UNIVERSIDAD DE MURCIA

FACULTAD DE BIOLOGÍA

Metabolism of cyanogenic compounds and development of molecular markers for bitterness in almond *[Prunus dulcis* (Miller) D. A. Webb].

Metabolismo de los compuestos cianogénicos y desarrollo de marcadores moleculares para el amargor en el almendro *[Prunus dulcis* (Miller) D. A. Webb].

D. Jorge Luis del Cueto Chocano

2016

Metabolism of cyanogenic compounds and development of molecular markers for bitterness in almond [*Prunus dulcis* (Miller) D. A. Webb].

Metabolismo de los compuestos cianogénicos y desarrollo de marcadores moleculares para el amargor en el almendro [*Prunus dulcis* (Miller) D. A. Webb].

D. Jorge Luis del Cueto Chocano

Ingeniero Agrónomo

Murcia, 2016

Directores:

Dra. Raquel Sánchez Pérez

Plant Biochemistry Section,

PLEN Depart. (Univ. of Copenhagen)

Dr. Federico Dicenta López-Higuera

Dpto. Mejora Vegetal (CEBAS-CSIC)

Agradecimientos

Quiero agradecer a todas aquellas personas e instituciones que me han apoyado y ayudado en la realización de esta tesis doctoral. Sin ellos no hubiera sido posible dar este paso tan importante en mi vida.

En primer lugar quiero dar gracias a mis padres, por darme una excelente educación y valores, por permitirme estudiar y apoyarme en todo momento. Gracias a mi madre por ser mucho más que una madre, más bien una amiga, por su amor, su trabajo y por estar siempre a mi lado. Gracias a mi padre por su exigencia y por ser un ejemplo todos los días. Gracias a mi hermano Javi por su amistad y por su gran humanidad.

Gracias a Raquel Sánchez Pérez por ser mi guía durante toda la tesis, por su ayuda y paciencia, por sus consejos, por sus grandes conocimientos y por su competencia. He aprendido mucho contigo.

Gracias a Federico Dicenta por apostar por mí para realizar esta tesis doctoral, por su ayuda y consejos y por su tranquilidad en los momentos difíciles. Gracias por permitirme estar en este lugar fantástico que es el Cebas, he sentido que he formado parte de un gran grupo tanto a nivel humano como de trabajo. Por algo somos "líderes mundiales".

Gracias al Consejo Superior de Investigaciones Científicas (CSIC) por la beca recibida y la financiación para mi estancia en la Universidad de Copenhague (Dinamarca) y al Ministerio de Economía y Competitividad por el proyecto "Mejora genética del almendro" (AGL2010-22197-C02-02) y a la Fundación Seneca por el proyecto "Breeding stone fruit species assisted by molecular tools", con los que he financiado las investigaciones conducentes a esta Tesis Doctoral.

Gracias a Manolo, Tere, Eva, Juan, Ángela, Eli, Bea y Amine, compañeros y amigos de fatigas en los que me he apoyado todos estos años y que me han ayudado muchísimo. Por tantas risas, tan buenos momentos, por las riñas siempre para mejorar. En definitiva he dado con grandes personas de las que he aprendido mucho.

Gracias a Pedro, Encarna, David y Pepe por su ayuda y simpatía durante todos estos años. Gracias también a todo el personal de la finca de Santomera.

Gracias a Pedro Díaz, Félix, Gregorio, Chu, Aingeru, Puerto, Cristina, Alfonso, Enas, Pepito, Ana, Silvia, María y Sandra por su amistad durante estos años.

Gracias a Elena por su gran ayuda e importancia en que haya podido llevar a cabo esta tesis doctoral.

Gracias a todo el personal del Cebas, especialmente a Cari, Juan, Jesús, Alberto, Irene, Pepi y Gonzalo, que me han hecho sentir como en casa.

Gracias al Dr. Birger Lindberg Møller por permitirme desarrollar mi estancia en la Universidad de Copenhague, por sus grandes conocimientos y por su simpatía. Ha sido un honor trabajar en su grupo.

Gracias a la Dra. Kirsten Jørgensen por su amabilidad, y por permitirme también estar en su fantástico departamento.

Gracias a todos mis compañeros de Dinamarca, por su grandísima ayuda y su gran educación, especialmente a Daniela, Martina, Lizzie, Adam, Sarah, Rubini, Irina, Fred, Stefan, Soren, Nanna, Pernille y Ditte.

Gracias a todo el equipo de la Universidad de Bari, a Francesca, Stefano y Titti por su gran colaboración en el tramo final de la tesis y por su predisposición.

Gracias a mis amigos los peligrosos que siempre han estado ahí…Tayo, Pellús, Rubén, Chato y José Luis y a mi tío Fernando por preocuparse siempre por mis estudios.

Gracias a mi familia de Cph con los que pasé momentos inolvidables…Aldo, Ali, Diego, Sofia, Annelies, Estelle, Marta, Dia, Regina, Sarah y Carmen.

Gracias a Bacôa por su gran ayuda, comprensión y apoyo durante mi último duro año final de tesis, te llevo en mi corazón siempre.

Dedicado especialmente a la memoria de mi abuela Amalia.

General Index **Page**

Tables Index **Page**

Figures Index **Page**

Resumen

Resumen

Introducción

El almendro (*Prunus dulcis* Miller) es una especie del género *Prunus* originario de Asia Central, que fue domesticado hace unos 4.000 ó 5.000 años. Desde el siglo III A.C. fue distribuido por la cuenca mediterránea por griegos, romanos y árabes, y en el siglo XVIII fue llevado a California. Es uno de los frutales de hueso más importantes tanto en producción como en superficie cultivada, siendo Estados Unidos el principal productor mundial. España ocupa el tercer puesto y la Región de Murcia es una de las principales productoras.

A pesar de que en cada región la producción del almendro puede estar limitada por diferentes factores, los problemas de cuajado de frutos derivados de su incompatibilidad floral y los daños por helada consecuencia de su floración temprana, son cuestiones importantes en la mayoría de los países productores. Con el objetivo de resolver estos problemas, a lo largo del tiempo se han desarrollado varios programas de mejora genética principalmente en Estados Unidos, Francia, España y Australia, que persiguen objetivos tales como la elevada productividad, la floración tardía y la auto-compatibilidad floral. De estos programas han surgido nuevas variedades que superan a las tradicionales, incrementando la productividad y la rentabilidad de las explotaciones.

Mientras que en EEUU y Australia, el cultivo del almendro ha sido tradicionalmente muy tecnificado y con elevadas producciones, en la cuenca mediterránea y otros países, se ha llevado a cabo de manera marginal, normalmente en secano, con pocos cuidados y bajas producciones. Recientemente, el notable incremento del precio de la almendra ha disparado el interés por el cultivo del almendro, mejorándose las técnicas de cultivo y aumentándose notablemente la superficie cultivada y la productividad de las explotaciones.

Aunque la mayoría de las variedades de almendro son dulces, todos los almendros silvestres de los que procede la especie cultivada, eran amargos. El sabor dulce debió aparecer como una mutación, que posteriormente fue conservada por los

agricultores para el aprovechamiento de las semillas comestibles. En relación con este carácter, podemos establecer tres tipos de almendras: dulces, amargas y ligeramente amargas.

El sabor amargo de la almendra es producido por los glucósidos cianogénicos prunasina y la amigdalina. Estos metabolitos secundarios son compuestos de defensa de muchas plantas, entre las que se encuentran las rosáceas. La biosíntesis de estos compuestos comienza con un aminoácido, la fenilalanina, que a través de dos citocromos da lugar al mandelonitrilo. Éste es glucosilado por la glucosiltransferasa-1 para producir prunasina (mono-glucósido). La prunasina es posteriormente glucosilada por la glucosiltransferasa-2 para producir amigdalina (di-glucósido).

Durante su degradación, los glucósidos cianogénicos son hidrolizados por las β-glucosidasas. Así, la amigdalina es hidrolizada por la amigdalina hidrolasa para dar lugar a la prunasina, que a su vez es hidrolizada por la prunasina hidrolasa produciendo mandelonitrilo. El mandelonitrilo es hidrolizado por la mandelonitrilo liasa, dando lugar a benzaldehído (responsable del sabor amargo) y cianuro (tóxico). Los sustratos y enzimas involucrados en estas rutas metabólicas son inertes por separado, y sólo cuando los tejidos son dañados, se ponen en contacto, produciéndose la liberación del cianuro.

Respecto a su localización, la prunasina se halla presente en las partes vegetativas del almendro (raíces, tallos y hojas) y en los frutos amargos en desarrollo, mientras que la amigdalina está principalmente presente en los frutos amargos maduros. En el fruto, la prunasina parece ser sintetizada en el tegumento de la semilla y es transportada a los cotiledones donde es transformada en amigdalina.

El sabor amargo de la semilla del almendro es un carácter monogénico, cuyo gen responsable (*Sk*, sweet kernel) se encuentra ubicado en el grupo de ligamiento 5 (GL5) del genoma del almendro, aunque su localización precisa y función aún no ha sido determinadas. La mayoría de variedades comerciales de almendro son heterocigóticas para el sabor amargo, por lo que cuando se cruzan en los programas de mejora, el 25% de los descendientes dará frutos amargos. Actualmente, el

mejorador debe de esperar 3 ó 4 años a que los árboles entren en producción para determinar el sabor de las almendras de los descendientes creados. Ello supone el establecimiento y cultivo durante este tiempo de árboles que finalmente serán eliminados por su amargor, años después. Por ello sería de gran interés disponer de marcadores moleculares para este carácter, que permitieran la selección asistida por marcadores durante el primer año mediante una sencilla PCR, ahorrando tiempo y dinero, y mejorando la eficiencia de los programas de mejora.

Los objetivos de la presente tesis son el estudio de los compuestos cianogénicos durante el ciclo vegetativo y el desarrollo de un marcador molecular para el sabor amargo del almendro. Para ello se han seguido las siguientes estrategias:

- Estudio de la evolución de los glucósidos cianogénicos durante el desarrollo de la flor.
- Saturación con marcadores moleculares de la región del grupo de ligamiento 5 donde se encuentra el locus *Sk*.
- Análisis del trascriptoma de almendros dulces y amargos mediante el estudio de la expresión diferencial de genes candidatos.
- Caracterización de las enzimas de la ruta del sabor amargo y determinación de las diferencias entre variedades dulces y amargas.

Evolución de los compuestos cianogénicos desde el inicio de la salida del letargo hasta la floración

Este primer capítulo analiza la relación entre la época de floración y la presencia de compuestos cianogénicos en la flor. Para ello hemos estudiado la evolución de los glucósidos cianogénicos en flores desde el inicio de la salida del letargo invernal hasta la apertura de la flor, en variedades dulces y amargas de distinta época de floración, mediante LC-MS (Liquid chromatography–mass spectrometry).

La prunasina y la amigdalina fueron detectadas por primera vez en las yemas florales de las cinco variedades estudiadas, siendo la amigdalina el compuesto más importante durante la salida del letargo y la prunasina justo antes de la floración. En general, el contenido de prunasina fue 60 veces mayor que el de amigdalina, siendo siempre mayores en la variedad amarga. No se observó ninguna relación entre el contenido de glucósidos cianogénicos y la época de floración de las variedades estudiadas.

Estos resultados muestran que la prunasina podría tener un papel importante en la evolución de la yema floral y sugieren dos posibles hipótesis para explicar la función de estos compuestos en el desarrollo de la flor. La primera sería la detoxificación del cianuro a través de las nitrilasas, cuyos productos finales son aminoácidos y amoniaco, es decir nitrógeno, que podrían utilizarse para el desarrollo de la flor. La segunda sería la activación, por el cianuro liberado, de especies reactivas de oxígeno, que impulsarían los procesos implicados en la apertura de la flor.

Estos glucósidos cianogénicos también fueron detectados en distintas partes de la flor: sépalos, pétalos, pistilos y polen, siendo la concentración mayor en el cultivar amargo en todos los tejidos. El polen fue el tejido con mayor concentración de prunasina.

Por primera vez fueron observados en las flores algunos compuestos derivados de la prunasina (amidas, anitrilos, ácidos, pentosas). Todos parecen estar implicados en una ruta alternativa sin liberación de cianuro (el cual dañaría a la planta) y con la producción final de nitrógeno, reutilizable por la propia planta. Esta teoría ha sido propuesta por Pičmanová et al., (2015), los cuales llevaron a cabo un estudio similar pero estudiando la germinación de la semilla del almendro.

Saturación del *Sk* locus

El objetivo del segundo capítulo fue desarrollar un marcador molecular ligado al sabor amargo de la semilla mediante la saturación del locus *Sk*, que permitiera la

selección asistida por marcadores en los programas de mejora. Para ello se desarrollaron tres estrategias complementarias: 1) saturación del locus *Sk* con nuevos marcadores tipo SSRs y CAPs basados en SNPs; 2) re-secuenciación de dos genotipos de almendro, uno dulce y otro amargo; 3) análisis del trascriptoma del tegumento de los dos genotipos mencionados, en dos tiempos del desarrollo del fruto, mediante RNA-seq.

El primer objetivo fue completar los trabajos de saturación del GL5 realizados por Sánchez Pérez et al. (2007, 2010) y Koepke et al. (2013). Para ello se utilizaron marcadores SSRs y CAPs en 550 descendientes de R1000 x Desmayo Largueta, que cubrieron una región de 800 kb. Gracias a la colinearidad entre el melocotón y el almendro, en el GL5 del melocotón se detectaron once genes candidatos en una región de 95.76 kb, seis de los cuales fueron factores de transcripción MYC, una glyoxal oxidasa, un citocromo P450, un alcohol O-acetiltransferasa, una metionina liasa y una hidrolasa. Con este trabajo hemos conseguido reducir la región del locus *Sk* desde los 3,6 Mb hasta los 95.76 kb.

En segundo objetivo fue realizar la re-secuenciación del genoma integrando los datos de los marcadores con el nuevo transcriptoma. Tomando el genotipo amargo como referencia, se anotaron 10 genes candidatos en el locus *Sk*.

El tercer objetivo consistió en profundizar en el estudio transcriptómico de estos genes candidatos mediante RNA-seq, revelando la existencia de dos grupos para el locus *Sk*. El primero estaba compuesto por tres miembros de la familia MYC de factores de transcripción y el segundo por enzimas que podrían tener una función en el metabolismo de la amigdalina (una hidrolasa, un citocromo P450 y una glioxal oxidasa). Este análisis transcriptómico mostró expresión diferencial de seis de los diez genes. Por último, en tres de los genes candidatos (dos factores de transcripción MYC y la glioxal oxidasa) fueron detectados SNPs. Algunas de estas mutaciones fueron "missense", dando lugar a una secuencia de aminoácidos diferente entre dulces y amargos, que podría cambiar la función de la proteína, y así ser responsable del sabor dulce o amargo.

Caracterización de enzimas del sabor amargo

El tercer capítulo de la tesis está dedicado al estudio de las enzimas implicadas en la ruta de los compuestos cianogénicos, para determinar su papel en el sabor dulce o amargo del almendro. Sabemos que todos los enzimas de la biosíntesis y degradación de los compuestos cianogénicos se encuentran tanto en los almendros amargos como en los dulces. La pregunta es ¿por qué las variedades amargas acumulan amigdalina mientras que las dulces no? Podríamos considerar dos estrategias, relacionadas con la ruta biosintética o la de degradación.

Trabajos anteriores realizados por Sánchez-Pérez et al (2012) demostraron que la acumulación de amigdalina en las semillas amargas podría depender del contenido de prunasina en el tegumento y de la localización (apoplasto o simplasto) de las prunasinas hidrolasas. En este tejido materno había sido detectada una acumulación de prunasina en los genotipos amargos, que no se observó en los dulces (Sánchez-Pérez et al., 2008).

Respecto a la ruta de degradación, nos centramos en las prunasinas hidrolasas, obteniendo las secuencias de putativas prunasinas hidrolasas en genotipos dulces y amargos, en tres tejidos diferentes (tegumento, nucela y cotiledón). Para ello, a partir de una prunasina hidrolasa identificada por Sánchez-Pérez et al. (2012), se desarrollaron cebadores y se llevaron a cabo RT-PCRs y clonaciones en *Escherichia coli*. Se observaron diferencias de un solo nucleótido (SNPs) entre algunas secuencias, pero estos SNPs no estuvieron relacionados con el sabor dulce o amargo. Estos genes se expresaron en *Agrobacterium tumefaciens* y, aunque se observó actividad prunasina hidrolasa y β-glucosidasa, tampoco se observaron diferencias entre dulces y amargos. La principal conclusión de este trabajo fue la caracterización de prunasinas hidrolasas en almendros dulces y amargos.

Por otro lado, respecto a las enzimas anabólicas relacionadas con el sabor amargo, cinco putativas glucosiltransferasas y un citocromo CYP71 fueron caracterizadas en el cotiledón de variedades amargas. Cuando el citocromo CYP7934 de eucalipto y el CYP71AN24 de almendro fueron agroinfiltrados en plantas de

tabaco, junto con las diferentes glucosiltransferasas, prunasina y/o amigdalina fueron sintetizadas. De hecho, tres glucosiltransferasas fueron capaces de producir prunasina y también tres produjeron amigdalina. Sorprendentemente, dos de ellas sintetizaron prunasina y amigdalina. Por primera vez se han caracterizado varios clones de la enzima anabólica glucosiltransferasa-2, que transforma la prunasina en amigdalina.

Nuestro estudio de los transcriptomas del tegumento de un genotipo dulce y otro amargo, mostró la expresión diferencial de genes candidatos para citocromos, glucosiltransferasas e hidrolasas. Respecto de los citocromos, de los tres genes candidatos analizados, dos se sobre-expresaron en el genotipo amargo. Respecto de las glucosiltransferasas, de los seis genes candidatos analizados, sólo uno se sobre-expresó en el genotipo amargo. Finalmente, respecto de las hidrolasas, sólo uno de los tres genes candidatos se expresó, pero de igual forma en dulces y amargos.

1. Introduction

1. INTRODUCTION

1.1. THE ALMOND

1.1.1. Taxonomy and origin

The almond [*Prunus dulcis* (Miller) D. A. Webb] is a deciduous tree of the *Rosaceae* family. Its taxonomic classification is as follows:

Division: *Spermatophyta*	Family: *Rosaceae*
Subdivision: *Angiospermae*	Subfamily: *Prunoidea*
Class: *Dicotyledoneae*	Genus: *Prunus*
Subclass: *Rosidas*	Subgenus: *Amygdalus*
Superorder: *Rosanae*	Species: *Prunus dulcis* (Miller) D.A. Webb

The almond species is described in the bibliography as *Prunus dulcis*, *Prunus amygdalus*, *Prunus communis* or *Amygdalus communis*. The term *A. communis* came from Linnaeus in 1753. Batsch later named the species *Prunus amygdalus* in 1801, and many botanists, from Spach (1843) to Grasselly (1976), identified the species as *Prunus dulcis*. Finally , the General Committee for botanical nomenclature decided upon *P. dulcis* (Miller) D.A. Webb as the official name for almond in 1964. *P. amygdalus* Batsch and *P. communis* (L.) Archangeli were accepted as synonyms.

The origin of the domesticated almond can be traced to 4,000-5,000 years ago in Central Asia (Grasselly, 1976). There are two theories to explain the origin of the cultivated almond.

Several Russian botanists have found evidence that the cultivated almond may have come from a wild ancestor from the mountains of Asia (Popov, 1929; Vavilov, 1930; Kovalev and Kostina, 1935). According to this theory, the wild ancestor could

be *Amygdalus communis* (syn. *Prunus communis* Archang) (Vavilov, 1930; Watkins, 1979; Denisov, 1988; Kester et al., 1990).

The other theory sustains that cultivated almond comes from natural hybridisations between the wild species *Prunus fenzliana* Fritsch, *Prunus bucharia* Korschinsky and *Prunus kuramica* Korschinsky (Evreinoff, 1958; Wilsie, 1966; Grasselly, 1976).

In a recent study, Ladizinsky (1999) indicated that the cultivated almond originated from a single specie: *P. fenzliana*.

In the beginning, all these wild species were bitter. Sweet almonds, likely due to a mutation, began to spread by human selection.

The origins and diffusion of the almond can be classified into three main stages (Figure 1.1):

A) *Central Asia:* Almond seems to have originated in the mountains of Central Asia (Grasselly, 1976), from several wild species such as *P. fenzliana*, *P. bucharica* and *P. kuramica.* These species could be found growing in an area that stretched from the Tian Shan Mountains in Western China to the deserts of Kurdistan, Turkestan, Afghanistan and Iran (Grasselly, 1976; Browick and Zohary, 1996). In these areas, almonds were grown on dry soils for years under subsistence agricultural practices. The crops were distributed via commercial routes throughout Persia, Mesopotamia and around the Mediterranean Sea. Almond was grown with other crops like olives and carob, and these crops were subjected to processes of natural and human selection, resulting in different ecotypes and local cultivars (Grasselly and Crossa-Raynaud, 1984; Kester et al., 1990).

B) *Mediterranean basin:* Almond adapted well to the Mediterranean climate (Grasselly, 1976) and was spread by the Greeks, Romans and Arabs throughout the Mediterranean basin starting from 300 B.C. (Bacarella et al., 1991. The Arabs expanded the cultivation of almonds in North Africa and southern Spain and Portugal (Gradziel, 2009). In these regions, orchards were established on hillsides to avoid frost. Plants were multiplied by seeds and cultivated in poor soils and marginal

conditions (Muncharaz, 2004). In the 19th century, thanks to human selection, new cultivars were selected from local populations (Felipe, 2000).

C) *California:* Almond culture was introduced in California by Spanish missionaries (Franciscans) in the mid 18th century and was fully established in the 19th century, when the first grafts and seeds of the French cultivar Languedoc were introduced (Kester et al., 1990; Gradziel and Kester, 1996). Nowadays, most of the American cultivars come from the cultivar Nonpareil. Furthermore, California is the top almond producer in the world, which is due to several reasons: the availability of water and efficient irrigation systems, good soil and a favourable climate, automated management systems and well-adapted cultivars and rootstocks (Gradziel, 2009).

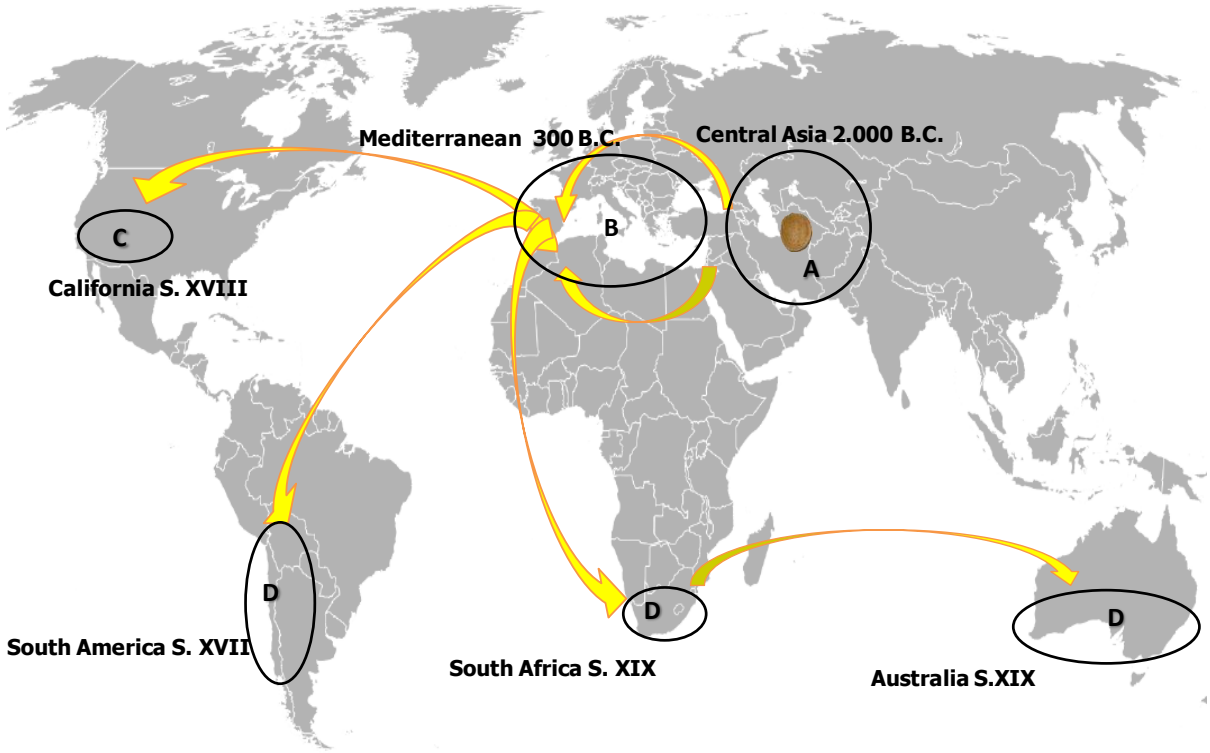


Figure 1.1. The origins and diffusion of the almond. Four stages: A) Central Asia; B) the Mediterranean basin; C) California; and D) South America, South Africa and Australia.

