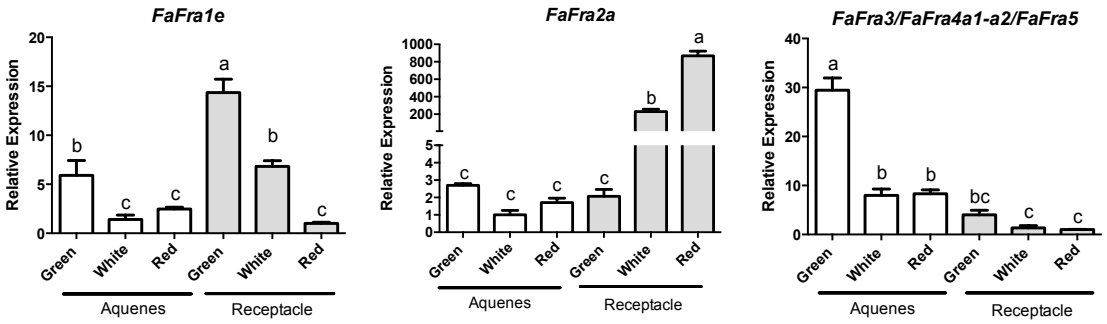


Figure 12. Expression analysis of the *FaFra1e*, *FaFra2a*, and *FaFra3* genes in the strawberry *F. xananassa* cv. Camarosa. *FaFra1e*, *FaFra2*, and *FaFra3* show different expression patterns, as determined by qPCR, in the achenes and receptacles of strawberry fruits during ripening. Primers used for *FaFra3* qPCR, might also amplify sequences *FaFra4a1*, *FaFra4a2* and *FaFra5*, as simulated with Lasergene software (DNASTAR, Inc.). For all genes, the expression level is relative to the sample with the highest Ct value. *FaRib413* was used as housekeeping gene to normalize the data. Different letters indicate significant differences, in achenes and receptacles, for each gene using ANOVA and the Tukey HSD test adjusted to a 95% significance level. Error bars indicate standard errors from three biological replicates.

Figure 13. *(Legend appears on following page.)*

Figure 13. Strawberry fruits (*F. xananassa* cv. Camarosa) used for RNA-Seq analysis and immunoblotting. a. Green, white, turning and red strawberries were collected to perform RNA-Seq analysis of aquenes and receptacles during fruit ripening. **b.** Western blot analysis of strawberry receptacles indicates that the FaFra1 protein level decreases from the green to the ripe fruit whereas the FaFra2 protein accumulates in the late stages of fruit ripening. This result is in accordance with the transcript levels observed by qPCR. FaFra1E, FaFra2A and FaFra3 recombinant proteins purified from *E. coli* are used as a control. Anti-FaFra2A antibodies recognize FaFra2A preferentially; therefore, 10X recombinant FaFra1E and FaFra3 proteins were loaded in comparison to the FaFra2A control. Equal amounts of total protein extracts from strawberry were loaded after their quantification by the modified Lowry method (Hartree, 1972).

RNAi Transient Down-Regulation of FaFra Genes

To shed light on FaFra proteins role in strawberry receptacle ripening, we set out to identify differentially expressed genes between pools of *FaFra-RNAi* transiently silenced and control receptacles. Massive RNA-sequencing provides an accurate quantification of differential transcript expression (Trapnell *et al.*, 2012; Sánchez-Sevilla *et al.*, 2014), thus an RNA-Seq experience was designed to further investigate FaFra proteins function. RNAi-mediated down-regulation of the *FaFra* genes in receptacle during fruit ripening was performed using the pBI-*Fraa1ei* vector, as previously described (Muñoz *et al.*, 2010). Half *Fragaria xananassa* (cv. Camarosa) fruits were infiltrated, and 7 days later, an impaired colour development was observed in fruits injected with *Agrobacterium* containing the pBI-*Fraa1ei* construct, while control fruits developed an uniform red colour (Figure 14a). As a positive control of the down-regulation experiments, pBI-*FaCHSi* was agroinfiltrated; injected fruits exhibited the lack of pigmentation also displayed by pBI-*Fraa1ei* infiltrated fruits (data not shown), which validated silencing of target genes. RNA was extracted from bulked pools of the half agroinfiltrated fruits and it was used for both qPCR analysis and RNA-Seq. To confirm down-regulation of *FaFra* genes we examined the expression levels of *FaFra1e*, *FaFra2a* and *FaFra3* genes by qPCR (Figure 14b). Previous studies of transient silencing with the same construct showed a general down-regulation of *FaFra* genes (Muñoz *et al.*, 2010). However, in this case we found that only *FaFra2* was down-regulated to a 50%, and simultaneously enhanced expression of *FaFra1e* and *FaFra3* genes was observed. This result was also extended to the protein content, as revealed by immunoblot assays on *Fra-RNAi* silenced and control receptacles (Figure 14c). In all the cases we found that the FaFra protein content was parallel to the level of transcripts.

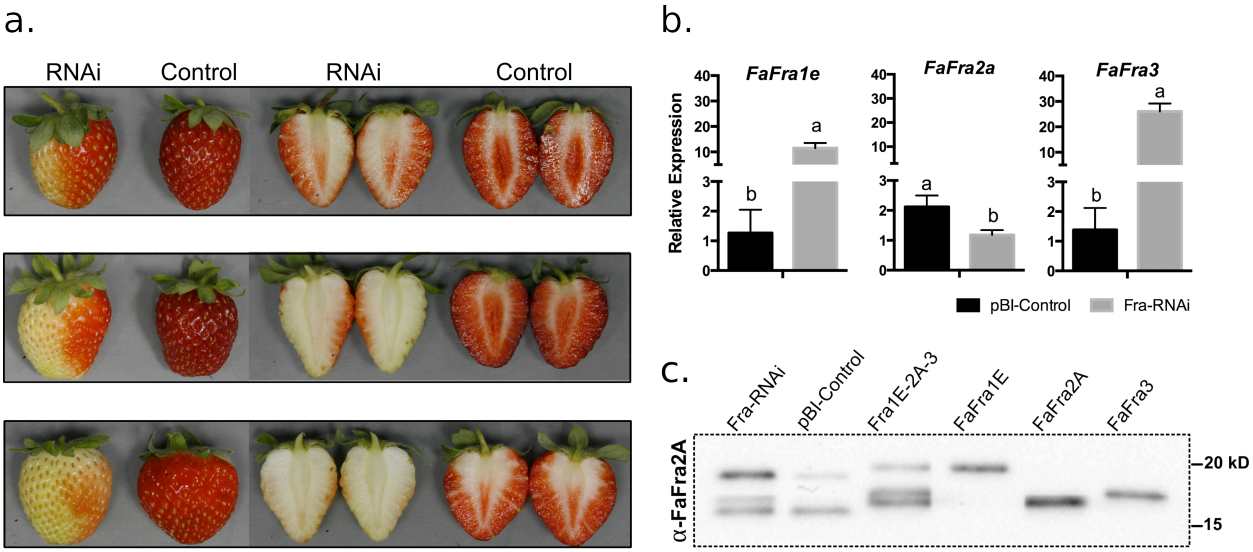


Figure 14. Transient RNAi-mediated silencing of the *FaFra* genes in strawberry fruits (*F. xananassa* cv. Camarosa). a. Fruits of *F. xananassa* cv. Camarosa after 7 days of infiltration with *A. tumefaciens* harboring the control vector pBI–Intron and the pBI-*Fraa1ei* for RNAi-mediated silencing of the *FaFra* genes. Loss of pigmentation demonstrates the effect on anthocyanin formation. Three independent experiments are shown. **b.** Relative gene expression levels in agroinfiltrated fruits show down-regulation of *FaFra2a* as determined by qPCR. *GAPDH* was used as housekeeping gene to normalize the data. Different letters indicate significant differences for each gene using ANOVA and the Tukey HSD test adjusted to a 95% significance level. Error bars indicate standard deviations from three biological replicates. **c.** Western blot analysis of strawberry receptacles indicates that the FaFra1 and FaFra3 proteins accumulate in the pBI-Fraa1ei treated fruits whereas the FaFra2 protein level remains similar in comparison to control fruits. This result is in accordance with the transcript levels observed by qPCR. FaFra1E, FaFra2A, FaFra3 recombinant proteins purified from E. coli, and the mix of the three of them (Fra1E-2A-3), are used as a control. Anti-FaFra2A antibodies recognize FaFra2A preferentially; therefore, 10X recombinant FaFra1E and FaFra3 proteins were loaded in comparison to the FaFra2A control. Equal amounts of total protein extracts from strawberry were loaded after their quantification by the modified Lowry method (Hartree, 1972).

The fact that the *F. vesca* sequenced genome is still incomplete, led us to first identify the expressed *FaFra* transcripts using a *de novo* assembly strategy. This has been of help to characterize the members of the FaFra gene family expressed in the sampled tissues and stages, i.e. leaf, root, achene, and receptacle (Figure 9). However, the *F. vesca* genome is suitable when global expression analysis of gene families is intended. Therefore, RNA-Seq data analysis to identify differentially expressed genes between pools of *FaFra-RNAi* transiently silenced and control receptacles was performed using the *F. vesca* genome as the input reference. RNA-Seq

reads were analysed using TopHat2 (Kim *et al.*, 2013) and Cufflinks (Trapnell *et al.*, 2012) softwares. First, TopHat2 enabled alignment of reads to the reference genome *Fragaria vesca* Whole Genome (v1.1) (http://www.rosaceae.org/), which provided the location from which the reads originated, and also annotation of the genes (CDS v1.0). After mapping the RNA-Seq reads to the reference genome (*F. vesca*), transcripts were assembled and their relative abundances calculated using Cufflinks. Differential gene expression between control and *FaFra-RNAi* pools was calculated using the ratio of the FPKM values of each gene in both pools. A total of 4426 predicted genes were differentially expressed between the pools. Of these, 2528 were up-regulated and 1898 were down-regulated. Among the 4426 differentially expressed genes, 4215 corresponded to annotated genes in the *F. vesca* gene model v.1.0 (Shulaev *et al.*, 2011) while 211 matched with not annotated genome regions. We focused our analysis first on the *FvFra* family members and then on those genes involved in the phenylpropanoid and flavonoid pathways.

Table 3. List of the 10 *FvFra* genes up-regulated and down-regulated in the Control and RNAi silenced receptacles of *Fragaria xananassa*.

Fvesca gene	Function	FPKM control	FPKM RNAi	log2 fold change
gene07080-v1.0-hybrid	FvFra1	3177,37	34797,40	3,45
gene07065-v1.0-hybrid	FvFra2A	5179,45	2706,01	-0,94
gene07086-v1.0-hybrid	FvFra2B	488,74	672,18	0,46
gene07082-v1.0-hybrid	FvFra3	162,77	1718,21	3,40
gene07085-v1.0-hybrid	FvFra4	1720,45	8494,59	2,30
gene07084-v1.0-hybrid	FvFra5	313,89	1046,64	1,74
gene05185-v1.0-hybrid	FvFra6	25,86	19,13	-0,43
gene32299-v1.0-hybrid	FvFra7	180,60	996,45	2,46
gene11094-v1.0-hybrid	FvFra8	12,50	26,23	1,07
gene07077-v1.0-hybrid	FvFra9	451,39	515,10	0,19
gene07076-v1.0-hybrid	FvFra10	4,74	6,80	0,52

Gene number according to the *Fragaria vesca* genome draft (www.Strawberrygenome.org).

Table 4. List of *F. vesca* genes up-regulated and down-regulated in the Control and RNAi silenced receptacles of *Fragaria xananassa*.

Fvesca gene	Function	FPKM control	FPKM RNAi	log2 fold change
gene09753-v1.0-hybrid	PAL	283,80	240,80	-0,24
gene23261-v1.0-hybrid	PAL	392,47	415,17	0,08
gene26265-v1.0-hybrid	C4H	8997,53	4111,24	-1,13
gene28093-v1.0-hybrid	C4H	694,78	560,68	-0,31
gene12222-v1.0-hybrid	C4H	0,57	6,95	3,61
gene32451-v1.0-hybrid	4CL	221,57	107,87	-1,04
gene09603-v1.0-hybrid	4CL	697,80	352,64	-0,98
gene26998-v1.0-hybrid	4CL	118,40	79,80	-0,57
gene12918-v1.0-hybrid	4CL	244,88	188,25	-0,38
gene05255-v1.0-hybrid	4CL	82,06	133,97	0,71
gene08621-v1.0-hybrid	4CL	8,86	21,62	1,29
gene12577-v1.0-hybrid	4CL	9,17	31,50	1,78
gene15877-v1.0-hybrid	4CL	2,75	21,77	2,98
gene26825-v1.0-hybrid,gene26826-v1.0-hybrid	CHS	6992,45	5049,06	-0,47
gene21346-v1.0-hybrid	CHI	8910,58	5875,72	-0,60
gene32055-v1.0-hybrid,gene32056-v1.0-hybrid	DFR	63,52	30,99	-1,04
gene29482-v1.0-hybrid,gene29483-v1.0-hybrid	DFR	208,97	92,58	-1,17
gene29484-v1.0-hybrid	DFR	111,44	85,20	-0,39
gene15174-v1.0-hybrid	DFR	2500,82	2134,04	-0,23
gene15176-v1.0-hybrid	DFR	2340,34	2157,00	-0,12
gene32066-v1.0-hybrid	DFR	34,43	66,70	0,95
gene31464-v1.0-hybrid	DFR	35,84	115,50	1,69
gene29478-v1.0-hybrid,gene29479-v1.0-hybrid	DFR	6,05	22,40	1,89
gene18711-v1.0-hybrid	DFR	0,72	5,47	2,92
gene14611-v1.0-hybrid	F3H	6288,94	3699,77	-0,77
gene12992-v1.0-hybrid	F3H	1,75	2,56	0,55
gene12991-v1.0-hybrid	F3H	91,80	134,50	0,55
gene19319-v1.0-hybrid	F3H	0,87	1,81	1,05
gene19320-v1.0-hybrid	F3H	0,63	2,07	1,72
gene20725-v1.0-hybrid	3-GT	291,72	25,55	-3,51
gene20726-v1.0-hybrid,gene20728-v1.0-hybrid	3-GT	55,53	7,53	-2,88
gene14947-v1.0-hybrid	3-GT	458,76	89,94	-2,35
gene07876-v1.0-hybrid	3-GT	43,11	19,51	-1,14
gene12591-v1.0-hybrid	3-GT	1996,46	1100,89	-0,86
gene26342-v1.0-hybrid,gene26343-v1.0-hybrid,gene26344-v1.0-hybrid	3-GT	54,81	35,25	-0,64
gene26345-v1.0-hybrid	3-GT	14,81	9,82	-0,59
gene24225-v1.0-hybrid,gene24226-v1.0-hybrid	3-GT	134,27	92,79	-0,53
gene08733-v1.0-hybrid	3-GT	145,12	101,81	-0,51
gene26352-v1.0-hybrid	3-GT	16,75	24,49	0,55
gene13733-v1.0-hybrid	3-GT	2,97	4,41	0,57
gene26353-v1.0-hybrid	3-GT	144,01	232,77	0,69
gene12684-v1.0-hybrid	3-GT	1,25	5,19	2,05
gene34393-v1.0-hybrid	3-GT	3,68	16,20	2,14
gene22597-v1.0-hybrid	3-GT	2,09	13,30	2,67
gene15107-v1.0-hybrid	3-GT	6,03	46,46	2,95

Gene number according to the *Fragaria vesca* genome draft (www.Strawberrygenome.org).

In Table 3, differential expression of the *F. ananassa* genes corresponding to the *FvFra* genes annotated in the *F. vesca* map is displayed. The observed ratios (log2 fold change) of differential expression ranged from 3.45 to -0.94, with positive and negative values indicating up- and down-regulation of the *Fra* genes (from red to blue). It is noteworthy the case of the two *Fra2* genes, the most expressed in the ripe receptacle. Whereas *FaFra2A* was down-regulated in the silenced fruits (Table 3), as also determined by qPCR (Figure 14), *FaFra2B* was up-regulated (Table 3); however, differences in FPKM values between the two *Fra2* genes was high, and this must be taken account. Together with the *FaFra2A*, *FaFra6* was found to be down-regulated as well. However, the expression level and the diminution ratio were much lower. In contrast, as previously observed by qPCR and western blotting (Figure 14), *FaFra1* was up-regulated in relation to the untransformed fruits. Also, as revealed by our qPCR and immunoblot assays (Figure 14), *FaFra3* was found to be up-regulated and, in fact, presented the most outstanding differential expression ratio within the family after *FaFra1*. *FaFra4* and *FaFra5* also presented an increased expression in the silenced fruits (Table 3); since the primers used for *FaFra3* analysis would also amplify these two genes, this increase could be reflected in the qPCR analysis (Figure 14). *FaFra7* also resulted in an increased expression in silenced receptacle; same was observed for *FaFra8-9-10*, although the effect on these genes might be less remarkable due to their low FPKM or log 2 fold ratios.

To study the effect of the *FaFra* silencing on the phenylpropanoid and flavonoid pathways, genes related to the anthocyanins biosynthesis were the principal targets of our analysis. Forty-five genes belonging to these pathways were differentially expressed between the two pools: Of these, 24 were down-regulated and 21 up-regulated in the transformed receptacle relative to the untransformed control (Table 4). The observed ratios (log2 fold change) of differential expression ranged from 3.61 to -3.51, with positive and negative values indicating up- and down-regulation of the genes (from red to blue). Phenylalanine ammonia-lyase (*PAL*) (gene09753) and chalcone synthase (*CHS*) (gene26825-26826) were down-regulated together with *FaFra2A*, as previously described (Muñoz *et al.*, 2010). Three cinnamate-4-hydroxylase (*C4H*) genes were also identified; of these, the two with higher expression levels were significantly decreased in silenced receptacles (gene26265, gene28093). Eight 4-coumarate CoA ligase (*4CL*), 4 up-regulated and 4-down regulated genes, were also identified; within this gene family the four presenting the higher expression (gene32451, gene09063, gene26998, gene12198) were significantly down-regulated. The suppression of chalcone isomerase (*CHI*) (gene21346) was found to be significant as well. Dihydroflavonol reductase (*DFR*) genes were also characterized within this analysis; in this case, the highest expressed genes (gene15174 and gene15176) were

also down-regulated, being in this case the diminution ratio lower than in other genes. Further analysis led to the determination of five differentially expressed flavanone 3-hydroxylase (*F3H*) genes. Again, the only down-regulated gene within this group (gene14611) corresponded to the highest expressed one in receptacle (over 3000 FPKM). Finally, 17 3-O-glucosyltransferase genes (*3-GT*) were identified and their differential expression evaluated. In this family, also the three genes with significant higher expression in the receptacle were down-regulated (gene12591, gene14947, gene20725).

It is known that *MYB* transcription factors play a role in the regulation of anthocyanin biosynthesis (Lepiniec *et al.*, 2006; Vogt *et al.*, 2010). In strawberry, *FaMYB1* and *FaMYB10* have been proposed to be involved in the control anthocyanin biosynthesis in fruits (Aharoni *et al.*, 2001; Medina-Puche *et al.*, 2013). Our analysis of their expression in the silenced *FaFra* fruits revealed that both *FaMYB1* (gene09407) and *FaMYB10* (gene31413) were down-regulated in parallel to the *FaFra2* gene (Figure 15).

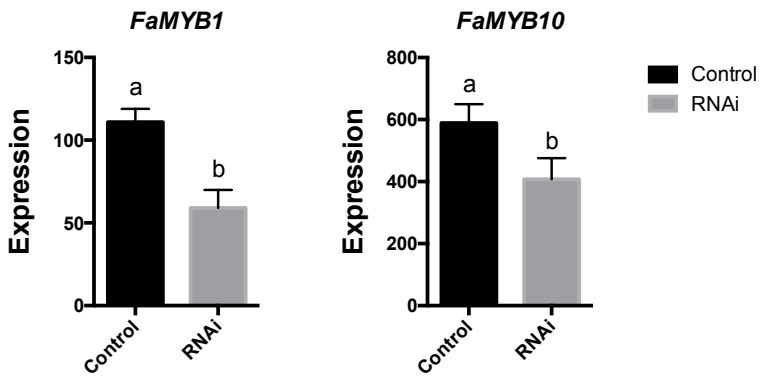


Figure 15. Expression of *FaMYB* transcription factors. RNA-Seq analysis of RNAi-mediated silencing of *FaFras*, shows down-regulation of *FaMYB1* and *FaMYB10* genes. Different letters indicate significant differences for each gene using ANOVA and the Tukey HSD test adjusted to a 95% significance level. Error bars indicate standard errors from three and five biological replicates for control and RNAi samples, respectively. Absolut expression levels are expressed as FPKM values.

In Vivo Localization of FaFra Proteins

Subcellular localization is an important information that can contribute to understand the function of a protein. It is known that the end products of the flavonoid metabolism mainly accumulate in vacuoles and are sometimes secreted to the cell wall (Saslowsky and Winkel-Shirley 2001), but they can also be found in other cellular locations, such as the cytoplasm and the nucleus (Saslowsky *et al.*, 2005). Regarding the biosynthetic pathway it has been reported to

take place primarily in the cytoplasm, where the enzymes of the pathway are organized as metabolons, or multiprotein complexes, at the endoplasmic reticulum (ER) (Winkel, 2004). However, there are also reports that point to specific localization of some enzymes of the pathway to the nucleus (Saslowsky *et al.*, 2005). Previous results (Muñoz *et al.*, 2010) and silencing phenotype (Figure 14) support a prominent role of FaFra in flavonoids biosynthesis in developing receptacle. To examine subcellular distribution of FaFra proteins, transient transformation of *Nicotiana benthamiana* leaves and transgenic *A. thaliana* plants, expressing FaFra1E, FaFra2A and FaFra3 proteins fused in their N- or C- terminus to GFP6, were produced. Only the plants transformed with *GFP6* fused to the N terminus of *FaFra* (pMDC43 vectors) displayed fluorescence, and were used for further analysis.

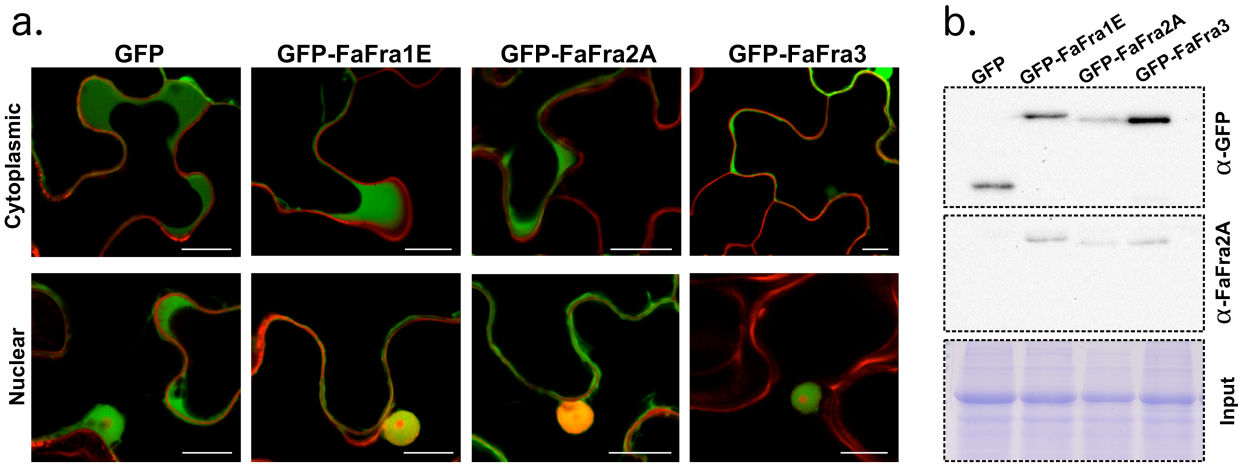


Figure 16. Co-localization of FaFra-GFP fusion proteins with propidium iodide in wild-type *Nicotiana benthamiana* plants. a. Transient expression of GFP, GFP-FaFra1E, GFP-FaFra2A and GFP-FaFra3 was achieved by agroinfiltration of wild type *N. benthamiana* leaves. Images of leaf epidermal cells are shown. Free GFP and all GFP-FaFra fusion proteins localized predominately to the nucleus and the cytoplasm, but not to the nucleolus. Proteins were visualized using confocal microscopy and propidium iodide staining after 5-10 min of uptake. Scale bars, 20 μm. **b.** Expression of all constructs was confirmed by western blot.