1.1.2. Almond production and economic relevance

Among the stone fruits, almond is in the fourth position behind peach, plum and apricot in terms of production and economic relevance in the world, with around 1 metric tonne (Mt) of kernel (about 3 Mt in-shell).

World surface area and production levels

Concerning almond production levels (Figure 1.2), according to the Food and Agriculture Organization of the United Nations (FAO), Spain was the third world producer in 2013 with 0.15 Mt of in-shell almonds (5.1% of the total global production). USA (California) was the first producer with 1.81 Mt, representing 62.2% of the total production in the world, followed by Australia (0.16 Mt, 5.5%). There are other important producer countries like Morocco (3.3%), Iran (3%), Syria (2.9%), Italy (2.5%), Tunisia (1.8%) and China (1.5%). This information from FAO can give us a general idea of world production levels, but it is not totally accurate. Indeed, the current percentages for almond kernel production are different for the USA and Australia (which together represent around 50% of total production) and for the rest of the almond-producing countries (around 30%).

Despite being the first producer in the world, USA was the second country in terms of cultivated surface area in 2013 according to FAO (Figure 1.3), with 315,590 ha, representing 19.1% of the global total. Spain was identified as the country with the largest amount of hectares dedicated to almond cultivation in the world, with 530,000 ha (32.1%). The 2013 FAO data also include the surface areas harvested in several other countries where almond is an important crop, shown here as percentages of the global total: Tunisia (11.5%), Morocco (9.1%), Iran (4.2%), Italy (4.1%) and Australia (1.7%).

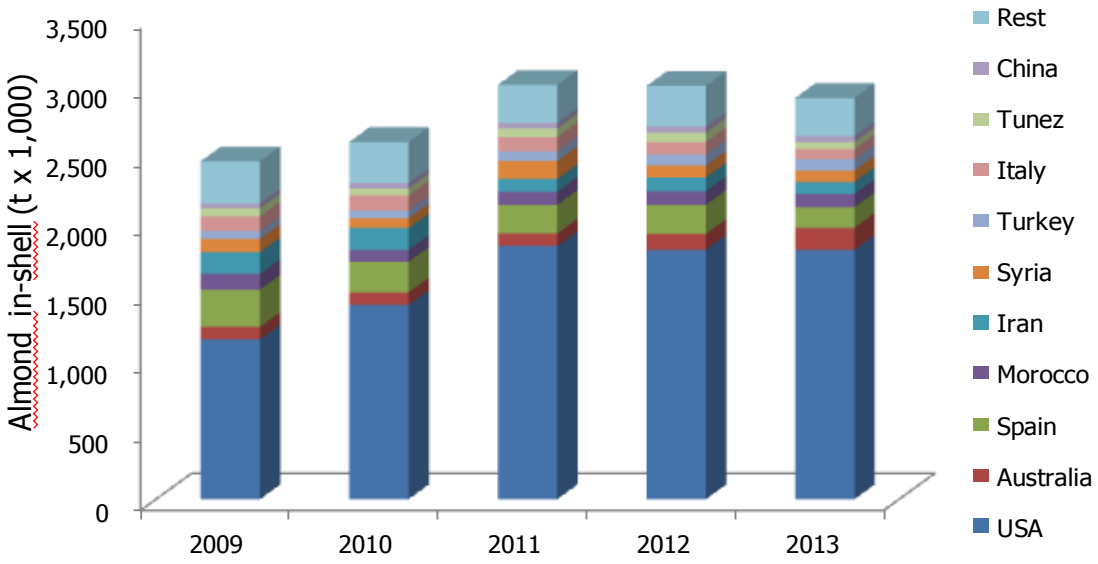


Figure 1.2. Production of in-shell almonds (t x 1000) in the most important producer countries, between 2009 and 2013. Source: FAO (http://faostat.fao.org).

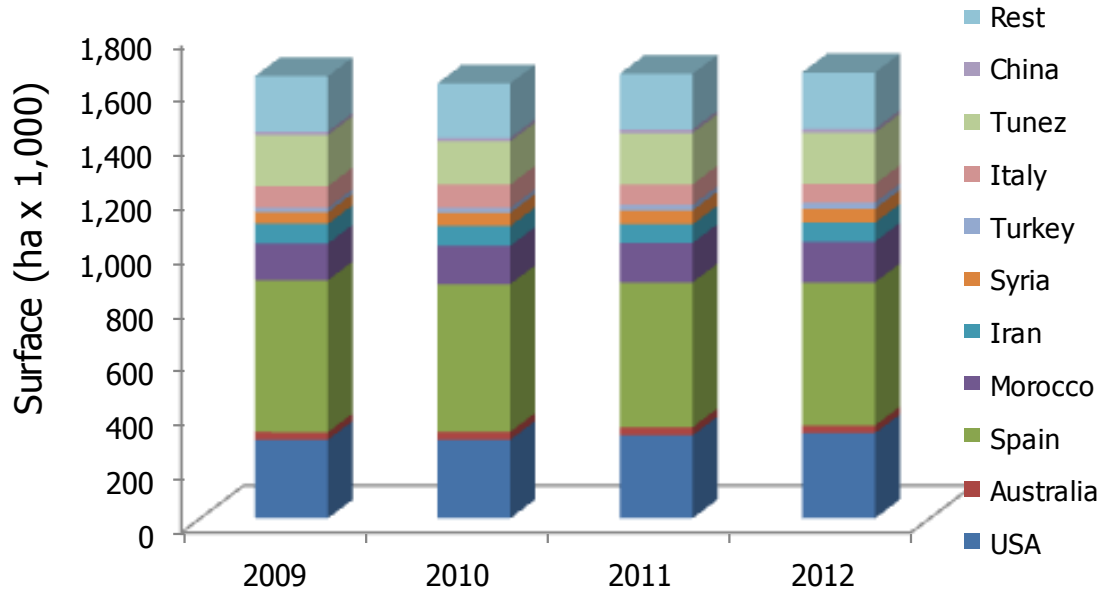


Figure 1.3. Surface area dedicated to almond cultivation (ha) in the most important producer countries in the world, between 2009 and 2012. Source: FAO.

Surface area and production levels in Spain

In Spain, Andalusia was the most important producer region in 2013 (with 13,000 t, representing 27.4% of national production), followed by Aragón (12,800 t, 26.9%) and Castilla-La Mancha (7,200 t, 15.1%) (Figure 1.4). Murcia was the fifth producer with 3.13 t x 1000, which represented 6.6% of the total. The production in Murcia was very low in 2013 due to drought. Andalusia was the region with the most cultivated surface area (25.5%), followed by Murcia (18.5%), Valencia (16.4%), Aragón (13.7%) and Castilla-La Mancha (11.5%) (Figure 1.5). It is important to keep in mind that Murcia is smaller than the other almond producing regions (it has only one province, whereas the other regions have between 3 and 8). In this context, the fact that Murcia is second in terms of cultivated surface area indicates the importance of almond cultivation in the region.

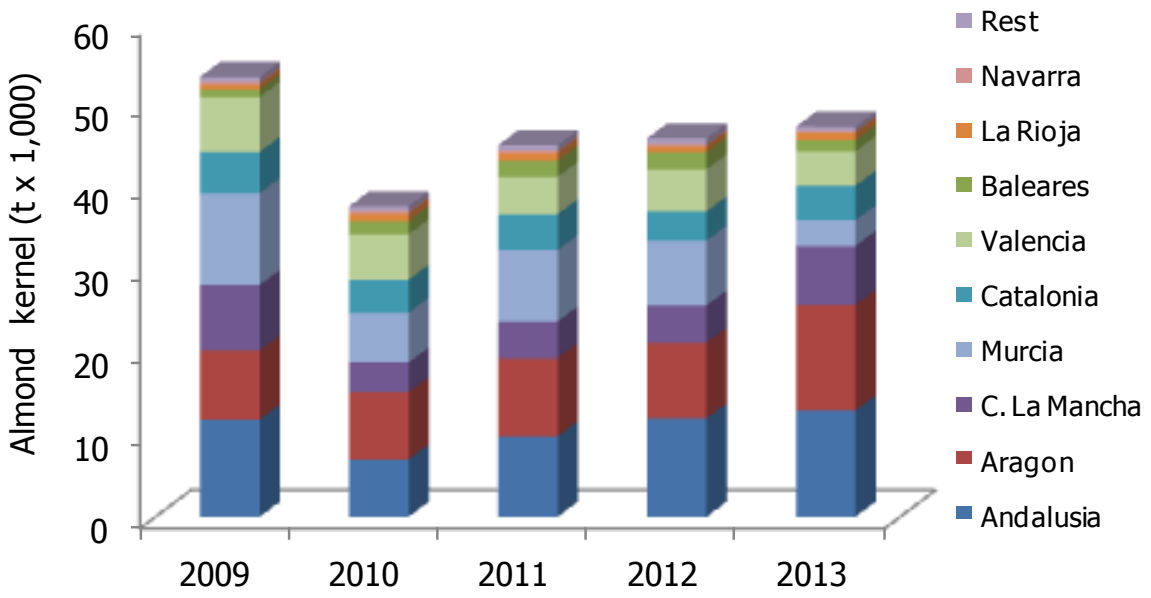


Figure 1.4. Production (almond kernel t x 1,000) in the most important producer regions in Spain, between 2009 and 2013. Source: Cooperativas Agroalimentarias de España (http://www.agro-alimentarias.coop).

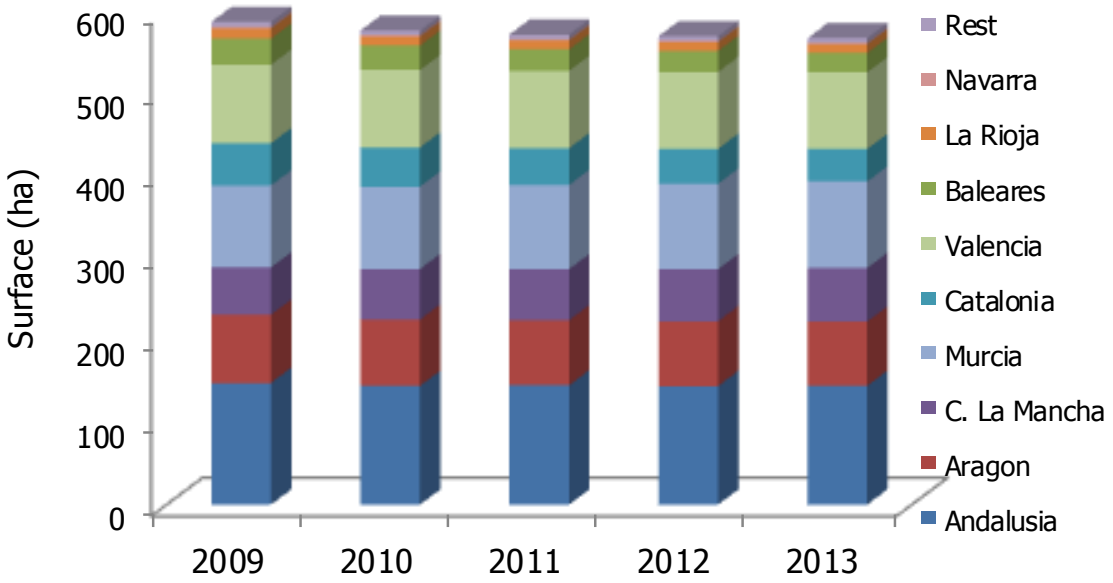


Figure 1.5. Surface area dedicated to almond cultivation (ha) in the most important producer regions in Spain, between 2009 and 2013. Source: Cooperativas Agroalimentarias de España (http://www.agro-alimentarias.coop).

Economic relevance

Despite the increase in world almond production in recent years, the demand for almonds is currently higher than the potential production. This situation is the consequence of the significant marketing campaigns of Californian producers on the one hand, and, on the other hand, difficulties in increasing the productive surface area in the main producer countries (USA and Australia) due to drought. The consumption of almonds has increased in important countries like China, India and South Korea as well as in Europe and in the domestic markets in both the USA and Australia. Due to such disequilibrium in the market, almond prices have increased significantly (tripling in Spain) and will probably remain high for the next 10 years.

1.1.3. Morphological and physiological characteristics

The almond species shows a strong ability to adapt to difficult growing conditions thanks to its deep and extensive root system. This characteristic makes it possible for almonds to grow in arid regions of Asia and North Africa, for instance (Kester et al., 1991; Ladizinsky, 1999). Furthermore, almond can resist cold winters and dry and hot summers. The almond tree is dormant during winter, although it is considered the earliest *Prunus* species to bloom due to its low chilling requirements and quick response to warm growing temperatures. This early flowering characteristic limits almond production to relatively warmer areas, because late frosts in colder areas can damage the fruits.

Almond has always been considered a marginal crop in the Mediterranean region. As a result, cultivation practices have tended to minimise fertilisers, water use and labour. Up until the 19th century, almond had not evolved much as it had only been propagated by seeds and only subjected to natural and human selection (Felipe, 2000).

Morphologically, almond is a deciduous tree with different shapes, from erect to opened habit. The leaves are alternate and lanceolate. Almond roots are deep and branched to survive in dry climates, although they are sensitive to root asphyxia. The main function of the roots is to anchor the tree in the soil and to store reserves.

Between one and three buds are found on each node of the axial leaves. These can be either vegetative or flower buds, which at the beginning are protected by scales. The terminal bud on a branch is always vegetative.

The almond flower is hermaphrodite and is composed of the calyx, with five green sepals, and the corolla, with five white or rose oval-shaped petals. In addition, the male part of the flower is made up of around 30 stamens whose anthers contain the pollen necessary for pollination. Finally, the female part of the flower is formed by the pistil, which contains the ovary with ovules, the style and the stigma. Almond is in general a self-incompatible species, so it requires cross-pollination. In fact,

inadequate cross-pollination has been one of the main causes of low productivity in most almond growing areas (Socias et al., 1992).

After successful pollination and therefore fertilization, the almond fruit, a drupe, starts to grow in the first months of the year. Growth continues until the summer, when maturity takes place. The external part of the fruit is made up of the exocarp and the mesocarp. The internal part of the fruit is composed of the endocarp (the shell), which is usually hard, dry and lignified. The mature seed is found inside the endocarp. This seed is made up of the embryo and the tegument (also called the seed coat). In the immature seed, the nucellus and endosperm can still be observed. Finally, the embryo is formed by two cotyledons and the embryonic axis, where the epicotyl and hypocotyl will, after germination, produce the shoots and the roots, respectively.

The differentiation of buds into leaves or flower buds takes place during the summer. Flowering is determined by chilling requirements to break dormancy and then heat requirements for bud development. Almond flowering in Spain starts (depending on the region and the cultivar) during January and continues into April. Fruit development starts after fecundation, and the almond harvest usually occurs between August and October.

1.1.4. Commercial uses and nutritional values

The edible part of the kernel is the seed, which is consumed as a nut that can either be natural or processed. Although almond consumption is largely based on sweet cultivars, bitter and slightly bitter almonds also have a place in the market. The nut can be commercialised in different forms: in-shell, with kernel with tegument and blanched. Almonds can also be used in the confectionery industry, for example for chocolates or nougat. Additionally, almond fruits can be used for cosmetics (almond kernel oil), for animal feed (hull) or as an energy source (shell). The bitter almond is toxic due to the presence of amygdalin (a compound that will be further

described later). Bitter almonds are used to make products like marzipan and amaretto liquors.

Almonds are an excellent source of vitamins, minerals, proteins and essential fatty acids (Saura-Calixto et al., 1981). Furthermore, they are one of the best natural sources of vitamin E (Sabate and Haddad, 2001) and of magnesium, phosphorous, folic acid and fibre (Vezvaei et al., 1995; Schirra, 1997). Almond oil (over 50% of the dry kernel) is mainly composed of oleic acid (monounsaturated), which has been shown to reduce the risk of cardiovascular diseases (Fulgoni et al., 2002; Lovejoy et al., 2002; Socias i Company et al., 2007).

1.1.5. Factors limiting production

In order to attain high production levels in any crop, it is essential to have thorough knowledge of the factors that affect cultivation. The most important factors affecting almond are the following:

Climate: Almond is a species that is well adapted to the Mediterranean conditions. It is resistant to drought, although low rainfall drastically affects the yield. Dormant buds are very resistant to low temperatures during the winter, although open flowers and young fruits are highly susceptible to late frosts and can be seriously damaged. Such damage can be avoided using extra-late flowering cultivars.

Soil: Almond can be cultivated in almost any type of soil, but permeable and deep soils with good nutrient and humidity levels are obviously better than clay soils. In good soils such as those mentioned above, the tress become bigger and more productive earlier than in bad soils.

Plant material: The cultivar must be well adapted to the local environmental conditions. Almond is a highly polymorphic species due to the fact that it has traditionally been propagated by seeds together with the fact that cross-pollination is necessary. This polymorphism has been an advantage for adaptation to new areas. More recently, with the selection and breeding of new cultivars adapted to each environment, growers have been able to cultivate the best-performing cultivars in

each area. Similarly, the rootstocks have been selected according to each type of soil and level of water availability.

Cultivation techniques: The application of modern cultivation techniques is very important to maintain a high yield year by year. In this regard we have to consider irrigation techniques, fertilisation, the correct application of pesticides, pruning, weed management, and so on.

1.1.6. Breeding programs

The first breeding programs date from the 20th century and were located in the USA (1923) and Russia (1932). Later on, other programs were carried out in France, Italy, Spain and Tunisia (Kester and Gradziel, 1996). Nowadays, the genetic breeding of this species is primarily being performed in almond breeding programs in the USA (California), Spain (Zaragoza, Reus and Murcia) and Australia (Adelaida).

Self-compatibility and late flowering are the main objectives of these current breeding programs. Other important traits for breeders are sweet kernels (versus bitter); early ripening; high productivity; high quality kernels (size, shape, no-doubles); pest and disease resistance; and drought resistance.

Following are the most important almond breeding programs today:

USA: The breeding program in the USA started at the University of California (Davis) in 1923, within the United States Department of Agriculture (USDA). Nowadays, the main objective of the almond breeding program at the University of California (Davis), which is led by Dr. Gradziel, is to obtain cultivars that are compatible with Nonpareil for pollination purposes as well as self-compatible cultivars with plague and disease resistance (Gradziel and Kester, 1994; 1998; 1999; Gradziel et al., 2001). This program has obtained cultivars like Solano, Sonora and Padre (Kester et al., 1984). Furthermore, the new cultivar Winters, which is a good pollinizer for Nonpareil, has recently been obtained by the USA program.

Russia: The second oldest almond breeding program was started in 1932 in the Nikitskij Botanical Garden in Yalta (Crimea) (Denisov, 1988). The main objectives

of this program were late flowering and the potential use of wild species for breeding. This program obtained the following cultivars: Desertny, Bumazhnoskorlupy, Krymskij, Nikitskij 2240, Mjagkoskorlupy, Sovetskij, Yaltinskij and Turkmenskij Otlichnij (Denisov, 1988).

France: In 1951, a collection of 400 cultivars was established at the National Institute of Agronomical Research (INRA) in Bordeaux, under the direction of Dr. Grasselly. In 1960, the breeding program was launched with the following objectives: late flowering, high production and good kernel quality. As a result of this early work, the Ferragnès and Ferraduel cultivars were released. Since 1972, self-compatibility has also been included as an objective in the French breeding program, resulting in the release of and Lauranne and Steliette (Grasselly, 1972), and, more recently, Mandaline.

Italy: The first breeding program in Italy was started in 1957 at the Agronomic Institute of Bari. Since 1970, the Experimental Institute for Fruit Tree Research in Rome has selected traditional Italian cultivars like Cristomorto for late flowering and Genco and Tuono for self-compatibility.

Spain – Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA), Zaragoza: This breeding program was started in 1974 in order to obtain self-compatible and late flowering cultivars (Felipe and Socias i Company, 1985). The first results were the new cultivars Guara, Aylés and Moncayo (Felipe and Socias et al., 1987). In 1999, 2006 and in 2008, a total of six more cultivars were added to the breeding program: Blanquerna; Cambra; Felisia (Socias and Felipe, 1999); Soleta; Belona (Socias i Company and Felipe, 2006); and Mardía (Socias i Company et al., 2008). A recent work has proven that Guara, of unknown origin, is actually the Italian cultivar Tuono (Dicenta et al., 2015).

Spain - Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Centro Mas de Bover, Reus: Since 1975, the IRTA breeding program has pursued the following objectives: late flowering, high yield, fruit quality and ease of pruning (Vargas et al., 1980; 1982; 1984; Vargas and Romero, 1988; Rovira et al., 1997; Vargas et al., 1997). The first cultivars released were Masbovera, Glorieta (self-incompatible) and

Francolí (self-compatible) (Vargas and Romero, 1992). More recently the self-compatible Constantí, Vairo and Marinada and the self-incompatible Tarraco were obtained (Vargas et al., 2006).

Centro de Edafología y Biología Aplicada del Segura (CEBAS-CSIC), Murcia: This program was started in 1971 (Egea et al., 1985; Egea and García, 1988). The objectives were to obtain late flowering and self-compatible cultivars that maintained the good characteristics of the native cultivars (productivity, quality kernel and hard shell). The first cultivars released were Antoñeta and Marta (García et al., 1996; Egea et al., 2000). Later, in 2007, the first extra-late flowering self-compatible cultivars, Penta and Tardona, were released (Dicenta et al., 2009). Tardona is the latest flowering almond cultivar release so far (Prudencio et al., 2016).

Australia: In 1995, the Australian Almond Breeding Program was started in collaboration with the University of Adelaide, the Almond Board of Australia and the research and development organisation Horticulture Australia Limited. Dr. Wirthensohn leads this project. The main objectives are self-compatibility and good kernel quality (Bertozzi et al., 1998; Sedgley and Collins, 2002). American, French, Australian and Spanish parentals have been used for breeding purposes.

1.2. CYANOGENIC GLUCOSIDES

1.2.1. Diversity of cyanogenic glucosides in nature

The cyanogenic glucosides (CNGlcs) are defence compounds present in more than 3,000 plant species (Bak et al., 2006; Sánchez-Pérez et al., 2008; Gleadow and Moller, 2014), including economically important crops such as sorghum (*Sorghum bicolour* Moench), cassava (*Manihot esculenta* Crantz) and rosaceous stone fruits. These compounds are also present in insects (Blum and Woodring, 1962; Conn, 1969). Cyanogenesis is the ability of plants to release hydrogen cyanide (HCN) to defend against herbivores or insects. Moreover, one or more sugar molecules and a ketone or an aldehyde are released in this process (Conn, 1980; Gleadow and Moller,

2014). CNGlcs are secondary metabolites stabilised by glycosylation. They are β-glucosides of α-hidroxinitriles (Conn, 1980; Zheng and Poulton, 1995; Morant et al., 2008).

Phytoanticipins are preformed defence compounds present in the plant before pathogen attack, and they are the first chemical barrier against such attacks (Bak et al., 2000). CNGlcs are phytoanticipins derived from the amino acids present in ferns, gymnosperms and angiosperms (mono and eucotyledonous) (Conn, 1980; Bak, 2006; Zagrobelny et al., 2008). CNGLcs are degraded by β-glucosidases, and both substrate and enzymes are inert when they are separated. When the plant is attacked and the tissues broken down, CNGlcs are degraded and thus provide immediate chemical defence against herbivores and pathogens (Morant et al., 2008).

Ferns and gymnosperms have aromatic CNGlcs derived from either tyrosine or phenylalanine, while angiosperms have aliphatic and aromatic CNGlcs derived from valine, leucine, isoleucine, tyrosine or phenylalanine (Bak et al., 2006). Linamarin and lotaustralin are glucosides of aliphatic α-hydroxynitriles, while dhurrin and taxiphyllin are glucosides of aromatic α-hydroxynitriles.

There are mono- and di-CNGlcs, depending on the presence of one or two sugars in the molecule (Table 1.1). Recently, for the first time, tri-CNGlcs have been found (Pičmanová et al., 2015). Prunasin is derived from phenylalanine. Sambunigrin, the epimer of prunasin, is also present in plants. Amygdalin is a prunasin-derived diglucoside. Linamarin and linustatin are mono- and diglucosides, respectively, derived from valine. Lotaustralin is derived from isoleucine and typically occurs with linamarin. Epiheterodendrin is derived from leucine. Tetraphyllin B is derived from the nonproteinaceous amino acid 2-cyclopentenyl glycine following an additional hydroxylation. Proteacin, triglochinin, and dhurrin are all derived from tyrosine. The sugar residues involved are glucose, xylose, and apiose (Gleadow and Møller, 2014). These compounds are present in species belonging to the *Rosaceae, Leguminosae, Graminae, Araceae, Compositae, Euphorbiceae* and *Passifloraceae* families (Conn, 1980).

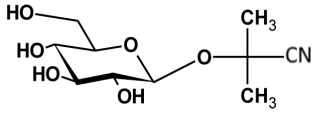
Table 1.1. Main cyanogenic glycosides and species involved.

Amygdalin	*Prunus Sp.*
Prunasin	*Prunus Sp.* and many *Rosaceae* and *Eucalyptus Sp.*
Sambunigrin	*Sambucus nigra, Acacia Sp.*
Prulaurasin	*Rosaceae*
Vicianin	*Vicia angustifolia L.* and other *Vicia Sp.*
Dhurrin	*Sorghum Sp.*
Taxiphyllin	*Taxus Sp.*
Zierin	*Zieria laevigata Sm.*
Linamarin	*Linum usitatissimum, L.; Phaseolus lunatus Sp.; Manihot Sp; Dimorphoteca Sp.; Lotus japonicus; Hevea brasiliensis*
Lotaustralin	*Linum usitatissimum, L.; Phaseolus lunatus Sp.; Manihot Sp; Dimorphoteca Sp.; Lotus japonicus; Hevea brasiliensis*
Acacipetalin	*Acacia Sp*.
Gynocardin	*Gynocardia odorota*

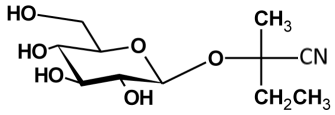
There are herbivores that are able to *de novo* synthesise CNGlcs, such as a few species of *Diploda* (millipedes), *Chilopoda* (centipedes), *Coleoptera* (beetles) and *Heteroptera* (true bugs). These herbivores all synthesise aromatic CNGlcs. On the other hand, more than 200 species within *Lepidoptera* (butterflies) synthesise aliphatic CNGlcs. Finally, *Zyganea* species are able to sequester the glucosides obtained from their food plants to use in their own defence against predators (Zagrobelny et al., 2004; Morant et al., 2008).

Following are the structures of the most important CNGlcs:

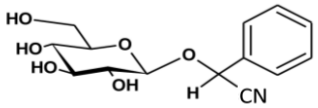
Linamarin: 2-Hydroxy-2-methylpropanenitrile-O-β-D-glucopyranoside



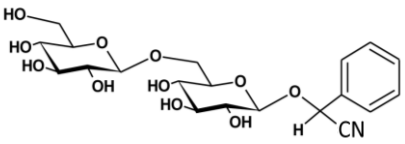
Lotaustralin: 2 (R)-Hydroxy-2-methylbutanenitrile-O-β-D-glucopyranoside



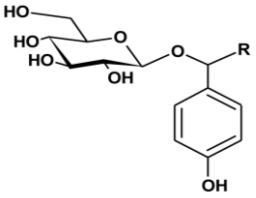
Prunasin: (R)-Mandelonitrile-O-β-D-glucopyranoside



Amygdalin: (R)-Mandelonitrile-O-[β-D-glucopyranosyl-(1→6)-β-D-glucopyranoside]



Dhurrin: (S)-4-Hydroxymandelonitrile-O-β-D-glucopyranoside



1.2.2. Cyanogenic glucosides in almonds: prunasin and amygdalin

In almond, there are two CNGlcs, prunasin and amygdalin, which are *de novo* synthesised in the almond kernel (Sánchez-Pérez et al., 2008). Bitterness in the almond kernel is determined by the CNGlc amygdalin (McCarty et al., 1952; Conn, 1980; Frehner et al., 1990; Møller and Seigler, 1991; Swain et al., 1992; Poulton and Li, 1994; Dicenta et al., 2002; Franks et al., 2008; Sánchez-Pérez et al., 2008). Amygdalin is a diglucoside of R-mandelonitrile, composed of two glucoses with β

orientation and joined by a 1,6 bond (Kuroki et al., 1987; Swain et al., 1992; Hu and Poulton, 1999). The other bitterness compound is the monoglucoside prunasin. Prunasin is converted into amygdalin during the fruit development stage of bitter genotypes (Sánchez-Pérez et al., 2008). Another compound related with almond bitterness, denominated 2,3-butanediol, was discovered for the first time by Wirthensohn et al. (2008) . This compound is the major volatile compound in non-bitter kernels.

Amygdalin is basically found in ripe bitter kernels in different concentrations, and it is not present (or is present in very low concentrations) in sweet kernels (Dicenta et al. 2002; Sánchez-Pérez et al., 2008; Wirthensohn et al., 2008; Arrázola et al., 2012). Prunasin is found in roots, leaves and unripe fruits (Frehner et al., 1990; Arrázola, 2002; Dicenta et al., 2002; Wirthensohn et al., 2008). There does not appear to be a relationship between the prunasin content in the vegetative parts and the amygdalin content in the kernels (Dicenta et al., 2002).

The almond kernel is composed of mother tissues (endocarp, mesocarp, tegument and nucellus) and by tissues from both progenitors (endosperm, cotyledon and the embryo axis). Amygdalin is only stored in the mature cotyledon of bitter kernels (Frehner et al., 1990; Sánchez-Pérez et al., 2008). Prunasin, however, is mainly accumulated in the tegument of bitter genotypes during development, while the prunasin levels in sweet genotypes are much lower.

During the development of bitter kernels, prunasin levels drop in parallel with increases in amygdalin levels, suggesting that the prunasin is converted into amygdalin (Sánchez-Pérez et al., 2008).

These CNGlcs are involved in a metabolic pathway (Figure 1.6), which can be divided into three routes: biosynthesis of prunasin and amygdalin, degradation of these compounds and the release of HCN by β-glucosidases, and the detoxification of HCN.

Biosynthesis: This route starts with the amino acid L-phenylalanine (Phe) (Mentzer et al., 1961), which through the two cytochromes P450 (CYP79 and CYP71) (Yamaguchi et al., 2014) and a UDP-glucosyltransferase (GT1) produces prunasin (cyanogenic monoglucoside) (Franks et al., 2008). Later, through another UDP-glucosyltransferase (GT2), prunasin is converted into amygdalin (cyanogenic diglucoside).

Degradation: Amygdalin can be degraded by a β-glucosidase called amygdalin hydrolase (AH) to produce prunasin, releasing one glucose molecule. This prunasin is subsequently degraded by another β-glucosidase called prunasin hydrolase (PH), releasing mandelonitrile and glucose (Kuroki et al., 1987; Li et al., 1992; Zheng, 1995; Zhou et al., 2002; Sánchez-Pérez et al., 2008; 2009; 2012). Finally, this mandelonitrile is degraded by mandelonitrile lyase 1 (MDL1) (Swain and Poulton, 1994; Zheng et al., 1995; Suelves et al., 1998; Hu et al., 1999), producing benzaldehyde (which gives the bitter taste) and cyanide (toxic) (Evreinoff, 1952).

Detoxification: There is a cyanide detoxification pathway through the nitrilases, providing a nitrogen source for the plant (Swain and Poulton, 1994b). This route consists of two steps involving two enzymes. In the first step, HCN is catalysed together with L-cysteine through β-cyanoalanine synthase, producing β-cyanoalanine. In the second step, nitrilases catalyse the production of asparagine or ammonia and aspartate from H_2O and β-cyanoalanine (Swain and Poulton, 1994).

Chapter 3 specifically addresses the enzymes involved in the biosynthesis and degradation of CNGlcs.

Although the main function of CNGlcs is chemical defence against pathogens and herbivores, other functions have been proposed, as follows (Gleadow and Moller, 2014): the control of germination (Swain and Poulton, 1994); the transport of carbon and nitrogen (Selmar, 1988); dormancy breaking of the buds via the release of HCN (Barros et al., 2012); the supply of nitrogen in the form of ammonia (Sánchez-Pérez et al., 2008); and the modulation of oxidative stress (Neilson et al., 2013).

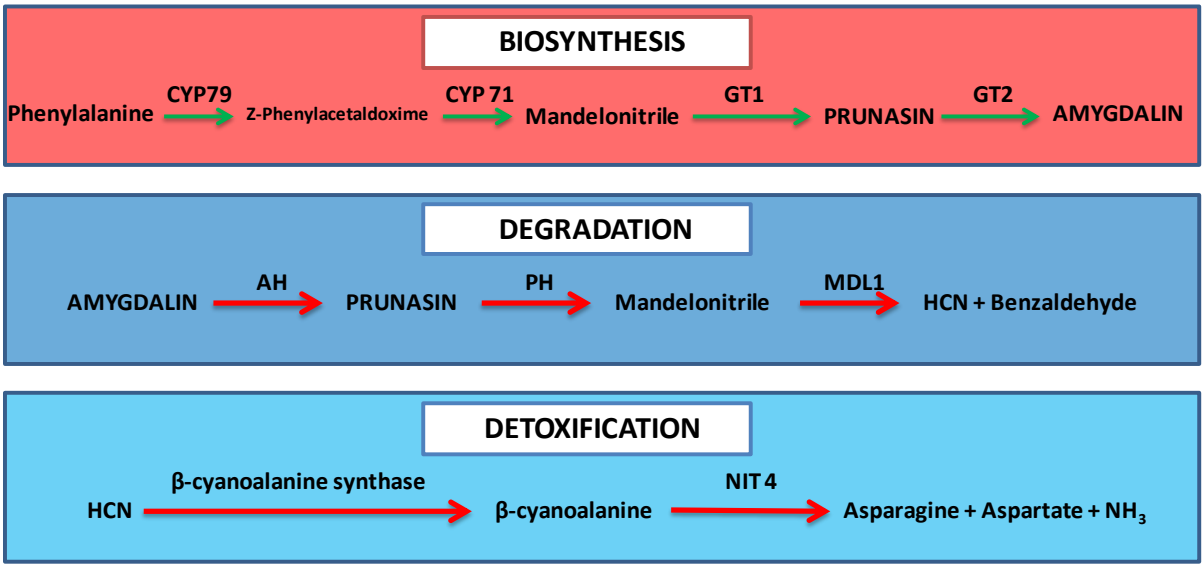


Figure 1.6. Metabolic pathway for the biosynthesis, degradation and detoxification of prunasin and amygdalin in almond. CYP79 and CYP71: cytochromes P450; GT1 and GT2: glucosyltransferases; AH: amygdalin hydrolase; PH: prunasin hydrolase; MDL1: mandelonitrile lyase 1; NIT4: nitrilase.

1.3. ALMOND BREEDING AND BITTERNESS

1.3.1. The genetic control of bitterness: inheritance

Wild almond species have bitter kernels, and the sweet taste might have appeared due to a mutation. Later, humans selected these sweet individuals and multiplied them, initially by seeds, and subsequently by grafting, thus stabilising the sweet taste in this species.

Bitterness is a monogenic trait, and the sweet allele (*Sweet kernel*, *Sk*) is dominant over the bitter (*sk*) (Heppner, 1923; 1926; Dicenta and García, 1993; Dicenta et al., 2007). Commercial sweet cultivars can be homozygous (*SkSk*) or heterozygous (*Sksk*) for this trait. Slightly bitter cultivars are always heterozygous (*Sksk*), but as indicated above, heterozygous cultivars can also be sweet. Furthermore, a slightly bitter taste can change to some extent year by year or even among the kernels of the same tree. So far, there is no explanation for this fact.

Finally, bitter genotypes are always homozygous recessive (*sksk*) (Grasselly and Crossa-Raynaud, 1980; Dicenta et al., 1991; Vargas et al., 2001).

Most commercial cultivars (like Marcona, Desmayo Largueta and Nonpareil) are heterozygous for bitterness. When two heterozygous cultivars are crossed in breeding programs, 25% of the descendants will be bitter and therefore discarded from the breeding program (Dicenta and García, 1993). The ratio of bitter seedlings can be 50% if for some special reason (resistance to disease, late flowering time, early ripening) a bitter genitor is used in combination with the heterozygous cultivar (Dicenta et al., 2007).

On the other hand, the use of a sweet homozygous progenitor (like Ramillete, Lauranne, Ferragnès, etc.) ensures a sweet taste in 100% of the progeny, regardless of the other genitor. A bitter seed can therefore produce an almond with sweet kernels if the male parent carries the dominant allele *Sk*. The use of slightly bitter heterozygous progenitors such as Garrigues increases the proportion of slightly bitter versus sweet among heterozygous seedlings.

The gene responsible for sweetness (*Sk*) is located in linkage group 5 (LG5) (Joobeur et al., 1998; Bliss et al., 2002; Sánchez-Pérez et al., 2007). This point will be more accurately explained in Chapter 5, which is dedicated to the genomics of bitterness.

1.3.2. Truths and lies on bitterness in almond

Among farmers there is much confusion with respect to the bitterness of almonds. Nevertheless, all sources of such confusion can be explained from a scientific point of view, based on the genetic control of bitterness as described above.

The influence of "bitter pollen" on the taste of the sweet cultivars: There is some confusion about the possible influence of pollen on whether the resulting kernel is sweet or bitter. The pollen of some species can affect the characteristics of the endosperm (xenia) or adjacent tissues (metaxenia). Swamy and Krishnamurthy

(1980), Kester and Asay (1975) and Kester and Gradziel (1996), however, have reported that the sweet or bitter taste of almond is not affected by the pollinizer cultivar.

Dicenta et al. (2000) studied the effect of the pollinizer on sweet or bitter almond flavour by tasting the seeds of 32 crosses between sweet, bitter and slightly bitter parents. These researchers concluded that the seed taste only depended on the female parent, and they did not detect any influence of the pollinizer on the seed bitterness. Sánchez-Pérez et al. (2012) studied the influence of the pollinizer on the amygdalin content of the seed, finding that when a sweet pollinizer was involved, the amygdalin content could be reduced by up to 21% with respect to the bitter × bitter cross. Nevertheless, given the significant differences in amygdalin content between truly sweet and bitter kernels, this 21% difference did not change the bitter flavour of the seeds.

The kernels of a given cultivar will thus have the same sweet or bitter flavour regardless of the pollinizer, and the influence of both progenitors is expressed in the following generation (Kester and Assay, 1975; Dicenta et al., 2000; Arrázola, 2002).

The presence of sweet and bitter kernels on the same tree: The presence of both sweet and bitter kernels on the same tree has sometimes been described. Furthermore, the appearance of bitter kernels on a previously sweet kernelled tree has also been reported. From the scientific point of view, these claims cannot be explained except for the unlikely event of a somatic mutation. Nowadays we know that the maternal control of the sweet or bitter trait (Heppner, 1923; Kester and Gradziel, 1996; Dicenta and García, 1993) implies that all almonds from a tree will always be sweet, bitter or slightly bitter throughout the lifetime of the tree. These claims are probably related to the coexistence on the same tree of branches of the cultivar with branches of the bitter rootstock.

The influence of the rootstock on the bitterness of the cultivar: It has also been suggested that the use of bitter kernelled rootstocks could transfer bitterness to sweet cultivars. However, it has been shown that the sweet or bitter almond taste is independent of the sweet or bitter genotype of the rootstock. This finding is

corroborated by the fact that despite the high number of potentially bitter rootstocks that exist in the field, resulting from the open pollination of the heterozygous *Sksk* Garrigues, there is a scarcity of bitter kernelled trees.

1.3.3. Strategies for the control of bitter almonds

The recent rise in almond prices has increased the harvest of abandoned bitter almonds, which are being introduced into the market. This is damaging the Spanish market with respect to the American market and has caused great alarm, prompting efforts to solve this problem on different levels: through research, regional and national administrations and grower associations. Although we have pointed out that there is a small market for bitter almonds, they must not be mixed with sweet almonds. In this context, the following points must be taken into account to protect the Spanish market:

- Even though it is not a rule, most of the bitter almonds in the field are early flowering. This is due to their probable origin: rootstocks from the open pollination of Garrigues or descendants of Marcona x Desmayo Largueta and other heterozygous early flowering cultivars.
- In monovarietal orchards with new cultivars (especially grafted on almond x peach GF677 or Garnem rootstocks), the presence of bitter trees is nil or very scarce. In case of failure of the graft of the cultivar and the growth of the rootstock, the fruits of these rootstocks are different from the almonds. Furthermore, in the case of Garnem the rootstock is detectable because the leaves are red. As a result, it is easy to detect the rootstocks and thus avoid harvesting bitter almonds.
- It is necessary to develop campaigns to pull up abandoned almond trees, which are isolated and often bitter, so that they are no longer harvested and introduced into the food chain.
- Finally, the processing companies should only buy almonds from reliable growers.

1.3.4. Strategies for early selection in breeding programs

Parental choice and descendant selection:

As mentioned before, bitter individuals frequently appear among the offspring in almond breeding programs despite the sweet taste of the parents used. This is the case when both parents are heterozygous (*Sksk*), so 25% of the offspring will inevitably be bitter (*sksk*). A first strategy for avoiding bitter seedlings among the offspring would therefore be to use at least one sweet homozygous parental (*SkSk*), which would guarantee 100% sweet descendants.

Nevertheless, this is not always possible if we want include a specific trait in our new cultivars. Moreover, as we have already noted, most commercial cultivars and many selections used in breeding programs are heterozygous and thus generate bitter descendants. The easiest way to eliminate these bitter seedlings would be to wait until the tree bears fruit and then taste just one kernel to determine if the kernels are sweet or bitter.

This solution, however, implies growing the descendants for several years, spending time and money. One alternative strategy would be to develop early selection methodologies, like molecular markers, to identify and eliminate the bitter genotypes during the first year, in the nursery, before taking them to the experimental orchards.

Early selection for bitterness:

The development of early selection strategies is very important for breeders. Such strategies make it possible to select descendants at an early stage, saving time and money and increasing the efficiency of breeding programs. These methods do not accelerate the time for obtaining a new cultivar, but they do increase the chances for success, making it possible to work with a higher number of seedlings.

The easiest way to develop a method for early selection would be to find a narrow correlation between a characteristic that is expressed early on (during the first year) and the trait that we want to evaluate (sweetness or bitterness).

Nowadays, this is not possible for bitterness since the bitter or sweet taste has not been correlated with any trait studied so far (Sánchez-Pérez et al., 2007).

The other strategy for early selection would be to develop molecular markers. DNA molecular markers have enormous advantages for their high reliability and easy application to large populations. The key issue is the need to develop a marker as close as possible to the bitterness gene, to ensure reliability. This early selection strategy is already a reality for floral incompatibility. As a result, in our group, annual self-incompatible descendants are removed during the first year in the nursery before taking the plants to the field to evaluate for other characteristics.

Although two molecular markers have already been found upstream and downstream of the *Sk* locus (Sánchez-Pérez et al., 2010), we have to deepen our knowledge in order to find a more accurate molecular marker for bitterness. This is one of the main objectives of this thesis, which will be discussed in depth in Chapter 5.

1.4. GENERAL OBJECTIVES

The final objective of this PhD thesis is to understand the metabolism of the cyanogenic compounds in almond and identify the gene responsible for bitterness, to develop a molecular marker that would make it possible to identify sweet and bitter seedlings during the first year in the nursery, thus increasing the efficiency of breeding programs. With this aim, several studies were carried out:

- The characterisation of the catabolic and anabolic enzymes related to bitterness metabolism in almond, involving an evaluation of the differences between sweet and bitter genotypes in different tissues (tegument, nucellus and cotyledon) to determinate their role in the bitterness of almonds.
- Analysis of the evolution of the CNGlcs prunasin and amygdalin from dormancy to flowering and the identification of other CNGlcs involved in bitterness and their expression in the flower tissues, all related to dormancy control and flowering time.
- The development of a molecular marker for bitterness in almond by saturating linkage group five (LG5) with SNPs and CAPS markers. This is the linkage group where the *Sk* locus is located, which is responsible for bitterness in almond.
- Analysis of the transcriptome of a sweet and a bitter almond through the study of the differential expression of candidate genes for the *Sk* locus.

2. Plant material

2. PLANT MATERIAL

The plant material consisted of different almond cultivars and offspring provided by the Almond Breeding Program of "Centro de Edafología y Biología Aplicada del Segura" (CEBAS-CSIC). The trees are located in the experimental station "Tres Caminos" of CEBAS-CSIC (Figure 2.1) located in Santomera (Murcia, South-East Spain, 38º 6' 34.28" N; 1º 2' 16.71" O). With an altitude of 130 meters, this farm is characterised by very hot summers and cool winters, with minimum temperatures usually higher than 0 ºC. The trees are drip irrigated and they are spaced differently, from 4x5 m (cultivars) to 3x1 m (offspring).

Figure 2.1. A general view of the experimental almond orchards (delineated by the yellow lines) located in the experimental station "Tres Caminos" of CEBAS-CSIC (Santomera, Murcia, South-East Spain).

2.1. ALMOND CULTIVARS AND SELECTIONS.

Achaak: A local cultivar from Sfax (Tunisia) with sweet taste. It is self-incompatible and is early flowering and maturing. The fruit is hard-shelled and semi-elongated, with a good quality kernel with high fat content. This cultivar is very productive.


Figure 2.2. Achaak tree and fruit.

Desmayo Largueta: A well-known sweet Spanish cultivar dating from the 19th century in Tarragona, which spread throughout the Mediterranean area. It is self-incompatible, with early flowering and late maturation times. The fruit is elongated and the shell consistency is hard. Due to its good quality, Desmayo Largueta has a high price in the market. Its productivity is high.

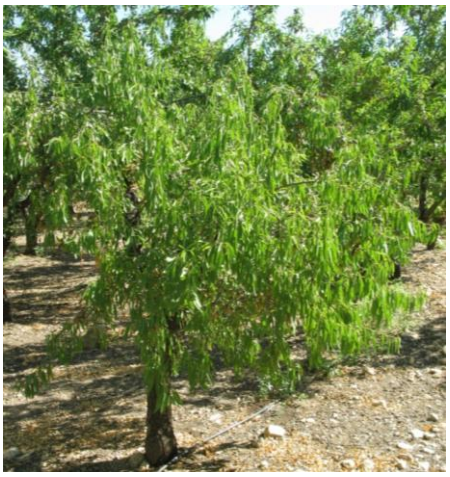



Figure 2.3. Desmayo Largueta tree and fruit.