In transient expression assays in *N. benthamiana* leaf epidermal cells, double labelling was performed using GFP and propidium iodide (PI) as reporters of GFP-FaFra fusions. PI is a nuclear acid binding dye and a fluorescent molecule that is used to visualize the nucleus and other DNA/RNA containing organelles in the cytoplasm (Suzuki *et al.*, 1997). It is commonly used for cell wall staining as well (Nakajima & Benfey, 2002; Petricka & Benfey, 2008). Confocal images of observed fluorescence are shown in Figure 16a. *N. benthamiana* leaves

expressing GFP6 alone, displayed fluorescence throughout the cytoplasm and the nucleus. PI (red) outlines cells and stains nuclei. Cell walls are visible in red from PI staining. Although, co-localization of GFP and PI is found in the nucleus, only PI is found to localize to the nucleolus, which indicates that GFP6 is not present is this organelle. The same pattern and distribution was observed for the GFP-FaFra proteins and PI co-localization. Expression of all constructs was confirmed by western blot (Figure 16b). *A. thaliana* plants expressing GFP6 alone displayed fluorescence in the cytoplasm and nucleus of cotyledon, hypocotyl and root cells (Figure 17a). Similar results were observed in GFP-FaFra transgenic plants, in which localization in the nucleus and cytoplasm was also visible in several independent lines (Figure 17a). Expression of all constructs was confirmed by western blot analysis (Figure 17b).

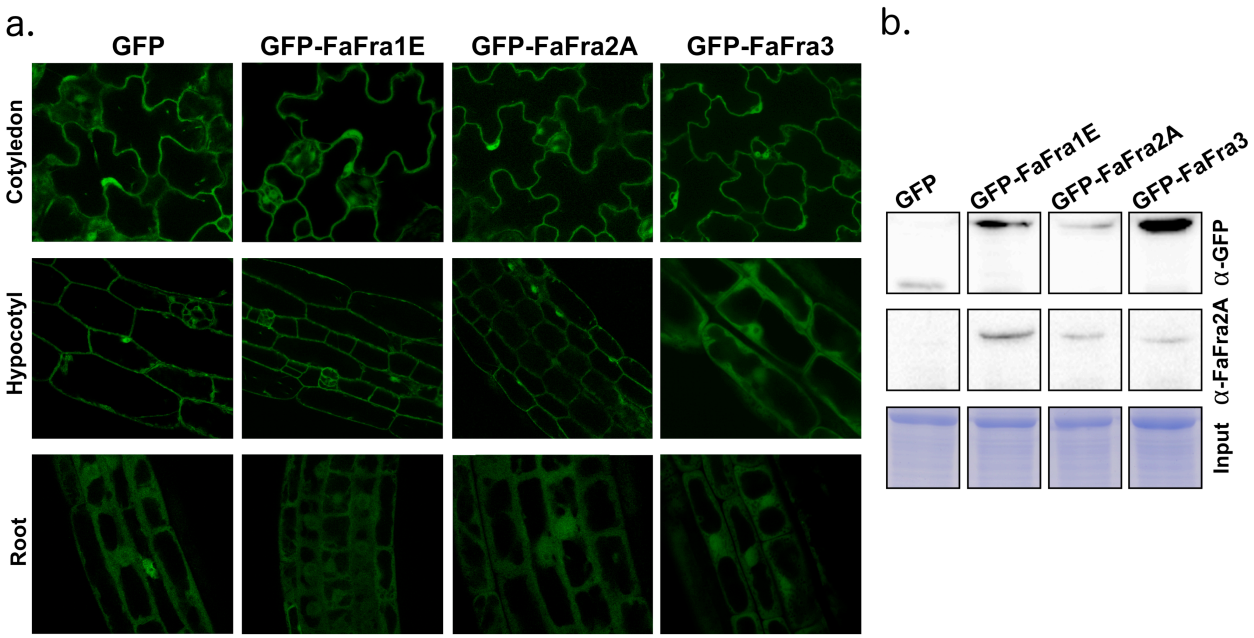


Figure 17. Subcellular localization of FaFra1E, FaFra2A and FaFra3 GFP fusion proteins in *A. thaliana* transgenic plants. a. Confocal microscopy images of GFP, GFP-FaFra1E, GFP-FaFra2A and GFP-FaFra3 proteins in cotyledon, hypocotyl and root cells of transgenic *A. thaliana* are depicted. Free GFP and all GFP-FaFra fusion proteins localized predominately to the nucleus and the cytoplasm. Scale bars, 20 μm. **b.** Expression of all constructs was confirmed by western blot.

DISCUSSION

Strawberries are rich in flavonoids, which are known by their antioxidant capacity and their beneficial effects in health (Giampieri *et al.*, 2012) and also play and important role in the quality and commercial value of fruits (Fait *et al.*, 2008; Tulipani *et al.*, 2008; Muñoz *et al.*, 2011). To date, the flavonoid biosynthetic pathway has been extensively studied (Hassan & Mathesius, 2012) but additional research is needed to better understand how the regulation of this pathway is coordinated through the many players, (hormones, developmental, environmental and transcription factors) that seem to be part of this highly integrated regulatory network. Previous studies have demonstrated that FaFra proteins are involved in the control of the flavonoid pathway (Muñoz *et al.*, 2010); however, their function at the molecular level is to be elucidated. The data that we show represent a step towards the understanding of the FaFra roles in strawberry plants, since the characterization of this family of proteins in this species is here presented.

Because the current reference genome for the *Fragaria* genus is still a draft and its annotation remains incomplete, we performed a *de novo* assembly of transcripts using Trinity (Grabherr *et al.*, 2010) for the identification of the most relevant expressed *FaFra* genes. Transcriptome analysis of *Fragaria xananassa* (cv. Camarosa) strawberry fruits by RNA-Seq allowed us to identify 15 *FaFra* sequences, 12 of them not characterized before. Up to now, only three members of the FaFra family had been previously described: *FaFra1*, *FaFra2a* and *FaFra3* (Musidlowska-Persson *et al.*, 2007; Muñoz *et al.*, 2010). In total, the obtained sequences were grouped into 10 putative *FaFra* genes that were successively named from *FaFra1* to *FaFra10* (Table 1). The corresponding *F. vesca* homologs were identified after a blast search in the strawberry genome and phytozome databases (www.Strawberrygenome.org; www.phytozome.net). Interestingly, gene annotations in these databases identified 21 putative *FvFras*; however our results indicate that a lower number of transcripts are expressed in strawberry. This finding, together with the fact that the *F. vesca* genome is still unfinished, raises the question whether there are in reality 21 *Fra* genes in *F. vesca*. In this regard, *Bet v 1* and the homologous gene from apple *Mal d 1* are members of gene families consisting of at least 20 and 15 members, respectively (Swoboda *et al.*, 1995; Atkinson *et al.*, 1996)

It is noteworthy to mention that the *FaFra2a* and *FaFra3*, described in previous studies (Musidlowska-Persson *et al.*, 2007; Muñoz *et al.*, 2010) were not identified in our analysis of the transcriptome, most probable because this family of genes presents high homology and reconstruction of full-length transcripts from short reads poses substantial computational

challenges, since sequences that are repeated in different genes introduce ambiguity in the transcriptome assembly. This ambiguity can be amplified by the fact that many *Fra* genes annotated in the *F. vesca* genome are found in tandem. But also, assembly of the strawberry transcriptome is further challenged considering that *Fragaria xananassa* is an octoploid species and, as such, a large allelic variation is expected. Our analysis allowed us to identify allelic variations for *FaFra1*, *FaFra2, FaFra4*, *FaFra7* and *FaFra9* that were named as *FaFra1c1-c2*, *FaFra2b*, *FaFra4a1-a2* and *FaFra4b1-b2*, *FaFra7a-b* and *FaFra9a1-a2* (see Figures 3-7); pairs were found at the same locus of the assembled transcriptome and more than 90% sequence homology between them was always observed.

Expression profiling of the *FaFra* genes was also explored. In general, the highest expression levels were observed for *FaFra1c1*, *FaFra1c2*, *FaFra2b* and *FaFra4a1* (see Table 2, Figure 10, Figure 11). Root was by far the tissue that accumulated the highest abundance of *FaFra* transcripts; especially *FaFra1* exhibited extremely high expression values (Figure 10b). Receptacles followed roots; in this organ *FaFra1c1* and *FaFra2b* showed complementary expression profiles since the transcripts levels of *FaFra1c1* decreased from the green to the red stage whereas the levels of *FaFra2b* were increased (Figure 10d). In achenes, the expression profiles of *FaFra1c1* and *FaFra1c2* were complementary along the development of this organ, where maximum transcript levels for one allele coincided with minimums for the other (Figure 10c, Figure 11). As a whole, the differential expression profiles observed here, most likely represent the possibility that FaFra proteins may bind different metabolites, since the different flavonoid compounds accumulate in a developmental manner (Halbwirth *et al.*, 2006; Fait *et al.*, 2008), similar to the variation observed in the *FaFra* transcripts. In the case of Bet v 1, it has been reported a specific binding with a natural compound (Seutter von Loetzen *et al.*, 2014). Our results favour this possibility for FaFra proteins, and we reason that this FaFra-flavonoid binding might be critical for the role played by this family of proteins in strawberry.

The identified *FaFra* sequences code for proteins of 159-163 amino acids (see Tabla 1, Figure 8), which is in agreement with reported FaFra and PR-10 proteins (Musidlowska-Persson *et al.*, 2007; Muñoz *et al.*, 2010, Fernandes *et al.*, 2013). The amino acid sequences of PR-10 proteins diverge considerably, but also share conserved features, such as their similar 3D structure (Liu & Ekramoddoullah, 2006). Four amino acids remain invariable across the PR-10 family (Liu & Ekramoddoullah, 2006) and these have been also found within the FaFra protein sequences (Gly-47, Gly-52, Asp-77 and Leu-152 in FaFra1E) (Figure 8). Twenty-five additional amino acids are conserved among all the FaFra proteins. Of these, Tyr-84 has been described to be as an essential amino acid in the binding of flavonoid compounds (Casañal *et al.*, 2013b; This

Thesis, Chapter 3). However, some variability was observed in the FaFra proteins at certain key positions involved in the binding of metabolites (Casañal *et al.*, 2013b; This Thesis, Chapter 3), which would support different ligand selectivity for the different members of the family. For instance, His-70, which is important in the stabilization of metabolites in the hydrophobic cavities of FaFra3 (Casañal *et al.*, 2013b; This Thesis, Chapter 3), is conserved among the whole family, except for FaFra8 where Asp has replaced His (Figure 8). Furthermore, Phe-59, which is present in FaFra1, FaFra2, FaFra6 and FaFra8-10, is switched to Leu in FaFra3-FaFra5 (group 3 in Figure 8) and FaFra7; and more interestingly, Lys-140, which is conserved in FaFra1, FaFra2A, FaFra4B and FaFra6 is replaced by Arg in FaFra2B, FaFra3, FaFra4A, FaFra5 and FaFra7-9; this additionally supports that different proteins of the family, such as FaFra2A-2B and FaFra4A-4B, might present different affinities by different ligands. Overall, the most divergent sequences (Figure 9), FaFra8, FaFra9 and FaFra10 show the highest variability in residues building the cavity, such as Val-39, Gly-60, Ser-63 or Leu-143, which are replaced by other amino acids in comparison to the rest of the sequences (Figure 8).

Gene silencing is a good tool of reverse genetics and it has proven to be very useful to reveal the role played by certain genes. Thus, we investigated the genes differentially expressed in *Fra-RNAi* transiently silenced *Fragaria xananassa* fruits (cv. Camarosa), by RNA-Seq analysis. Our study confirmed that the down-regulation of *FaFra2A* and *FaFra6* transcript levels results in reduced pigment formation in strawberry fruits (see Figure 14a), so that the encoded proteins might be involved in processes leading to the formation and/or accumulation of anthocyanins. Although the target genes to be silenced were *FaFra* genes, and we expected co-down-regulation of most of them, as it had been described before for *Fragaria xananassa* cv. Elsanta fruits (Muñoz *et al.*, 2010), we observed that whereas some genes were down-regulated (*FaFra2A*, *FaFra6*), others were up-regulated (*FaFra1*, *FaFra2B*, *FaFra3*, *FaFra4*, *FaFra5*, *FaFra7*, *FaFra8, FaFra9, FaFra10*). This observation suggests that the observed phenotype (lack of colour) in the Camarosa silenced fruits might be mainly assigned to the down-regulation of the *FaFra2A* and *FaFra6* genes, and most probably to *FaFra2A*, whose expression in the red stage is the highest among the members of the family (see Table 1 and Table 2). The increased mRNA levels detected for other *FaFra* transcripts do not have a straightforward explanation. It might be the result of a complex regulatory network controlling the *FaFra* genes at a transcriptional level, largely unknown that supports a functional redundancy for the different members of the family. Thus, a self-regulatory mechanism might be involved in which, the induced silencing of *FaFra2A* and *FaFra6* would in turn cause an increase in gene transcription. It is also important to remark that the silencing performed here is through a transient assay and

that its extension, in regards to the number of cells within the receptacle that might be certainly transformed, is in fact undetermined.

RNAi-mediated silencing of *FaFra* triggered the expected down-regulation of the phenylalanine ammonia-lyase (*PAL*) transcript levels and it also caused the reduction of the chalcone synthase (*CHS*) gene, as it was previously reported in fruits of other variety (cv. Elsanta) (Muñoz *et al.*, 2010). The first evidence of the relationship between the FaFra proteins and the flavonoid biosynthetic pathway was discovered by Emanuelsson and co-workers (Hjerno *et al.*, 2006) whose research was motivated by the observation that white strawberries are well tolerated by individuals presenting adverse allergenic reactions to normal fruits; this finding lead to the proteome comparison of white and red varieties of strawberry. They found that the amount of FaFra proteins, reported as allergenic, was diminished in the white varieties (Hjerno *et al.* 2006). In a search for strawberry genotypes with low Fra levels, they found that the total content of FaFra proteins was always lower in colourless (white) strawberry varieties when compared to red ones (Hjerno *et al.* 2006). Therefore, the ripe colourless fruits that were tolerated by individuals affected by allergy were found to be virtually free from the strawberry allergen.

Among the proteins that exhibited reduced accumulation levels in white varieties were chalcone synthase (CHS), flavanone 3- hydroxylase (F3H), and dihydroflavonol reductase (DFR) (Hjerno *et al.*, 2006). Interestingly, our results show that the reduced *FaFra2A* and *FaFra6* gene expression achieved in the transiently transformed Camarosa fruits not only affects the expression of *PAL* and *CHS* genes, but also some additional flavonoid pathway genes with high expression in ripe receptacle; these include not only *F3H* and *DFR*, but also others such as cinnamate-4-hydroxylase (*C4H*), 4-coumarate CoA ligase (*4CL*), chalcone isomerase (*CHI*) and 3-O-glucosyltransferases (*3-GT*). More striking, the concerted transcriptional down-regulation of all these genes was accompanied to the down-regulation of two regulatory genes *FaMYB1* and *FaMYB10*. Regulation of flavonoid and anthocyanin formation by basic helix-loop-helix bHLH and MYB-R2R3-type transcription factors has been extensively reported in the literature (Lepiniec *et al.*, 2006; Vogt *et al.*, 2010; Schaart *et al.*, 2013). Specifically in strawberry, FaMYB1 and FaMYB10 have been identified as regulators of the flavonoid metabolism in fruits, as it has been shown that both, *FaMYB1* and *FaMYB10*, are primarily expressed in the red ripe strawberry fruit when anthocyanin levels reach their maximum (Aharoni *et al.*, 2001; Medina-Puche *et al.*, 2014). However, while FaMYB10 is a positive regulator, FaMYB1 has been reported as a repressor (Aharoni *et al.*, 2001), and both have been found to be down-regulated in *FaFra* silenced fruits. Whether this unexpected result, in which both activator (*FaMYB10*)

and repressor (*FaMYB1*) are co-down-regulated in *FaFra* silenced fruits, is connected to the joint regulation of all *FaFra* genes above suggested, deserves to be explored.

Since PR-10 proteins are able to bind different ligands (Mogensen *et al.*, 2002; Markovic-Housley *et al.*, 2003; Fernandes *et al.*, 2013; Casañal *et al.*, 2013b; Seutter von Loetzen *et al.*, 2014) and this interaction seems to be critical for their functional role, it has been hypothesized that FaFra proteins may also act as flavonoid binders within the cell (Muñoz *et al.*, 2010), being this interaction relevant for their role in the flavonoid biosynthetic pathway. We envisage two possibilities for the potential FaFra-flavonoid complex as an active player within this pathway. It is known that the structural enzymes, including PAL, CHS, and C4H co-localize to the endoplasmic reticulum in different plant species (Winkel, 2004). These enzymes are organized as metabolons at the cytosolic side of the membrane where synthesis of many flavonoids takes place (Zhao *et al.*, 2010). Therefore, one possibility is that FaFra proteins might interact with this multiprotein complex channelling intermediates to the pathway. However, the extensive transcriptional changes found in *FaFra2A* silenced fruits, especially the effect in regulatory genes, points to a wider regulatory function. An example for the mechanism that might take place, would be the well known PYR/PYL/RCAR abscisic acid (ABA) receptor, which belong to the START superfamily (Melcher *et al.*, 2009; Miyazono *et al.*, 2009; Nishimura *et al.*, 2009; Santiago *et al.*, 2009). The formation of the ABA-receptor complex triggers a down-stream signalling pathway that includes a first interacting protein with this complex, whose final effects are dramatic expression changes. This second possibility cannot be pulled out for FaFra proteins, and therefore the search of a protein interactor partner for the FaFra-flavonoid complex need to be considered. In the literature there are some reports on potential FaFra interacting proteins (Puehringer *et al.*, 2003) that might be good potential candidates.

To shed light on the function of FaFra proteins, subcellular localization of FaFra1E, FaFra2 and FaFra3 was undertaken. Our results indicate that the FaFra proteins localized to the nucleus and the cytoplasm. However, the same localization was observed for free GFP, preventing an unequivocal determination of the FaFra localization. Yet, the fact that no specific localization signals have been identified in the FaFra proteins, together with the observation that PR-10 proteins are generally cytosolic (Fernandes *et al.*, 2013), suggest that the FaFra might be cytosolic as well. Nevertheless, localization to the ER cannot be excluded, since it might be mediated by the presence of secondary proteins and interactor partners.

Together, this work provides valuable insight for understanding the Fra family of proteins as key regulators of the anthocyanin pathway and their biological role in the secondary metabolism

in plants. For this reason, we have focused our efforts on the determination of the natural ligands of the FaFra proteins and their structural characterization to help gain insight into their specific physiological function and mechanisms of action at the molecular level.

Chapter 2

Purification, Crystallization and Preliminary X-ray Analysis of the Strawberry Allergens FaFra1E and FaFra3 in the Presence of Catechin

The work presented in this chapter has been published as part of the following article:

Casañal, A., Zander, U., Dupeux, F., Valpuesta, V., & Márquez, J. A. 2013. Purification, crystallization and preliminary X-ray analysis of the strawberry allergens Fra a 1E and Fra a 3 in the presence of catechin. *Crystallization communications. Acta Cryst (2013) F69, 510-514*, 1–5.

INTRODUCTION

Food is the most common origin of allergenic responses and food allergies are the subject of intense research as they affect up to 6% of young children and 3%-4% of adults worldwide (Herman, 2003; Wang & Sampson, 2011). Food allergy to vegetables, fruits and berries is often caused by proteins homologous to Bet v 1, the major allergen in birch tree pollen (Hjerno *et al.*, 2006; Musidlowska-Persson *et al.*, 2007). The strawberry FaFra proteins, which are highly expressed during the late steps of fruit development, (Muñoz et al., 2010) show a high degree of sequence similarity to Bet v 1 and have been implicated in allergic reactions to strawberry. Indeed, strawberries varieties showing decreased expression of FaFra proteins were well tolerated by people allergic to normal fruits (Karlsson *et al.*, 2004; Hjerno *et al*, 2006).

Both Bet v 1 and Fra proteins belong to the ubiquitous family of plant pathogenesis-related proteins (PR-10), which are implicated in the response of plants against pathogenic infections and abiotic stress (Marković-Housley *et al.*, 2003). However, although the allergenic properties of PR-10 proteins have been widely studied, their physiological function is still poorly understood (Mogensen *et al.*, 2007). PR-10 proteins share a common fold with the START and PYR/PYL/RCAR proteins, which is characterized by the presence of a central hydrophobic cavity (Iyer *et al.*, 2001; Mogensen *et al.*, 2002; Radauer *et al.*, 2008; Santiago *et al.*, 2012). Proteins with this fold are widespread in eukaryotes and participate in a variety of processes such as non-vesicular lipid transport and steroid hormone synthesis in mammals (Soccio & Breslow, 2003), and hormone signalling in plants (Ma *et al.*, 2009; Melcher *et al.*, 2009; Miyazono *et al.*, 2009; Nishimura *et al.*, 2009; Park *et al.*, 2009, Santiago *et al.*, 2009). Elucidation of the structures of these proteins has contributed to gain insight into their mechanisms of action at the molecular level, which in all cases involves the binding of specific ligands into their hydrophobic cavity.

Several isoforms of the FaFra protein have been described in strawberry including FaFra1E and FaFra3 (Musidlowska-Persson *et al.*, 2007; Muñoz *et al.*, 2010). In addition to their properties as food allergens, it has been shown that FaFra proteins play an important role in the control of flavonoid biosynthesis and are thus required for the development of colour during fruit ripening (Hjerno *et al.*, 2006; Muñoz *et al.*, 2010). Flavonoids are among the most important bioactive secondary metabolites in plants. They are responsible for the colour and flavour in flowers, fruits and other plant organs (Halbwirth *et al.*, 2006; Fait *et al.*, 2008), and they also present antioxidative and anticarcinogenic activities for humans when consumed in the diet (Ghasemzadeh and Ghasemzadeh, 2011). Suppression of the expression of *FaFra* genes in

strawberry fruits leads to decreased expression of the *Phenylalanine Ammonia-Lyase* (*PAL*) and *Chalcone Synthase* (*CHS*) genes, that code for two major enzymes in the flavonoid biosynthetic pathway, and to a decrease in the accumulation of the main flavonoids responsible for red colour of fruits, as well as other intermediates of the pathway (Muñoz *et al.*, 2010). The molecular mechanism through FaFra proteins control flavonoid biosynthesis is yet unknown. However, the fact that these proteins are predicted to have cavities for the binding of small ligands, suggests that FaFra proteins might bind metabolic intermediates of the flavonoid pathway. Here, we report the cloning, expression, purification, crystallization and preliminary X-ray analysis of the FaFra1E and FaFra3 allergens of strawberry, the latter in the presence of catechin, a natural flavonoid compound. This work could contribute to the structural analysis of these proteins, which would shed light on the molecular function of FaFra proteins and potentially other members of the PR-10 proteins.

MATERIALS AND METHODS

Cloning

The coding regions of *FaFra1e* and *FaFra3* (EMBL accession numbers CAJ85645 and GQ148819) were amplified by PCR using as a template the plasmids *pEntry-Fraa1ei* and *pEntry-Fraa3i* (Muñoz *et al.*, 2010). *FaFra1e* and *FaFra3* open reading frames were PCR-amplified using the primers FwF1- GGGCCATGGCGGGTGTTTACACTTATGAAAACGAG/ RvF1-CCCGGATCCTTAGTTGTATTCGCTGGGG and FwF3- GGGCCATGGCGGGTGTGTTCACATACGAATCCG/ RvF3- CCCGGATCCTTAGTTGTATTCCTCAGGATGGG, respectively. The forward and reverse primers contained NcoI and BamHI restriction sites. The amplified sequences were digested with NcoI/BamHI enzymes and cloned into pETM11 (Dummler *et al.*, 2005). The expression constructs, named F1-pETM11 and F3-pETM11 included an N-terminal His6-tag and the TEV cleavage sequence. After TEV cleavage, only three foreign amino acids (Ala-Met-Ala) remained at the N-terminal end of both proteins. DNA sequencing confirmed that the recombinant vectors encoded the expected sequences.