Lauranne: A sweet, self-compatible French cultivar obtained in 1978 at INRA (France) from the cross Ferragnès x Tuono. It shows late flowering and intermediate maturation times. The fruit is elongated and semi-hard shelled.

Figure 2.4. Lauranne tree and fruit.

Penta: A sweet Spanish cultivar obtained at CEBAS-CSIC from the cross between S5133 (selection of CEBAS-CSIC) and Lauranne. It is self-compatible, with extra-late flowering and early maturity times. The fruit is elongated and hard-shelled.

Figure 2.5. Penta tree and fruit.

Ramillete: A Spanish cultivar from Murcia, which is self-incompatible and has a sweet taste. It has early flowering and maturation times. The fruit is elongated and the shell consistency is hard. This cultivar has high productivity.

Figure 2.6. Ramillete tree and fruit.

D05-187: A bitter selection from the open pollination of D00-349 (CEBAS-CSIC selection from S5133 x Marta). It is self-compatible with extra-late flowering and early maturation times.

Figure 2.7. D05-187 tree and fruit.

R1000: A self-compatible and extra-late flowering time selection obtained from Tardy Nonpareil x Tuono at INRA (Avignon, France). It has intermediate productivity and the fruit is sweet, small and hard-shelled. It has an early maturation time.

Figure 2.8. R1000 tree and fruit.

S3067: A self-compatible and bitter selection obtained from Garrigues x Tuono at CEBAS-CSIC, with early flowering and intermediate maturation times.

Figure 2.9. S3067 tree and fruit.

2.2. R1000 x DESMAYO LARGUETA OFFSPRING

R1000 x Desmayo-Largueta (1997): Offspring from the cross between the French selection R1000 and the Spanish Desmayo Largueta. The seeds obtained were germinated at CEBAS-CSIC, and in 1997, plants were taken to the field in Santomera (Murcia, Spain). A total of 167 seedlings were selected for this study (Figure 2.10).

Figure 2.10. R1000 x Desmayo Largueta (1997) offspring in Santomera (Murcia, Spain).

R1000 x Desmayo Largueta (2010): Another 450 seedlings from the same cross were obtained and planted in 2010 in the Santomera (Murcia, Spain) experimental field, to increase the number of seedlings of this population (Figure 2.11).

Figure 2.11. R1000 x Desmayo Largueta (2010) offspring in Santomera (Murcia, Spain) in different stages. From top to bottom: plantation, adult tree flowering, seedlings heavily pruned for germoplasm conservation.

3. Cyanogenic glucosides in buds from dormancy to flowering

3. Evolution of cyanogenic glucosides from dormancy to flowering

3.1. INTRODUCTION

The cyanogenic glucosides (CNGlcs) prunasin and amygdalin have been broadly studied in relation with their defence function (Bak et al., 2006; Morant et al., 2008; Gleadow and Møller, 2014) and with the bitterness of the almond kernel (McCarty et al., 1952; Conn, 1980; Frehner et al., 1990; Møller and Seigler, 1991; Swain et al., 1992; Poulton and Li, 1994; Dicenta et al., 2002; Franks et al., 2008; Sánchez-Pérez et al., 2008). Moreover, other functions have also been also described, like the control of seed germination (Swain and Poulton, 1994; Pičmanová et al., 2015); the transport of carbon and nitrogen (Selmar, 1988); the supply of nitrogen in the form of ammonia (Sánchez-Pérez et al., 2008); and the modulation of oxidative stress (Neilson et al., 2013). These functions, however, had never been analysed in relation with flower bud dormancy.

Dormancy is an adaptive mechanism of some plant species in temperate climates that is controlled by the accumulation of chill hours during winter. The time when dormancy is overcome determines the flowering and leafing time, as well as seed germination, enabling the plant to take advantage of weather conditions that are favourable for development. More specifically, flowering will only happen when dormancy is broken (Fennell, 1999).

Flowering time is one of the most important agronomic traits in almond breeding programs, since late flowering can prevent the loss of crops due to late frosts (Dicenta et al., 2005). Flowering time in almond is mainly determined by the chilling requirements of each cultivar for breaking dormancy; heat requirements are less important (Egea et al., 2003). In the case of low chilling requirements (early-flowering cultivars), temperatures below zero in late winter or spring can cause yield loss due to frost (Scorza and Okie, 1990). The mechanism of dormancy is also present in the rest of *Prunus* species such as apricot (*Prunus armeniaca* L.) (Ruiz et

al., 2007); sweet cherry (*Prunus avium* L.) (Alburquerque et al., 2008); peach (*Prunus persica* L.) (Weinberger et al., 1950); and plum (*Prunus domestica* L.) (Okie et al., 2008).

Climate change is forcing not only fruit breeders to develop new cultivars adapted to the new conditions, but also scientists to develop new treatments to compensate for the loss of chilling hours that plants need in order to maintain high production levels. Nowadays there are artificial treatments to break dormancy and thus modify the flowering time using cyanamide, a compound that contains hydrogen cyanide. Since the degradation of the CNGlcs in almond (amygdalin and prunasin) releases hydrogen cyanide, this reaction could be involved in breaking dormancy in this species.

In this chapter, the evolution of amygdalin, prunasin and some derivate compounds has been studied between dormancy and flowering in the flower buds of five almond cultivars with different bitterness genotypes and flowering times. The objective is to determine if these compounds are related with the breaking of dormancy in almond and, consequently, with the flowering time.

3.2. PLANT MATERIAL

Five almond cultivars, described in detail in Section 2.1, were used in this study. These cultivars were selected according to their different flowering times and kernel bitterness. The cultivars used were Achaak (very early flowering time, sweet taste); Desmayo Largueta (very early, sweet); S3067 (early, bitter); Lauranne (late, sweet); and Penta (extra-late, sweet) (Table 3.1).

3.3. METHODOLOGY

Chilling and heat requirements for breaking dormancy and flowering

The orchards were visited weekly to follow the flowering time of the cultivars studied. Furthermore, every two weeks, three branches of 40 cm were collected from each cultivar and taken to a growth chamber in controlled conditions (light period of

16 h at 25 ºC and 40% relative humidity and a darkness period of 8 h at 20 ºC and 60% relative humidity).

Table 3.1. List of cultivars studied classified by their flowering time and kernel bitterness.

Cultivar	Flowering time	Bitterness (genotype)
Achaak	Very early	Sweet (*Sk*/--)
Desmayo Largueta	Very early	Sweet (*Sksk*)
S3067	Early	Bitter (*sksk*)
Lauranne	Late	Sweet (*SkSk*)
Penta	Extra-late	Sweet (*SkSk*)

The branches were placed in jars with a 5% saccharose and 1% aluminium sulphate solution, which was changed after 5 days, for a period of 10 days. The developmental state of the flower buds was then determined, and the date of dormancy breakage was established when 50% of the flower buds were in the B-C state. The full flowering date was determined in the field when 50% of the flowers of the tree had fully opened (F state) (Figure 3.1).

Figure 3.1. Phenological states considered in our trial, according to Felipe (1977). Left: undeveloped flower buds in state A; centre: flower buds in state B-C; right: opened flower in state F.

The calculation of chilling requirements was performed in Chill Units (CU), according to Richardson's method (1974), as a function of the number of hours accumulated at each range of temperature starting from November 15th.

Heat requirements were calculated as Growing Degree Hours (GDH), which is the hourly temperature minus -4,5 ºC. The heat requirements of each cultivar were calculated as the number of GDH accumulated between the end of the dormancy and the date when 50% of the flowers were open (F_{50}).

Cyanogenic glucoside content in flower buds

Between 5th November 2013 and 24th March 2014, four to six branches were collected from each cultivar every two weeks in the experimental orchard of CEBAS-CSIC. The branches were then brought to the laboratory, where the flower buds and flowers were removed and kept at -80 ºC. The number of buds and flower tissues (petals, sepals, pistils and pollen) was different depending on the type of sample (Table 3.2) according to the phenological stages as described by Felipe (1977).

Table 3.2. Number of buds and flowers analysed for each type of sample.

Flower buds (state A)	15
Flower buds (state B-C)	15
Flower (state F)	10
Sepals	10 flowers
Petals	10 flowers
Pistils	15 flowers
Pollen	30 flowers

Three technical replicates were used for each sample and the average and standard deviations of each value were calculated. Due to methodological problems in some samples only one or two replicates were available and so, in the first case, the standard deviation was zero. These samples were:

Batch 1: Desmayo, S3067 (2 replicates).

Batch 2: Desmayo (2 replicates).

Batch 4: Achaak, Desmayo, Lauranne, S3067 (2 replicates), Penta (1 replicate).

Batch 7: Desmayo (2 replicates).

Batch 8: Desmayo (2 replicates).

Batch 9: S3067 (1 replicate).

The concentration of the cyanogenic glucosides in the flower buds and flower tissues was analysed by LC-MS. To accomplish this, 100 mg of the samples were ground in liquid nitrogen and placed in 1.5 ml threaded tubes with 400 µL 85% methanol. The tubes were then boiled for 5 min and put in ice. Next, the samples were centrifuged for 5 min x 20,000 x g, and the supernatant was collected and kept at -20 ºC. A total of 20 µL of the supernatant was mixed with 70 µL of water and 10 µL of 500 µM internal standard linamarin and then filtered through an ELISA filter plate by centrifugation (5 min x 3,000 rpm) (Annex 8.16).

Analytical LC-MS was carried out using an Agilent 1100 Series LC (Agilent Technologies, Germany) coupled to a Bruker HCT-Ultra ion trap mass spectrometer (Bruker Daltonics, Bremen, Germany). Samples were analysed with the Bruker Daltonics program Data Analysis 4.0 (Annex 8.16).

3.4. RESULTS AND DISCUSSION

Table 3.3 shows the chill units (CU) accumulated each month according to the Richardson model, from 15th November 2013 until 12th March 2014. In this period a total of 939 CUs were accumulated. Table 3.4 shows the CUs required for breaking dormancy and the GDH requirements for flowering time for each of the five cultivars under study. As expected, Achaak was the earliest flowering cultivar and Penta was the latest, in accordance with the chilling requirements of each variety.

Table 3.3. Chill units (CU) accumulated each month and in total during the period of study according to the model of Richardson et al. (1974), from 15th November 2013 until 12th March 2014.

		November		December		January		February		March		Total	
Temperatures	**CU**	**Hours**	**CU**	**Hours**	**CU**	**Hours**	**CU**	**Hours**	**CU**	**Hours**	**CU**	**Hours**	**CU**
<1.5	0	0	0	0	0	0	0	0	0	0	0	0	0
1.5-2.4	0.5	1	0.5	0	0	0	0	0	0	0	0	1	0.5
2.5-9.1	1	184	184	263	263	141	141	113	113	67	67	768	768
9.2-12.4	0.5	110	55	261	131	275	138	214	107	42	21	902	451
12.5-15.9	0	76	0	188	0	224	0	209	0	82	0	779	0
16-18	0.5	12	-6	23	-11.5	54	-27	54	-27	61	-30.5	204	-102
>18	-1	2	-2	9	-9	50	-50	82	-82	36	-36	179	-179
Total		**385**	**232**	**744**	**373**	**744**	**202**	**672**	**111**	**288**	**22**	**2833**	**939**

Table 3.4. Chilling (CU) and heat (GDH) requirements for breaking dormancy and flowering. [a] Days from the start of the chilling accumulation (15th November 2013) until dormancy breaking. [b] Days from dormancy breaking until flowering time.

Cultivar	**Chilling requirements (CU)**			**Heat requirements (GDH)**		
	Breaking dormancy	**Days[a]**	**CU**	**Flowering time**	**Days[b]**	**GDH**
Achaak	30th Nov	15	231	25th Jan	51	10,310
Desmayo L.	6th Dec	21	306	30th Jan	54	11,730
S3067	12th Dec	27	391	13th Feb	58	8,767
Lauranne	25th Dec	40	533	28th Feb	62	10,517
Penta	02nd Feb	79	819	15th Mar	41	6,819

3.4.1. Prunasin and amygdalin in flower buds

The two main cyanogenic glucosides, prunasin and amygdalin, were detected in the flower buds of the early and late flowering cultivars during the entire developmental period from dormancy to flowering (Figures 3.2 and 3.3). Amygdalin reached the highest levels at the time of dormancy breaking, mainly in Desmayo and S3067 cultivars, although this result is not reliable due to a very high standard deviation (Figure 3.3). After this, the amygdalin levels dropped off to nearly zero. After dormancy breaking, prunasin started to accumulate and reached its maximum levels just before flowering time (Figure 3.2). This suggests that prunasin could play an important role in flower bud development after dormancy is broken. The prunasin levels increased and amygdalin decreased from the point dormancy was broken until just before flowering time and then decreased. This could suggest that amygdalin may be used to break dormancy and prunasin to promote the flower opening. To the best of our knowledge, this is the first time that the analysis of cyanogenic glucosides has been performed in flower development in almonds.

The level of prunasin was highest in the bitter cultivar (S3067) reaching 1.200 μmoles/100 mg, followed by Desmayo (0.900 μmoles/100 mg). The prunasin content in flower buds of the bitter S3067 was around 60 times higher than amygdalin content. The prunasin levels in Achaak and Lauranne (very early and late flowering, both sweet) were similar (around 0.400 μmoles/100 mg). The lowest prunasin levels were found in Penta (extra-late flowering) with 0.200 μmoles/100 mg.

The amygdalin content was always very low, from 0.050 μmoles/100 mg in Desmayo, to 0.020 μmoles/100 mg in S3067 and nearly zero in Achaak, Lauranne and Penta.

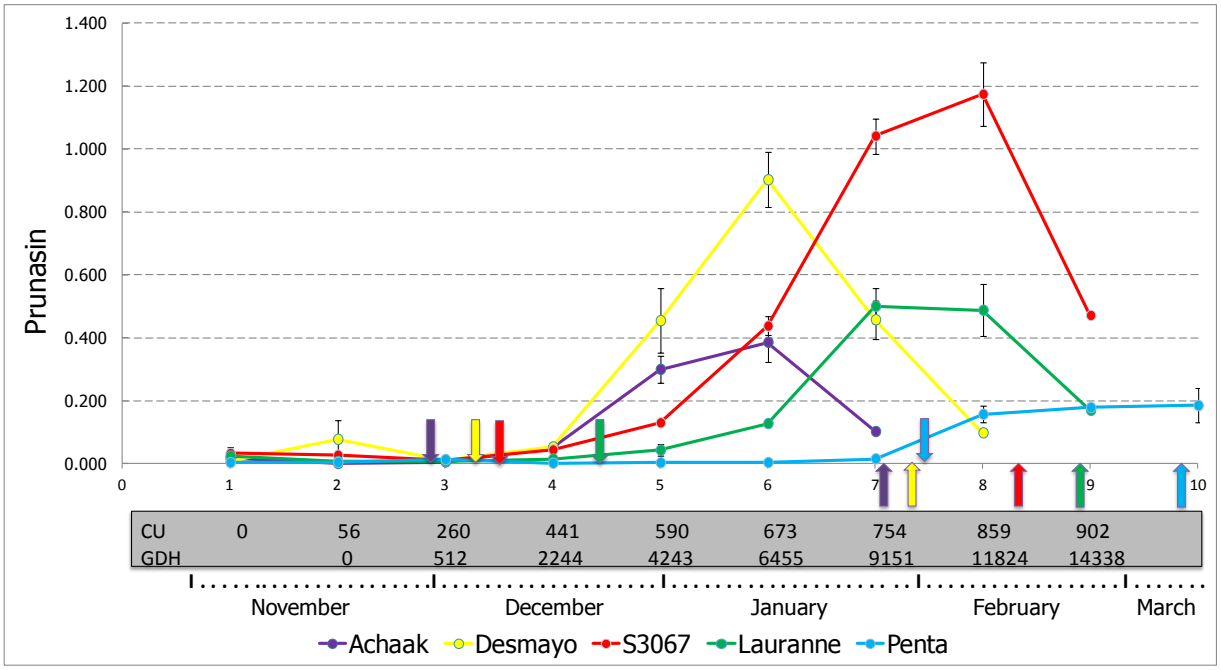


Figure 3.2. Prunasin content (µmoles / 100 mg FW) in flower buds during development in five almond cultivars. Down arrows indicate dormancy breaking date and up arrows indicate flowering time.

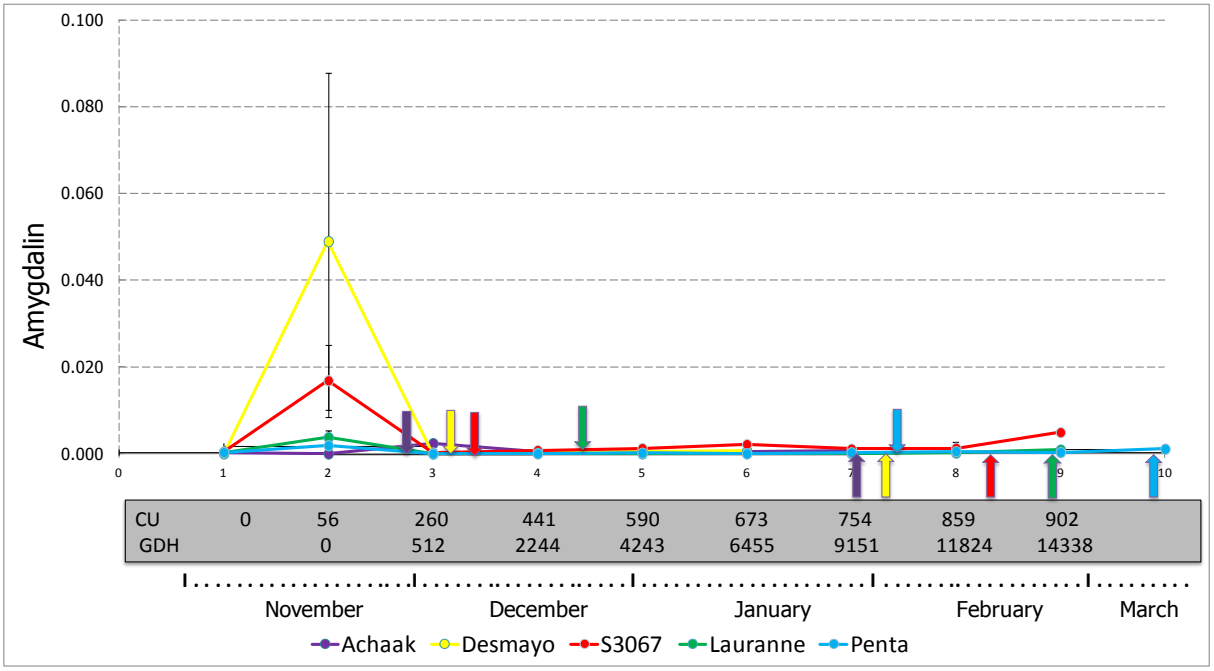


Figure 3.3. Amygdalin content (µmoles / 100 mg FW) in flower buds during development in five almond cultivars. Down arrows indicate dormancy breaking date and up arrows indicate flowering time.

No differences were thus observed in the cyanogenic glucoside content of early versus late flowering cultivars. In other words, if the prunasin peak is lower or higher, it does not mean that a cultivar will have an earlier or later flowering time.

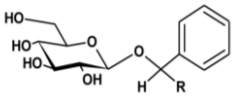
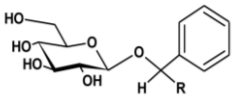
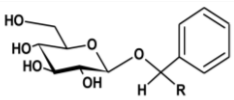
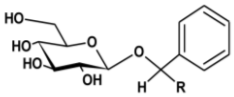
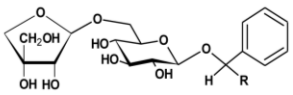
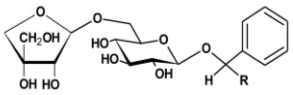
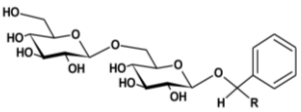
Cyanogenic compounds have also been detected in flowers of other species. Within the *Prunus* genus, prunasin has been detected in the flowers of *P. avium*, but not amygdalin (Nahrstedt et al., 1972). Matsuoka et al. (2011) found both prunasin and amygdalin in the flowers of *Prunus yedoensis*. According to Schappert et al. (2000), the cyanogenic glucoside levels of *Turnera ulmifolia* L. decreased to zero when the plant began to flower, so the plant lost its cyanogenic potential around flowering time. Cyanogenic glucosides have also been detected in flower tissues in species such as *Grevillea*; *Linum usitatissimum* L. (flax); *Lotus japonicus* L.; *Ryparosa kurrangii* B.L. Webber (rainforest tree); and *Eucalyptus camphora* R. T. Baker (Lamont et al., 1994; Niedzwiedz-Siegien et al., 1998; Forslund et al., 2004; Webber and Woodrow, 2008; Neilson et al., 2011). In the case of *L. japonicus*, there are two cyanogenic glucosides, linamarin and lotaustralin, that are present in the flower tissues. In this species, the reproductive organs are only cyanogenic when the specific β-glucosidase named BGD3 is expressed (Lai et al., 2015).

At least five different cyanogenic glucosides have been found in *E. camphora* flower buds, three of which have been identified: prunasin; sambunigrin (mandelonitrile β-D-glucoside, the (*S*)-epimer of prunasin); and amygdalin (Neilson et al., 2006). The only cyanogenic glucoside that has been found in the leaves of *Eucalyptus cladocalyx* is prunasin (Gleadow and Woodrow, 2000). In general, the cyanogenic glucosides have a higher cyanogenic capacity in the early stages of tree development and decrease in the mature stages (Loyd and Gray, 1970; Wheeler et al., 1990; Okolie and Obasi, 1993; Dahler et al., 1995). The other three phenylalanine-derived cyanogenic diglucosides (eucalyptosin A, eucalyptosin B and eucalyptosin C) have been identified in *E. camphora* (Neilson et al., 2011). The total cyanogenic glucoside content of this species has been found to reach maximum levels in flower buds, as occurred in our study.

3.4.2. Putative derivatives of prunasin in flower buds

In addition to cyanogenic glucosides, structurally related compounds were also found in the flower buds of the five cultivars studied (Table 3.5, Figures 3.4 and 3.5). These putative derivatives of cyanogenic glucosides (prunasin acid, amide, anitrile) and diglucosides (prunasin pentoside and prunasin anitrile pentoside) were found in levels much lower than those of prunasin but similar to those of amygdalin, except for prunasin anitrile pentoside, which was very abundant.

Table 3.5. Prunasin, amygdalin and prunasin derivative structures detected in the study.

Compound	Chemical name	Molecular formula	m/z $[M+Na]^+$	r.t. [min]	Estructure	
Prunasin	(R)-Mandelonitrile-O-β-D-glucopyranoside	$C_{14}H_{17}NO_6$	318	7		R = CN
Prunasin amide	(R)-Mandelamide-O-β-D-glucopyranoside	$C_{14}H_{19}NO_7$	336	4.4		R = $CONH_2$
Prunasin acid	(R)-Mandelic acid-O-β-D-glucopyranoside	$C_{14}H_{18}O_8$	337	5.7		R = COOH
Prunasin anitrile	Benzyl- O-β-D-glucopyranoside	$C_{13}H_{18}O_6$	293	6.5		R= H
Prunasin apioside	Mandelonitrile-O-[β-D-Apiofuranosyl-(1→X)-β-D-glucopyranoside]	$C_{19}H_{25}NO_{10}$	450	6.9		R = CN
Prunasin anitrile apioside	Benzyl-O-[β-D-Apiofuranosyl-(1→X)-β-D-glucopyranoside]	$C_{18}H_{26}O_{10}$	425	6.8		R= H
Amygdalin	(*R*)-Mandelonitrile-*O*-[β-D-glucopyranosyl-(1→6)-β-D-glucopyranoside]	$C_{20}H_{27}NO_{11}$	480	6.6		R = CN

Prunasin acid (Figure 3.4.A) was not detected during the whole development period, except for two tiny peaks observed at the beginning (before chill accumulation), and at the end, during the flowering of the cultivars Desmayo Largueta (0.080 µmoles/100 mg) and S3067 (0.160 µmoles/100 mg). The prunasin acid levels in Achaak, Lauranne and Penta were almost zero.

Prunasin amide (Figure 3.4.B) was not found during breaking dormancy (only in S3067), but its levels gradually increased till flowering time. The prunasin amide levels were similar in all the cultivars (between 0.01 and 0.02 µmoles/100 mg), and the maximum content was found in the sweet Achaak (0.03 µmoles/100 mg).

Prunasin anitrile (Figure 3.4.C) reached the highest content during breaking dormancy in the five cultivars. Later it decreased and started to increase right before flowering. The bitter cultivar S3067 had the highest level of this compound (0.025 µmoles/100 mg).

The diglucoside prunasin pentoside, a mixture of two prunasin pentosides (Pičmanová et al., 2015) (Figure 3.5.D), was quantified relatively in relation to the area prunasin because the synthesised standard was not available. Its levels were lower than the prunasin ones being Penta the cultivar with the highest content, although this result is not reliable due to the very high standard deviation.

Finally, prunasin anitrile pentoside (Figure 3.5.E) was also quantified compared to prunasin like for prunasin pentoside. All the cultivars showed high levels of prunasin anitrile pentoside at the beginning, during breaking dormancy. The cultivar Penta showed the highest concentration. Afterward, they were decreasing until they were hardly detectable during the flowering. It therefore seems that this compound was important for breaking the dormancy of the flower bud.

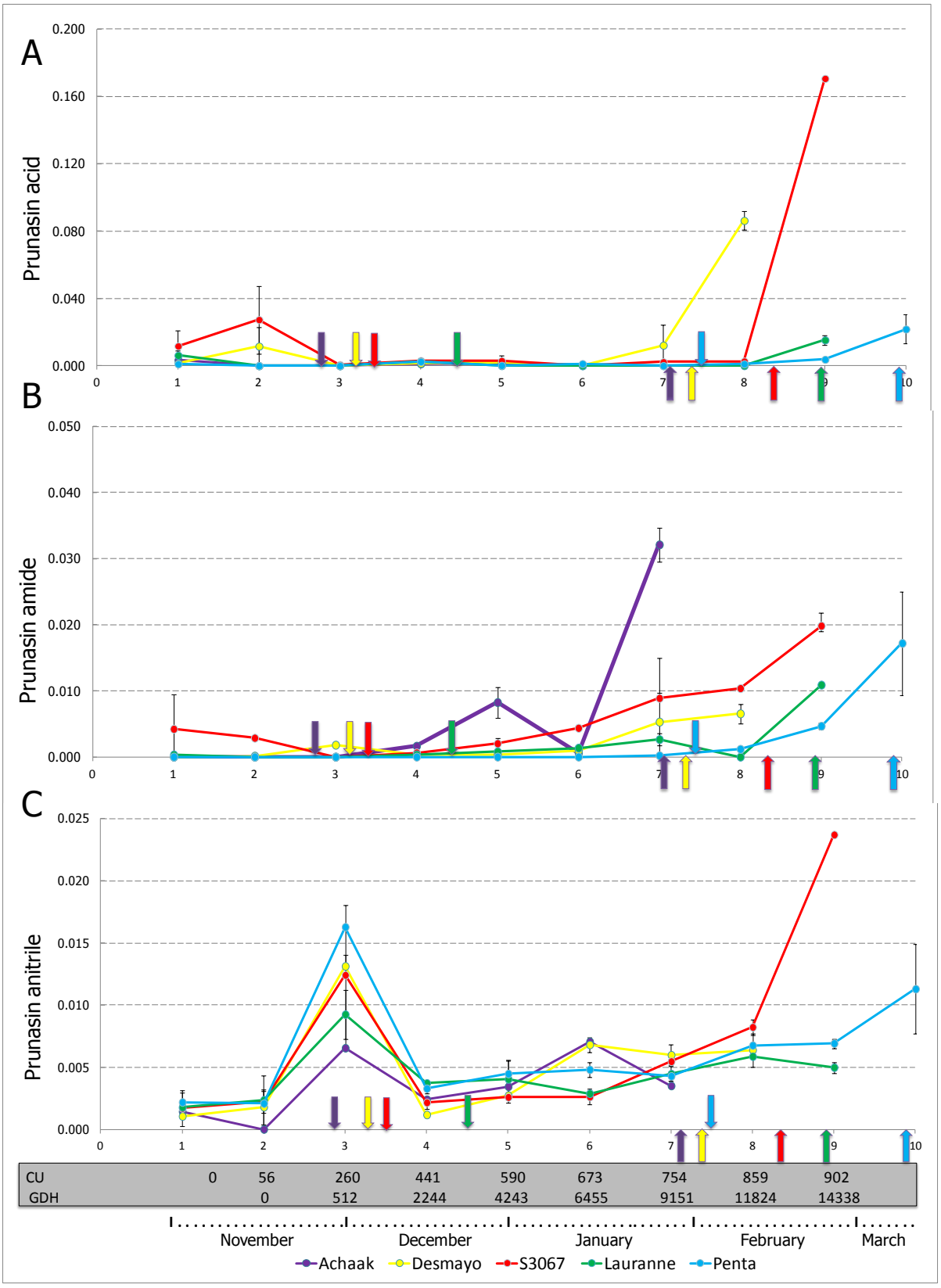


Figure 3.4. Prunasin derivative (A. prunasin acid, B. prunasin amide and C. prunasin anitrile) content (µmoles / 100 mg FW) in flower buds during development in five almond cultivars. Down arrows indicate dormancy breaking date and up arrows indicate flowering time.

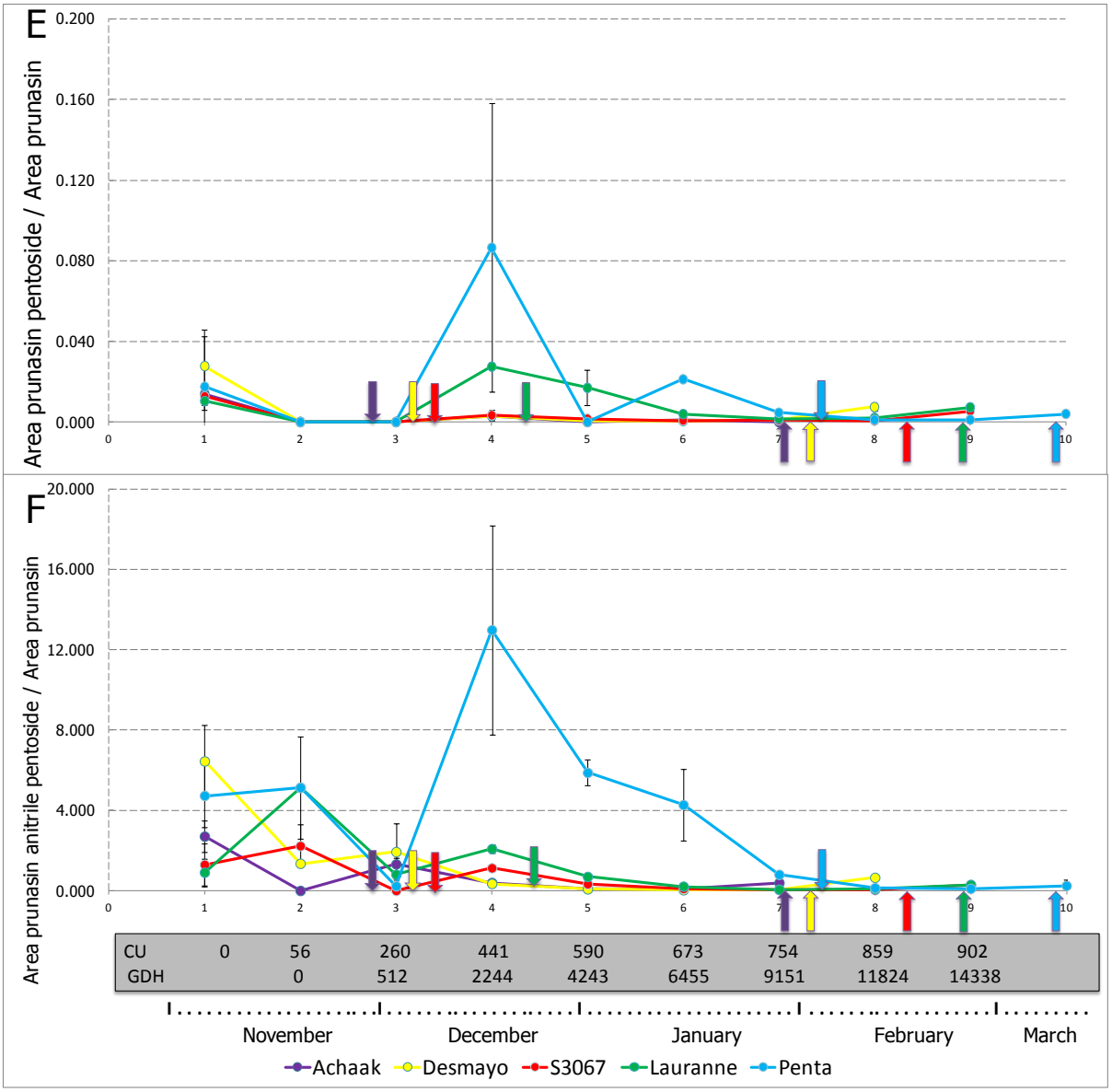


Figure 3.5. Prunasin derivative (D. prunasin pentoside and E. prunasin anitrile pentoside) content (expressed as area of prunasin pentioside / area prunasin and area of prunasin anitrile pentioside / area of prunasin) in flower buds during development in five almond cultivars. Down arrows indicate dormancy breaking date and up arrows indicate flowering time.

In this study, all of the before mentioned compounds (prunasin acid, amide, anitrile, pentioside and anitrile pentioside) are being described in almond flowers for the first time.

Putative cyanogenic glucoside derivatives and di-/tri glucosides of cyanogenic glucosides have been detected by LC-MS in cassava, sorghum and almond seeds (Pičmanová et al., 2015). These researchers detected prunasin and amygdalin putative derivatives (amides, acids and anitriles) during the germination of the bitter seeds of S3067 almond. In general, the levels of the putative derivatives of prunasin and amygdalin were much lower than the levels of prunasin and amygdalin themselves. Prunasin amide, acid and anitrile were found in low levels in the seeds, roots, shoots and leaves of the seedlings of all the cultivars studied in different stages of germination. Prunasin acid was the most abundant derivative in seeds, shoots and leaves, and prunasin anitrile was most abundant in roots. A significant amount of the prunasin derivatives was measured in these seeds at the beginning of germination (Pičmanová et al., 2015). Similarly, Fukuda et al. (2003) detected minor components related to cyanogenic glucosides in *P. persica* seeds. These compounds were amygdalinic acid, prunasin acid, benzyl β-gentiobioside and prunasin anitrile.

We are presenting further evidence in support of the conclusions drawn by Pičmanová et al. (2015), i.e., that cyanogenic glucosides occur together with their structural derivatives: amides, acids and anitriles. These authors suggested that these derivatives could play a role in the recycling of reduced nitrogen. The authors also proposed an alternative endogenous turnover pathway in which cyanogenic glucosides are converted into non-cyanogenic glucosides, without any release of HCN. Hypothetically, amides, acids and anitriles are produced from cyanogenic glucosides in this turnover pathway, with a concomitant release of NH_3 and CO_2. In this form, reduced nitrogen and carbon originating from the nitrile group of CNGlcs could be utilised in primary metabolism.

In the "conventional" turnover, amygdalin and prunasin are degraded and converted into benzaldehyde and HCN, which is further detoxified through β-cyanoalanine synthase with β-cyanoalanine that is converted into asparagine or aspartate and NH_3.

Two possible routes have been proposed for the alternative turnover pathway (Figure 3.6) (Pičmanová et al., 2015). In the first route, amygdalin is hydrolysed to

prunasin and further to prunasin amide and/or acid and NH_3. Prunasin acid is converted into prunasin anitrile with the release of CO_2. In the second route, amygdalin is hydrolysed to prunasin and prunasin is converted directly into the corresponding anitrile, with the release of NH_3 and CO_2. Both the NH_3 and CO_2 produced in these proposed pathways may be channelled into primary metabolism.

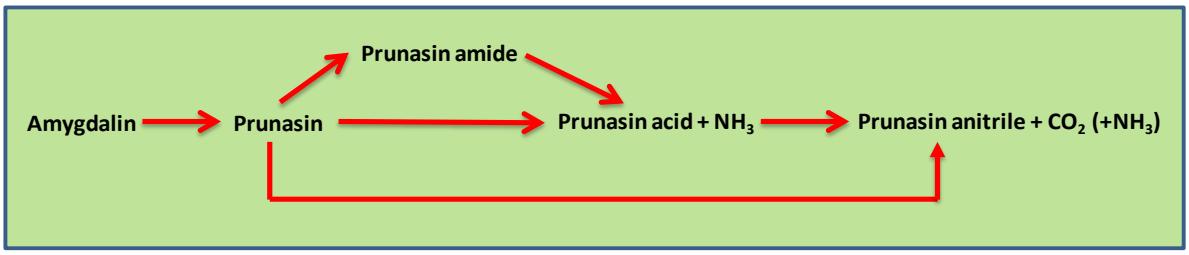


Figure 3.6. Alternative recycling pathways of amygdalin and prunasin without release of cyanide.

3.4.3. Cyanogenic glucosides in the flower tissues

In our study, cyanogenic glucosides were not only detected in flower buds but also in different parts of the flower. Petals, sepals, pistils and pollen all contained prunasin, amygdalin and some of the putative derivatives described previously (Table 3.6).

Prunasin was the most abundant cyanogenic compound in the flower tissues studied, especially in the pollen. The presence of amygdalin, prunasin amide and prunasin anitrile was almost negligible. In the petals and pistils, only prunasin was of some importance, especially in the bitter genotype. As previously mentioned, pollen was the tissue with the highest prunasin levels, especially in the bitter cultivar, but also in Penta and Desmayo Largueta. The remaining compounds were present in lower levels, which were generally not significant with the exception of prunasin acid in Penta. Sepals showed very low values for all the compounds, including prunasin (0.185 µl), although there were slightly larger amounts in the bitter cultivar.

The amygdalin content in almond pollen has been shown to be significant in other studies (1890 ppm) (London-Shafir et al., 2003). According to these authors, amygdalin would inhibit inefficient pollinators, thus allowing for more efficient

pollination by honeybees, which may be adapted to tolerate higher levels of amygdalin toxicity up to a certain point. Nevertheless, it is important to note that pollen is a tissue with minimal water content, which may explain the high level of cyanogenic glucosides and their derivatives in the pollen.

Table 3.6. Prunasin, amygdalin, prunasin acid, amide and anitrile content (µmoles / 100 mg FW) in the pistils, petals, pollen and sepals of the five cultivars.

		Prunasin		Amygdalin		Prunasin acid		Prunasin amide		Prunasin anitrile	
		nmol/µl	SD	nmol/µl	SD	nmol/µl	SD	nmol/µl	SD	nmol/µl	SD
Achaak	Pistil	0.015	0.013	0.000	-	0.003	0.001	0.001	0.001	0.004	0.003
	Pollen	1.415	-	0.012	-	1.223	-	0.300	-	0.277	-
	Petal	0.085	-	0.001	-	0.007	-	0.003	-	0.001	-
	Sepal	0.052	0.012	0.001	0.001	0.015	0.002	0.016	0.006	0.011	0.001
Desmayo	Pistil	0.210	0.216	0.000	-	0.003	0.004	0.002	0.001	0.012	0.011
	Pollen	2.398	0.839	0.017	0.004	0.278	0.393	0.644	0.004	0.182	0.081
	Petal	0.529	-	0.004	-	0.020	-	0.016	-	0.013	-
	Sepal	0.134	-	0.001	0.001	0.069	0.097	0.013	0.018	0.038	0.025
S3067	Pistil	0.870	0.587	0.002	0.001	0.025	0.018	0.023	0.012	0.057	0.035
	Pollen	3.802	3.123	0.246	0.331	0.318	-	0.233	0.120	0.199	0.134
	Petal	0.712	0.046	0.002	0.001	0.214	0.275	0.038	0.004	0.022	0.011
	Sepal	0.387	0.210	0.003	0.003	0.074	0.091	0.013	0.000	0.038	0.006
Lauranne	Pistil	0.097	0.086	0.000	-	0.001	0.001	0.001	0.001	0.007	0.002
	Pollen	0.734	1.024	0.013	0.010	0.778	1.007	0.624	0.653	0.189	0.116
	Petal	0.242	0.034	0.001	0.001	0.001	-	0.001	-	0.003	-
	Sepal	0.160	0.012	0.001	0.001	0.004	0.004	0.004	0.001	0.012	0.002
Penta	Pistil	0.135	0.141	0.002	0.001	0.150	0.249	0.074	0.124	0.014	0.014
	Pollen	2.574	2.003	0.016	0.018	4.693	3.292	0.802	0.434	0.097	0.032
	Petal	0.053	0.033	0.001	-	0.012	0.012	0.008	0.011	0.005	0.002
	Sepal	0.184	0.118	0.001	0.001	0.029	0.029	0.005	0.003	0.024	0.009

In a previous study, prunasin was also detected in the sepals, petals, pistils and pollen of bitter and sweet almonds (Abarrategui, 2010). In abarrategui`s study the amygdalin level was almost zero in all the cultivars except in the pollen of the bitter cultivars. Moreover, the sepals, petals and pistils contained larger amounts of prunasin in the bitter cultivars. In Abarrategui's study, the prunasin level in the pollen of the bitter genotype (S3067) was 2.600 µmoles/100 mg (2010), and in our study it was 3.800 µmoles/100 mg. In fact, our data were highly variable, ranging from 1.590 µmoles/100 mg in mid-February to 6.010 µmoles/100 mg at the end of the same month. The level of prunasin therefore increased four times during the pollen maturation period.

Lai et al. (2015) detected cyanide release in the early and mature stages of *L. japonicus* flowers. This cyanide was derived specifically from the keel and enclosed reproductive organs due to the presence of cyanogenic glucosides. The sepals, wings, buds and pods contained the cyanogenic glucosides linamarin and lotaustralin, although they did not release any cyanide because the β-glucosidases were not active in these tissues.

Future studies will be focused on the characterisation of β-glucosidases in almond flower tissues that could explain the levels detected of the cyanogenic glucosides prunasin and amygdalin.

3.4.4. Other functions of cyanogenic glucosides

Cyanogenic glucosides are formed by amino acids that must be transported through the plant to the sites where the cyanogenic glucosides will be synthesized. A supply of nitrogen is important for the biosynthesis of cyanogenic glucosides in the young plant tissues, which are weaker than mature tissues and need more protection against pathogens and herbivores. On the other hand, when the defence factor is not so important for the plant, nitrogen derived from cyanogenic glucosides can be reused for primary metabolism (Jenrich et al., 2007).

Gleadow et al. (2000) showed that up to 20% of the nitrogen in a plant is stored in cyanogenic glucosides and that the highest nitrogen levels are found in both young and reproductive tissues. These levels decreased when the tissues matured. In spring, coinciding with the flowering period, there was an important allocation of nitrogen to the reproductive tissues, to the detriment of the leaves, to form cyanogenic glucosides. These compounds decreased gradually during fruit development (buds-flowers-fruits).

The correlation between prunasin and amygdalin in our results suggests that cyanogenic glucosides have an important role in flower bud development (amygdalin during dormancy breaking and prunasin in the flowering period). More specifically, prunasin seems to play a role in the development and opening of flowers since it

increases before flowering. Amygdalin was present in the flower buds during dormancy breaking. Theoretically, this could be explained by the fact that the nitrogen released by prunasin and amygdalin degradation is utilised during flower development. Robinson (1929) was the first who suggested that cyanogenic compounds could be a nitrogen source. In the cyanogenic glucoside degradation pathway, one of the final products is cyanide, which is released and immediately detoxified into asparagine or aspartate and NH_3. These final products would provide a supply of nitrogen for the plant physiological processes.

Finally, the cyanogenic diglucosides may play additional roles like nitrogen transport, pollinator attraction and seedling germination. For example, one study found the highest levels of diglucosides in the flower buds and expanded leaves of *E. camphora* trees (Neilson et al., 2011). The authors concluded that diglucosides are synthesised in the expanded leaves and then transported to the developing flower buds. The cyanogenic diglucoside levels were much lower in the immature fruits, suggesting that nitrogen was remobilised and used during the flower development period (Neilson et al., 2011).

3.4.5. HCN factor

The cyanide released from cyanogenic glucoside degradation could have an effect on dormancy breaking and flower opening (Tanaka et al., 1983; Barros et al., 2012). According to Gleadow et al. (2000), cyanide is consumed during flower bud development and is associated with a decrease in the cyanogenic glucoside levels.

Interestingly, cyanide is also involved in breaking seed dormancy and in germination (Roberts, 1973; Roberts and Smith, 1977) by inducing the formation of Reactive Oxygen Species (ROS). The ROS activate a cascade involving the Ethylene Response Factor 1 (ERF1), producing germination-associated proteins (Oracz et al., 2007). This process would be the same in flowering, i.e., the cyanide induces the ROS to stimulate flowering (Taylorson et al., 1973; Bogatek et al., 1991; Barros et al., 2012; Flematti et al., 2013.)

3.5. CONCLUSIONS

- In our study, prunasin and amygdalin were detected for the first time in flower buds during the flower development period in almonds. Furthermore, these cyanogenic glucosides were found to play an important role in the flower bud development.

- Amygdalin was present in flower buds during dormancy breaking, whilst prunasin had a role in the development and opening of the flowers. The nitrogen released by the degradation of the prunasin and amygdalin most certainly promotes flower development.

- Pollen was the tissue with the highest levels of amygdalin and prunasin and their putative derivatives.

- Finally, prunasin acid, amide, anitrile, prunasin pentoside and prunasin anitrile pentoside were detected for the first time in the flowers of sweet and bitter cultivars. These derivative compounds may be produced from cyanogenic glucosides in an alternative turnover pathway.

4. Fine mapping of the Sk locus

4. Fine mapping of the *Sk* locus

4.1. INTRODUCTION

Sweet kernel taste is one of the most important agronomic traits desired in the almond breeding programs. As we explained in the point 1.3.1 of the Introduction chapter, the almond bitterness is a monogenic trait, being the sweet allele dominant over the bitter one (Heppner, 1923; 1926; Dicenta and García, 1993; Dicenta et al., 2007). Despite this characteristic was already mentioned in Roman times by Plinio, the responsible gene has not been identified so far. Nowadays, we know that the *Sweet kernel* (*Sk*) locus is placed in the linkage group 5 (LG5), but its precise localization and function remains still unknown (Joobeur et al., 1998; Bliss et al., 2002; Sánchez-Pérez et al., 2007).

Molecular markers

A molecular marker for a given trait is a specific sequence of DNA very close to the gene responsible for this trait. The ideal molecular marker should be efficient, easy to use and cheap, with a highly polymorphism, codominant and reproducible. Some of the most useful markers, like SSRs (microsatellites), SNPs (Single Nucleotide Polymorphisms) and CAPs (Cleaved Amplified Polymorphisms) are based on the Polymerase Chain Reaction (PCR) technique (Mullis et al., 1986).