Protein Expression and Purification

F1-pETM11 and F3-pETM11 constructs were introduced into *E.coli* One Shot BL21-DE3 competent cells (Invitrogen) by the heat shock method and grown overnight at 37°C on solid Luria-Bertani (LB) medium supplemented with 50 μg/ml kanamycin. Cells were inoculated in 2

l of LB medium containing 50 µg/ml kanamycin and grown at 37°C with shaking at 150 rpm until an OD_{600} of 0.6-0.8 was reached. Protein expression was then induced by addition of 1 mM IPTG. The cells were incubated overnight at 20°C, harvested by centrifugation at 10000g for 15 min at 4°C and stored at -80°C before purification.

Cell pellets were resuspended in 180ml lysis buffer (30 mM Tris pH 7.5, 500 mM NaCl, 15 mM Imidazole, 1 mM β-mercaptoethanol) containing 20 mg/ml DNAse I (Roche) and one EDTA-free protease cocktail inhibitor tablet (Roche) and lysed with a microfluidizer (Microfluidics). The lysate was centrifuged at 35000g and 4°C for 45 min. The clear supernatant was incubated for 2h in 25 ml of nickel-nitrilotriacetic acid (Ni-NTA) agarose column (Qiagen) equilibrated with lysis buffer. Unbound proteins were removed by washing with five column volumes of buffer A1 (30 mM Tris pH 7.5, 300 mM NaCl, 15 mM Imidazole, 1 mM β-mercaptoethanol) and buffer W (30 mM Tris pH 7.5, 300 mM NaCl, 30 mM Imidazole, 1 mM β-mercaptoethanol). The bound proteins were finally eluted with buffer B (30 mM Tris pH 7.5, 300 mM NaCl, 250 mM Imidazole, 1 mM β-mercaptoethanol). The 6xHis tag of the purified proteins was removed by digestion with the TEV protease. During digestion, samples were extensively dialyzed against buffer A2 (30 mM Tris-HCl pH 7.5, 300 mM NaCl and 1 mM β-mercaptoethanol). The dialyzed samples were kept at 4°C until TEV cleavage was complete (typically overnight). The samples were incubated with Ni-NTA to remove the undigested proteins, TEV protease and other contaminants. The correct size and purity of the recombinant proteins were verified by SDS-PAGE (Figure 1). FaFra1E and FaFra3 purified fractions were pooled, dialyzed in buffer C (30 mM Tris pH 7.5, 150 mM NaCl, 1 mM β-mercaptoethanol) to remove imidazole, concentrated to 60 mg/ml by ultrafiltration with Amicon Ultra-15-3K filter units (Millipore) and flash-frozen in liquid nitrogen for storage at -80°C.

Protein Characterization: Size-Exclusion Chromatography- Multiple Angle Laser Light Scattering (SEC-MALLS)

Size exclusion chromatography (SEC) combined with multi-angle laser light scattering (MALLS) and refractometry (RI) is a powerful method for measuring the absolute molecular mass of macromolecules and macromolecular complexes (Wyatt *et al.*, 1998, Gerard, *et al.*,2007). The determination of the molecular mass variation across the chromatographic peak also provides an estimate of the dispersity of the compound. SEC was performed with an S200 Superdex column (GE Healthcare) equilibrated with buffer containing 20mM Tris-HCl pH 7.5, 150 mM NaCl, and 1mM β-mercaptoethanol. Proteins were injected at a concentration of 200 µM. All separations were performed at 20°C with a flow rate of 0.5 ml/min. On-line MALLS

detection was performed with a DAWN-EOS detector (Wyatt Technology Corp., Santa Barbara, CA) using a laser emitting at 690 nm and by refractive index measurement using an RI2000 detector (Schambeck SFD). Weight- averaged molar masses (Mw) were calculated using the ASTRA software (Wyatt Technology Corp., Santa Barbara, CA) as described previously (Wyatt *et al.*, 1998). (Figure 2).

Crystallization

Initial crystallization conditions for FaFra1E and FaFra3-catechin were identified at the High Throughput Crystallization Laboratory of the EMBL Grenoble Outstation (https://htxlab.embl.fr) (Dimasi *et al.*, 2007).

Crystallization experiments were carried out at 293 K in a 96-well plate using the sitting-drop vapour diffusion method and the Hampton Research screens Crystal Screen I & II, Crystal Screen Lite & PEG/Ion, MembFac & Natrix, QuickScreen & Grid screens Ammonium Sulfate, Sodium Malonate-Sodium Formate, Grid screens PEG 6K, PEG LiCl, MPD & Screen MME and The Classic II. Droplets of 200 nl volume (with a 1:1 protein:precipitant ratio) were set with a 16 channels Cartesian PixSys robot (Cartesian Technologies) and equilibrated against 80 µl reservoir solution (Dimasi *et al.*, 2007). For these experiments FaFra1E and FaFra3 proteins were diluted in buffer C (30 mM Tris pH 7.5, 150 mM NaCl, 1 mM β-mercaptoethanol) to a concentration of 26 mg/ml. Both proteins were assayed in the presence and absence of the natural flavonoid compound (+)-catechin. Interestingly, FaFra1E produced crystals only in the absence of (+)-catechin and FaFra3 exclusively in its presence. In both cases, crystals of FaFra1E and FaFra3-catechin appeared within 48 h after setting up the crystallization experiments (Figure 3). The FaFra1E protein produced crystals in two different crystal forms. Rod shaped FaFra1E crystals (Figure 3a) were obtained using 0.2 M ammonium sulfate, 0.1 M sodium cacodylate pH 6.5 and 15% PEG 8000; FaFra1E crystals in the shape of hexagonal prisms (Figure 3b) were obtained in 0.2 M ammonium sulphate, 0.1 M HEPES pH 7.5 and 25% PEG 3350 as precipitant. The FaFra3 protein produced flat hexagon-like crystals (Figure 3c) that were obtained by adding 5mM (+)-catechin, diluted in buffer C supplemented with 10% DMSO, to 26 mg/ml (1.5 mM) Fra3 solution in 2.9 M sodium malonate pH 7.0. Further optimization of this condition was needed to obtain diffraction quality crystals. The final FaFra3-catechin crystallization protocol was as follows: Purified FaFra3 protein was diluted to a concentration of 26 mg/ml. To compensate for the limited solubility of (+)-catechin in aqueous solutions, solid (+)-catechin was added in excess to the protein solution as a powder. The sample was incubated at 4°C overnight in an overhead shaker. After centrifugation at 14000*g*

the supernatant was used to set up hanging-drop crystallization experiments at room temperature by mixing equal amounts of protein- and precipitant solution (2.25M sodium malonate, pH 7.0). Crystals acquired their final size within 2 days.

Data Collection

For X-ray data collection, FaFra1E crystals were mounted on CryoLoops (Hampton Research), soaked in cryoprotectant solution (15% glycerol) and flash cooled directly in a liquid nitrogen stream. The crystallization condition for FaFra3-catechin crystals was directly compatible with cryofreezing, so no addition of cryoprotecting agents was required and the crystals were directly flash-frozen in the liquid nitrogen stream prior to data collection.

FaFra1E and FaFra3-catechin diffraction experiments were performed, respectively, at the ID14-4 and the ID14-1 beam lines of the European Synchrotron Radiation Facility (ESRF) equipped with ADSC Q315r CCD and ADSC Q210 CCD detectors (McCarthy *et al.*, 2009). XDS was used for data reduction and integration (Kabsch, 2010). After conversion to *CCP4* format using Combat (Winn *et al.*, 2011), the data were scaled using SCALA (Evans, 2006). Matthews coefficient and solvent-content estimations were also performed using the *CCP4* software (Matthews, 1968; Kantardjieff & Rupp, 2003; Winn *et al.*, 2011) .

RESULTS AND DISCUSSION

Cloning, expression, purification and identification of crystallization conditions for FaFra1E and FaFra3 have been successfully undertaken. FaFra1E and FaFra3 open reading frames were cloned into pETM11 expression vectors. The resultant plasmids encoded FaFra1E and FaFra3 proteins fused to an N-terminal His6-tag. Recombinant proteins showed very high expression levels in *E. coli* BL21(DE3). FaFra1E and FaFra3 were purified to homogeneity by a two-step Ni-NTA metal-affinity, as described in the Materials and Methods section. SDS-PAGE analysis of the purified samples (Figure 1) indicated a migration consistent with the expected molecular weights and a high degree of purity. The typical yield for both proteins was 75-150 mg per litre of culture. SEC-MALLS experiments (Figure 2) indicated an estimated molecular weight of 18,0 kDa and 17.5 kDa for FaFra1E and FaFra3 respectively, which is in good agreement with their expected molecular masses, indicating that both proteins are monomeric in solution. FaFra1E diffraction quality crystals in the shape of rods (Figure 3a) were obtained by the sitting-drop vapour diffusion technique. The crystals diffracted reproducibly to 2.2Å resolution (Figure 4a). The hexagonal prims-like FaFra1E crystals (Figure 3b) diffracted to 6Å resolution. However,

efforts to obtain better diffracting crystals for this second condition were not successful. FaFra3 crystals were obtained after optimization by the hanging-drop method and only in the presence of catechin, a natural flavonoid compound (Figure 3c). The crystals diffracted to 3.0Å resolution (Figure 4b).

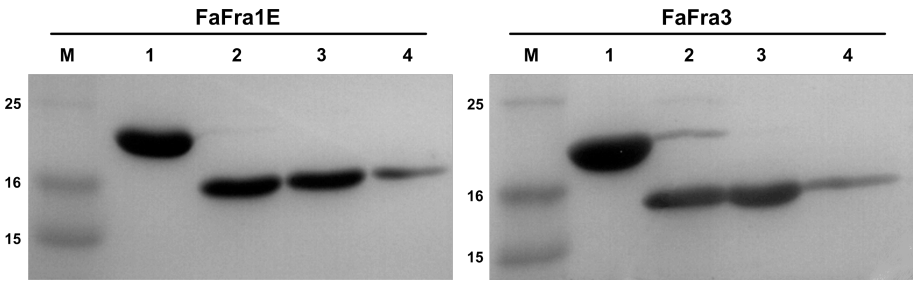


Figure 1. 12% SDS-PAGE analysis of the purified FaFra1E (left) and FaFra3 (right). Lane M, molecular-mass marker (labelled in kDa); lane 1, purified FaFra1E and FaFra3; lane 2, the same samples after TEV cleavage; lanes 3-4,the same samples as in 2 after reverse purification with Ni-NTA. As it can be observed, the samples used for crystallization (lanes 3 & 4) were highly pure and had a molecular weight close to the expected value (18 kDa).

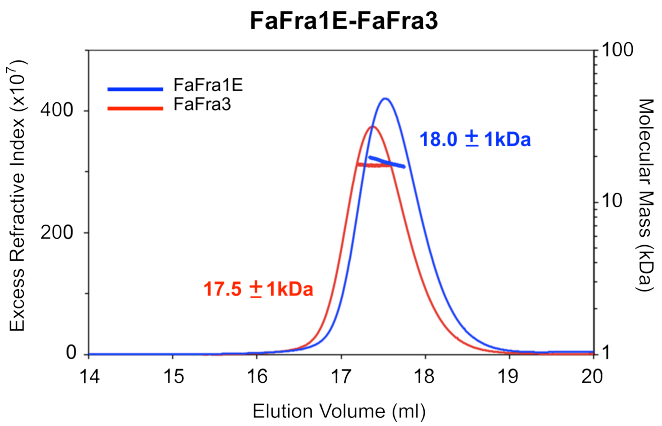


Figure 2. SEC-MALLS analysis of purified FaFra1E (blue) and FaFra3 (red). The experiments were performed at protein concentrations of 200 mM, as described in Materials and Methods. Both the SEC elution profiles (monitored through the excess refractive index, which is proportional to the protein concentration) and the molecular size calculated by MALLS (blue and red crosses shown on the peaks for each species) indicate that FaFra1E and FaFra3 proteins are monomeric in solution (expected masses of the monomeric forms are 17.8 kDa and 17.5 kDa respectively).

A complete dataset was collected for both FaFra1E (rods) and FaFra3-catechin crystals, with good completeness and crystallographic statistics (see Table 1). FaFra1E and FaFra3-catechin crystals belonged to space groups $P2_12_12_1$ and $C222_1$, with unit cell parameters a=70.02, b=74.42, c=84.04 and a=137.91, b=206.61, c=174.7, respectively. Matthews coefficients were calculated as 3.08 $Å^3Da^{-1}$ and 7.11 $Å^3Da^{-1}$. These values would be consistent with two and five molecules per asymmetric unit and an estimated solvent content of 60.04% and 82.71% for FaFra1E and FaFra3-catechin, respectively (Matthews, 1968). After data processing with XDS (Kabsch, 2010) and scaling with SCALA (Evans, 2006), the resulting datasets extended to a resolution of 2.2Å and 3.0Å, with *R*p.i.m. of 0.034 and 0.03 and a completeness of 99.9% and 99.5% for FaFra1E and FaFra3-catechin respectively. Data collection statistics are summarized in Table 1.

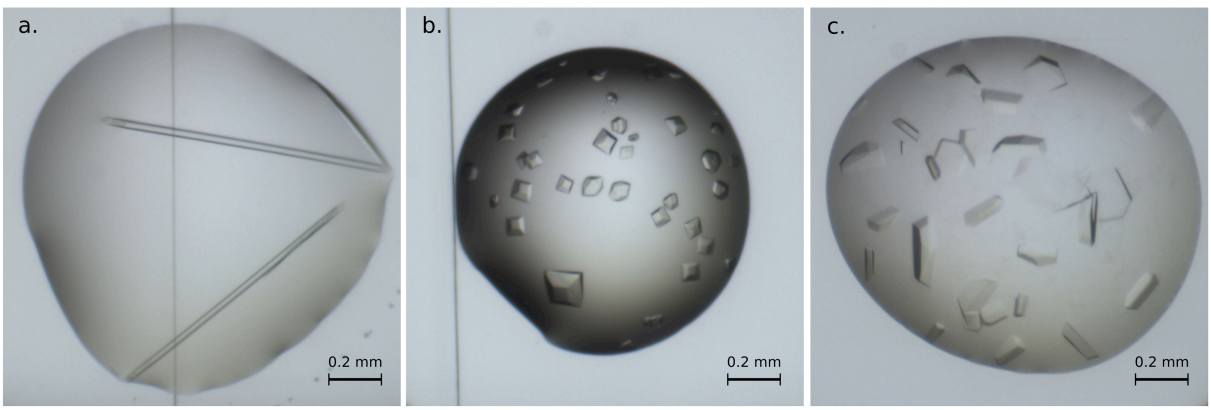


Figure 3. FaFra1E and FaFra3-catechin crystals. Two different crystallization conditions were identified for FaFra1E; crystals were obtained in the shapes of rods (a) and trigonal prisms (b). Crystals with the shape of hexagonal prisms were obtained for the FaFra3 only in the presence of catechin (c).

In conclusion, the results presented here provide methods for the large-scale production and purification of the FaFra proteins, which seem to be monomeric in solution. Conditions for producing crystals for X-ray diffraction experiments have been established for both proteins. The crystallographic datasets obtained so far are currently being analysed in order to obtain atomic structures of FaFra1E and FaFra3-catechin. Interestingly the fact that crystallization of both proteins is very sensitive to the presence of the natural flavonoid catechin suggests that ligand binding might induce conformational changes in the protein, which might be related to their function. The determination of FaFra1E and FaFra3-catechin structures would be an important step towards the understanding of their mechanisms of action, not only on the

control of the secondary metabolism in plants, but also to determine the origin for their immunogenic properties.

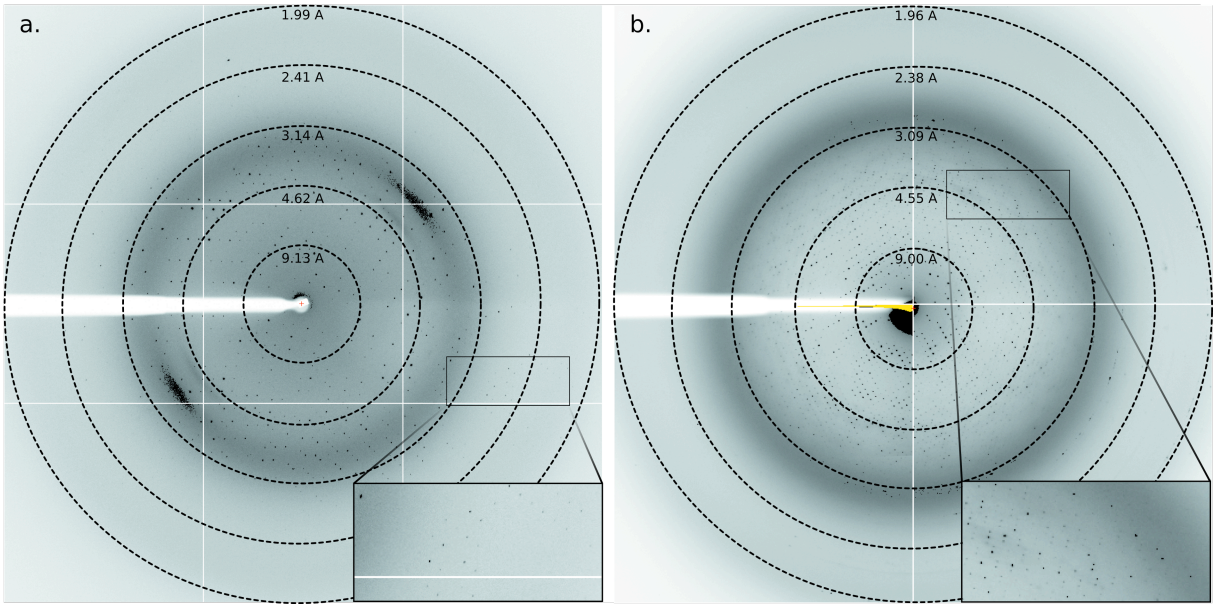


Figure 4. Diffraction patterns of FaFra1E and FaFra3-catechin crystals. The concentric circles show the resolution limits. FaFra1E (a) and FaFra3-catechin (b) crystals diffracted to a resolution of 2.2 Å y 3.0 Å, respectively. The high-resolution spots are highlighted in boxes.

Table 1. Crystallographic data-collection statistics for FaFra1E and FaFra3-catechin. Values in parentheses are for the highest resolution shell.

Data Collection	**FaFra1E**	**FaFra3-Catechin**
X-ray source	ID14-4	ID14-1
Space group	$P2_12_12_1$	$C222_1$
Unit cell a, b, c	70.02 74.42 84.04	137.91 206.61 174.7
α, β, γ	90 90 90	90 90 90
Matthews coefficient ($Å^3$ Da^{-1})	3.08	7.11
Solvent content	60.04	82.71
Resolution	36.03 – 2.2 (2.32- 2.2)[1]	30- 3.0 (3.16- 3.0)[1]
No Observations (overall/ unique)	165127/ 24107	371478/ 49904
Average redundancy	7.2 (7.3)	7.4 (7.5)
*R*p.i.m.	0.034 (0.183)	0.03 (0.23)
Completeness (%)	99.9 (100)	99.5 (100)
I/σ(I)	14.0 (4.2)	21.7 (3.6)

Chapter 3

The Strawberry Pathogenesis-Related 10 (PR-10) FaFra Proteins Control Flavonoid Biosynthesis by Binding to Metabolic Intermediates.

The work presented in this chapter has been published as part of the following article:
Casañal, A., Zander, U., Muñoz, C., Dupeux, F., Luque, I., Botella, M. A., Schwab, W., Valpuesta, V., & Márquez, J. A. 2013. The Strawberry Pathogenesis-related 10 (PR-10) Fra a Proteins Control Flavonoid Biosynthesis by Binding Metabolic Intermediates. *Journal of Biological Chemistry*, *288*(49), 35322–35332.

INTRODUCTION

The family Pathogenesis Related 10 (PR-10) proteins comprises a large number of sequences widely distributed among seed plants (Liu & Ekramoddoullah, 2006). However, their function is still poorly understood. PR-10 proteins were initially characterized by their increased expression levels in response to infection by plant pathogens and under abiotic stress conditions. Today, a large number of PR-10 genes have been identified in different species showing a diversity of expression patterns under both normal growth and stress conditions (Liu & Ekramoddoullah, 2006; Radauer at al., 2008). Some PR-10 proteins, like the white birch Bet v 1 and the apple Mal d 1, are highly abundant in pollen and fruits, respectively and are responsible for allergic reactions, including seasonal and food allergies (Spangfort *et al.*, 1997; Spangfort *et al.*, 1999; Mirza *et al.*, 2000; Gajhede at al., 2008; Zaborsky *et al.*, 2010; Geroldinger-Simic *et al.*, 2011). PR-10 proteins belong to the START superfamily. These proteins, adopt a helix-grip fold with an internal cavity capable of binding hydrophobic ligands (Ponting & Aravind, 1999; Iyer *et al.*, 2001; Soccio & Breslow, 2003; Radauer *et al.*, 2008). Only two protein families within the START superfamily have been extensively characterized at a functional and structural level: The Lipid Transport Proteins (LTPs), which are involved in non vesicular transport of lipids in eukaryotic cells (Ponting & Aravind, 1999; Lev, 2010) and the plant PYR/PYL/RCAR proteins that function as intracellular receptors for the plant hormone abscisic acid (Ma *et al.*, 2009; Melcher *et al.*, 2009; Miyazono *et al.*, 2009; Nishimura *et al.*, 2009; Park *et al.*, 2009; Santiago *et al.*, 2009). Some PR-10 proteins including Bet v 1, the mung bean Cytokinin Specific Binding (CSBP) protein and the *Prunus* LlPR10 protein, have been found to bind to a series of artificial and natural hydrophobic molecules, including cytokinins and phytosteroids (Mogensen *et al.*, 2002; Marković-Housley *et al.*, 2003; Pasternak *et al.*, 2006; Fernandes *et al.*, 2008; Radauer at al., 2008; Fernandes *et al.*, 2009; Kofler *et al.*, 2012; Fernandes *et al.*, 2013). However, the functional relevance of these interactions remains unclear. Recently, three new members of the PR-10 family, FaFra1E, FaFra2A and FaFra3, have been identified and shown to play an important role in the control of phenylpropanoids and flavonoids biosynthesis in strawberry fruits (Karlsson *et al.*, 2004; Hjerno *et al.*, 2006; Muñoz *et al.*, 2010).

Flavonoids and phenolic compounds are among the most important secondary metabolites in plants. In addition to color and flavor development they participate in many aspects of plant biology, including UV protection, as antioxidants, auxin transport regulators, and defense compounds against pathogens (Halbwirth *et al.*, 2006; Fait *et al.*, 2008; Vogt, 2010; Fraser *et al.*, 2011; Hassan & Mathesius, 2012). Thus, injury by pathogens or pests induces the accumulation

of flavonoids and other phenolic compounds with antimicrobial activity (Ahuja *et al.*, 2012). Flavonoids are also exuded by plant roots and act as signals that modify the transcriptional activity of nodulation genes in nitrogen-fixing bacteria, thereby promoting symbiotic association (Kobayashi *et al.*, 2004; Treutter, 2005). Other flavonoids have been implicated in pollen germination, seed resistance to pests and numerous other processes (Treutter, 2005; Mahajan *et al.*, 2011). The effect of dietary flavonoids in human health is also a subject of study due to their antioxidative and anticarcinogenic activities (Ghasemzadeh & Ghasemzadeh, 2011). Flavonoids are synthesized via the phenylpropanoid and flavonoid pathways (Figure 1) (Halbwirth *et al.*, 2006; Vogt, 2010; Fraser *et al.*, 2011). The first step in the phenylpropanoid pathway is catalyzed by the enzyme Phenylalanine Ammonia-Lyase (PAL) and leads to the production of cinnamic acid from L-phenylalanine. PAL is the gateway enzyme to the synthesis of phenolic and flavonoid compounds as well as many other secondary metabolites (Vogt, 2010). In Arabidopsis and other species, *PAL* gene expression is responsive to developmental and environmental clues like wounding, pathogen infection or UV radiation among others (Blount *et al.*, 2000; Olsen *et al.*, 2008Huang *et al.*, 2010; Moura *et al.*, 2010; Naoumkina *et al.*, 2010). Another important step in the synthesis of flavonoids is the production of naringenin, which is the first product in the pathway with a flavan structure and from which many other flavonoids are derived (Figure 1). This step is catalyzed by the enzyme Chalcone Synthase (CHS). Many of the final products of the flavonoid biosynthesis pathway accumulate as O-glycosyl derivatives at the position 3 of the C ring of the flavan nucleus and are accumulated in the vacuole or secreted through the plasma membrane into the apoplastic space (Zhao & Dixon, 2010). A number of flavonoids that account for colour of the fruit and contribute significantly to its taste are produced in the strawberry fruit in a developmental-specific pattern (Fait *et al.*, 2008; Munoz *et al.*, 2011). Proanthocyanidins (condensed tannins) are mostly produced in the young fruits that make them bitter while anthocyanins, mostly pelargonidin-3-O-glucoside and cyanidin-3-O-glucoside (Figure 1), that confer color, are abundant in the later stages of fruit maturation.