SSRs (Simple Sequences Repeats): They are short tandems repetitions of 1-6 nucleotides of known sequence and that can be amplified by a PCR (Litt and Luty, 1989). SSRs are polymorphic, abundant, codominant and reproducible. They require a low quantity of DNA and allow a high automation, with a reduced cost per analysis. However, microsatellites need the nucleotide repetition sequence information. These molecular markers have been applied to molecular characterization of *Prunus* species (Cipriani et al., 1999; Decroocq et al., 2003; Martínez-Gómez et al., 2003; Testolin 2004; Sánchez Pérez et al., 2005) and to the construction of linkage maps (Aranzana et al., 2003; Sánchez-Pérez et al., 2007; 2010; Salazar et al., 2013).

SNPs (Single Nucleotide Polymorphisms): These abundant molecular markers are based on variations of only one nucleotide. They are detected by sequencing DNA after a PCR. A SNP can be either non-synonymous (results in an amino acid sequence change) or synonymous (does not affect the amino acid sequence) (Sunyaev et al., 1999). These molecular markers have a very high polymorphism, high reproducibility, they are easy to use and the cost of the analysis is low. SNPs have been applied to saturate linkage maps in order to locate relevant traits in the genome and to find differences between cultivars. Nowadays, highly saturated linkage maps are constructed with SNPs arrays by genotyping by sequencing (GBS) (Guajardo et al., 2015). In *Prunus*, so far there are two SNPs arrays: the cherry 6K (Peace et al., 2012) and the peach 9K SNP array (Verde et al., 2012). Genotyping-by-sequencing has been used in many *Prunus* species with different families, but the efficiency of this high-throughput sequencing can be different from one to another family (Guajardo et al., 2015).

CAPs (Cleaved Amplified Polymorphisms): These molecular markers are based on SNPs. Sometimes a sequence with different SNPs creates a new cleavage position for a restriction enzyme which would give different fragment sizes in the electrophoresis (Akopyanz et al., 1992; Konieczny and Ausubel, 1993). Approximately, 30-40% of SNPs can be recognized by restriction enzymes and therefore CAPs can be developed. These molecular markers have a high reproducibility, they require low quantities of DNA (50-100 ng per reaction) and they are codominant, but they are not very polymorphic. These molecular markers have been applied to fine mapping, location of genes and allelic diversity studies.

Linkage maps

A linkage map is a representation of the genome. The linkage maps are useful to identify and locate genes that control important qualitative and quantitative traits in a wide range of species (Tanskley et al., 1989). The molecular markers can be located together with the traits in the linkage maps.

It is worth to indicate that there is a high synteny between the maps previously developed due to the low barriers between the different species of the *Prunus* genus and the high homology between loci. So the markers identified in previous maps were useful for developing other maps for the same or different species (Cipriani et al., 1999; Sosinski et al., 2000, Testolin et al., 2000; Cantini et al., 2001; Martínez-Gómez et al., 2003).

Some of the most important maps made in peach and almond are the following:

Year	**Authors**	**Species**
1994	Arús et al.	Texas (almond) x Earlygold (peach)
1995	Viruel et al.	Ferragnès (almond) x Tuono (almond)
1995	Fooland et al.	54P455 (peach) x Padre (almond)
1998	Joobeur et al.	Texas (almond) x Earlygold (peach)
1998	Jaúregui et al.	Ferragnès (almond) x Tuono (almond)
2001	Ballester et al.	Felisia (almond) x Bertina (almond)
2002	Bliss et al.	54P455 (peach) x Padre (almond)
2003	Aranzana et al.	Texas (almond) x Earlygold (peach)
2004	Dirlewanger et al.	Texas (almond) x Earlygold (peach)
2005	Verde et al.	(*P. persica x P. ferganensis*) x *P. persica*.
2005	Silva et al.	Texas (almond) x Earlygold (peach)
2007	Sánchez-Pérez et al.	R1000 (almond) x Desmayo (almond)
2010	Tavassolian et al.	Nonpareil (almond) x Lauranne (almond)
2011	Fernández i Martí et al.	Vivot (almond) x Blanquerna (almond)
2012	Font i Forcada et al.	Vivot (almond) x Blanquerna (almond)
2013	Fernández i Martí et al.	Vivot (almond) x Blanquerna (almond)

Fine mapping

Fine mapping is a methodology used to identify the locus of a gene by saturating the region with molecular markers. In recent years some works have identified genes by this method. In peach, Bouderi et al. (2009) fine mapped the *D* locus responsible for fruit acidity, delimiting this locus to an interval of 0.4 cM, and Pirona et al. (2013) found candidate genes of a major locus controlling maturity time, shortening the interval from 3.56 Mb to 220 kb. In Myrobolan plum (*Prunus cerasifera*) the *Ma* locus was fine mapped from 70 to 32 kb, finding three candidate genes (Claverie et al., 2011). Very recently, a QTL for powdery mildew resistance in cucumber (*Cucumis sativus* L.), named Pm1.1, was fined mapped (Xu et al., 2016). The QTL analysis delimited the region to 41.1 kb, in which eight genes were predicted. After the gene expression analysis only two candidate genes were detected.

The Sk *locus*

The first studies on this locus data from 1997, when Werner and Creller determined that the loci *G* (for 'skin hairiness' in nectarine) and *Sk* (sweet kernel in peach) were linked to a distance of 12 cM. Later, Joobeur (1998) assigned the bitterness of the seed *Sk* to LG5 in the map TxE (almond x peach), what it was confirmed by Bliss et al. (2002). As previously mentioned in the Introduction chapter, Sánchez-Pérez et al. (2007) developed an almond linkage map with 56 SSRs, in which the *Sk* locus was located in LG5, covering 50 cM out of 400 cM of this map (R1000 x Desmayo Largueta) (Figure 4.1). Out of the 8 SSRs mapped on LG5, PceGA025 was the closest marker to the *Sk* locus.

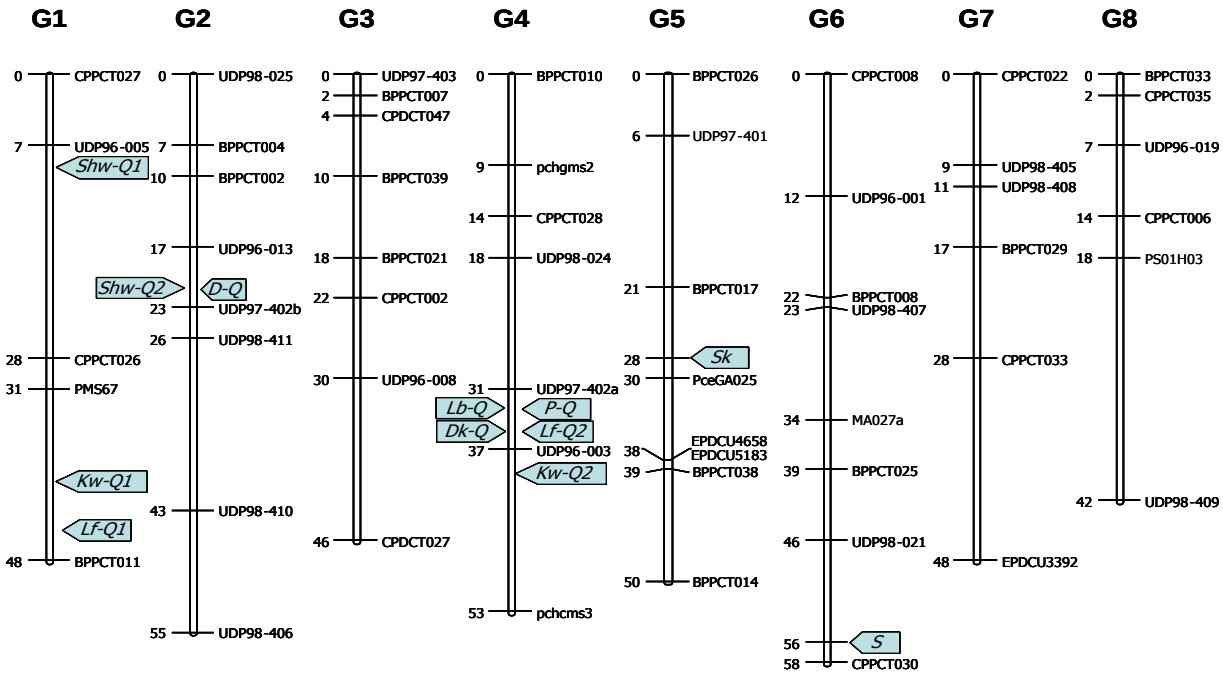


Figure 4.1. Molecular linkage map constructed with the JoinMap software using 56 SSRs, with the R1000 x Desmayo Largueta (R x D) F1 progeny (Sánchez-Pérez et al., 2007).

Later, Sánchez-Pérez et al. (2010) saturated the LG5 of RxD map with 12 new SSRs markers, covering 58.1 cM. Six of them, named UDA-045, EPDCU2584, CPDCT028, BPPCT037, PceGA025 and CPDCT016, were detected close to the *Sk* locus. When compared these markers of LG5 in almond with those of the LG5 of peach genome v.1.0, the physical distance between BPPCT017 and BPPCT038 was about 3.6 Mb (Sánchez-Pérez et al., 2010, Figure 4.2). The most interesting result of this work was the development of the haplotypes for the bitterness *Sk/sk* gene and its flanking SSR markers of some of the almond accessions studied. However, these markers were only useful when a pre-study of the parental genotype for these SSRs was previously done. So, it is necessary to develop universal *Sk* markers, efficient for any offspring or cultivar considered.

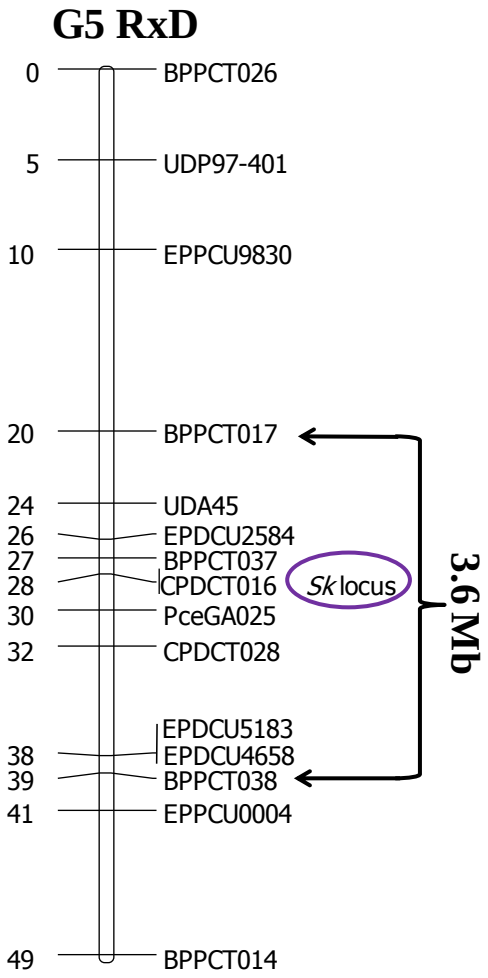


Figure 4.2. Saturation of linkage group 5 in the R x D population with JoinMap software v. 4.1 and localization of the *Sk* locus (Sánchez-Pérez et al., 2010).

Koepke et al. (2013) evaluated the polymorphism of this 3.6 Mb region using four almond accessions from CEBAS-CSIC, two bitters (S3067 and D05-187) and two sweets (Ramillete and Lauranne). Flanking the *Sk* locus, between the BPPCT017 (placed in G5 at 11Mb) and BPPCT038 (placed in G5 at 14.6 Mb) markers, 228 non-synonymous SNPs were associated with the *Sk* gene (Table 4.1). A total of 311,497 polymorphisms were identified in LG5 and 56,155 between BPPCT017 and BPPCT038. After removing the no homozygous polymorphisms for both (sweet/bitter) cultivars and within both bitter accessions this figure was reduced to 6,304, of which 228 caused codon-changing mutations.

Table 4.1. Number of polymorphisms comparing two sweets (Ramillete and Lauranne) and two bitter (S3067 and D05-187) almond genotypes in the linkage group five (Koepke et al., 2013).

Region / Filter	**Number polymorphisms**
A. Chromosome 5 (G5)	311,497
+ 3.6 Mb	56,155
+ fitting genetic patterns	6,304
+ with codon change	228 (120 genes)
+ stop codon	8

According to this background, the objective of this chapter is to develop a molecular marker linked to the kernel bitterness of almond by saturating the *Sk* locus located in LG5, which could be used for MAS (marker assisted selection) in almond programs all over the world.

To achieve this goal, three synergic strategies were planned: a) Fine mapping saturating the *Sk* locus with new SSRs and CAPS based on SNPs; b) Resequencing of a sweet (Lauranne) and a bitter (S3067) almond genotypes; c) RNA-seq analysis of the tegument of Lauranne and S3067 at two different times during fruit development.

4.2. PLANT MATERIAL

For this chapter, 550 seedlings of R1000 x Desmayo Largueta (RxD) population (described in paragraph 2.2) were studied. Moreover, the bitter homozygous S3067 and the sweet homozygous Lauranne were used in the genome and transcriptome analysis.

4.3. METHODOLOGY

The phenotypic characterization of the bitter, sweet or slightly bitter flavor of kernels of each seedling of R1000 x Desmayo Largueta was performed by tasting two seeds for each tree.

4.3.1 SNPs analysis and development of CAPS markers

SNPs analysis

In order to validate eleven SNPs out of the 228 SNPs identified in two sweets and two bitters almond genotypes (Koepke et al., 2013), a series of primers (Annex 8.17) were designed to obtain amplicons of 200-400 bp by "Primers 3 plus" software (Rozen and Skaletsky, 2000). Sometimes, the optimal temperature conditions for some primer were defined using a gradient thermocycler (SimpliAmp™ Thermal Cycler), performing a "Gradient PCRs" (Lopez and Prezioso, 2001), or/and "Touchdown PCRs" applying the protocol of R.H. Don et al. (1990).

Once the primers were designed, a screening on parental (R1000 and Desmayo) and six individuals of the F_1 population was carried out by Polymerase Chain Reaction (PCR). In particular, the proofreading Phusion High-Fidelity DNA Polymerase (New England Biolabs) was used as described in Annex 8.17.

The PCR products were purified, by using the "Nucleo Spin Extract II" kit (Macherey Nagel, Germany) or "QIAquick Gel Extraction kit" (Qiagen), depending on the size and the DNA content amplified. Once purified, they were sent to sequence to "Macrogen Europe" (Netherlands) or "Eurofins Genomics" (Germany).

The electropherograms obtained were analyzed by using the software "CLC Sequence Viewer 7", aiming to identify Single Nucleotide Polimorphisms (SNPs) between the two parents and the offspring of the RxD F_1 population.

Development of CAPs

In some SNPs detected, a restriction enzyme analysis was applied to convert them into CAPs by using the software "CAPs Designer". Indeed, digestion reactions

with the endonucleases were performed according to the protocol of Neff et al. (1998). The digestion mix used for every restriction enzyme, performed for each sample, is showed in Annex 8.18.

All samples were incubated in a water bath at specific temperature and time of incubation for every enzyme (*New England Biolabs*) (Annex 8.17). The digested fragments were visualized by Agarose or MetaPhor Agarose gel electrophoresis.

All CAPs were used to genotype 550 individuals of RxD F_1 population. Marker segregation data were analysed through the JoinMap 4.1 software, in order to develop the genetic map.

4.3.2 Microsatellites assay

SSR markers UDA045 and CPDCT028 (previously described in Sánchez-Pérez et al., 2007) were tested by "DreamTaq™ DNA Polymerase" (Thermo Scientific™) (Annex 8.19) in our 550 seedlings of RxD population.

On the other hand, SSR markers EPDCU2584 and BPPCT037 (Sánchez-Pérez et al., 2007) were analyzed by Capillary Electrophoresis Technique, designing with "Primer3plus: a Forward primer modified by adding a M13* tail (5'TGTAAAACGACGGCCAGT3') of 18 bp to the 5' end (Schuelke, 2000), and a reverse primer not labeled. Moreover, a universal M13* primer that was labeled with Fam (blue) or Hex (green) fluorescent dyes (Sigma Genosys) was added. Amplification reactions were carried out using a "DreamTaq™ DNA Polymerase" (Thermo Scientific™) as described in Annex 8.19.

Amplification products (1.2 µl) were added to 15 µl HiDi formamide (Applied Biosystems, Foster City, CA) and 0.3 µl GeneScan™-500 ROX™ Size Standard and, after a denaturation at 94 °C for 5', they were run on a 36 cm electrophoretic capillary (ABI-3500 Genetic Analyzer, Applied Biosystems, HITACHI, Foster City, CA, USA). The electropherograms transferred to the *Workstation*, were analyzed with the software "GeneMapper v. 5.0".

4.3.3. Fine mapping of the *Sk* locus

All the segregation phenotypic and genotypic data were uploaded in a LOC excel file and analyzed with the *JoinMap* v4.1 software (Van Ooijen, 2011) in order to produce a new genetic map. The logarithm of the odds (LOD) score threshold value was set to 3.0. The order and distance between the markers were determined using *JoinMap*'s *Maximum Likelihood* mapping algorithm for *cross-pollinated population (CP)*. Default mapping parameters were assumed with the following modifications: spatial sampling thresholds reduced to 0.050, 0.025, 0.015, 0.010 and 0.005.

4.3.4. Genome resequencing and transcriptome analysis

Genome resequencing

For the genome resequencing and the transcriptome analysis a homozygous sweet cultivar (Lauranne) and a homozygous bitter cultivar (S3067) were used.

Lauranne and S3067 were resequenced by two different NGS platforms: 1) *Illumina HiSeq* with pair-end reads (100 bp) and mate pairs (5 kb) at Genome Quebec (Canada) and 2) *Pacific Bio (PacBio)* with a 13 kb library size at Washington State University (USA).

The results of the two NGS platforms generated the data shown in Table 4.2. If we consider that almond genome size is about 240 Mb (Alioto et al., 2016), then the total coverage would be between 52x and 65x.

Table 4.2. Resequencing of two almond genomes by Illumina and PacBio, and coverage obtained by these two NGS platforms, in a sweet (Lauranne) and a bitter (S3067) cultivar.

Cultivar	Data Type	Amount	Coverage (x)	Genome Size
S3067	Illumina (100 bp)	14.6 Gb	60.42	
	Pac Bio	1.1 Gb	4.56	
	Total	15.7 Gb	64.98	~ 240 Mb
Lauranne	Illumina (100 bp)	11 Gb	45.6	
	Pac Bio	1.6 Gb	6.61	
	Total	12.7 Gb	52.21	

A *de novo* assembly using AllPath was done and the statistics obtained are shown in Table 4.3.

Table 4.3. Statistics of sweet and bitter genome assemblies.

	Bitter	**Sweet**
Combined length (bp)	144,664,161	131,238,080
Percent GC content	37.24	37.32
Number of sequences	5,035	5,843
Mean sequence length	28,731.71	22,460.74
N50	122,189	93,377
N90	16,255	12,366
Longest sequence	989,688	491,740

When compared this results to the very recent "Draft of the Almond Genome" (Alioto et al., 2016) we could observed that the estimated size of the almond in our work (144-131 Mb) was about 40% lower than the one estimated (240 Mb) by Alioto et al. (2016). This was due to the fact that between 40-45% of our sequences were containing repetitiveness, making the assembly very difficult. Right now we are sending new nuclear DNA to make another round of PacBio sequencing in order to make a better assembly of the almond genome.

Transcriptome analysis

Four transcriptomes from teguments (seed coats) belonging to Lauranne and S3067, at two different times (18th March and 16th April, 2014), were sent to sequence by *Illumina Hi-Seq* 100 bp (Macrogen) using a mRNA True Seq Library (non-stranded one). Enough reads were obtained to do a *de-novo* assembly of the almond transcriptome. Prior to further analysis, a quality check was performed on the raw sequencing data (Table 4.4), removing low quality portions while preserving

the longest high quality part of a NGS read. The minimum length established was 35 bp and the quality score 25, which increases the quality and reliability of the analysis.

Table 4.4. Statistics about the transcriptome assembly before and after cd-hit-est clustering.

Scaffold	Raw	Clustered
Total genes	100,668	75,743
Total transcripts	213,096	138,585
Percent GC	41.11	40.92

Having together the two datasets (genome and transcriptome) the workflow was performed by different steps by Sequentia Biotech (Figure 4.3).

The schema of this procedure is presented below, and it is explained in detail in the Annex 8.20.

Step 1. Almond reference genome.

Step 2. Genome annotation.

- Step 2.1 Transcript clustering.
- Step 2.2 Structural annotation (PASA).
- Step 2.3 Functional analysis.

Step 3. Differential expression analysis (RNAseq).

- Step 3.1 Data pre-processing.
- Step 3.2 Statistical analysis.
- Step 3.3 Differential expression (DE) analysis.
- Step 3.4. Network analysis.

Step 4. Variant calling

- Step 4.1 Mapping and pre-processing.
- Step 4.2 Variant calling.
- Step 4.3 Variant filtering.
- Step 4.4 Variant annotation.

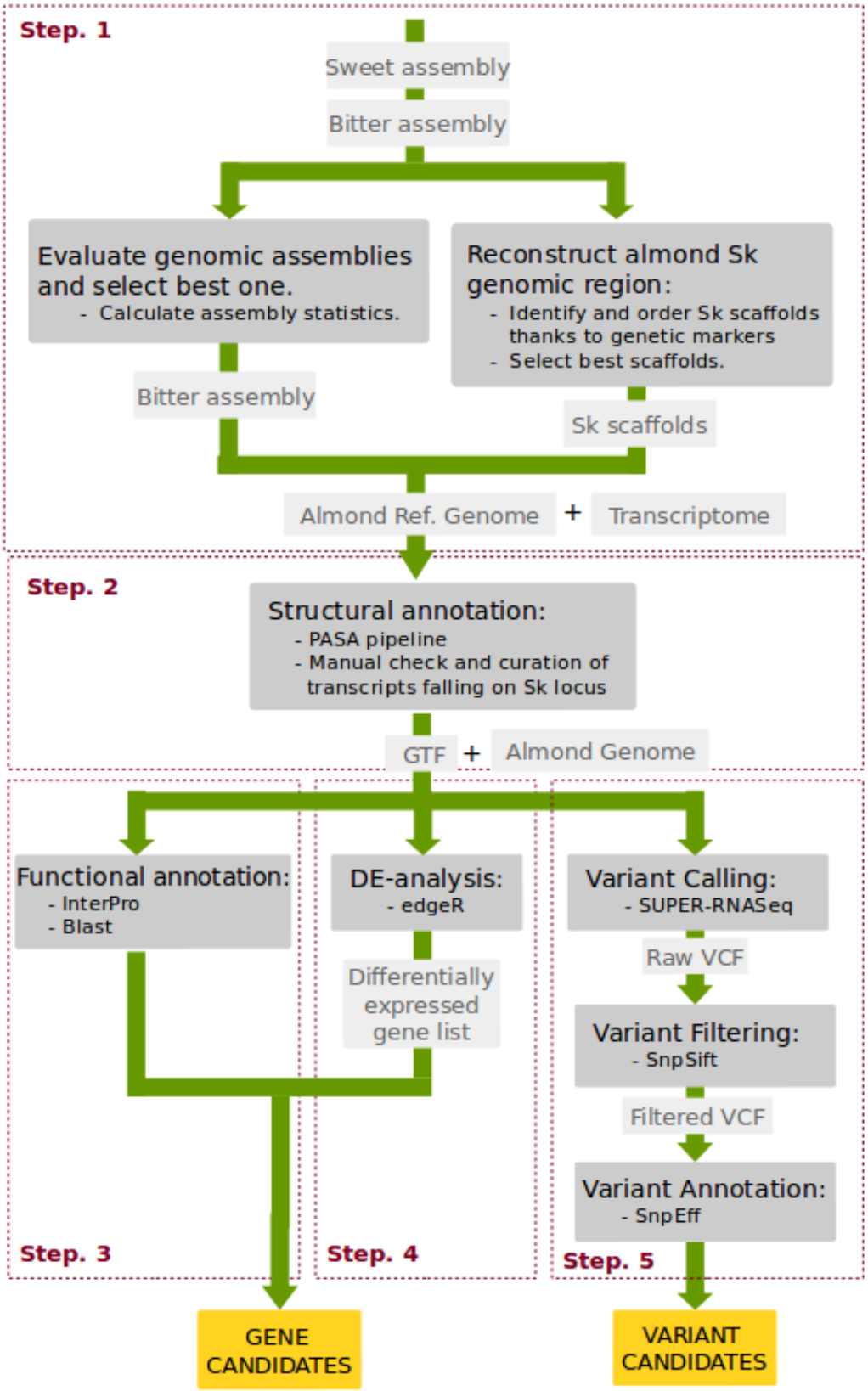


Figure 4.3. Workflow of the methodology used to find gene and variant candidates for the *Sk* locus in almond.