The strawberry *FaFra* transcripts are present in most plant organs, however they show maximal expression levels in open flowers, fruits and roots, depending on the member of the family (Muñoz *et al.*, 2010). The first evidence of the involvement of FaFra proteins in the control of the synthesis of flavonoids was provided by Emanuelsson and co-workers, who reported that fruits of colourless cultivars showed very low levels of FaFra1 protein expression in contrast to red-colored fruits (Hjerno *et al.*, 2006).

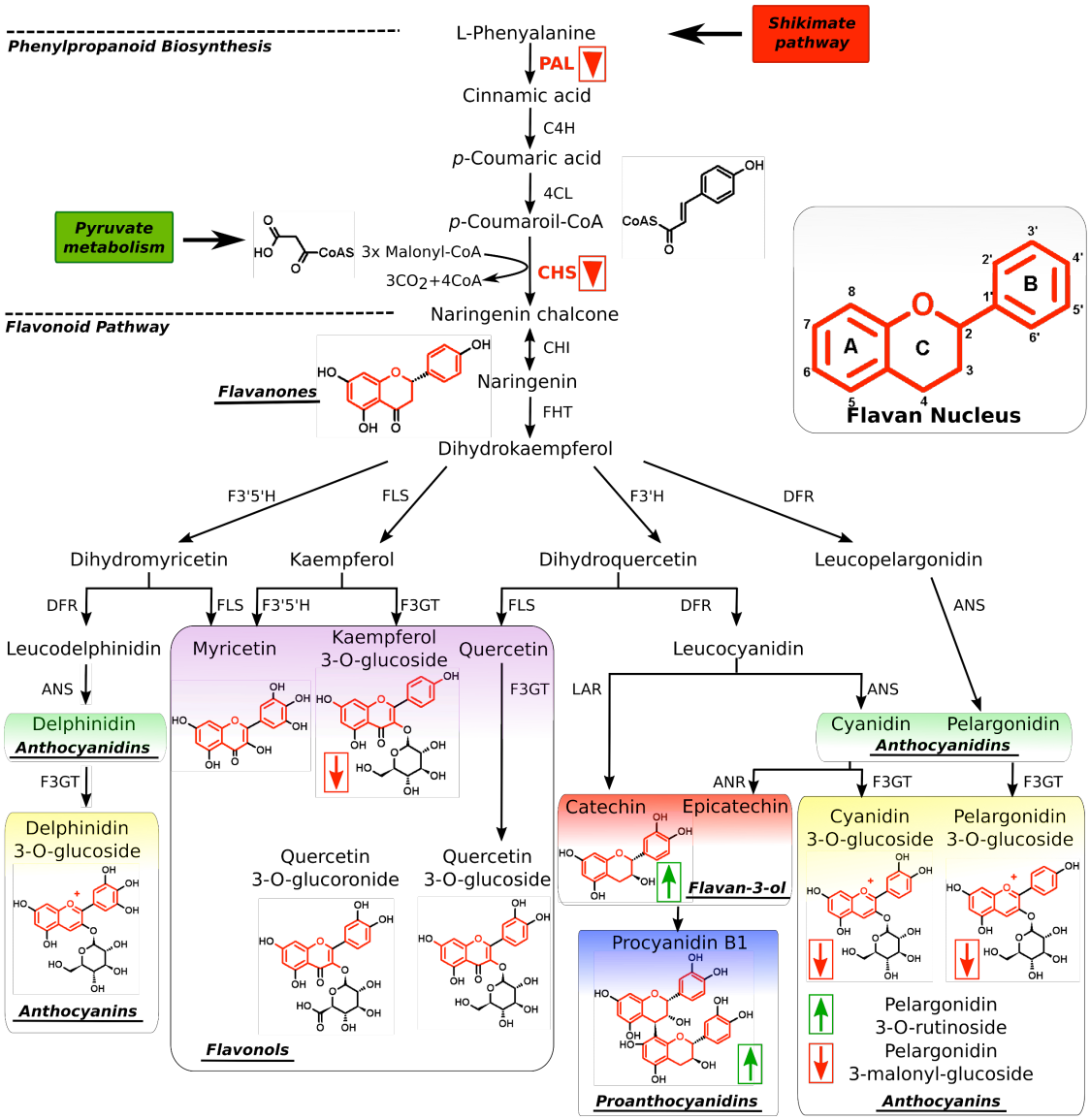


Figure 1. The Phenolic Compounds Biosynthesis Pathway. Schematic representation of the phenylpropanoid and flavonoid biosynthesis pathway. Major families of flavonoid compounds including are highlighted. Flavonoids are characterized by the presence of the flavan nucleus with A B and C rings, as indicated (inset). Final products of the flavonoid pathway, such as pelargonidin 3-O-glucoside, are often glycosylated at the position 3 of the C ring of the flavan nucleus. Suppression of FaFra protein expression affects the expression of phenylalanine ammonia lyase (*PAL*) and chalcone synthase (*CHS*) genes (red down-pointing triangles) and alters phenolic compound accumulation with an increase in the levels of catechin and a decreased accumulation of anthocyanins (as indicated by arrows) (Muñoz *et al.*, 2010).

Later, *FaFra-RNAi* silencing experiments showed that suppression of the expression of FaFra proteins in strawberry fruits led to decreased accumulation of the main flavonoids responsible for the red color of fruits, including cyanidin 3-O-glucoside and pelargonidin 3-O-glucoside (see Figure 1), while other aspects of fruit maturation were un-affected (Muñoz *et al.*, 2010). Other flavonoids like kaempferol 3-O-glucoside and pelargonidin 3-malonyl-glucoside also showed decreased levels in silenced fruits, while other intermediate metabolites of the flavonoid pathway such as catechin and proanthocyanidins accumulated at higher levels (Figure 1). Interestingly, silencing of FaFra proteins in strawberry also produced a decrease in the expression levels of genes encoding for *PAL* and *CHS* in the fruits, showing that FaFra proteins are required for the expression of structural genes in the flavonoid biosynthesis pathway and suggesting that their function may be regulatory (Muñoz *et al.*, 2010). These studies indicated that the FaFra proteins participate in the control of flavonoid biosynthesis in strawberry fruits. Recently, a solution structure of the FaFra1 protein has been described (Seutter von Loetzen *et al.*, 2012). However, the molecular basis for the function of the Fra proteins remains unknown.

In this work we show that strawberry FaFra proteins bind natural flavonoids, providing a basis for their function in the control of flavonoid metabolisms. Moreover, we present crystallographic structures of FaFra1E and FaFra3 in complex with catechin. The analysis of these structures shows that flavonoid binding is associated with conformational changes in critical loop regions providing for the first time a molecular basis for the function of Fra proteins in the control of flavonoid biosynthesis.

MATERIALS AND METHODS

Cloning Expression and Purification

Isolation of *FaFra1e* and *FaFra3* cDNAs ORFs from strawberry has been described previously (Muñoz *et al.*, 2010). *FaFra1e* and *FaFra3* ORFs were PCR-amplified and cloned into pETM11 (Dummler *et al.*, 2013) to produce expression constructs F1-pETM11 and F3-pETM11 respectively. These constructs include an N-terminal His6-tag followed by the TEV cleavage sequence. After TEV cleavage only three amino acids (Ala-Met-Ala) remain at the N-terminal end of the proteins. Purification of FaFra proteins was carried out as described by Casañal *et al.* (Casañal *et al.*, 2013a). Briefly, *E. coli* BL21 (DE3) cells were transformed with either F1-pETM11 or F3-pETM11 constructs and grown in 2 l of LB medium containing 50 µg/ml kanamycin to an OD at 600 nm of 0.6-0.8. At this point, 1 mM IPTG was added and the cells were harvested after overnight induction at 20°C and stored at -80°C before purification. The

cells were re-suspended in 180 ml lysis buffer (30 mM Tris pH 7.5, 500 mM NaCl, 15 mM imidazole, 1 mM β-mercaptoethanol and protease cocktail inhibitor) and lysed with a microfluidizer (Microfluidics). A cleared lysate was obtained after centrifugation at 20000 rpm for 45 min. The protein extract was incubated in 25 ml of nickel-nitrilotriacetic acid (Ni-NTA) agarose and washed with 125 ml of lysis buffer. The bound protein was eluted with a buffer containing 30 mM Tris pH 7.5, 300 mM NaCl, 250 mM imidazole and 1 mM β-mercaptoethanol. The 6xhis tag of the purified proteins was removed by digestion with the 6xhis-tagged version of the TEV protease. The cleaved samples were incubated with Ni-NTA to remove the undigested proteins the TEV protease and other contaminants. The correct size of the recombinant proteins was verified by SDS-PAGE. Purified FaFra1E and FaFra3 were extensively dialyzed against sample buffer (30 mM Tris-HCl pH 7.5, 150 mM NaCl and 1 mM β-mercaptoethanol), concentrated to 60 mg/ml and flash-frozen in liquid nitrogen for storage at -80°C.

Isothermal Titration Calorimetry (ITC)

ITC experiments were performed using an ITC200 Micro-calorimeter (MicroCal, Inc.). Prior to the experiments protein solutions were dialyzed against sample buffer (20 mM HEPES pH 7.5, 1 mM β-mercaptoethanol). Ligands were dissolved in the dialysis buffer and the pH was carefully adjusted to pH 7.5. All solutions were filtered and degassed. ITC measurements were performed with the three FaFra isoforms and the following flavonoid compounds: (+)-catechin, quercetin-3-O-glucuronide, myricetin and pelargonidin-3-O-glucoside. A number of other phenolic compounds were tested with individual FaFra isoforms according to their specific spatial and developmental expression profiles during fruit ripening (Kosar *et al.*, 2004; Aaby *et al.*, 2007; Muñoz *et al.*, 2010; Muñoz *et al.*, 2011). This included quercetin-sophoroside with FaFra1E, cyanidin-3-O-glucoside and naringenin with FaFra2A and procyanidin-B2 and quercetin-3-O-glucoside with FaFra1E and FaFra3 respectively. Early precursors of the phenylpropanoid pathway (L-phenylalanine, coumaric acid, cinnamic acid and caffeic acid) were tested with FaFra2A. Thermograms were recorded at 25°C and typically involved 26 injections of 1.5 µl, with 3 s intervals between injections with flavonoid compounds as injectants and proteins loaded in the calorimeter cell. Initial protein concentrations varied from 20 µM to 100 µM, while injectant concentrations were between 0.4 mM and 3 mM. Higher injectant concentrations were not possible due to limited solubility of the ligands in aqueous solutions. The heat associated with the binding process was calculated as the difference between the heat of reaction and the corresponding heat of dilution, as obtained from independent titrations of the ligand into buffer.

The resulting binding isotherms were analysed and fitted through non-linear least-squares to an appropriate thermodynamic model with the MicroCal Origin Software (MicroCal, Inc.).

Crystallization, X-Ray Data Collection and Structure Solution

FaFra1E, FaFra2A and FaFra3 proteins were assayed for crystallization by the vapor diffusion method the High Throughput Crystallization Laboratory of the EMBL Grenoble Outstation (https://embl.fr/htxlab) (Dimasi *et al.*, 2007) in the presence and absence of the flavonoid compounds indicated in the previous section. The FaFra1E protein produced crystals in two different crystal forms. Crystal form A was obtained at a protein concentration of 26 mg/ml and using 0.2 M ammonium sulfate, 0.1 M sodium cacodylate pH 6.5, 15% PEG 8000 as precipitant and at an incubation temperature of 20ºC. Crystal form B was obtained by diluting purified FaFra1E to a concentration of 26 mg/ml in sample buffer and adding solid (+)-catechin powder in excess. This solution was incubated at 4°C overnight in an overhead shaker and centrifuged at 14.000 g in a bench top centrifuge. The supernatant was used for crystallization. Crystals were obtained in 0.1M tri-sodium citrate dihydrate, 0.1M tris hydrochloride pH 8.5 and 5% PEG 400. The FaFra3 protein produced crystals only in the presence of (+)-catechin. The best diffracting crystals were obtained by adding (+)-catechin powder in excess to a 26 mg/ml FaFra3 solution, as described for FaFra1E crystal form B. FaFra3-catechin crystals were optimized by the vapor diffusion method by mixing 1 µl of protein and 1 µl of precipitant solution (2.25 M sodium malonate, pH 7.0) equilibrating against a reservoir containing 0.5 ml of precipitant solution. Crystals in the shape of flat hexagonal prisms appeared within two days. Crystals were flash frozen in liquid nitrogen using 15% glycerol as cryo-protectant. The FaFra2A protein did not produce any crystals, neither alone nor in the presence of any of the mentioned flavonoids. X-ray diffraction data were collected at the ID14-4, ID 23-2 and ID14-1 beamlines of the ESRF for FaFra1E crystal form A, FaFra3-catechin and FaFra1E crystal form B, respectively. Crystallographic data reduction and scaling was carried out with the software XDS (Kabsch, 2010). Initial phases were obtained using Phaser (McCoy *et al.*, 2007) by the molecular replacement method using the yellow lupine LIPR-10.2B protein (3E85) (Fernandes *et al.*, 2009) from the protein data bank (Berman *et al.*, 2000) as search model. Successive rounds of automatic refinement and manual building were carried out with RefMac5 (Murshudov *et al.*, 2011) and Coot (Emsley *et al.*, 2010). The three structural models have been deposited in the Protein Data Bank (Berman *et al.*, 2003) with codes 4C9C, 4C94 and 4C9I.

RESULTS

The FaFra Proteins Bind Natural Flavonoids with Different Specificities

Silencing of *FaFra* genes in strawberry fruits leads to alterations in the accumulation of specific metabolites of the flavonoid pathway, including increased levels of catechin and proanthocyanidins and decreased levels of some anthocyanins and flavonols (Muñoz *et al.*, 2010). These compounds are structurally related, as they all contain a flavan nucleus (Figure 1). We reasoned that FaFra proteins might exert their function through binding to a flavonoid compound containing a flavan nucleus. To test this hypothesis we performed binding studies with a number of natural compounds from the major branches of the flavonoid biosynthesis pathway (Figure 1, see Experimental Procedures section).

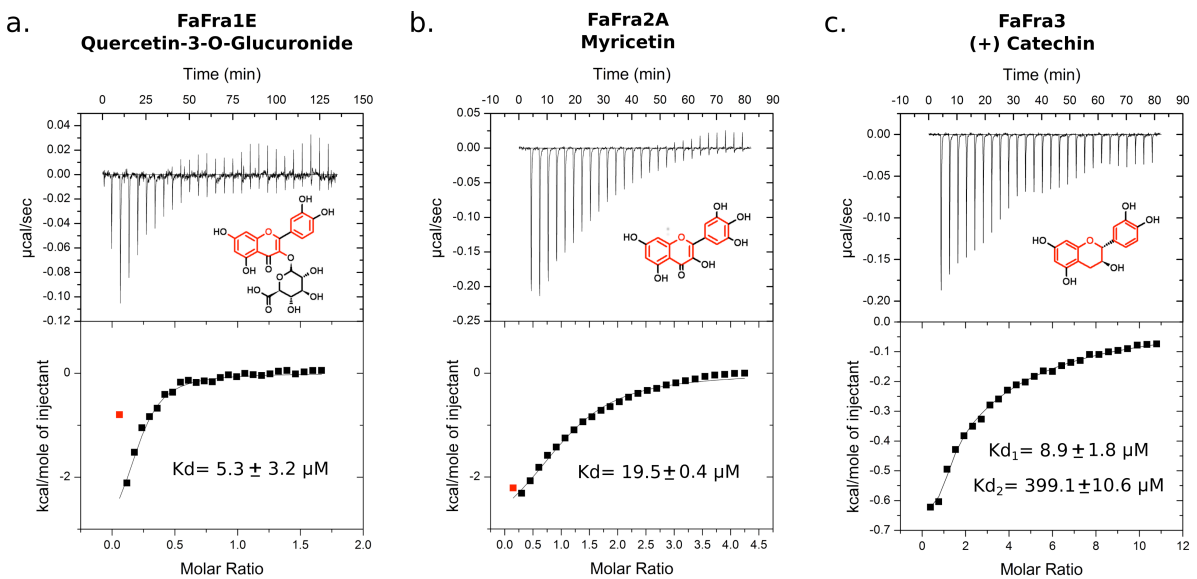


Figure 2. FaFra proteins bind natural flavonoids. Isothermal Titration Calorimetry (ITC) representative experiments demonstrating binding of different flavonoids to FaFra1E (a), FaFra2A (b) and FaFra3 (c) are shown. Concentrated solutions of quercetin-3-O-glucuronide, myricetin and (+)-catechin (as indicated) were injected into the ITC cell containing the respective protein solutions. The values of the dissociation constants (Kd) are indicated (averaged values over several independent experiments with standard deviations between brackets). The thermograms for the FaFra3 (+)-catechin pair (c) indicate the presence of two distinct ligand binding sites with high (Kd_1) and low high (Kd_2) affinities respectively.

ITC experiments allowed the identification of three specific interactions between FaFra proteins and different flavonoid compounds and with affinities in the low micromolar range (Figure 2). Quercetin-3-O-glucuronide binds to FaFra1E with a dissociation constant (Kd) of 5.3 μM, while FaFra2A binds myricetin with a Kd of 19.5 μM. In both cases the binding stoichiometry is 1:1. Similarly, ITC experiments demonstrated binding between (+)-catechin and FaFra3. In this case the thermograms showed a complex binding curve indicating the presence of two distinct binding sites, which were later confirmed by the crystallographic model (see below). For FaFra3, thermograms were fitted to a two-site binding model, which indicated a Kd of 8.9 μM and a second site with very low binding affinity (Kd in the higher micromolar range, Figure 2). ITC experiments failed to detect binding between FaFra2A, the most abundant isoform in the fruit and early precursors in the phenylpropanoid pathway, like L-phenylalanine, coumaric acid, cinnamic acid and caffeic acid. These results demonstrate that FaFra proteins can bind metabolites of the flavonoid pathway with affinities in the low μM range and with different selectivity.

The Structure of apo-FaFra1E Shows Conformational Flexibility in Loop Regions Surrounding the Ligand Binding Cavity

We attempted to crystallize FaFra1E, FaFra2A and FaFra3 proteins both in the presence and absence of flavonoids. For this purpose the three FaFra proteins were expressed in *E.coli* and purified to homogeneity (see experimental procedures). SEC-MALLS experiments showed an apparent Mr of 18 kDa, 28 kDa and 17.5 kDa for FaFra1E, FaFra2A and FaFra3 respectively. In the case of FaFra1E and FaFra3 these values are in good agreement with the expected molecular masses indicating that both proteins are monomeric in solution. The estimated molecular weight for FaFra2A is slightly higher than expected. FaFra2A did not produce crystals in any of the conditions tested.

The FaFra1E protein crystallized in two different crystal forms. The two crystal forms belong to the same space group (P $2_12_12_1$), but show different unit cell dimensions and molecular arrangement. Crystal form A contains two protein molecules per asymmetric unit while crystal from B contains 6 independent copies per asymmetric unit. Since FaFra1E is monomeric in solution, the inter-subunit contacts in both crystal forms are likely to be induced by the crystallization process. The first crystals diffracted to 2.2 Å resolution while the second crystal form diffracted to 3.1 Å (see Table 1 for crystallographic data collection and refinement statistics).

Initial phases for crystal form A of FaFra1E were obtained by the molecular replacement method using the yellow lupine LIPR-10.2B protein (3E85) (Fernandes *et al.*, 2009) as search model. The final model was refined to a resolution of 2.2 Å with an R_{work} and R_{free} of 0.21 and 0.25 respectively. The refined model contains 2 molecules of FaFra1E in the asymmetric unit. Both molecules show a similar structure and comprise amino acids 2 to 160. The region corresponding to the β3-β4 loop (amino acids 61-63 and 61-65 in chains A and B, respectively) is disordered and could not be modelled (Figure 3). The three amino acids from the N-terminal purification tag that remain after TEV cleavage are also visible in the structure. As can be appreciated in Figure 3a, the structure of FaFra1E conforms to the START fold (Iyer *et al.*, 2001; Radauer *et al.*, 2008) and consists of a 7-stranded β-sheet, two short helical segments between strands 5 and 7 and a long alpha helix at the C-terminus. The β-sheet adopts a slightly curved shape with the long C-terminal alpha helix (α3) juxtaposed against the concave side, which creates a cavity in the middle of the protein. Two short helical segments (α1 and α2) close the cavity on one of its sides, while the other extremity is open and accessible to the solvent. The open end of the cavity is surrounded by loops α2-β2, β3-β4, β5-β6, and β7-α3, designated here respectively as L3 (aa 35-39), L5 (aa 60-65), L7 (aa 89-96), and L9 (aa 125-130), and by the N-terminal half of the last alpha helix (α3). A number of well-ordered water molecules are found inside the cavity. The overall configuration of the FaFra1E backbone is similar to those of other START and PR-10 proteins, like Bet v 1 from *Betula berrucosa* (Spangfort *et al.*, 1997; Gajhede at al., 2008) or LIPR-102.B from yellow lupin (Fernandes *et al.*, 2009). The major differences are found in the conformation of the loops surrounding the open end of the cavity.

Initial phases for FaFra1E crystal form B were obtained by the molecular replacement method, using the refined model of crystal form A. The final model was refined to a resolution of 3.1 Å with an R_{work} and R_{free} of 0.21 and 0.26. The refined model contains 6 molecules of FaFra1E in the asymmetric unit. None of these molecules contained catechin inside the cavity. However, a molecule of catechin was found at the surface of molecule F between two FaFra1E subunits in adjacent asymmetric units. The nature of the interaction suggests that this catechin molecule has been captured as a consequence of the crystallization process. As expected, the six molecules in crystal form B show a very similar structure, which is also similar to that of crystal form A (see Figure 3b). However, a high degree of conformational variation is observed in the loop regions L3, L5 and L7. In particular loop L5, shows a distinct conformation in every of the 6 molecules found in the asymmetric unit.

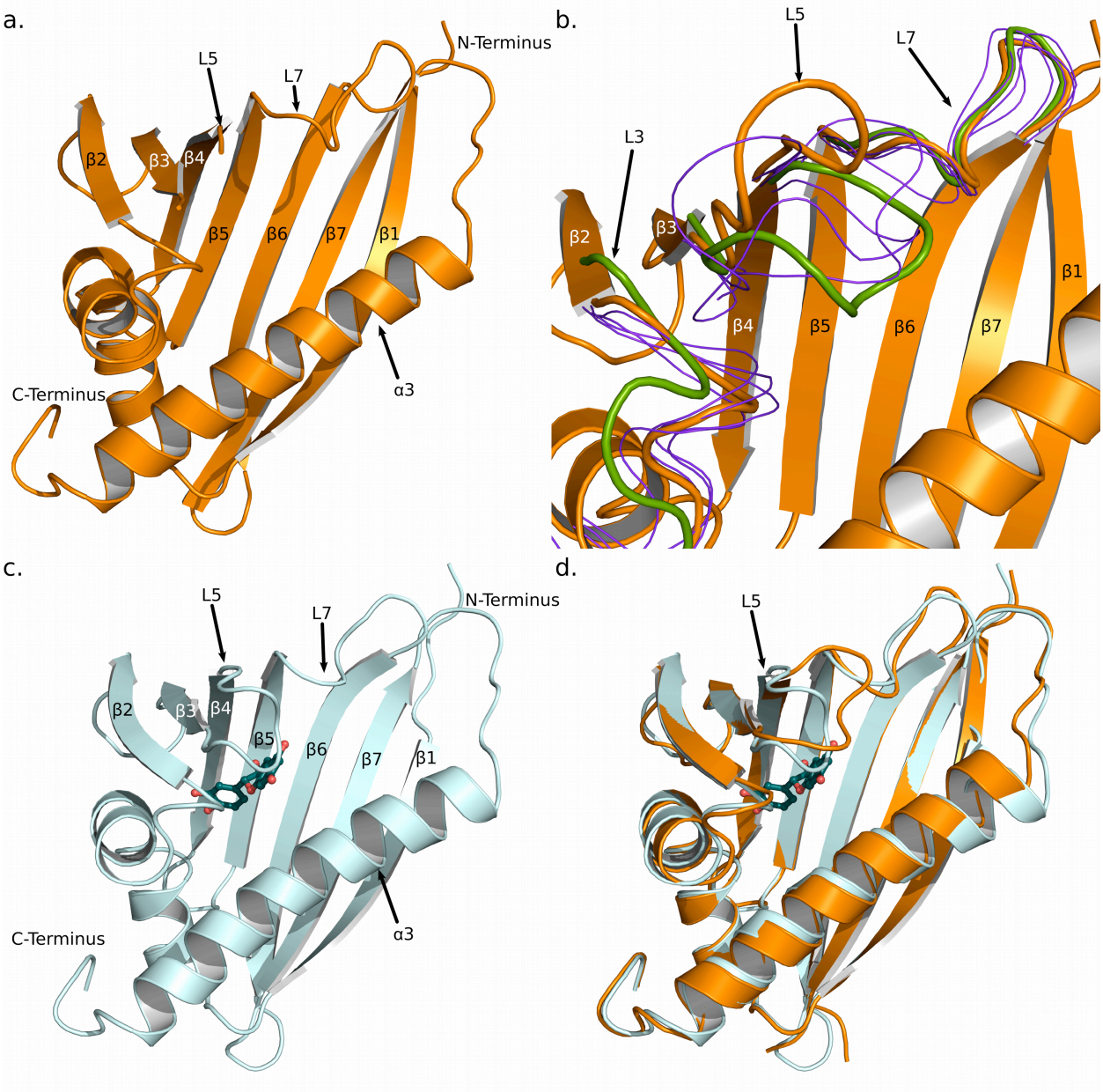


Figure 3. Structure of the FaFra1E and FaFra3-catechin complex. **a.** FaFra1E, crystal form A, shown in ribbon representation. **b.** Alternative loop conformations in FaFra1E, crystal form B. The backbone atoms of the six independent molecules in the asymmetric unit have been superposed. Chain F is shown in ribbon representation. The different conformations of the loops L3, L5 and L7 are represented in different colors: orange: open conformation (molecule F), green: closed conformation (molecule C) and blue: intermediate conformations (molecules A, B, D & E). **c.** FaFra3-(+)-catechin complex, the protein and ligand are shown in ribbon and ball-and-stick representations respectively. **d.** Superposition of FaFra1E (orange) and FaFra3-(+)-catechin (cyan) structures.

These conformations range from a "closed" conformation, represented by molecule C, in which loop L5 is folded over the central cavity of the protein and approaches the C-terminal alpha helix (in green in Figure 3b), to an "open" conformation represented by molecule F, in which loop L5 is at a maximal distance from the C-terminal alpha helix (in orange in Figure 3b). When these two molecules are superimposed, the positions of the Cα atoms of serine 63, in the middle of loop L5 are more than 9 Å apart. This indicates that loops L3, L5 and L7 surrounding the central cavity of FaFra show considerable conformational flexibility and that loop L5 is capable of adopting both open and closed conformations.

Table 1. Crystallographic data collection and refinement statistics.

	FaFra1E (Crystal form A)	FaFra1E (Crystal form B)	FaFra3-Catechin
Data Collection			
Space group	$P2_12_12_1$	$P2_12_12_1$	$C222_1$
Unit cell a, b, c	70.02 74.42 84.04	81.73 82.46 224.77	137.91 206.61 174.7
α, β, γ	90 90 90	90 90 90	90 90 90
Resolution	36.03- 2.2 (2.32- 2.2)[1]	30- 3.1 (3.27- 3.1)[1]	30- 3.0 (3.16- 3.0)[1]
No Observations (overall/ unique)	165127/ 24107	104880/27085	371478/ 49904
Average redundancy	7.2 (7.3)	3.9 (3.9)	7.4 (7.5)
*R*p.i.m.	0.034 (0.183)	0.078 (0.307)	0.03 (0.23)
Completeness	99.9 (100%)	96.5 (98)	99.5 (100)
I/σ(I)	14.0 (4.2)	7.9 (2.6)	21.7 (3.6)
Refinement			
Resolution Range (Å)	36.03 – 2.2	29.87- 3.1	29.67-3.0
Rwork	0.21	0.21	0.18
Rfree	0.25	0.26	0.20
No. of non-H atoms			
Protein	2427	7567	6346
Solvent	157	23	12
Ramachandran plot (%)	98.4/1.6/0.0	96.9/3.1/0.0	98.3/5.6/0.1
R.m.s. deviations			
Bond Length	0.019	0.012	0.027
Angles	1.961	1.531	2.692
Average B-factors			
Protein	39.83	58.45	72.45
Ligand	-	-	92.94

[1]Highest Resolution Range

Structure of the FaFra3 Protein in Complex with Catechin

In the case of FaFra3, crystals were obtained only in the presence of catechin, a natural flavonoid present in strawberries. The structure of the FaFra1E protein was used to obtain initial phases by the molecular replacement method. The final FaFra3 model was refined to a resolution of 3.0 Å with an R_{work} and R_{free} of 0.18 and 0.20 respectively (see Table 1 for

crystallographic and refinement statistics). The crystallographic model contains 5 protein chains and six molecules of catechin. The 5 protein chains in the asymmetric unit show a similar configuration. The 159 amino acids of the FaFra3 sequence could be modelled in all the chains except for amino acids 61 to 65 (corresponding to loop L5) of chain D that were not modelled due to weak density. SEC-MALLS analysis indicated that FaFra3 is a monomer in solution. Hence, the inter-subunit interactions between FaFra3 subunits are likely to be induced by the crystallization process. For simplicity we will refer to chain A for the rest of the text.

As expected, FaFra3 shows a typical START fold (Figure 3c) and high structural similarity with FaFra1E (both structures can be superimposed with an RSMD of 0.82 Å over 148 Cα atoms). The major differences between FaFra3 and FaFra1E are found in the loops surrounding the entry to the cavity (Figure 3d, see below). All the molecules of FaFra3 in the asymmetric unit showed additional density inside the cavity. This density could be easily interpreted as (+)-catechin (Figure 4a). The sixth (+)-catechin molecule in the FaFra3 crystals is not found inside the cavity, but at a surface location, at the interface between three FaFra3 proteins in adjacent asymmetric units. The presence of this additional molecule is in agreement with the ITC results that identified the presence of a low affinity binding site. In the crystal structure this second catechin molecule is stabilized by contacts to three different protein molecules, however FaFra3 is monomeric in solution. Hence, this second site is unlikely to play a major physiological role.

The catechin molecule inside the FaFra3 binding cavity is stabilized both by polar and hydrophobic interactions (Figure 4a) and shows the same conformation in the 5 copies within the asymmetric unit. The flavan nucleus is oriented with its long axis approximately parallel to the helix α3, and with the B ring pointing towards the bottom of the cavity. Interestingly, as in the case of the structurally related PYR/PYL/RCAR abscisic acid receptor proteins (Miyazono *et al.*, 2009; Nishimura *et al.*, 2009; Park *et al.*, 2009; Santiago *et al.*, 2009), many of the polar interactions are mediated by water molecules (Figure 4b).

The two hydroxyl groups in the C ring are involved in a hydrogen bond network with water molecules and residues Asp28, and His70. Similarly, the hydroxyl group in position 6 of the A ring, at the opposite end of the catechin molecule, is stabilized by a series of hydrogen bond interactions with a water molecule, the hydroxyl group of Ser63 and backbone atoms in residues Gln37, Ala38 and Gly60. The hydroxyl group in position 3 of the central ring of catechin makes hydrogen bond interactions with His70 and with a water molecule, which in turn is hydrogen bonded to Tyr84. The aromatic A ring is pinned between Leu59 and the guanidinium group of

Arg139. Finally, the non-polar groups of catechin are surrounded by hydrophobic side chains including Ile31, Val39, Leu 59, Leu 143. These interactions stabilize the catechin molecule in the FaFra3 cavity and restrict the rotation around the single bond linking rings C and B, whose central planes are oriented at an angle of approximately 90 degrees.

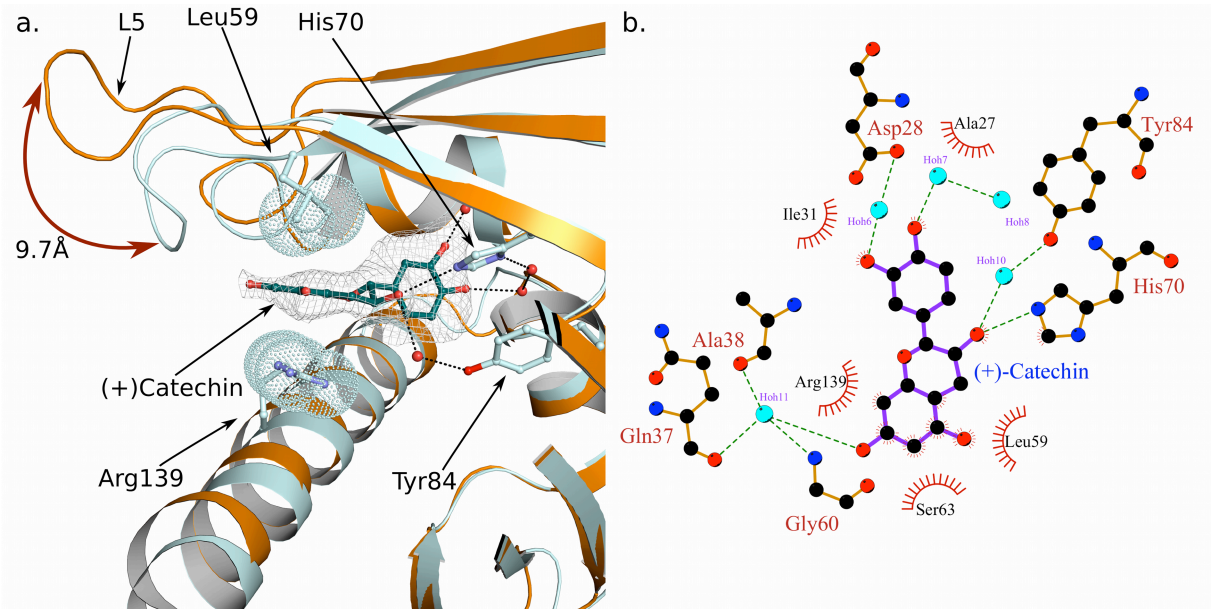


Figure 4. Binding of (+)-catechin to the FaFra3 cavity involves both polar and hydrophobic interactions and a closed conformation of loop L5. a. Structural superposition of FaFra1E (orange) and FaFra3 (cyan) around the loop 60-64. Protein chains are shown in ribbon representation. The catechin molecule and the side chains of the binding residues 59, 70, 84 and 139 are shown in ball and stick representation. Electron density around catechin is shown (omit map). The Van der Waals spheres of the distal atoms of the side chains of residues 59 and 139 are indicated by dot representation. The closed conformation of FaFra3, stabilized by interactions with the ligand, can be appreciated. **b.** Ligplot (Wallace *et al.*, 1995) representation of the molecular interactions between (+)-catechin and FaFra3. Green dashed lines indicate hydrogen bonds. Hydrophobic interactions are indicated by red hemispheres.

Binding of Catechin to FaFra a 3 Involves a Closed Conformation of Loop L5

The most notable difference between the FaFra1E and FaFra3 structures presented here are the presence of the ligand and the conformation of the loop L5, at the edge of the binding cavity (see Figure 3). In the FaFra3 structure the loop L5 wraps around one end of the catechin molecule and approaches the C-terminal alpha helix (α3) on the opposite side restricting the access to the cavity. At the same time, helix α3 of FaFra3 shows a slight bend at its center

towards loop L5, while loop L3 approaches loop L5 and the ligand (see Figure 3d). This "closed" conformation is stabilized by interactions with the catechin molecule. Namely, polar interactions between the hydroxyl in position 6 of the A ring of catechin, Ser63 in loop L5 and backbone atoms in residues Gln37, Ala38 of loop L3 and the hydrophobic interaction with Leu59 in loop L5 described above. Finally, the bending of helix α3 contributes to the packing of the guanidinium group of Arg139 against the A ring of catechin. This conformation produces a very compact structure with the catechin molecule enclosed inside the cavity. The cavity also shows a significant reduction in volume as compared to the FaFra1E structures (1646.4 A^3 for FaFra3, 2204.8 A^3 for molecule A in the FaFra1E crystal form A). Molecule C in crystal form B of FaFra1E, is the one that more closely resembles the FaFra3 catechin complex with loop L5 approaching helix α3 (Figure 2B). However, in this case, loop L3 adopts an open conformation and the bend in helix α3 is not present. This suggests that loops L3, L5 and L7 of FaFra a proteins may display a high level of conformational flexibility, but the presence of a ligand would be necessary to promote coordination between critical secondary structural elements leading to a fully closed conformation.

DISCUSSION

FaFra proteins are members of the pathogenesis related 10 family and are required for the normal accumulation of flavonoids and the development of colour in strawberry fruits (Karlsson *et al.*, 2004; Muñoz *et al.*, 2010). However, their function, as that of PR-10 proteins in general, is still not clearly understood. The data presented here demonstrates that FaFra proteins bind natural flavonoids, providing for the first time mechanistic insight on the function of these proteins in the control of flavonoid biosynthesis.

ITC experiments show that FaFra proteins can bind metabolites of the flavonoid pathway with affinities in the low μM range and with different selectivity. The three ligands identified in this study have been shown to be present in fruits as well as other parts of the strawberry plant (Kosar *et al.*, 2004; Aaby *et al.*, 2007; Hanhineva *et al.*, 2008; Sultana & Anwar, 2008; Muñoz *et al.*, 2011) and accumulate at the same organ and developmental stage where the highest expression levels of the FaFra proteins occur (Hjerno *et al.*, 2006; Muñoz *et al.*, 2010). The structure of the FaFra3-catechin complex provides details on the mechanism of ligand binding and stabilization. The catechin molecule adopts a linear disposition with its long axis approximately parallel to the axis of the long C-terminal α-helix of FaFra3. It is stabilized by polar interactions with the side chains of Asp28, Ser63, His70, Tyr84 and Arg139 and backbone

atoms of residues Gln37, Ala38 and Gly60, while the non-polar groups of catechin are surrounded by hydrophobic side chains including Ile31, Val39, Leu 59, Leu 143. The amino acids facing the cavity in the three FaFra proteins are generally conserved. However, variability is observed between the three proteins at certain positions (Figure 5), which could explain their different selectivity toward ligands. For example key amino acids, such as Leu59 and Arg139, involved in FaFra3 (+)-catechin interaction are replaced by Phe and Lys in both FaFra1E and FaFra2A.

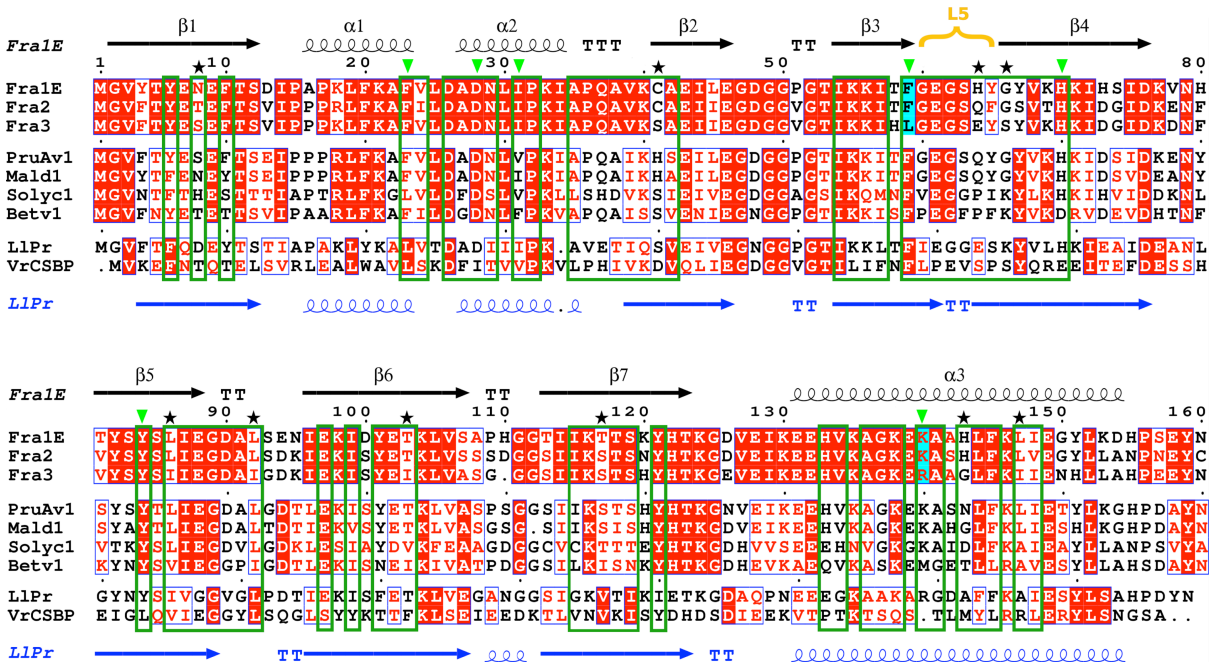


Figure 5. Multiple sequence alignment of the strawberry FaFra1E, FaFra2A and FaFra3 proteins and other related Bet v 1/START proteins. The secondary-structural elements correspond to the FaFra1E (black) and LIPR-10.2B (blue) structures. The L5 loop has been highlighted in yellow. Residues oriented towards the cavity and residues involved in catechin binding (for FaFra3) are indicated by dark green boxes and light green triangles, respectively. Positions showing sequence variations and that are either important for catechin binding (cyan) or facing the cavity (black stars) are also indicated. The sequence alignment was performed using ClustalW (EBI server), and the figure was generated by ESPript (Gouet *et al.*, 2003).

Comparison of the structures of the apo forms of FaFra1E and the FaFra3-catechin complex indicates that FaFra proteins show considerably flexibility in the loop regions surrounding the cavity (loops L3, L5 and L7) and that ligand-binding induces important conformational changes. FaFra3 adopts a more compact structure with a closed conformation of loop L5 that traps the

catechin molecule inside the cavity. Interestingly, loop L5 in FaFra proteins is structurally equivalent to the Ω1 loop of the mammalian START proteins (Tsujishita & Hurley, 2000) and the β3-β4 loop of the plant PYRL/PYL/RCAR hormone receptors (Santiago *et al.*, 2009), which also adopt closed conformations upon ligand binding (see Figure 6). In the case of the mammalian START proteins these conformational changes are thought to play a role in lipid extraction and solubilisation (Kudo *et al.*, 2000; Tsujishita & Hurley, 2000; Roderick *et al.*, 2002), while in the plant abscisic acid receptors the closed conformation stabilizes the hormone inside the cavity and promotes interaction between the receptor and protein phosphatases of the class 2C, leading to the activation of the ABA signalling pathway (Melcher *et al.*, 2009; Nishimura *et al.*, 2009; Santiago *et al.*, 2009). This suggests that ligand-induced conformational changes are a conserved feature in the START protein superfamily and might also play an important role in the function of other members of the PR-10 family.

The structural analysis presented here together with the molecular mechanisms previously described for LPTs and PYR/PYL/RCAR proteins suggest two possible functions for the FaFra proteins at molecular level. FaFra proteins could act as transporters or "chemical chaperones" binding to flavonoid intermediates and making them available to processing enzymes. The key enzymes of the phenylpropanoid and flavonoid biosynthesis pathways, including PAL, CHS and C4H have been shown to co-localise to the endoplasmic reticulum in different plant species. These enzymes form multi-protein complexes at the cytosolic side of the membrane where synthesis of many flavonoid and phenylpropanoid compounds occurs (Winkel, 2004; Lepiniec *et al.*, 2006; Zhao & Dixon, 2010). This association in multi-protein complexes has been proposed to help sequester unstable or toxic intermediates and to control the metabolic flux among the multiple branches of the pathway thereby determining which compounds are synthesized preferentially (Winkel, 2004). FaFra proteins might form part of these complexes, contributing to limit diffusion of intermediates and making them available to downstream processing enzymes. FaFra proteins could also be involved in the transport of flavonoids from the ER to other cellular membranes, like the tonoplast or the plasma membrane. Indeed, anthocyanins and other conjugated flavonoids like glycosylated catechin and epicatechins are translocated to and accumulated into the vacuole through the action of specific membrane transport proteins, while other flavonoid compounds are secreted to the apoplast through the plasma membrane, especially in roots. (Marinova *et al.*, 2007; Zhao & Dixon, 2009; Zhao & Dixon, 2010; Zhao *et al.*, 2011). In this case FaFra proteins might have a function analogous to that of the mammalian START proteins that act as cytosolic transporters of lipids shuttling between different cellular membranes (Soccion & Breslow, 2003; Lev, 2010).

Another possibility is that FaFra proteins might play a role as regulatory components involved in intracellular signalling. In maize, Arabidopsis and other species, the genes coding for enzymes involved in phenylpropanoid and flavonoid biosynthesis are regulated at a transcriptional level though the activity of MYB and bHLH type transcription factors (Lepiniec *et al.*, 2006; Vogt, 2010; Verdier *et al.*, 2012). Moreover, it has been recently shown that a blockage in downstream flavonoid processing enzymes results in transcriptional inhibition of PAL and that this inhibition is dependent on the accumulation of flavonoids, demonstrating that the expression of structural genes is mediated by a metabolic intermediate downstream of naringenin (Yin *et al.*, 2012). The capacity of FaFra proteins to bind specific flavonoids, suggests that they could play a role as signalling components, monitoring the metabolic flux through different branches of the pathway and influencing the expression level of specific regulatory genes. This would be consistent with the effect of FaFra silencing on the transcriptional activity of PAL and CHS genes and the altered accumulation of certain flavonoids (Muñoz *et al.*, 2010).

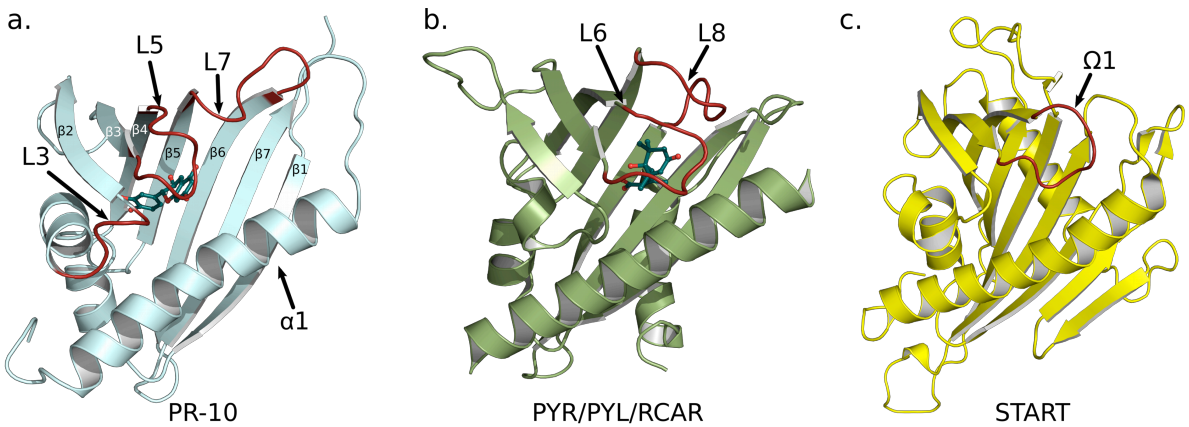


Figure 6. Function of PR-10, PYR/PYL/RCAR and START proteins involve conformational changes in loop regions a. FaFra3 in ribbon representation with bound (+)-catechin. The flexible loops L3, L5 and L7 are shown in red. **b.** The abscisic acid receptor PYR1 (3K90) in ribbon representation. The gating loops undergoing conformational changes hormone binding and receptor activation are depicted in red. **c.** The START domain of the CERT protein (3H3S) in ribbon representation. The Ω1 loop is shown in red.