4.4. RESULTS AND DISCUSSION

4.4.1. Fine mapping of *Sk* locus

4.4.1.1. SNPs analysis

Out of the 228 codon-changing mutations encoded by 120 genes (Koepke et al., 2013), only eight genes generated a non-sense amino acid change (stop codon). Apart from these eight genes, three more without stop codon, but interesting for annotation, were include in this study. In total, 11 genes were analyzed in 167 individuals (RxD population) within the 3.6 Mb region, according to the peach genome (Table 4.5). Their functions, and the position of the amino acid changed (to generate or not a stop-codon), are also shown.

Table 4.5. Genes analyzed in the F1 RxD population within the range of 3.6 Mb where the *Sk* locus is placed (referred to the peach genome), and the position of the amino acid changed (in parenthesis). *Genes with stop codon mutation.

Genes	Function
Mis-sense mutations	
ppa005388m (intron)	Predicted hydrolase/acyltransferase
ppa018792m (S11T)	Multi-trans-membrane protein within sugar efflux transporter for intercellular exchange family
ppa001981m (R465L)	Predicted NAD synthase, contains CN hydrolase domain
Bitter non sense	
ppa008772m (Y175*)	Sequence specific DNA binding transcription factor activity
ppa024141m (E334*)	Protein binding with unknown function
Sweet non sense	
ppa001291m (E407*)	Regulation of transcription, DNA dependent
ppa006138m (Y275*)	5´AMP activated protein kinase, gamma subunit
ppa024207m (Y131*)	β-glucosidase, hydrolyzing O glycosyl compounds
ppa003453m (E258*)	GTPase activator protein of rab like small GTPases
ppa004278m (E45*)	Protein, O-Linked N-Acetylglucosamine Transferase, OGT
ppa023181m (Y31*)	Protease inhibitor/seed storage/LTP family

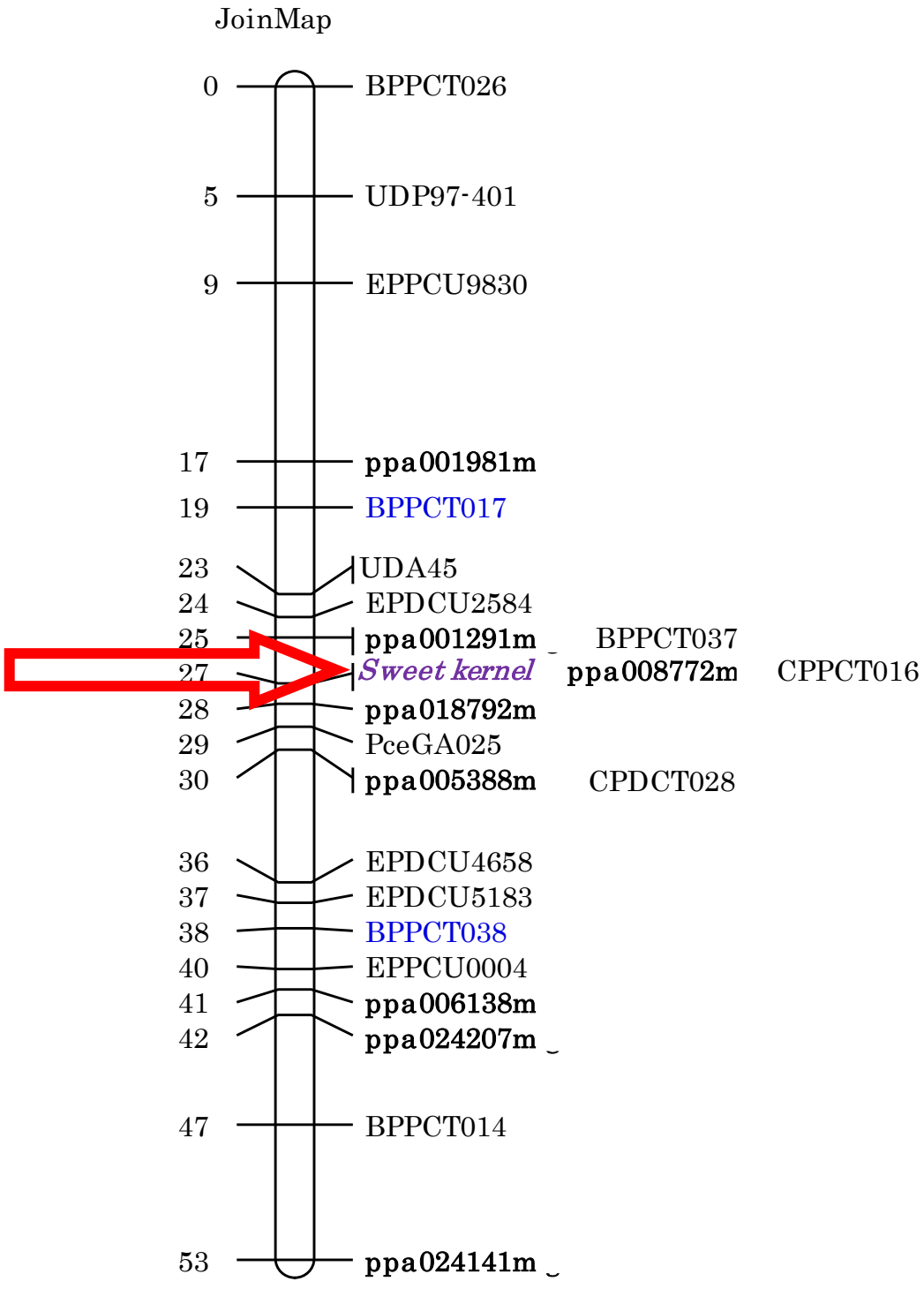


Figure 4.4. Saturation of LG5 of RxD population with JoinMap v4.1 and localization of the *Sweet kernel* locus. Red arrow indicates the *Sk* locus. In bold, the new genes identified in this linkage group. *Analyzed only in 96 seedlings.

Out of 11 candidate genes analyzed, eight (in bold in the Figure 4.4) in the G5 of the RxD map were successfully localized. They were ppa001981m, ppa006138m, ppa024207m, ppa024141m, ppa001291m, ppa008772m, ppa018792m and ppa005388m, being the last four the closest to the *Sk* locus. Markers with an asterisk were only analyzed in 96 individuals of the RxD (Figure 4.4).

Moreover, when compared these results with the peach physical map only for ppa006138m, ppa024207m and ppa024141m, the position of the candidate genes did not match. In fact, their positions were outside of the 3.6 Mb region (Figure 4.2; 4.4). One reason could be that the analysis of these genes was only performed in 96 individuals of the RxD population. The other reason could be that the synteny between almond and peach in this region is not as conserved as in other regions, suggesting that bitterness in almond and peach could not have the same origin.

4.4.1.2. Development of CAPS markers and microsatellite analysis

In order to shorten the distance between the markers and the *Sk* locus in the RxD map, and to find out if those discrepancies above mentioned between almond and peach were consistent, CAPS and *Sk*-linked SSR markers were used to genotype the RxD population, increasing the number of seedlings from 167 to 550.

Consequently, new CAPs markers between EPDCU2584 and CPDCT028 were developed (Figure 4.5), covering a physical region of about 800 kb (Figure 4.6). Inside this 800 kb region, several SNPs were detected and two new CAPs markers (ppa001838m and ppa006282m) were found above the other two CAPs previously developed (ppa005388m and ppa018792m). These four CAPs were analysed in the 550 individuals of RxD population.

CAPs developed in ppa001838m, ppa006282m, ppa005388m and ppa018792m were located to 2.9, 3.8, 4.6 and 6.2 cM from the UDA045 marker. CAPs ppa005388m was the closest marker to the *Sk* locus (0.1 cM), followed by ppa006282m (0.7 cM), ppa001838m (1.6 cM) and ppa018792m (1.7 cM). The heredity of the studied markers is shown in Table 4.6. According to the genotype

alleles, we only were able to detect sweet phenotypes but no bitter with the CAPs developed, while with the SSRs EPDCU2584 and BPPCT037 we could detect both sweet and bitter phenotypes.

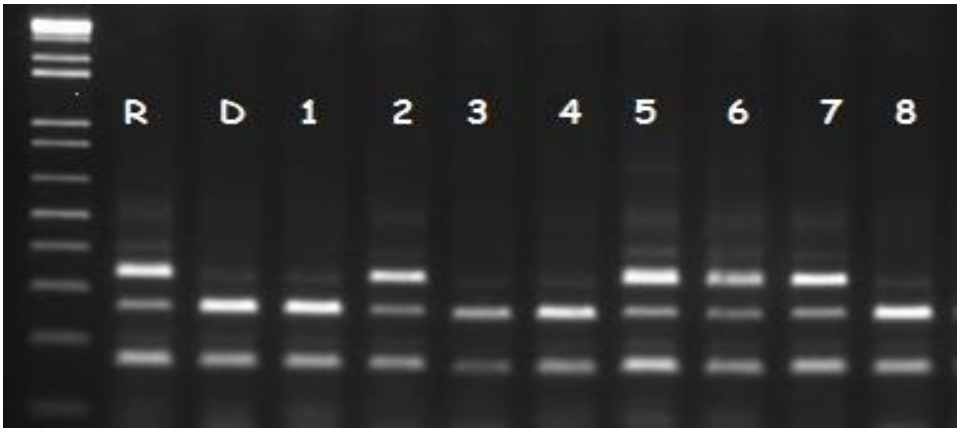


Figure 4.5. Segregation of CAP ppa018792m for R1000, Desmayo Largueta and 8 RxD individuals by MetaPhor Agarose gel electrophoresis.

Table 4.6. Segregation of genotypes (molecular markers) and phenotypes (bitter, slightly bitter and sweet), as percentage of seedlings in a population of 550 descendants of RxD.

	Genotype		**Phenotype**		
Marker	**alleles**	**%**	**Bitter**	**Slight. bitter**	**Sweet**
UDA045	nn	47	3	-	97
	np	53	54	-	46
EPDCU2584	hh	23	1	-	99
	hk	47	3	1	96
	kk	29	97	-	3
BPPCT037	ac	23	1	-	99
	ad	24	5	1	94
	bc	24	2	-	98
	bd	29	97	-	3
ppa001838m	nn	53	51	-	49
	np	47	1	-	99
ppa006282m	lm	48	1	-	99
	ll	52	55	-	45
ppa027182m	lm	51	53	-	47
	ll	49	1	-	99
ppa005388m	lm	53	54	-	46
	ll	47	1	-	99
CPDCT028	nn	46	1	-	99
	np	54	54	1	45
ppa018792m	lm	53	54	-	46
	ll	47	2	-	98

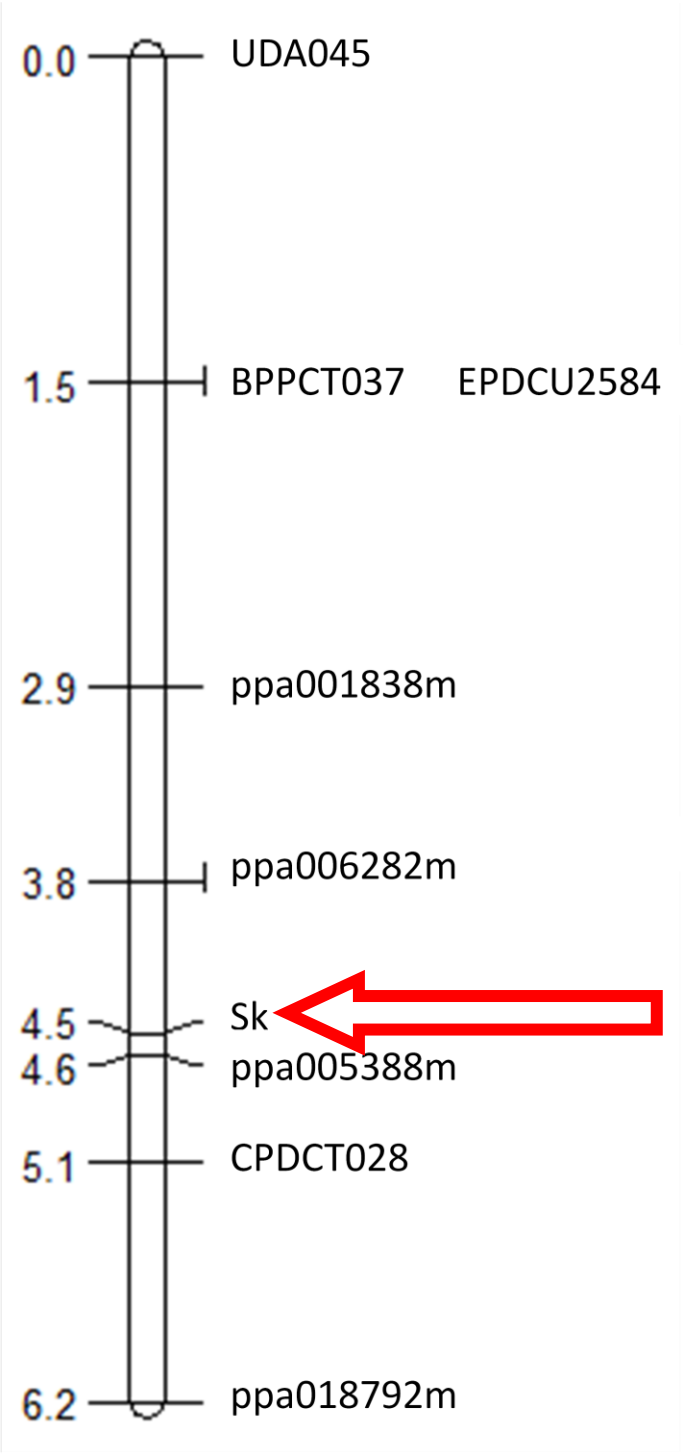


Figure 4.6. Saturation of LG5 with CAPs in the RxD population with JoinMap v.4.1 and localization of the *Sk* locus (marked with a red arrow).

Notably, mapping of new developed CAPs and the four previously reported SSRs revealed full overlap with the peach physical map (Figure 4.7), thus substantiating the notion that *Prunus* genomes are co-linear (Dirlewanger et al., 2004). In fact, after having analysed the segregation of the new and old markers in the RxD population, we found some discrepancies in the phenotyping and genotyping when the markers were previously analysed in the smaller population.

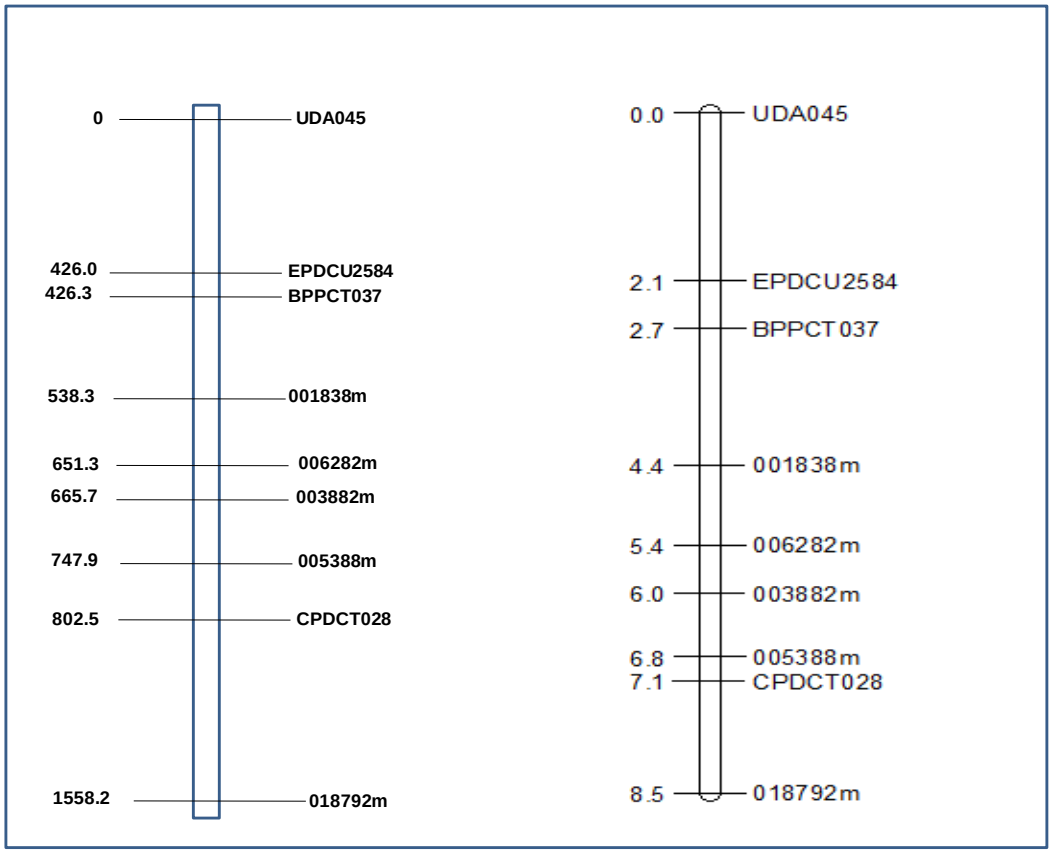


Figure 4.7. Co-linearity in LG5 between the physical map of in peach (left, in Kb) and the genetic map almond (right, in cM). Marker distances are indicated in Kb and cM, respectively.

Thank to the good collinearity between peach an almond, we found in the peach map eleven candidate genes in a region of 95.76 kb, between ppa006282m and ppa005388m markers. Six of them were transcription factors of the MYC family, the others were a glyoxal oxidase, a cytochrome P450, an alcohol O-acetyltransferase, a methionine lyase and a hydrolase (Table 4.7). Therefore, we fine mapped the *Sk* locus from 3.6 Mb to 95.76 kb.

Table 4.7. Candidate genes and predicted functions in 95.76 kb region flanking the *Sk* locus.

Genes	Predicted functions
ppa006282m	Alcohol O-acetyltransferase
ppa005470m	Methionine lyase
ppa003882m	Cytochrome P450
ppa011942m	Transcription factor MYC
ppa023406m	Glyoxal oxidase
ppa022201m	Transcription factor MYC
ppa025417m	Transcription factor MYC
ppa027182m	Transcription factor MYC
ppa015634m	Transcription factor MYC
ppa005343m	Transcription factor MYC
ppa005388m	Hydrolase

4.4.2. Genome resequencing and transcriptome analysis

In order to check if the candidate genes found in peach were also present in the almond genome, the almond genome resequencing and a transcriptome analysis were performed in a sweet (Lauranne) and a bitter (S3067) genome. Tegument tissue was chosen because it is the only maternal tissue of the seed, and bitterness is controlled by the female parental (Kester and Asay, 1975). Furthermore, the content of prunasin (precursor of the amygdalin) is very different in sweet and bitter genotypes (Sánchez-Pérez et al., 2008).

4.4.2.1. Bitter assembly as the reference genome

After alignments, all scaffolds of sweet and bitter genotypes falling within the *Sk Prunus* region in scaffold 5 were extracted. Moreover, the scaffolds covering as much as possible the *Sk* locus with no overlap were selected (Figure 4.8, Table 4.8).

Therefore, the final almond reference genome was obtained by joining together the bitter assembly and the scaffolds covering the *Sk* locus. The bitter assembly was proposed as the reference genome for the *Sk* gene in the GL5.

4.4.2.2. Genome annotation

In total, 15 transcripts were annotated on *Sk* locus, clustered within 10 genes (Table 4.9). The PASA_cluster_23031 had a sequence match with a *Pseudomonas* gene suggesting that it might be a contamination of bacterial DNA that was carried out to the genome assembly step, so we did not consider it in the following analyses.

4.4.2.3. Differential expression analysis: RNA-seq

Using RNASeq data, a precise quantification of gene expression levels on almond individual samples was carried out. All changes in gene expression levels among sweet and bitter samples were captured, compared and analysed, with a particular focus on *Sk* region annotations. The transcriptome of the teguments of the sweet Lauranne and the bitter S3067 in March and April were compared. By this way 3,168 genes differentially expressed were found in the sweet genotype between March and April and 3,817 genes in the bitter genotype between March and April.

If we consider the genes differentially expressed in sweet almond at both times, we came down to 1,382 in total. Among them, only nine genes were annotated in the *Sk* locus (Table 4.10).

Table 4.8. Almond scaffolds covering *Sk* locus.

Scaffold	Assembly	Size (kb)	Molecular marker
Unscaf_bitter_scaffold_176	Bitter	91,675	ppa001838
Unscaf_bitter_scaffold_403	Bitter	91,675	ppa003882
Unscaf_bitter_scaffold_1398	Bitter	23,271	-
Unscaf_sweet_scaffold_483	Sweet	357,641	ppa005388

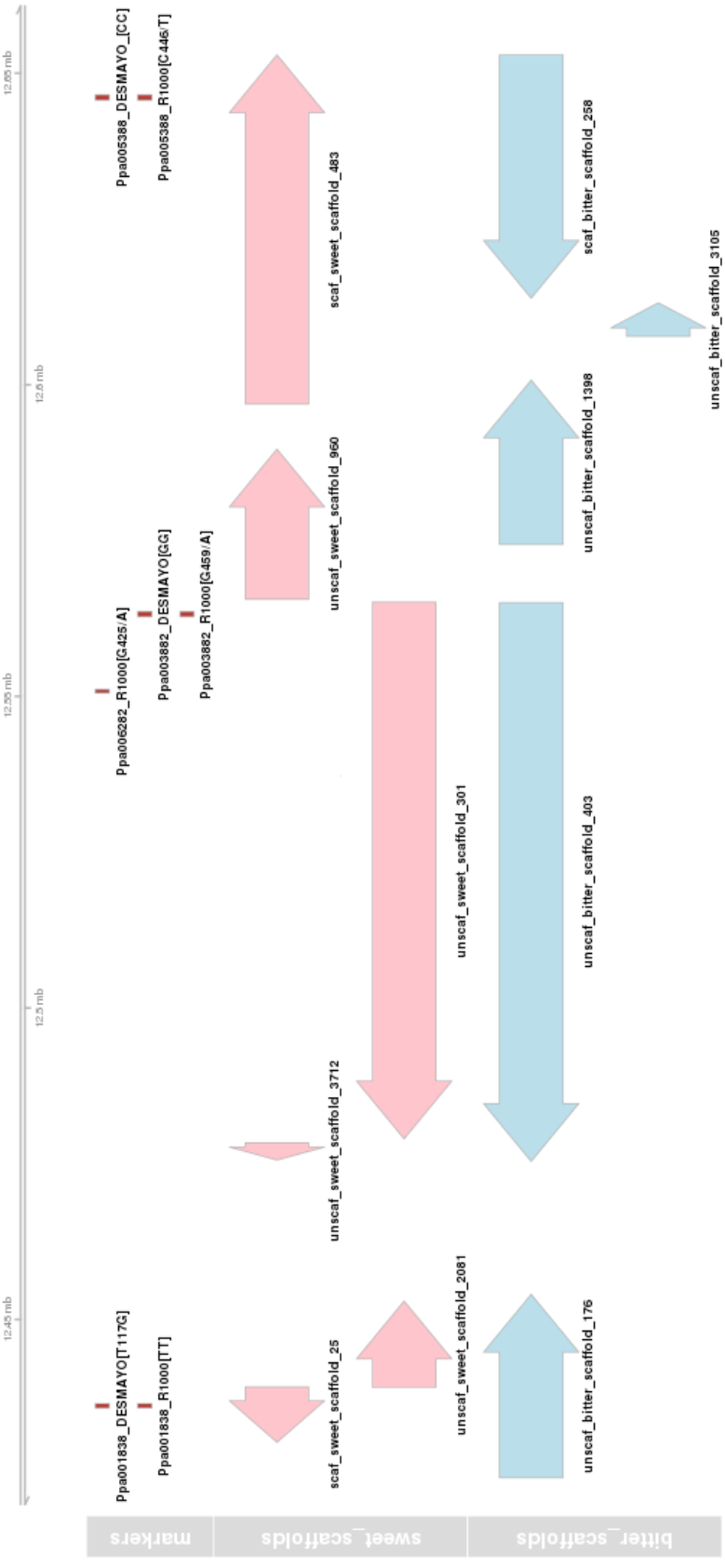


Figure 4.8. Representation of the mapping of the sweet and bitter scaffolds on the putative *Sk* locus using the *P. persica* genome as reference. On the top, the position of the molecular markers analysed is shown. In the middle and the bottom (sweet_and bitter_scaffolds) the position and orientation of the sweet and bitter scaffolds mapped on the region are shown. The names of the selected scaffolds are listed in Table 4.8.