Close homologues of the FaFra proteins have been found in apple, peach and tomato, some of which are also expressed to high levels during fruit ripening (Beuning *et al.*, 2004; Botton *et al.*, 2009). The amino acids involved in the FaFra3-catechin interaction are also highly conserved in these proteins (see Figure 5), which suggest that these proteins might also have the capacity to bind structurally-close flavonoids. However, other PR-10 proteins show more divergent amino

acids sequences in the cavity region (see Figure 5) and might bind other ligands. The phenylpropanoid and flavonoid biosynthesis pathway is responsible for the production of a large proportion of secondary metabolites in plants and shows a high degree of variability among species. It is not only involved in development of colour in fruits and flowers but it is also important for many other biological functions in plants, including defense against pathogens, insect attraction and pollination and UV protection among others. The structural analysis of the FaFra proteins suggests that PR-10 proteins, which are widespread in plants, might function in the control of flavonoid or other secondary metabolic pathways in plants.

Conclusions

1. Transcriptome analysis of strawberry (*Fragaria xananassa*) fruit along development identifies 10 putative *FaFra* genes, and a number of allelic variations.

2. Expression profiling of *FaFra* transcripts shows differential expression of the members between the different strawberry tissues, which suggests different functions and possibly different ligand binding specificities for the FaFra proteins.

3. Silencing of *FaFra2A* and *FaFra6* is in concert with down-regulation of genes encoding key enzymes of the flavonoid pathway and *MYB* transcription factors; this contribution supports a regulatory role of the FaFra proteins in the function of the flavonoid biosynthetic pathway.

4. The FaFra1-3 proteins selectively bind different natural flavonoids, with affinities within the micromolar range.

5. Binding of natural flavonoids to the hydrophobic cavity of FaFra induces conformational changes that stabilize the protein-ligand complex. These conformational changes recapitulate those observed in the START/PYR-PYL proteins and might be of functional relevance for the FaFra proteins.

6. Amino acid differences in the binding site of the central cavity of the FaFra proteins may explain their selectivity towards different metabolites. Our results provide for the first time mechanistic insights on the function of FaFra in the control of flavonoid biosynthesis.

Bibliography cited

Aaby, K., Ekeberg, D., & Skrede, G. (2007a). Characterization of Phenolic Compounds in Strawberry (Fragaria ×ananassa) Fruits by Different HPLC Detectors and Contribution of Individual Compounds to Total Antioxidant Capacity. *Journal of Agricultural and Food Chemistry*, *55*(11), 4395–4406.

Aaby, K., Skrede, G., & Wrolstad, R. E. (2005). Phenolic Composition and Antioxidant Activities in Flesh and Achenes of Strawberries (Fragaria ×ananassa). *Journal of Agricultural and Food Chemistry*, *53*(10), 4032–4040.

Aaby, K., Wrolstad, R. E., Ekeberg, D., & Skrede, G. (2007b). Polyphenol Composition and Antioxidant Activity in Strawberry Purees; Impact of Achene Level and Storage. *Journal of Agricultural and Food Chemistry*, *55*(13), 5156–5166.

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst (2010). D66, 213-221* [1–9.

Aharoni, A., De Vos, C. H., Wein, M., Sun, Z., Greco, R., Kroon, A., et al. (2001). The strawberry FaMYB1 transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. *The Plant journal : for cell and molecular biology*, *28*(3), 319–332.

Ahuja, I., Kissen, R., & Bones, A. M. (2012). Phytoalexins in defense against pathogens. *Trends in Plant Science*, *17*(2), 73–90.

Ali Ghasemzadeh. (2011). Flavonoids and phenolic acids: Role and biochemical activity in plants and human. *Journal of Medicinal Plants Research*, *5*(31).

Alvarez-Suarez, J. M., Dekanski, D., Ristić, S., Radonjić, N. V., Petronijević, N. D., Giampieri, F., et al. (2011). Strawberry Polyphenols Attenuate Ethanol-Induced Gastric Lesions in Rats by Activation of Antioxidant Enzymes and Attenuation of MDA Increase. (I. V. Lebedeva, Ed.)*PLoS ONE*, *6*(10), e25878.

Aron, P. M., & Kennedy, J. A. (2008). Flavan-3-ols: Nature, occurrence and biological activity. *Molecular Nutrition & Food Research*, *52*(1), 79–104.

Atkinson, R. G., Perry, J., Matsui, T., Ross, G. S., & Macrae, E. A. (1996). A stress-, pathogenesis-, and allergen-related cDNA in apple fruit is also ripening-related. *New Zealand Journal of Crop and Horticultural Science*, *24*(1), 103–107.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Research*, *28*(1), 235–242.

Berman, H., Henrick, K., & Nakamura, H. (2003). Announcing the worldwide Protein Data Bank. *Nature structural biology*, *10*(12), 980.

Beuning, L., Bowen, J., Persson, H., Barraclough, D., Bulley, S., & Macrae, E. (2004). Characterisation of Mal d 1-related genes in Malus. *Plant Molecular Biology*, *55*(3), 369–388.

Biesalski, H. K., Dragsted, L. O., Elmadfa, I., Grossklaus, R., Müller, M., Schrenk, D., et al. (2009). Bioactive compounds: Definition and assessment of activity. *Nutrition*, *25*(11-12), 1202–1205.

Blount, J. W., Korth, K. L., Masoud, S. A., Rasmussen, S., Lamb, C., & Dixon, R. A. (2000). Altering expression of cinnamic acid 4-hydroxylase in transgenic plants provides evidence for a feedback loop at the entry point into the phenylpropanoid pathway. *Plant Physiology*, *122*(1), 107–116.

Bombarely, A., Merchante, C., Csukasi, F., Cruz-Rus, E., Caballero, J. L., Medina-Escobar, N., et al. (2010). Generation and analysis of ESTs from strawberry (Fragaria xananassa) fruits and evaluation of their utility in genetic and molecular studies. *BMC Genomics*, *11*(1), 503.

Botton, A., Andreotti, C., Costa, G., & Ramina, A. (2009). Peach (Prunus persica L. Batsch) Allergen-Encoding Genes Are Developmentally Regulated and Affected by Fruit Load and Light Radiation. *Journal of Agricultural and Food Chemistry*, *57*(2), 724–734.

Carbone, F., Preuss, A., DE VOS, R. C. H., D'Amico, E., Perrotta, G., Bovy, A. G., et al. (2009). Developmental, genetic and environmental factors affect the expression of flavonoid genes, enzymes and metabolites in strawberry fruits. *Plant, Cell & Environment*, *32*(8), 1117–1131.

Casado-Díaz, A., Encinas-Villarejo, S., Santos, B. de L., Schilirò, E., Yubero-Serrano, E. M., Amil-Ruíz, F., et al. (2006). Analysis of strawberry genes differentially expressed in response to Colletotrichum infection. *Physiologia Plantarum*, *128*(4), 633–650.

Casañal, A., Zander, U., Dupeux, F., Valpuesta, V., & Marquez, J. A. (2013a). Purification, crystallization and preliminary X-ray analysis of the strawberry allergens Fra a 1E and Fra a 3 in the presence of catechin. *Crystallization communications. Acta Cryst (2013) F69, 510-514*, 1–5.

Casañal, A., Zander, U., Muñoz, C., Dupeux, F., Luque, I., Botella, M. A., Schwab, W., Valpuesta, V., & Marquez, J. A. (2013b). The Strawberry Pathogenesis-related 10 (PR-10) Fra a Proteins Control Flavonoid Biosynthesis by Binding Metabolic Intermediates. *Journal of Biological Chemistry*, *288*(49), 35322–35332.

Clough, S. J., & Bent, A. F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *The Plant Journal : for cell and molecular biology*, *16*(6), 735–743.

Cotelle, N. (2001). Role of Flavonoids in Oxidative Stress. *Current topics in medicinal chemistry*, *1*(6), 569–590.

Csukasi, F., Donaire, L., Casañal, A., Martínez-Priego, L., Botella, M. A., Medina-Escobar, N., et al. (2012). Two strawberry miR159 family members display developmental-specific expression patterns in the fruit receptacle and cooperatively regulate Fa-GAMYB. *New Phytologist*, *195*(1), 47–57.

Curtis, M. D. (2003). A Gateway Cloning Vector Set for High-Throughput Functional Analysis of Genes in Planta. *Plant Physiology*, *133*(2), 462–469.

Diamanti, J., Capocasa, F., Balducci, F., Battino, M., Hancock, J., & Mezzetti, B. (2012). Increasing Strawberry Fruit Sensorial and Nutritional Quality Using Wild and Cultivated Germplasm. (C. Peace, Ed.)*PLoS ONE*, *7*(10), e46470.

Dimasi, N., Flot, D., Dupeux, F., & Márquez, J. A. (2007). Expression, crystallization and X-ray data collection from microcrystals of the extracellular domain of the human inhibitory receptor expressed on myeloid cells IREM-1. *Acta Crystallographica Section F Structural Biology and Crystallization Communications*, *63*(3), 204–208.

Dümmler, A., Lawrence, A.-M., & de Marco, A. (2005). Simplified screening for the detection of soluble fusion constructs expressed in E. coli using a modular set of vectors. *Microbial cell factories*, *4*, 34.

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Research papers. *Acta Cryst (2010). D66, 486-501,* 1–16.

Evans, P. (2006). Scaling and assessment of data quality. *Acta Cryst. D62, 72-82,* 1–11.

Fait, A., Hanhineva, K., Beleggia, R., Dai, N., Rogachev, I., Nikiforova, V. J., et al. (2008). Reconfiguration of the Achene and Receptacle Metabolic Networks during Strawberry Fruit Development. *Plant Physiology*, *148*(2), 730–750.

Fernandes, H., Bujacz, A., Bujacz, G., Jelen, F., Jasinski, M., Kachlicki, P., et al. (2009). Cytokinin-induced structural adaptability of a Lupinus luteus PR-10 protein. *FEBS Journal*, *276*(6), 1596–1609.

Fernandes, H., Michalska, K., Sikorski, M., & Jaskolski, M. (2013). Structural and functional aspects of PR-10 proteins. *FEBS Journal*, *280*(5), 1169–1199.

Fernandes, H., Pasternak, O., Bujacz, G., Bujacz, A., Sikorski, M. M., & Jaskolski, M. (2008). Lupinus luteus Pathogenesis-Related Protein as a Reservoir for Cytokinin. *Journal of Molecular Biology*, *378*(5), 1040–1051.

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

Gajhede, M., Osmark, P., Poulsen, F. M., Ipsen, H., Larsen, J. N., Joost van Neerven, R. J., et al. (1996). X-ray and NMR structure of Bet v 1, the origin of birch pollen allergy. *Nature structural biology*, *3*(12), 1040–1045.

Gerard, F. C. A., Ribeiro, E. de A., Albertini, A. A. V., Gutsche, I., Zaccai, G., Ruigrok, R. W. H., & Jamin, M. (2007). Unphosphorylated RhabdoviridaePhosphoproteins Form Elongated Dimers in Solution †. *Biochemistry*, *46*(36), 10328–10338.

Geroldinger-Simic, M., Zelniker, T., Aberer, W., Ebner, C., Egger, C., Greiderer, A., et al. (2011). Birch pollen–related food allergy: Clinical aspectsand the role of allergen-specific IgE and IgG. *Journal of Allergy and Clinical Immunology*, *127*(3), 616–622.e1.

Giampieri, F., Tulipani, S., Álvarez-Suarez, J. M., Quiles, J. L., Mezzetti, B., Battino, M. (2012). The strawberry: Composition, nutritional quality, and impact on human health. *Nutrition*, *28*(1), 9–19.

Giampieri, F., Alvarez-Suarez, J. M., Mazzoni, L., Romandini, S., Bompadre, S., Diamanti, J., et al. (2013). The potential impact of strawberry on human health. *Natural Product Research*, *27*(4-5), 448–455.

Giovannoni, J. J. (2004). Genetic Regulation of Fruit Development and Ripening. *The Plant Cell Online*, *16*, S170–S180.

Gouet, P. (2003). ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Research*, *31*(13), 3320–3323.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652.

Grierson, D. (2013). Ethylene and the Control of Fruit Ripening. In *The Molecular Biology and Biochemistry of Fruit Ripening* (pp. 43–73). Blackwell Publishing Ltd.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*(8), 1494–1512.

Halbwirth, H., Puhl, I., Haas, U., Jezik, K., Treutter, D., & Stich, K. (2006). Two-Phase Flavonoid Formation in Developing Strawberry (Fragaria ×ananassa) Fruit. *Journal of Agricultural and Food Chemistry*, *54*(4), 1479–1485.

Hanhineva, K., Rogachev, I., Kokko, H., Mintz-Oron, S., Venger, I., Kärenlampi, S., & Aharoni, A. (2008). Non-targeted analysis of spatial metabolite composition in strawberry (Fragaria×ananassa) flowers. *Phytochemistry*, *69*(13), 2463–2481.

Harborne, J. B., & Williams, C. A. (2000). Advances in flavonoid research since 1992. *Phytochemistry*, *55*(6), 481–504.

Hartree, E. F. (1972). Determination of protein: a modification of the Lowry method that gives a linear photometric response. *Analytical Biochemistry*, *48*(2), 422–427.

Hassan, S., & Mathesius, U. (2012). The role of flavonoids in root-rhizosphere signalling: opportunities and challenges for improving plant-microbe interactions. *Journal of Experimental Botany*, *63*(9), 3429–3444.

Herman, E. M. (2003). Genetically modified soybeans and food allergies. *Journal of Experimental Botany*, *54*(386), 1317–1319.

Hichri, I., Barrieu, F., Bogs, J., Kappel, C., Delrot, S., & Lauvergeat, V. (2011). Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *Journal of Experimental Botany*, *62*(8), 2465–2483.

Hirakawa, H., Shirasawa, K., Kosugi, S., Tashiro, K., Nakayama, S., Yamada, M., et al. (2014). Dissection of the Octoploid Strawberry Genome by Deep Sequencing of the Genomes of Fragaria Species. *DNA Research*, *21*(2), 169-81.

Hjernø, K., Alm, R., Canbäck, B., Matthiesen, R., Trajkovski, K., Björk, L., et al. (2006). Down-regulation of the strawberry Bet v 1-homologous allergen in concert with the flavonoid biosynthesis pathway in colorless strawberry mutant. *Proteomics*, *6*(5), 1574–1587.

Hoffmann, T., Kalinowski, G., & Schwab, W. (2006). RNAi-induced silencing of gene expression in strawberry fruit (Fragaria ×ananassa) by agroinfiltration: a rapid assay for gene function analysis. *The Plant Journal*, *48*(5), 818–826.

Huang, J., Gu, M., Lai, Z., Fan, B., Shi, K., Zhou, Y. H., et al. (2010). Functional Analysis of the Arabidopsis PAL Gene Family in Plant Growth, Development, and Response to Environmental Stress. *Plant Physiology*, *153*(4), 1526–1538.

Iannetta, P. P. M., Laarhoven, L.-J., Medina-Escobar, N., James, E. K., McManus, M. T., Davies, H. V., & Harren, F. J. M. (2006). Ethylene and carbon dioxide production by developing strawberries show a correlative pattern that is indicative of ripening climacteric fruit. *Physiologia Plantarum*, *127*(2), 247–259.

Iyer, L. M., Koonin, E. V., & Aravind, L. (2001). Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. *Proteins: Structure, Function, and Bioinformatics*, *43*(2), 134–144.

Jaakola, L. (2013). New insights into the regulation of anthocyanin biosynthesis in fruits. *Trends in Plant Science*, *18*(9), 477–483.

Jia, H. F., Chai, Y. M., Li, C. L., Lu, D., Luo, J. J., Qin, L., & Shen, Y. Y. (2011). Abscisic Acid Plays an Important Role in the Regulation of Strawberry Fruit Ripening. *Plant Physiology*, *157*(1), 188–199.

Kabsch, W. (2010). XDS. Research papers. *Acta Cryst. D66, 125-132,* 1–8.

Karlsson, A. L., Alm, R., Ekstrand, B., Fjelkner-Modig, S., Schiott, A., Bengtsson, U., et al. (2004). Bet v 1 homologues in strawberry identified as IgE-binding proteins and presumptive allergens. *Allergy*, *59*(12), 1277–1284.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes inthe presence of insertions, deletions and genefusions. *Genome Biology*, *14*(4), R36.

Kobayashi, H., Graven, Y. N., Broughton, W. J., & Perret, X. (2004). Flavonoids induce temporal shifts in gene-expression of nod-box controlled loci in Rhizobium sp. NGR234. *Molecular Microbiology*, *51*(2), 335–347.

Koes, R. E., Quattrocchio, F., & Mol, J. N. (1994). The flavonoid biosynthetic pathway in plants: function and evolution. *BioEssays*, *16*(2), 123–132.

Kofler, S., Asam, C., Eckhard, U., Wallner, M., Ferreira, F., & Brandstetter, H. (2012). Crystallographically Mapped Ligand Binding Differs in High and Low IgE Binding Isoforms of Birch Pollen Allergen Bet v 1. *Journal of Molecular Biology*, *422*(1), 109–123.

Kosar, M., Kafkas, E., Paydas, S., & Baser, K. H. C. (2004). Phenolic Composition of Strawberry Genotypes at Different Maturation Stages. *Journal of Agricultural and Food Chemistry*, *52*(6), 1586–1589.

Kris-Etherton, P. M., Lefevre, M., Beecher, G. R., Gross, M. D., Keen, C. L., & Etherton, T. D. (2004). Bioactive Compounds In Nutrition And Health-Research Methodologies For Establishing Biological Function: The Antioxidant and Anti-inflammatory Effects of Flavonoids on Atherosclerosis. *Annual Review of Nutrition*, *24*(1), 511–538.

Kudo, N., Kumagai, K., Tomishige, N., Yamaji, T., Wakatsuki, S., Nishijima, M., et al. (2008). Structural basis for specific lipid recognition by CERT responsible for nonvesicular trafficking of ceramide. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(2), 488–493.

Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, *227*(5259), 680–685.

Lebel, S., Schellenbaum, P., Walter, B., & Maillot, P. (2010). Characterisation of the Vitis vinifera PR10 multigene family. *BMC Plant Biology*, *10*(1), 184.

Lepiniec, L., Debeaujon, I., Routaboul, J.-M., Baudry, A., Pourcel, L., Nesi, N., & Caboche, M. (2006). Genetics and biochemistry of seed flavonoids. *Annual Review of Plant Biology*, *57*, 405–430.

Lev, S. (2010). Non-vesicular lipid transport by lipid-transfer proteins and beyond. *Nature Publishing Group*, *11*(10), 739–750.

Liu, J.-J., & Ekramoddoullah, A. K. M. (2006). The family 10 of plant pathogenesis-related proteins: Their structure, regulation, and function in response to biotic and abiotic stresses. *Physiological and Molecular Plant Pathology*, *68*(1-3), 3–13.

Ma, Y., Szostkiewicz, I., Korte, A., Moes, D., Yang, Y., Christmann, A., & Grill, E. (2009). Regulators of PP2C Phosphatase Activity Function as Abscisic Acid Sensors. *Science*.

Mahajan, M., Ahuja, P. S., & Yadav, S. K. (2011). Post-Transcriptional Silencing of Flavonol Synthase mRNA in Tobacco Leads to Fruits with Arrested Seed Set. (R. P. Niedz, Ed.) *PLoS ONE*, *6*(12), e28315.

Manning, K. (1991). Isolation of nucleic acids from plants by differential solvent precipitation. *Analytical Biochemistry*, *195*(1), 45–50.

Marinova, K., Pourcel, L., Weder, B., Schwarz, M., Barron, D., Routaboul, J. M., et al. (2007). The Arabidopsis MATE Transporter TT12 Acts as a Vacuolar Flavonoid/H+-Antiporter Active in Proanthocyanidin-Accumulating Cells of the Seed Coat. *The Plant Cell Online*, *19*(6), 2023–2038.

Marković-Housley, Z., Degano, M., Lamba, D., Roepenack-Lahaye, von, E., Clemens, S., Susani, M., et al. (2003). Crystal Structure of a Hypoallergenic Isoform of the Major Birch Pollen Allergen Bet v 1 and its Likely Biological Function as a Plant Steroid Carrier. *Journal of Molecular Biology*, *325*(1), 123–133.

Martin, L. B. B., Fei, Z., Giovannoni, J. J., & Rose, J. K. C. (2013). Catalyzing plant science research with RNA-seq. *Frontiers in plant science*, *4*, 66.

McCarthy, A. A., Brockhauser, S., Nurizzo, D., Theveneau, P., Mairs, T., Spruce, D., et al. (2009). A decade of user operation on the macromolecular crystallography MAD beamline ID14-4 at the ESRF. Research papers. *J. Synchrotron Rad. 16, 803-812*, 1–10.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). *Phaser* crystallographic software. Research papers. *J. Appl. Cryst. 40, 658-674*, 1–17.

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., et al. (2013). Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research*, *41*(W1), W597–W600.

Medina-Puche, L., Cumplido-Laso, G., Amil-Ruiz, F., Hoffmann, T., Ring, L., Rodriguez-Franco, A., et al. (2014). MYB10 plays a major role in the regulation of flavonoid/phenylpropanoid metabolism during ripening of Fragaria xananassa fruits. *Journal of Experimental Botany*, *65*(2), 401–417.

Melcher, K., Ng, L. M., Zhou, X. E., Soon, F. F., Xu, Y., Suino-Powell, K. M., et al. (2009). lock mechanism for hormone signalling by abscisic acid receptors. *Nature*, *462*(7273), 602–608.

Merchante, C., Vallarino, J. G., Osorio, S., Araguez, I., Villarreal, N., Ariza, M. T., et al. (2013). Ethylene is involved in strawberry fruit ripening in an organ-specific manner. *Journal of Experimental Botany*, *64*(14), 4421–4439.

Mirza, O., Henriksen, A., Ipsen, H., Larsen, J. N., Wissenbach, M., Spangfort, M. D., & Gajhede, M. (2000). Dominant Epitopes and Allergic Cross-Reactivity: Complex Formation Between a Fab Fragment of a Monoclonal Murine IgG Antibody and the Major Allergen from Birch Pollen Bet v 1. *The Journal of Immunology*, *165*(1), 331–338.

Miyazono, K.-I., Miyakawa, T., Sawano, Y., Kubota, K., Kang, H.-J., Asano, A., et al. (2009). Structural basis of abscisic acid signalling. *Nature*, *462*(7273), 609–614.

Mogensen, J. E. (2002). The Major Birch Allergen, Bet v 1, Shows Affinity for a Broad Spectrum of Physiological Ligands. *Journal of Biological Chemistry*, *277*(26), 23684–23692.

Mogensen, J. E., Ferreras, M., Wimmer, R., Petersen, S. V., Enghild, J. J., & Otzen, D. E. (2007). The Major Allergen from Birch Tree Pollen, Bet v 1, Binds and Permeabilizes Membranes †. *Biochemistry*, *46*(11), 3356–3365.

Moura, J. C. M. S., Bonine, C. A. V., de Oliveira Fernandes Viana, J., Dornelas, M. C., & Mazzafera, P. (2010). Abiotic and Biotic Stresses and Changes in the Lignin Content and Composition in Plants. *Journal of Integrative Plant Biology*, *52*(4), 360–376.

Muñoz, C., Hoffmann, T., Escobar, N. M., Ludemann, F., Botella, M. A., Valpuesta, V., & Schwab, W. (2010). The Strawberry Fruit Fra a Allergen Functions in Flavonoid Biosynthesis. *Molecular Plant*, *3*(1), 113–124.

Muñoz, C., Sánchez-Sevilla, J. F., Botella, M. A., Hoffmann, T., Schwab, W., & Valpuesta, V. (2011). Polyphenol Composition in the Ripe Fruits of Fragaria Species and Transcriptional Analyses of Key Genes in the Pathway. *Journal of Agricultural and Food Chemistry*, *59*(23), 12598–12604.

Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., et al. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst. D67, 355-367*, 1–13.

Musidlowska-Persson, A., Alm, R., & Emanuelsson, C. (2007). Cloning and sequencing of the Bet v 1-homologous allergen Fra a 1 in strawberry (Fragaria ananassa) shows the presence of an intron and little variability in amino acid sequence. *Molecular Immunology*, *44*(6), 1245–1252.

Nakajima, K., & Benfey, P. N. (2002). Signaling in and out: control of cell division and differentiation in the shoot and root. *The Plant Cell*, *14 Suppl*, S265–76.

Naoumkina, M. A., Zhao, Q., Gallego-Giraldo, L., Dai, X., Zhao, P. X., & Dixon, R. A. (2010). Genome-wide analysis of phenylpropanoid defence pathways. *Molecular Plant Pathology*, *11*(6), 829–846.

Nishimura, N., Hitomi, K., Arvai, A. S., Rambo, R. P., Hitomi, C., Cutler, S. R., et al. (2009). Structural Mechanism of Abscisic Acid Binding and Signaling by Dimeric PYR1. *Science*, *326*(5958), 1373–1379.

Olsen, K. M., Lea, U. S., Slimestad, R., Verheul, M., & Lillo, C. (2008). Differential expression of four Arabidopsis PAL genes; PAL1 and PAL2 have functional specialization in abiotic environmental-triggered flavonoid synthesis. *Journal of Plant Physiology*, *165*(14), 1491–1499.

Osorio, S., Castillejo, C., Quesada, M. A., Medina-Escobar, N., Brownsey, G. J., Suau, R., et al. (2007). Partial demethylation of oligogalacturonides by pectin methyl esterase 1 is required for eliciting defence responses in wild strawberry (*Fragaria vesca*). *The Plant Journal*, *54*(1), 43–55.

Park, S. Y., Fung, P., Nishimura, N., Jensen, D. R., Fujii, H., Zhao, Y., et al. (2009). Abscisic Acid Inhibits Type 2C Protein Phosphatases via the PYR/PYL Family of START Proteins. *Science*.

Pasternak, O., Bujacz, G. D., Fujimoto, Y., Hashimoto, Y., Jelen, F., Otlewski, J., et al. (2006). Crystal Structure of Vigna radiata Cytokinin-Specific Binding Protein in Complex with Zeatin. *The Plant Cell Online*, *18*(10), 2622–2634.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, *29*(9), e45.

Ponting, C. P., & Aravind, L. (1999). START: a lipid-binding domain in StAR, HD-ZIP and signalling proteins. *Trends in biochemical sciences*, *24*(4), 130–132.

Poustka, F., Irani, N. G., Feller, A., Lu, Y., Pourcel, L., Frame, K., & Grotewold, E. (2007). A Trafficking Pathway for Anthocyanins Overlaps with the Endoplasmic Reticulum-to-Vacuole Protein-Sorting Route in Arabidopsis and Contributes to the Formation of Vacuolar Inclusions. *Plant Physiology*, *145*(4), 1323–1335.

Puehringer, H. M., Zinoecker, I., Marzban, G., Katinger, H., & Laimer, M. (2003). MdAP, a novel protein in apple, is associated with the major allergen Mal d 1. *Gene, 321*, 173-183.

Radauer, C., Lackner, P., & Breiteneder, H. (2008). The Bet v 1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. *BMC Evolutionary Biology*, *8*(1), 286.

Roderick, S. L., Chan, W. W., Agate, D. S., Olsen, L. R., Vetting, M. W., Rajashankar, K. R., & Cohen, D. E. (2002). Structure of human phosphatidylcholine transfer protein in complex with its ligand. *Nature structural biology*.

Rosado, A., Hicks, G. R., Norambuena, L., Rogachev, I., Meir, S., Pourcel, L., et al. (2011). Sortin1-Hypersensitive Mutants Link Vacuolar-Trafficking Defects and Flavonoid Metabolism in Arabidopsis Vegetative Tissues. *Chemistry & Biology*, *18*(2), 187–197.

Rousseau-Gueutin, M., Gaston, A., Aïnouche, A., Aïnouche, M. L., Olbricht, K., Staudt, G., et al. (2009). Molecular Phylogenetics and Evolution. *Molecular Phylogenetics and Evolution*, *51*(3), 515–530.

Salvatierra, A., Pimentel, P., Moya-León, M. A., & Herrera, R. (2013). Increased accumulation of anthocyanins in *Fragaria chiloensis* fruits by transient suppression of *FcMYB1* gene. *Phytochemistry*, *90*(C), 25–36.

Santiago, J., Dupeux, F., Round, A., Antoni, R., Park, S.-Y., Jamin, M., et al. (2009). The abscisic acid receptor PYR1 in complex with abscisic acid. *Nature*, *462*(7273), 665–668.

Saslowsky, D. E., Warek, U., & Winkel, B. S. J. (2005). Nuclear Localization of Flavonoid Enzymes in Arabidopsis. *Journal of Biological Chemistry*, *280*(25), 23735–23740.

Saslowsky, D., & Winkel-Shirley, B. (2001). Localization of flavonoid enzymes in Arabidopsis roots. *The Plant Journal*, *27*(1), 37–48.

Sánchez-Sevilla, J. F., Cruz-Rus, E., Valpuesta, V., Botella, M. A., & Amaya, I. (2014). Deciphering gamma-decalactone biosynthesis in strawberry fruit using a combination of genetic mapping, RNA-Seq and eQTL analyses. *BMC Genomics*, *15*(1), 1–15.

Scalzo, J., Politi, A., Pellegrini, N., Mezzetti, B., & Battino, M. (2005). Plant genotype affects total antioxidant capacity and phenolic contents in fruit. *Nutrition*, *21*(2), 207–213.

Schaart, J. G., Dubos, C., Romero De La Fuente, I., van Houwelingen, A. M. M. L., DE VOS, R. C. H., Jonker, H. H., et al. (2013). Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (Fragaria ×ananassa) fruits. *New Phytologist*, *197*(2), 454–467.

Schaefer, A. L., Greenberg, E. P., Oliver, C. M., Oda, Y., Huang, J. J., Bittan-Banin, G., et al. (2008). A new class of homoserine lactone quorum-sensing signals. *Nature*, *454*(7204), 595–599.

Sels, J., Mathys, J., De Coninck, B. M. A., Cammue, B. P. A., & De Bolle, M. F. C. (2008). Plant pathogenesis-related (PR) proteins: A focus on PR peptides. *Plant Physiology and Biochemistry*, *46*(11), 941–950.

Seutter von Loetzen, C., Hoffmann, T., Hartl, M. J., Schweimer, K., Schwab, W., Rösch, P., & Hartl Spiegelhauer, O. (2014). Secret of the major birch pollen allergen Bet v 1: identification of the physiological ligand. *Biochemical Journal*, *457*(3), 379–390.

Seutter von Loetzen, C., Schweimer, K., Schwab, W., Rösch, P., & Hartl Spiegelhauer, O. (2012). Solution structure of the strawberry allergen Fra a 1. *Bioscience Reports*, *32*(6), 567–575.

Seymour, G. B., Østergaard, L., Chapman, N. H., Knapp, S., & Martin, C. (2013). Fruit Development and Ripening. *Annual Review of Plant Biology*, *64*(1), 219–241.

Shulaev, V., Sargent, D. J., Crowhurst, R. N., Mockler, T. C., Folkerts, O., Delcher, A. L., et al. (2010). The genome of woodland strawberry (Fragaria vesca). *Nature Genetics*, *43*(2), 109–116.

Soccio, R. E. (2003). StAR-related Lipid Transfer (START) Proteins: Mediators of Intracellular Lipid Metabolism. *Journal of Biological Chemistry*, *278*(25), 22183–22186.

Solovchenko, A. (2003). Significance of skin flavonoids for UV-B-protection in apple fruits. *Journal of Experimental Botany*, *54*(389), 1977–1984.

Spangfort, M. D., Gajhede, M., Osmark, P., Poulsen, F. M., Ipsen, H., Larsen, J. N., et al. (1997). Three-Dimensional Structure and Epitopes of Bet v 1. *International Archives of Allergy and Immunology*, *113*(1-3), 243–245 AB.

Spangfort, M. D., Mirza, O., Holm, J., Larsen, J. N., Ipsen, H., & Løwenstein, H. (1999). The structure of major birch pollen allergens-epitopes, reactivity and cross-reactivity. *Allergy*, *54 Suppl 50*, 23–26.

Sultana, B. & Anwar, F. (2008). Flavonols (kaempeferol, quercetin, myricetin) contents of selected fruits, vegetables and medicinal plants. *Food Chemistry*, *108* (3), 879-884.

Suzuki, T., Fujikura, K., Higashiyama, T., & Takata, K. (1997). DNA Staining for Fluorescence and Laser Confocal Microscopy. *Journal of Histochemistry & Cytochemistry*, *45*(1), 49–53.

Swoboda, I., Jilek, A., Ferreira, F., Engel, E., Hoffmann-Sommergruber, K., Scheiner, O., et al. (1995). Isoforms of Bet v 1, the major birch pollen allergen, analyzed by liquid chromatography, mass spectrometry, and cDNA cloning. *The Journal of biological chemistry*, *270*(6), 2607–2613.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*, *28*(10), 2731–2739.

Thompson, E. P., Wilkins, C., Demidchik, V., Davies, J. M., & Glover, B. J. (2010). An Arabidopsis flavonoid transporter is required for anther dehiscence and pollen development. *Journal of Experimental Botany*, *61*(2), 439–451.

Tohge, T., Watanabe, M., Hoefgen, R., & Fernie, A. R. (2013). Shikimate and phenylalanine biosynthesis in the green lineage. *Frontiers in plant science*, *4*, 62.

Trainotti, L. (2005). Different ethylene receptors show an increased expression during the ripening of strawberries: does such an increment imply a role for ethylene in the ripening of these non-climacteric fruits? *Journal of Experimental Botany*, *56*(418), 2037–2046.

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., & Pachter, L. (2012). Differential analysis of gene regulation at transcript resolution with rNA-seq. *Nature Biotechnology*, *31*(1), 46–53.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). nbt.1621. *Nature Biotechnology*, *28*(5), 516–520.

Treutter, D. (2005). Significance of flavonoids in plant resistance and enhancement of their biosynthesis. *Plant biology (Stuttgart, Germany)*, *7*(6), 581–591.

Tsujishita, Y., & Hurley, J. H. (2000). Structure and lipid transport mechanism of a StAR-related domain. *Nature structural biology*, *7*(5), 408–414.

Tulipani, S., Alvarez-Suarez, J. M., Busco, F., Bompadre, S., Quiles, J. L., Mezzetti, B., & Battino, M. (2011a). Strawberry consumption improves plasma antioxidant status and erythrocyte resistance to oxidative haemolysis in humas. *Food Chemistry*, *128*(1), 180–186.

Tulipani, S., Marzban, G., Herndl, A., Laimer, M., Mezzetti, B., & Battino, M. (2011b). Influence of environmental and genetic factors on health-related compounds in strawberry. *Food Chemistry*, *124*(3), 906–913.

Tulipani, S., Mezzetti, B., Capocasa, F., Bompadre, S., Beekwilder, J., de Vos, C. H. R., et al. (2008). Antioxidants, Phenolic Compounds, and Nutritional Quality of Different Strawberry Genotypes. *Journal of Agricultural and Food Chemistry*, *56*(3), 696–704.

van Loon, L. C., Rep, M., & Pieterse, C. M. J. (2006). Significance of inducible defense-related proteins in infected plants. *Annual Review of Phytopathology*, *44*, 135–162.

Verdier, J., Zhao, J., Torres-Jerez, I., Ge, S., Liu, C., He, X., et al. (2012). MtPAR MYB transcription factor acts as an on switch for proanthocyanidin biosynthesis in Medicago truncatula. *Proceedings of the National Academy of Sciences*, *109*(5), 1766–1771.

Vikram, A., Jayaprakasha, G. K., Jesudhasan, P. R., Pillai, S. D., & Patil, B. S. (2010). Suppression of bacterial cell-cell signalling, biofilm formation and type III secretion system by citrus flavonoids. *Journal of Applied Microbiology*, *109 (2)*, 515-527.

Vogt, T. (2010). Phenylpropanoid Biosynthesis. *Molecular Plant*, *3*(1), 2–20.

Voinnet, O., Rivas, S., Mestre, P., & Baulcombe, D. (2003). An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *The Plant Journal : for cell and molecular biology*, *33*(5), 949–956.

Wallace, A. C., Laskowski, R. A., & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein engineering*, *8*(2), 127–134.

Wang, J., & Sampson, H. A. (2011). Food allergy. *Journal of Clinical Investigation*, *121*(3), 827–835.

Winkel, B. S. J. (2004). Metabolic Channeling In Plants. *Annual Review of Plant Biology*, *55*(1), 85–107.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., et al. (2011). *Acta Cryst. D67, 235-242*, 1–8.

Wyatt, P. (1998). Submicrometer Particle Sizing by Multiangle Light Scattering following Fractionation. *Journal of colloid and interface science*, *197*(1), 9–20.

Yin, R., Messner, B., Faus-Kessler, T., Hoffmann, T., Schwab, W., Hajirezaei, M. R., et al. (2012). Feedback inhibition of the general phenylpropanoid and flavonol biosynthetic pathways upon a compromised flavonol-3-O-glycosylation. *Journal of Experimental Botany*, *63*(7), 2465–2478.

Zaborsky, N., Brunner, M., Wallner, M., Himly, M., Karl, T., Schwarzenbacher, R., et al. (2010). Antigen Aggregation Decides the Fate of the Allergic Immune Response. *The Journal of Immunology*, *184*(2), 725–735.

Zhang, X., Henriques, R., Lin, S.-S., Niu, Q.-W., & Chua, N.-H. (2006). Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. *Nature Protocols*, *1*(2), 641–646.

Zhang, Y., Seeram, N. P., Lee, R., Feng, L., & Heber, D. (2008). Isolation and Identification of Strawberry Phenolics with Antioxidant and Human Cancer Cell Antiproliferative Properties. *Journal of Agricultural and Food Chemistry*, *56*(3), 670–675.

Zhao, J., & Dixon, R. A. (2009). MATE Transporters Facilitate Vacuolar Uptake of Epicatechin 3'-O-Glucoside for Proanthocyanidin Biosynthesis in Medicago truncatula and Arabidopsis. *The Plant Cell Online*, *21*(8), 2323–2340.

Zhao, J., & Dixon, R. A. (2010). The 'ins' and “outs” of flavonoid transport. *Trends in Plant Science*, *15*(2), 72–80.

Zhao, J., Huhman, D., Shadle, G., He, X. Z., Sumner, L. W., Tang, Y., & Dixon, R. A. (2011). MATE2 Mediates Vacuolar Sequestration of Flavonoid Glycosides and Glycoside Malonates in Medicago truncatula. *The Plant Cell Online*, *23*(4), 1536–1555.

Zifkin, M., Jin, A., Ozga, J. A., Zaharia, L. I., Schernthaner, J. P., Gesell, A., et al. (2012). Gene Expression and Metabolite Profiling of Developing Highbush Blueberry Fruit Indicates Transcriptional Regulation of Flavonoid Metabolism and Activation of Abscisic Acid Metabolism. *Plant Physiology*, *158*(1), 200–224.

Resumen

MOTIVACIÓN Y OBJETIVOS

La fresa (*Fragaria xananassa*) es un fruto altamente consumido y apreciado por su delicado sabor, aroma y valor nutritivo (Tulipani *et al.*, 2008; Giampieri *et al.*, 2012). El cultivo de fresa se ha extendido por todo el mundo ocupando en la actualidad una superficie de 254.000 hectáreas y una producción aproximada de 4 millones de toneladas con una distribución global del 42.8%, 29.2%, 18.4%, 8.8% and 0.8% en América, Europa, Asia, África y Oceanía, respectivamente (FAOSTAT, 2012; http://faostat3.fao.org). España ocupa el primer lugar mundial dentro de los exportadores de fresas y el cuarto en relación a producción (FAOSTAT, 2012), siendo la provincia de Huelva, comunidad autónoma de Andalucía, el principal productor, en donde se concentra aproximadamente el 95% de la producción nacional. España ocupa un puesto privilegiado no sólo en producción sino también en exportación y desarrollo de programas de mejora de variedades, lo que genera retornos importantes tanto en la comercialización de fruta en fresco como los generados por "royalties". En los últimos años, los programas de mejora se han centrado en caracteres tales como productividad y morfología, entre los que destacan el tamaño de fruto, la dureza y el color. Sin embargo, en la actualidad el consumidor está demandando frutos con mayor valor en compuestos saludables (conocidos como nutracéuticos) debido a los efectos beneficiosos de éstos en la salud, lo que influye en toda la cadena de distribución y producción. Este hecho ha producido un giro en la selección de nuevas variedades buscando, además de los aspectos agronómicos, una mejora en la calidad organoléptica y nutricional. Estos esfuerzos buscan, entre otras cosas, un incremento en la sostenibilidad y competitividad de este cultivo en España, que le permita seguir liderando las exportaciones y e incrementando su producción.

En este escenario, nuestro grupo lleva varios años liderando proyectos de investigación basados tanto en la identificación de genes implicados en la regulación del desarrollo y maduración del fruto de fresa como en la caracterización de herramientas biotecnológicas dirigidas a la mejora de dichos frutos. Uno de los principales intereses del grupo y objetivo de esta tesis doctoral, es el estudio y determinación de la función biológica de los alérgenos Fra de fresa (*Fragaria xananassa*), que se sabe están involucrados en el control de la ruta de biosíntesis de flavonoides (Muñoz *et al.*, 2010). Los flavonoides son compuestos que, además de ser responsables del color y aroma del fruto de fresa (Fait *et al.*, 2008; Muñoz *et al.*, 2011), son actualmente de gran interés biotecnológico debido a las propiedades nutricionales, farmacéuticas y medicinales que presentan para el consumo humano (Hichri *et al.*, 2011; Giampieri *et al.*, 2013). La identificación de estos alérgenos fue resultado de los estudios de Emanuelsson y colaboradores que, motivados por la observación de que ciertas variedades blancas de fresa eran bien toleradas por individuos

que presentaban reacciones alérgicas a variedades rojas, realizaron un estudio proteómico de variedades comerciales de ambos tipos de frutos (Karlsson *et al.*, 2004; Hjerno *et al.*, 2006). Finalmente, pusieron de manifiesto que las variedades blancas de fresa presentaban un contenido en la proteína FaFra1, potencial alérgeno de fresa, significativamente inferior al de las variedades rojas. Posteriormente, en nuestro laboratorio se identificaron dos genes codificantes de nuevas isoformas del alérgeno Fra (FaFra2 y FaFra3) y se estableció una relación entre la expresión de dichos genes y la ruta de biosíntesis de flavonoides que sugería un papel regulador de las proteínas FaFra en esta ruta biosintética (Muñoz *et al.*, 2010).

Las proteínas Fra de fresa pertenecen a la familia de proteínas PR-10 ("pathogenesis-related 10 proteins"), denominadas de esta forma por su relación con la respuesta de las plantas frente a patógenos (Markovic-Housley *et al.*, 2003). A su vez, las proteínas Fra, también forman parte de la superfamilia de proteínas Bet v 1, cuyo principal representante es el alérgeno del polen de abedul (que da nombre al grupo), y que engloba una variedad de alérgenos, no sólo de polen, sino también de frutas y alimentos (Markovic-Housley *et al.*, 2003; Fernandes *et al.*, 2013). Sin embargo, aunque las propiedades alergénicas de este tipo de proteínas han sido ampliamente estudiadas, sus funciones y mecanismos de actuación en las plantas que las producen permanecen prácticamente desconocidos (Mogensen *et al.*, 2007). Para investigar la función biológica desempeñada por los alérgenos FaFra en el fruto de fresa y poder establecer el papel que estas proteínas juegan en la regulación de la biosíntesis de flavonoides y, finalmente, obtener información sobre la funcion de las proteínas de tipo PR-10 ampliamente distribuidas en el reino vegetal, en este trabajo se abordaron los siguientes objetivos:

I. Caracterización de la familia multigénica *Fra* de fresa (*Fragaria xananassa*), para profundizar en el estudio de su efecto en el control de la ruta de biosíntesis de flavonoides.

II. Búsqueda de potenciales ligandos naturales de las proteínas FaFra mediante el empleo de técnicas bioquímicas y biofísicas.

III. Análisis estructural de las proteínas FaFra para obtener detalles sobre su mecanismo de actuación a nivel molecular.

INTRODUCCIÓN

La Fresa

La fresa pertenece a la familia Rosaceae y al género *Fragaria*. En *Fragaria*, existen cuatro grupos básicos de fertilidad que se asocian principalmente con el número de ploidía o número de cromosomas. La especie silvestre más común, *F. vesca* L., presenta 14 cromosomas, es diploide y sirve de modelo experimental para el género (Shulaev *et al.*, 2011). La fresa cultivada es, sin embargo, una especie octoploide de 56 cromosomas que procede del cruce de las especies *F. chiloensis* y *F. virginiana*, ambas especies octoploides (Rousseau-Gueutin *et al.*, 2009). Uno de los cultivares más ampliamente extendido en todo el mundo es "Camarosa", ideal para inviernos suaves, como es el caso de Huelva en España. Siendo esta la variedad donde hemos realizado mayoritariamente nuestras investigaciones.

De forma general, los frutos pueden clasificarse como climatéricos o no-climatéricos en función del patrón de respiración y la producción de etileno que tienen lugar a lo largo del proceso de maduración (Grierson, 2013). El fruto de fresa sirve como modelo para el estudio de los frutos no-climatéricos; en estos, no se observa un aumento drástico en la tasa de respiración durante la maduración ni tampoco una elevada producción de etileno (Iannetta *et al.*, 2006). La fresa es, además, un fruto muy particular ya que, a diferencia de otros frutos botánicamente definidos como resultado de la expansión del ovario, la fresa es en realidad un receptáculo floral engrosado, compuesto de una médula central o corazón, un receptáculo carnoso, la epidermis y un conjunto de haces vasculares que conectan los aquenios, que son los verdaderos frutos de la fresa. Los aquenios, que son una combinación de tejido del ovario y de la semilla y se originan en la base de cada pistilo, se disponen de forma helicoidal en la superficie del fruto y se encuentran insertados en la capa epidérmica del receptáculo (Perkins-Veazie, 1995).

Todo esto hace que el estudio del fruto de fresa sea especialmente interesante tanto desde el punto de vista genético como fisiológico.

Composición Fenólica del Fruto de Fresa

La fresa se caracteriza por ser un fruto con un alto contenido en metabolitos secundarios, muchos de ellos considerados como compuestos "bioactivos", ya que se ha demostrado que presentan efectos beneficiosos para la salud humana (Diamanti *et al.*, 2012; Giampieri *et al.*, 2013). En particular, las fresas son ricas en minerales, vitamina C, folato y compuestos fenólicos.

Los compuestos fenólicos no sólo poseen propiedades antioxidantes y anti-carcinogénicas (Scalzo *et al.*, 2005, Tulipani *et al.*, 2008; Giampieri *et al.*, 2013) sino que además, juegan un papel esencial en la biología del fruto, puesto que no sólo son los principales responsables del color y sabor de la fresa, sino que también desempeñan funciones tan importantes como la defensa frente a patógenos y a condiciones ambientales adversas, como puede ser la exposición a la radiación ultravioleta (Aaby *et al.*, 2005, 2007).

Los compuestos fenólicos se sintetizan a partir de L-fenilalanina a través de las rutas de biosíntesis de fenilpropanoides y flavonoides (Tohge *et al.*, 2013). Entre la gran variedad de compuestos fenólicos presentes en la fresa, los mayoritarios son los flavonoides. La síntesis de estos compuestos tiene lugar en dos fases durante el desarrollo del fruto de fresa, lo cual está íntimamente relacionado con la función que cada uno de estos metabolitos desempeña a lo largo de la maduración (Halbwirth *et al.*, 2006; Aaby *et al.*, 2007; Fait *et al.*, 2008). Durante las primeras etapas del desarrollo del fruto, los flavonoides principalmente acumulados en la fresa son de tipo flavan-3-ols y proantocianidinas; estos metabolitos, son responsables del sabor astringente de los frutos inmaduros. En las etapas más tardías del desarrollo, cuando el fruto comienza a madurar, se incrementa la síntesis de otros compuestos flavonoides, entre ellos, antocianinas y flavonoles, que contribuyen a la coloración de frutos y flores (Halbwirth *et al.*, 2006, Fait *et al.*, 2008). Las antocianinas son, cuantitativamente, los polifenoles más importantes en el fruto de fresa maduro y están principalmente representados por derivados glucosilados de pelargonidina y cianidina; estos metabolitos son los pigmentos que confieren a la fresa su color característico (Tulipani *et al.*, 2008; Muñoz *et al.*, 2011).

De este modo, la composición fenólica de la fresa no sólo afecta a su valor nutricional sino también a la calidad final del fruto, lo que actualmente se traduce en la existencia de muchas líneas de investigación enfocadas al estudio del desarrollo y maduración del fruto de fresa, incluyendo la síntesis de flavonoides, lo que hace que éste sea un ámbito científico muy relevante.

Proteínas PR-10

Las proteínas relacionadas con patogénesis PR ("Pathogenesis-Related proteins") están ampliamente distribuidas en el reino vegetal y se clasifican en 17 familias (PR1-PR17) en función de su estructura primaria y las propiedades biológicas y bioquímicas que presentan (van Loon *et al.*, 2006; Sels *et al.*, 2008; Fernandes *et al.*, 2013). En un inicio, las proteínas PR se clasificaron en base a la respuesta que inducían frente a patógenos. Sin embargo, más tarde, este

término se amplió y se incluyeron en esta familia aquellas proteínas no sólo relacionadas con la defensa a patógenos, sino también aquellas relacionadas con la respuesta a condiciones que mimetizaban el efecto frente a agentes patogénicos, como el estrés abiótico o el daño mecánico (Sels *et al.*, 2008; Lebel *et al.*, 2010). A pesar de que la familia de proteínas PR ha sido ampliamente estudiada, la función de la mayoría de ellas sigue siendo desconocida y, en particular, el papel biológico de las proteínas PR-10 no está bien definido.

Entre las proteínas de tipo PR-10 se encuentra un grupo de alérgenos que pertenece a la superfamilia de proteínas Bet v 1 (Markovic-Housley *et al.*, 2003). Las proteínas Bet v 1 se caracterizan por presentar una estructura muy conservada que encierra una cavidad central que es capaz de unir ligandos hidrofóbicos (Mogensen *et al.*, 2002; Markovic-Housley *et al.*, 2003; Seutter von Loetzen *et al.*, 2013). Actualmente en fresa se conocen tres miembros de la familia Bet v 1: FaFra1, FaFra2 y FaFra3 (Hjerno *et al.*, 2006; Muñoz *et al.*, 2010). Se sabe que las proteínas FaFra están relacionadas con la formación de compuestos coloreados en la fresa (Muñoz *et al.*, 2010), lo que principalmente depende de la producción de ciertos flavonoides como el cianidin-3-O-glucósido y el pelargonidin-3-O-glucósido. Estudios previos desarrollados en nuestro laboratorio (Muñoz *et al.*, 2010), permitieron establecer una relación directa entre las proteínas FaFra y la ruta de biosíntesis de flavonoides. El silenciamiento transitorio (mediado por RNAi) de los genes *FaFra* en frutos de fresa (cv. Elsanta), puso de manifiesto que el contenido de aquellos metabolitos responsables del color del fruto (cianidin-3-O-glucósido y pelargonidin-3-O-glucósido) se veía reducido en los frutos silenciados. Además de la acumulación de estos compuestos, también se vio afectada la de otros como el kaempferol-3-O-glucósido y el pelargonidin-3-malonil-glucósido que mostraron niveles reducidos en comparación a los frutos control. Sin embargo, no sólo se observó la disminución en la acumulación de estos compuestos sino que también otros metabolitos, como la catequina y ciertas proantocianidinas, aumentaron sus niveles en los frutos silenciados (Muñoz *et al.*, 2010). De forma paralela, se observó que el silenciamiento de los genes *FaFra* tenía un efecto a nivel transcripcional sobre genes fundamentales de la ruta de biosíntesis de flavonoides, específicamente, fenilalanin-amonio-liasa (*PAL*) y chalcona sintasa (*CHS*) aparecían co-silenciados junto a los alérgenos FaFra (Muñoz *et al.*, 2010). Estos resultados indicaban la posibilidad de un papel regulador de las proteínas FaFra en la ruta de biosíntesis de flavonoides. Sin embargo, se carece de información sobre el mecanismo molecular por el cual estas proteínas podrían estar ejerciendo su función.

La estructura conservada de las proteínas PR-10 y su habilidad para unir distintos tipos de ligandos en su cavidad hidrofóbica central indican que la función de la proteínas FaFra en la ruta de biosíntesis de flavonoides podría estar íntimamente relacionada con su unión a un

compuesto flavonoide y que, cabría la posibilidad de que cada uno de los miembros de la familia pudiesen unir distintos intermediarios de la ruta. Debido a la presencia de cavidades conservadas en las proteínas PR-10 y su capacidad de unir ligandos, se han propuesto varias funciones para esta familia de proteínas y, en particular para las proteínas de la familia Bet v 1, se han planteado principalmente dos posibilidades: un posible papel como transportadores celulares de metabolitos y una posible función como componentes esenciales en cascadas de señalización (Mogensen *et al.*, 2002; Liu y Ekramoddoullah, 2006; Radauer *et al.*, 2006; Mogensen *et al.*, 2007).

Mediante la caracterización de la familia multigénica *FaFra* y el estudio bioquímico, biofísico y estructural de las proteínas FaFra en presencia y ausencia de intermediarios de la ruta de biosíntesis de flavonoides, este trabajo tiene como objetivo principal aclarar el papel de las proteínas FaFra de fresa en la regulación de la ruta de biosíntesis de antocianos. Este trabajo debería también contribuir al el entendimiento general del la función biológica de las proteínas PR-10, que se encuentran ampliamente distribuidas en el reino vegetal.

RESULTADOS Y DISCUSIÓN

I. Caracterización de la Familia Multigénica *Fra* de *Fragaria xananassa*

Recientemente ha sido publicado el genoma de la especie *Fragaria vesca* (*F. vesca*) que sirve como modelo para el género *Fragaria* (Shulaev *et al.*, 2011). Esto ha generado un conjunto de herramientas muy valiosas que, en nuestro caso, han sido muy útiles para profundizar en el estudio de la familia de genes *Fra* de *Fragaria xananassa* (*FaFra*). Por otro lado, los grandes avances alcanzados en los últimos años en el campo de la secuenciación masiva de RNA (RNA-Seq), han hecho que el RNA-Seq se haya convertido en un método muy efectivo para obtener grandes cantidades de datos transcriptómicos de muchos organismos y distintos tipos de tejidos y que, además, se considere como una potente herramienta para estimar tanto la abundancia de los genes expresados en una circunstancia determinada como su expresión diferencial en distintas condiciones de estudio (Grabherr *et al.*, 2010; Trapnell *et al.*, 2010).

En base a estos avances científicos y tecnológicos y para caracterizar la familia multigénica *FaFra*, se plantearon dos estrategias metodológicas principales, ambas basadas en la secuenciación masiva de RNA mensajero (RNA-Seq) extraído de plantas de fresa. La primera de ellas tenía como objetivos principales la identificación de las secuencias *FaFra* expresadas en fresa y la determinación de sus perfiles de expresión. Para ello, se recurrió a una experiencia de

RNA-Seq que nuestro laboratorio ha desarrollado en plantas pertenecientes al cultivar Camarosa, no solo a lo largo de la maduración de los frutos de fresa en los estadios verde, blanco, viraje blanco-rojo y rojo (en aquenios y receptáculos por separado) sino también en tejidos vegetativos (hoja y raíz). El análisis de datos de RNA-Seq puede llevarse a cabo mediante dos metodologías principales: 1) Alineamiento de las lecturas a un genoma de referencia; 2) Ensamblado del transcriptoma *de novo* (Martin *et al.*, 2013). Aunque el genoma de la especie *F. vesca* (diploide) sirve de modelo para el género *Fragaria*, y el genoma de la especie cultivada *Fragaria xananassa* (octoploide) presenta un alto grado de conservación con el de la especie salvaje (Bombarely *et al.*, 2010), el hecho de que el genoma de *F. vesca* es aún incompleto y su anotación necesita mejoras, nos llevó a analizar los datos de RNA-Seq de nuestra experiencia de maduración y tejidos vegetativos de fresa utilizando un ensamblado del transcriptoma *de novo*, utilizando el método Trinity (Grabherr *et al.*, 2010). El análisis del transcriptoma realizado por nuestro grupo, en colaboración con el Dr. José. F. Sánchez Sevilla del centro IFAPA de Churriana (Málaga), ha permitido obtener una ingente cantidad de datos de expresión de genes. El análisis de los datos de RNA-Seq ha resultado en la identificación de 10 genes putativos principales que conforman la familia *Fra* (*Fragaria xananassa*) y ha permitido establecer sus correspondientes homólogos en *F. vesca*. Los genes *FaFra*, se han nombrado de forma consecutiva desde *FaFra1* a *FaFra10* (Tabla 1, Figuras 4-8, Capítulo 1). Además, se han identificado variedades alélicas para muchos de ellos, en particular, *FaFra1c1-c2*, *FaFra2b*, *FaFra4a1-a2* y *FaFra4b1-b2*, *FaFra7a-b* y *FaFra9a1-a2*. A su vez, las proteínas FaFra se han clasificado en cuatro grupos principales, de acuerdo a su nivel de homología con los alérgenos FaFra1, FaFra2 y FaFra3 previamente publicados (Muñoz *et al.*, 2010); el cuarto grupo, sin embargo, engloba las secuencias FaFra8, FaFra9 y FaFra10 que conforman el conjunto más divergente de proteínas FaFra (Figuras 9 y 10, Capítulo 1). Adicionalmente, se ha obtenido información detallada sobre los niveles de expresión de los genes *FaFra* tanto en tejidos vegetativos de fresa como a lo largo de los estadios de maduración del fruto, siendo los niveles de mensajeros observados distintos para cada una de los genes identificados. De forma general, los transcritos más representados corresponden a los genes *FaFra1c1*, *FaFra1c2*, *FaFra2b* y *FaFra4a1* (Tabla 2, Figura 11-12, Capítulo 1). El tejido donde la expresión de *FaFra* resultó ser más importante fue la raíz, seguido de receptáculo, donde los genes más representativos fueron *FaFra1c1* y *FaFra2b*; estos genes, mostraron perfiles de expresión complementaria ya que, mientras que *FaFra1c1* disminuía a lo largo de maduración, *FaFra2b* aumentaba su expresión (Figura 11d, Capítulo 1). Además, los niveles de expresión de estos genes, se corresponden con los niveles de proteína observados mediante Western-blot en los frutos de fresa (Figura 3, Capítulo 1).

La segunda experiencia de RNA-Seq, se diseñó con el fin de identificar aquellos genes que se expresaban de forma diferencial en receptáculos control y receptáculos donde se había silenciado de forma transitoria los genes *FaFra* inyectados con soluciones de *Agrobacterium tumefaciens* portadoras de construcciones *RNAi* del gen diana (Hoffmann *et al.*, 2006; Muñoz *et al.*, 2010). En este caso, el análisis de datos de RNA-Seq se llevó a cabo mapeando las lecturas obtenidas al genoma de referencia *F. vesca* utilizando los programas TopHat2 para el mapeado de secuencias (Kim *et al.*, 2013) y Cufflinks para la cuantificación de lecturas y determinación de la expresión diferencial de genes entre los frutos control y los frutos *Fra-RNAi* silenciados (Trapnell *et al.*, 2012). Como resultado, 4426 genes se vieron afectados, de los cuales 2528 resultaron sobre-expresados y 1898 reprimidos. Enfocando nuestra atención en la familia de genes *Fra* y en los genes involucrados en la biosíntesis de flavonoides, concluimos que el silenciamiento de *FaFra2A* (principal gen *FaFra* en receptáculo rojo) y *FaFra6*, afectaba a genes primordiales de la ruta de biosíntesis de flavonoides tales como fenilalanina-amonio liasa (*PAL*), chalcona sintasa (*CHS*), cinamato-4-hidroxilasa (*C4H*), 4-cumarato-CoA-ligasa (*4CL*), chalcona isomerasa (*CHI*), dihidroflavonol-reductasa (*DFR*), flavanona-3-hidroxilasa (*F3H*) y 3-O-glucosiltransferasa (*3-GT*), algunos de los cuales se verían sobre-expresados y otros silenciados, siendo los reprimidos los más importantes en cuanto a expresión en receptáculo. Asimismo, los factores de transcripción de tipo R2R3-Myb, *FaMYB1* y *FAMYB10*, aparecieron co-silenciados junto a *FaFra2A* y *FaFra6*. Nuestros resultados, de este modo, confirman el papel esencial regulador que ejercen los genes *FaFra* en la ruta de biosíntesis de flavonoides.

Finalmente, para obtener información sobre la localización subcelular de las proteínas FaFra, se llevaron a cabo ensayos de microscopía confocal en plantas de *Nicotiana benthamiana* (expresión transitoria) y de *Arabidopsis thaliana* (generación de líneas transgénicas) expresando proteínas FaFra a fusionadas a la proteína GFP (Fra-GFP). Los ensayos de localización subcelular de las proteínas Fra-GFP en *N. benthamiana* y *A. thaliana*, indicaron que las proteínas FaFra, que presentan una localización similar a la proteína GFP libre (nuclear y citoplasmática), podrían tener una localización citoplasmática, al igual que ocurre en la mayoría de las proteínas de tipo PR-10 (Fernandes *et al.*, 2013) sin poder descartar que su localización celular tenga lugar en otro orgánulo (como retículo endoplasmático) y que ésta pudiese depender de su interacción con otras proteínas presentes en la fresa.

II. y III. Búsqueda del Ligando Natural de las Proteínas FaFra y Análisis Estructural

La proteínas de la superfamilia START/Bet v 1 se caracterizan por presentar un plegamiento común formado por siete láminas beta antiparalelas y tres hélices alfa que encierran una cavidad central que es capaz de alojar ligandos hidrofóbicos (Mogensen *et al.*, 2002; Radauer *et al.*, 2008; Fernandes *et al.*, 2013). Nuestra hipótesis inicial planteaba que las proteínas FaFra, que presentan distintos perfiles de expresión a lo largo de la maduración del fruto y tejidos vegetativos de fresa, debían ejercer su función reguladora en la ruta de biosíntesis de flavonoides a través de la unión específica de ligandos de dicha ruta. De este modo, para obtener información sobre el mecanismo de actuación de las proteínas FaFra a nivel molecular, se llevó a cabo la clonación, expresión, purificación y caracterización de las proteínas FaFra1E, FaFra2 y FaFra3. La secuencias codificantes (CDS) de las proteínas FaFra se amplificaron mediante PCR y se clonaron en el vector de expresión pETM11 (Dummler *et al.*, 2005), que introduce una etiqueta hexa-histidina (6xHis-tag) y que presenta un sitio específico de reconocimiento de la proteasa TEV ("tobacco etch virus", virus jaspeado del tabaco) que permite la eliminación de dicha etiqueta de histidina tras la purificación de la proteína. Para llevar a cabo la expresión de las proteínas, las construcciones *FaFra*-pETM11 obtenidas se clonaron en *E. coli* BL21 (DE3). Para inducir la expresión de las proteínas se empleó IPTG (isopropil-β-D-1-tiogalactopiranósido) y la purificación se llevó a cabo mediante cromatografía de afinidad de metales inmovilizados, utilizando una resina de ácido nitrilotriacético de níquel (Ni-NTA) en dos pasos: en el primero se purificó la proteína haciendo uso de su etiqueta de histidina y en el segundo se eliminó la proteasa TEV utilizada para cortar dicha etiqueta. Tras el corte de la cola de histidina, tres aminoácidos adicionales (Ala- Met- Ala) quedaron unidos en el extremo N-terminal de las proteínas FaFra. La caracterización de las proteínas heterólogas obtenidas se realizó mediante cromatografía de exclusión por tamaño acoplada a dispersión multiangular de luz láser (SEC-MALLS). Los experimentos de SEC-MALLS indicaron masas moleculares de 18 kDa, 28 kDa y 17.5 kDa para FaFra1E, FaFra2 y FaFra3, respectivamente. En el caso de FaFra1E y FaFra3, los valores obtenidos concordaban con las masas moleculares esperadas, lo que demostraba que ambas proteínas eran monoméricas en solución. Sin embargo, la masa molecular estimada para FaFra2 era superior a lo esperado (17.5 kDa), lo que al final resultó bastante importante ya que, los ensayos posteriores de cristalización, no permitieron obtener cristales de esta proteína en ninguna de las condiciones ensayadas.

Con el objeto de identificar el ligando natural de estas proteínas, se desarrollaron una serie de ensayos de interacción de ligando mediante calorimetría isotérmica de titulación (ITC), en los

que se analizaron un conjunto de compuestos pertenecientes a la ruta de biosíntesis de fenilpropanoides y flavonoides, disponibles comercialmente y muchos de ellos potenciales ligandos candidatos por tratarse de metabolitos ampliamente representados en el receptáculo de fresa y por el hecho de que la acumulación de algunos de ellos en la fresa se había visto afectada en los frutos *FaFra-RNAi* previamente silenciados en nuestro laboratorio (Muñoz *et al.*, 2010). Los ensayos de interacción con ligandos mediante ITC revelaron que, efectivamente, las proteínas FaFra son capaces no sólo de unir intermediarios de la ruta de flavonoides, sino que presentan la capacidad de interaccionar con diferentes metabolitos de la ruta con distinta afinidad en el rango micromolar. Específicamente, se demostró que FaFra1E, FaFra2 y FaFra3 unian quercetin-3-O-glucurónido, miricetina y catequina, respectivamente (Figura 2, Capítulo 3). Además, se sabe que estos metabolitos se acumulan en la fresa (Kosar *et al.*, 2004; Aaby *et al.*, 2007; Hanhineva *et al.*, 2008; Sultana y Anwar, 2008; Muñoz *et al.*, 2011), en estadíos de desarrollo y tejidos similares en los que se registran los valores máximos de expresión de los correspondientes genes *FaFra*. De forma paralela al análisis ITC se llevaron a cabo la cristalización y determinación estructural de las proteínas FaFra1E y FaFra3, lo que ha generado una gran cantidad de información estructural y mecanística. Tras la obtención de cristales y experimentos de diffracion de rayos X en el synchotron de Grenoble, se determinaron tres estructuras cristalográficas por el método de reemplazamiento molecular (MR): dos correspondientes a FaFra1E en su forma apo y una estructura para el complejo FaFra3-catequina, con resoluciones de 2.2 Å, 3.1 Å y 3.0 Å, respectivamente (Tabla 1, Capítulo 3). Las estructuras obtenidas para FaFra1E nos permitieron determinar que, efectivamente, el plegamiento esperado estaba conservado en las proteínas FaFra (Figura 3, Capítulo 3). Además, se identificaron regiones altamente desordenadas (giros L3, L5 y L7), que indicaban un alto grado de flexibilidad conformacional para estas proteínas. La determinación y análisis estructural del complejo FaFra3-catequina, resultó esencial para para establecer definitivamente la capacidad de las proteínas FaFra de unir flavonoides naturales asi como para comprender el mecanismo de actuación de esta familia de proteínas. La unión del metabolito en la cavidad central de FaFra3, daba lugar a un cambio conformacional en la proteína esencial para la estabilidad del complejo formado (Figura 4, Capítulo 3). Los giros altamente flexibles observados en las formas apo de FaFra1E, adoptaban en este complejo una conformación única, curvándose hacia la cavidad central para adoptar una conformación cerrada que estabiliza el ligando en el interior de la cavidad. El plegamiento global se veía afectado también por el cambio conformacional de la hélice α3, cuya contribución daba lugar a una estructura general más compacta de la proteína en presencia de ligando. Uno de los principales resultados de este análisis ha sido la identificación de los residuos aminoacídicos involucrados en la unión de

ligandos en la cavidad central de los alérgenos así como los residuos esenciales implicados en la formación de dicha cavidad. Esto ha permitido caracterizar no sólo aquellos amino ácidos fundamentales en la unión de la catequina a FaFra3 y aquellos conformando la cavidad, sino también aquellos residuos variables entre las distintas proteínas (Figura 5, Capítulo 3), que podrían ser la respuesta a la diferencia en especificidad de ligando observada para estas proteínas.

En resumen, los experimentos de interacción de las proteínas FaFra1E, FaFra2 y FaFra3 con distintos compuestos flavonoides de la ruta de biosíntesis de antocianos, ha permitido identificar un ligando natural y diferente para cada una de estas proteínas, lo que sugiere que cada una de las proteínas FaFra podría estar desempeñando una función diferente en el control de esta ruta metabólica. La cristalización y determinación estructural de FaFra1E y FaFra3 en complejo con el ligando flavonoide identificado (catequina), ha revelado una información muy valiosa sobre el posible mecanismo de actuación de estas proteínas. Los resultados obtenidos, ponen de manifiesto una vez más la relación existente entre las proteínas FaFra y la ruta de biosíntesis de flavonoides y sugieren dos hipótesis principales sobre la función de FaFra en esta ruta metabólica a nivel molecular. Por un lado, las proteínas FaFra podrían funcionar como transportadores de metabolitos, trasladando compuestos flavonoides desde su lugar de síntesis (retículo endoplasmático) hasta su destino final, generalmente la vacuola o la pared celular. Se sabe que las principales enzimas de la ruta de biosíntesis de fenilpropanoides y flavonoides co-localizan en la cara citosólica del retículo endoplasmático de muchas especies vegetales diferentes (Winkel, 2004). De este modo, FaFra podría ser un componente esencial de estos complejos multiproteicos responsables de la biosíntesis de flavonoides, donde estarían facilitando la disponibilidad de los distintos intermediarios metabólicos, para su uso por las enzimas procesadoras de la ruta, en el momento necesario. Por otro lado, las proteínas FaFra parecen ejercer un papel regulador de la ruta de biosíntesis de flavonoides a nivel transcripcional y podrian funcionar como componentes tempranos de una cascada de señalización. El hecho de que el silenciamiento de FaFra afecte tanto a la expresión de genes esenciales de la ruta de biosíntesis de flavonoides (*PAL*, *CHS*, *C4H*, *4CL*, *CHI*, *DRF*, *F3H*, *3-GT*) como a la expresión de factores de transcripción involucrados en el control de la misma (*FaMYB1*, *FaMYB10*), sumado al hecho de que estas proteínas son capaces de unir distintos compuestos flavonoides, sugiere que las proteínas FaFra podrían estar regulando el flujo de metabolitos en la ruta, lo que se traduciría en una regulación transcripcional indirecta de los genes principales de la ruta. De forma global, este estudio propone un papel para las proteínas PR-10 de control del metabolismo secundario en plantas que no había sido descrito hasta el momento.

CONCLUSIONES GENERALES

1. Análisis del transcriptoma de fresa (*Fragaria xananassa*) a lo largo del desarrollo y maduración ha permitido la identificación de 10 genes *FaFra* putativos, y un número de variedades alélicas.

2. El análisis del perfil de expresión de los transcritos *FaFra* indica que cada uno de los miembros de la familia presenta un patrón de expresión diferencial en los distintos tejidos vegetativos estudiados y a lo largo de la maduración del fruto, lo que apunta hacia la posibilidad de que las proteínas FaFra puedan presentar especificidad de unión de ligando y funciones diferenciales.

3. El silenciamiento de los genes *FaFra2A* y *FaFra6* en la fresa conlleva la represión de genes que codifican enzimas principales de la ruta de biosíntesis de flavonoides y de factores de transcripción de tipo MYB. Este resultado, respalda el papel de regulación que las proteínas FaFra ejercen en la ruta de biosíntesis de flavonoides.

4. Las proteínas FaFra1E, FaFra2 y FaFra3 son capaces de unir de forma selectiva distintos flavonoides naturales con diferente afinidad en el rango micromolar.

5. La unión de flavonoides naturales en la cavidad hidrofóbica de las proteínas FaFra induce una serie de cambios conformacionales que estabilizan el complejo FaFra-ligando. Estos cambios conformacionales podrían tener relevancia funcional.

6. La variabilidad de secuencia observada en el sitio de unión de la cavidad central en las proteínas FaFra, podría ser la explicación a la distinta selectividad de ligandos determinada para cada una de las proteínas. Por primera vez, nuestros resultados proporcionan detalles mecanísticos sobre la función de las proteínas FaFra en el control de la ruta de biosíntesis de flavonoides.