Table 4.9. List of the loci annotation in the *Sk* locus in the almond assembly. The columns report the name of the locus, the name of the scaffold and the description of the genes.

Locus	Scaffold	Description
align_id:103783/c50965_g1_i4	scaf_sweet_scaffold_483	*Prunus mume* protein PHLOEM PROTEIN 2-LIKE A1-like (LOC103341947)
align_id:25184/c33639_g1_i1	scaf_sweet_scaffold_483	Transcription factor MYC/MYB N-terminal
gene_curated_1	scaf_sweet_scaffold_483	Transcription factor MYC/MYB N-terminal
gene_curated_2	scaf_sweet_scaffold_483	Transcription factor MYC/MYB N-terminal
gene_curated_3	unscaf_bitter_scaffold_403	Cytochrome P450
PASA_cluster_23031	scaf_sweet_scaffold_483	Alpha/beta hydrolase fold-1
PASA_cluster_23048	scaf_sweet_scaffold_483	*Prunus persica* clone JGIBIXB-13/11
PASA_cluster_27122	unscaf_bitter_scaffold_1398	Glyoxal oxidase, N-terminal
PASA_cluster_27128	unscaf_bitter_scaffold_1398	Mediator of RNA polymerase II transcription subunit 10b
PASA_cluster_27129	unscaf_bitter_scaffold_1398	*Malus x P. domestica* uncharacterized ncRNA

Table 4.10. Analysis of the differentially expressed genes (reported as the logFC: fold-change in logarithmic scale) annotated in the *Sk* locus, and correlation cluster. NDE= Non differentially expressed.

		Lauranne vs S3067		
Locus	**Description**	**18 March**	**16 April**	**Correlation Cluster**
Gene_curated_1	ppa022201m.g (MYC)	1,20	1,05	Cluster2
align_id:25184|c33630_g1_i1	ppa025417m.g (MYC)	0,75	-1,81	Cluster1
Gene_curated_2	ppa005343m.g (MYC)	-1,04	-2,68	Cluster1
PASA_cluster_23031	ppa005388m.g (Hydrolase)	0,71	-0,30	
PASA_cluster_27122	ppa023406m.g (Glyoxal Oxidase)	2,30	4,96	
Gene_curated_3	ppa003882m.g (Cytochrome P450)	-2,58	-5,13	Cluster2
PASA_cluster_23048	Prunus persica clone JGIBIXB-13I11	NDE	NDE	
PASA_cluster_27128	ppa011942m.g (NUT2)	NDE	NDE	Cluster1
PASA_cluster_27129	*Malus x domestica* uncharacterized LOC103409623 ncRNA	NDE	NDE	

The results of the RNA-seq are shown in Table 4.10. Nine candidate genes were found in the *Sk* locus, including three not differentially expressed.

Among the other six genes, three MYC transcription factors were differentially expressed (corresponding to *P. persica* ppa005343m, ppa022201m, ppa025417m).

The first MYC (ppa02201m) was up-regulated in the sweet genotype in both times. The second MYC (ppa025417m) was up-regulated in March and down-regulated in April in the sweet genotype. The third MYC (ppa005343m) was down-regulated in the sweet genotype in both times.

An additional gene, corresponding to *P. persica* ppa003882m encoding a putative cytochrome P450, had the highest variation, being strongly down regulated in the sweet genotype.

Another gene named ppa005388m encoded a hydrolase and it was up-regulated in March and down-regulated in April in the sweet genotype.

Finally, the last gene found in the *Sk* region corresponding to *P. persica* ppa023406m, that encode a putative glyoxal oxidase, was up-regulated in the sweet genotype (Figure 4.9).

Since the three transcription factors of the MYC family showed a subtle difference of the expression between sweet and bitter almonds, we perform a network analysis to identify the genes that might be regulated by them, involved in the amygdalin metabolism. We found that the three MYC genes belonged to two different clusters (Cluster 1 and 2) (Table 4.10). Gene_curated_1 (corresponding to *P. persica* ppa022201m) belongs to Cluster 2 and it is strongly co-regulated with gene_curated_3 (encoding a putative cytochrome P450) with a Pearson correlation of -0.93. The MYC transcription factors, gene_curated_2 and "align_id:25184|c33630_g1_i1" belong to the Cluster 1. They show a correlation of 0.76 and they are in the same group of PASA_cluster_27128 (NUT2), which was not differentially expressed.

Applying the explained filters in the step 4.3 of the Annex 8.20, the number of variants decreased from 451,744 to 350,577 (Figure 4.10).

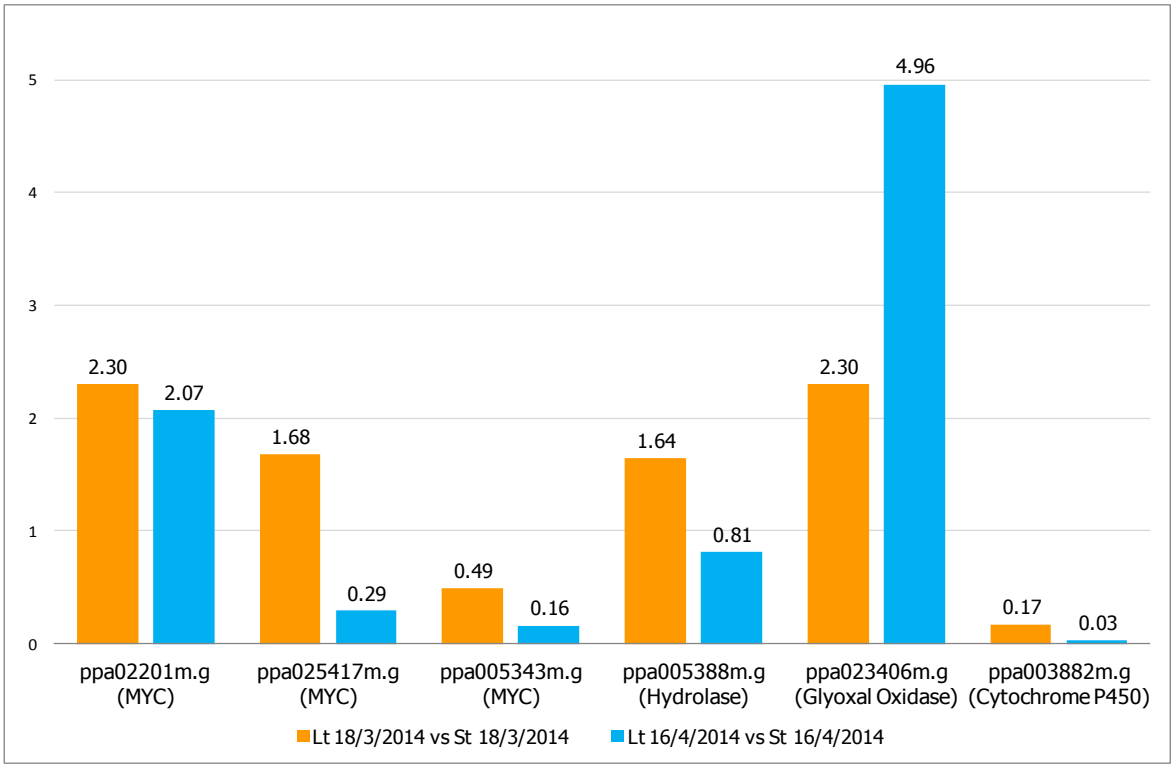


Figure 4.9. Comparison of the expression values (FPKM values) between the closest candidate genes found next to the *Sk* locus. Teguments of a sweet (Lauranne, Lt) and a bitter genotype (S3067, St) were compared in two different times (March: 18/03/14) and April: 16/04/14).

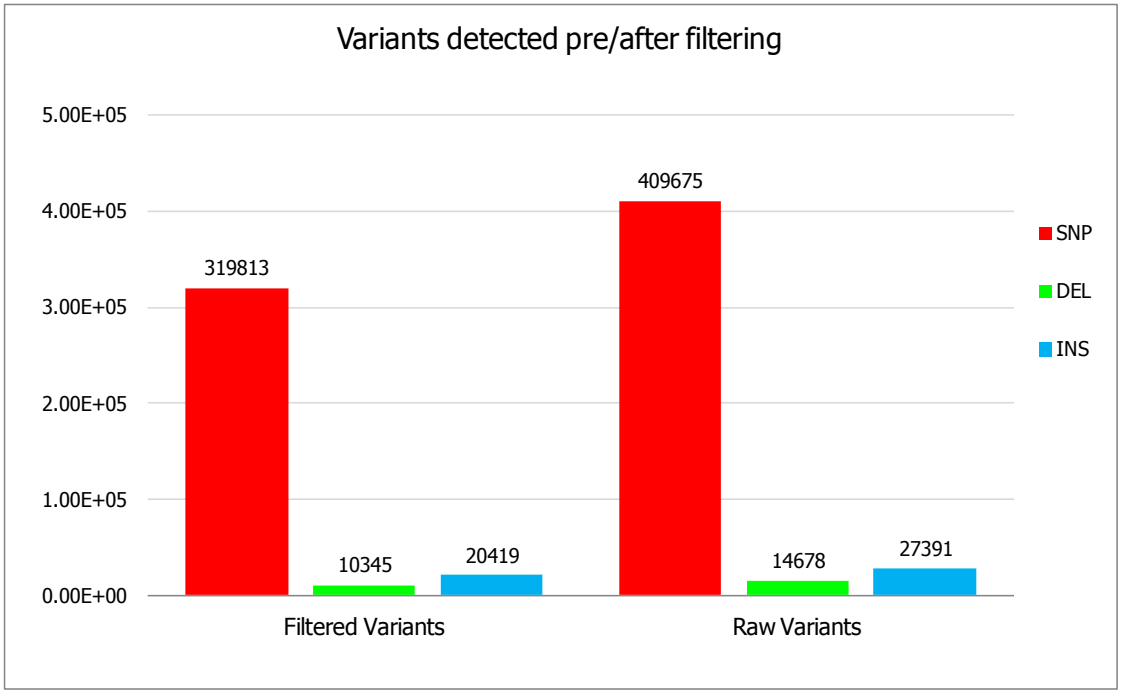


Figure 4.10. Variant calling results. The histograms represent the total number of raw variant detected by *SUPER* (Simply Unified Pair-End Read, see Annex 8.20) and the total variant obtained after the mentioned filters.

The results of the co-regulation analysis were intersected with the identification of candidate genes involved in amygdalin metabolism (see paragraph 5.4.3) to identify candidates that might be regulated by the transcription factors in the *Sk* locus.

4.4.2.4. Variant Calling. Identification of SNPs and INDELS

The variant annotation analysis was performed in those 63 variants detected inside the *Sk* locus.

Table 4.11 shows a summary of the potential variants candidates to be significantly affecting the considered candidate genes. Two MYC transcription factors and the putative glyoxal oxidase were differentially expressed. The putative subunit of RNA polymerase II was not differentially expressed. These variants were synonymous (the amino acid does not change) and miss-sense (the amino acid changes), but no non-sense variations (the amino acid changes into a stop codon).

The gene (ppa022201m) encoding a MYC transcription factor had two variants, one was synonymous and the other a missense. Interestingly, the missense mutation could change the amino acid residue from the positively charged lysine (in sweet) to the negatively charged glutamate (in bitter) that might have an effect on the function of the protein.

In the gene (ppa025417m) encoding the other MYC transcription factor a missense mutation was observed. In this case the aromatic phenylalanine (in sweet) was changed by leucine (in bitter).

As regards the gene (ppa023406m) encoding the glyoxal oxidase of the three observed SNPs, one was synonymous, one leaded to a change from alanine (in bitter) to valine (in sweet) while the third leaded to a change from the aliphatic glycine (in bitter) to the positively charged arginine (in sweet).

Table 4.11. Variant calling analysis done in the DE genes annotated in the *Sk* locus. The table shows the following terms: Gene_ID: The ID of the gene affected by the variant; Description: Gene description; Scaffold: Scaffold where the variant is included; Pos: Position of the variant in the scaffold; Sweet: Genotype for sweet phenotype in the given position; Bitter: Genotype for bitter phenotype in the given position; Type: Type of variant; Effect: Amino acid change due to the variant.

Gene_ID	Description	Scaffold	Pos	Sweet	Bitter	Type	Effect
gene_curated_1	(ppa022201m) Transcription factor MYC	scaf_sweet_scaffold_483	19320	T	C	synonymous	Ile->Ile
gene_curated_1	(ppa022201m) Transcription factor MYC	scaf_sweet_scaffold_483	19351	A	G	missense	Lys->Glu
align_id:25184| c33630_g1_i1	(ppa025417m) Transcription factor MYC	scaf_sweet_scaffold_483	24418	T	C	missense	Phe->Leu
PASA_cluster_23031	(ppa05388m) 2-hydroxymuconate semialdehyde hydrolase	scaf_sweet_scaffold_483	70341	T	C	missense	Lys->Glu
PASA_cluster_27128	(ppa011942m) Mediator of RNA polymerase II transcript.subunit 10b	unscaf_bitter_scaffold_1398	3769	G	A	missense	Asn->Asp
PASA_cluster_27122	(ppa023496m) Glyoxal_Oxidase	unscaf_bitter_scaffold_1398	12983	T	C	missense	Ala->Val
PASA_cluster_27122	(ppa023496m) Glyoxal_Oxidase	unscaf_bitter_scaffold_1398	14058	C	G	synonymous	Ala->Ala
PASA_cluster_27122	(ppa023496m) Glyoxal_Oxidase	unscaf_bitter_scaffold_1398	14695	C	G	missense	Gly->Arg

4.5. CONCLUSIONS

- A population of 550 seedlings of R1000 x Desmayo was phenotyped, increasingly the previous analysis from 167 to 550 individuals.

- Ten new CAPs and four SSRs were located in the GL5. These markers have allowed the saturation the *Sk* locus, fine mapping the region from 3.6 Mb to 95.76 kb. Acording to the peach genome, in this region we found eleven candidate genes, of which six of them were transcription factors of the MYC family, the others were a glyoxal oxidase, a cytochrome P450, an alcohol O-acetyltransferase, a methionine lyase and a hydrolase.

- The resequencing of the almond has allowed transporting these candidate genes from the peach to the almond.

- The bitter assembly was proposed as the reference genome for the *Sk* gene in the GL5, annotating 10 genes inside the *Sk* locus.

- The transcriptome analysis in almond has shown candidate genes differentially expressed in sweet and bitter almonds, being able to find six candidate genes involved in the *Sk* locus: three MYC transcription factors, a P450 cytochrome, a hydrolase and a glyoxal oxidase.

- Several SNPs were detected in three candidate genes present in the *Sk* region of 95.76 kb, two MYC transcription factors and the putative glyoxal oxidase. The gene (ppa022201m) encoding a MYC transcription factor had two variants, one was synonymous and the other a missense. In the gene (ppa025417m) enconding MYC transcription factor a missense mutation was observed. Finally, regarding to the gene (ppa023406m) enconding a glyoxal oxidase, out of the three SNPs observed, one was synonymous and the other two were missense.

5. Characterization of enzymes involved in almond bitterness

5. CHARACTERIZATION OF ENZYMES INVOLVED IN ALMOND BITTERNESS

5.1. INTRODUCTION

The metabolism of bitterness in almonds can be divided in three parts, accurately described in Figure 1.6 of the Introduction chapter and in Figure 5.1:

- Biosynthesis: Cyanogenic glucosides prunasin and amygdalin are synthesized from the amino acid phenylalanine through two cytochromes (CYPs) named CYP79 and CYP71 and two udp-glucosyltransferases (UGTs) named UGT1 and UGT2).
- Degradation: These compounds can be degraded by β-glucosidases (BGAs) amygdalin hydrolase (AH) and prunasin hydrolase (PH) and a mandelonitrile lyase (MDL1), liberating glucose, benzaldehyde (bitter) and hydrogen cyanide (toxic).
- Detoxification: Hydrogen cyanide is detoxified through β-cyanoalanine synthase (CAS) and nitrilases (NIT4) supplying a nitrogen source for the plant.

In almond the majority of the enzymes above mentioned have not been characterized yet, except for the UGT1 (UGT85A19, Franks et al., 2008), MDL1 (Suelves and Puigdomènech, 1998) and PH (Sánchez-Pérez et al., 2012). UGT1 was suggested to be associated with bitterness because UGT1 was mainly accumulated in the bitter genotypes at the end of kernel development, showing a three-fold higher mandelonitrile glucosyltransferase activity in the bitter compared to the non-bitter genotypes. Moreover, MDL1 was found in the fruit and flower samples, finding that the presence of this protein was not the limiting factor in the production of hydrogen cyanide. Finally, two PHs were identified in almond (*Ph691* and *Ph692*) (Sánchez-Pérez et al., 2012). However, no *Ph* gene has been expressed and its protein activity tested so far.

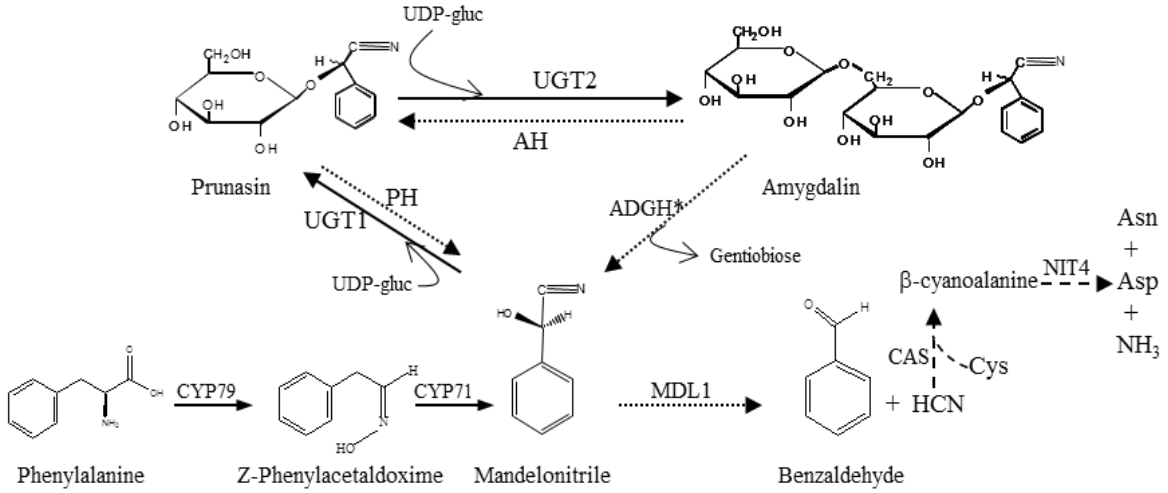


Figure 5.1. The metabolic pathway for synthesis and catabolism of the cyanogenic glucosides prunasin and amygdalin in almonds. Biosynthetic enzymes (solid lines): CYP79 and CYP71, Cytochrome P450 monooxygenases, UGT1, UDPG-mandelonitrile glucosyltransferase; UGT2, UDPG-prunasin glucosyltransferase; Catabolic enzymes (dotted lines): AH, amygdalin hydrolase; PH, prunasin hydrolase; MDL1, mandelonitrile lyase; ADGH*, amygdalin diglucosidase (putative); Detoxification enzymes (dashed lines) CAS: β-cyanoalanine synthase; NIT4: nitrilases. Source: adapted from Sánchez-Pérez et al. (2008).

On the other hand, in other *Prunus* species, enzymes related to the amygdalin pathway were studied. AH, PH and MDL were characterized in *Prunus serotina* seeds (Kuroki et al., 1987; Li et al., 1992; Zheng et al., 1995; Hu et al., 1999; Zhou et al., 2002) showing the microheterogeneity of the PH regarding aglycone specificity (Zhou et al., 2002). A total of five PHs and two AHs were described. Recently, two cytochromes P450 involved in the synthesis of prunasin and amygdalin were characterized in Japanese apricot (*Prunus mume*) (Yamaguchi et al., 2014), named CYP79D16 and CYP71AN24, responsible for the first and second reaction of the amygdalin pathway, respectively.

Cytochromes P450 constitute a supergene family found in both eukaryotes and prokaryotes, where they catalyze a vast array of different reactions. In plants, P450s are involved in the biosynthesis of many different secondary metabolites like phenylpropanoids (lignin, phytoalexins, anthocyanins, etc.) terpenoids (gibberellins,

sesquiterpenoids, etc.), cyanogenic glucosides, as well as in the detoxification of many xenobiotics, e.g. herbicides (Halkier et al., 1995).

As already mentioned throughout this thesis, the bitter flavor is produced by the presence of amygdalin in the almond kernel, so it would be logical to think that the difference between sweet and bitter almonds would be related to differences in any of these routes (biosynthesis, degradation or detoxification), which would allow or not the accumulation of amygdalin in the seed. However, Sánchez-Pérez et al. (2008) observed that all enzymes and substrates necessary to produce or degrade prunasin and amygdalin were present in both bitter and sweet almonds. Then, why do only bitter almond kernels accumulate amygdalin in high amounts?

From the genetic point of view, sweetness is dominant over bitterness. Since the original trait in wild almonds is the bitter, the change of just one allele could have prevented the accumulation of amygdalin. The question would be if this would prevent the synthesis or it would favour the degradation of these compounds. In the literature, there are two theories to explain why an almond is sweet or bitter: the first one relies on the biosynthesis and the second one on the degradation.

In the first theory, Frehner et al. (1990) proposed the anabolic enzyme glucosyltransferase as the responsible for the accumulation of amygdalin. However, since we know that the bitter is recessive, an anabolic enzyme cannot be directly the responsible for the bitterness. However, other factors such as a transcription factor could modulate the biosynthetic enzymes. Now we know that there are two glucosyltransferases involved in the amygdalin biosynthesis (Figure 5.1). The first one, named UGT1, was present both in bitter and in sweet genotypes (Franks et al., 2008). The second one, called UGT2, has not been characterized in any *Prunus* species yet.

In fact, until now, no glucosyltransferase has been identified responsible for the second glycosylation in other cyanogenic diglucosides like linustatin, neolinustatin, amygdalin and dhurrin-6-glucoside (Pičmanová et al., 2015). Interestingly, the UGT94D1 was identified in *Sesamum indicum* to produce (+)

sesaminol 2-O-β-D-glucoside with a β-(1-6) glucosidic bond (Noguchi et al., 2008), like amygdalin has.

In the second theory, Sánchez-Pérez et al. (2008; 2009; 2012) focused on the catabolic enzymes, suggesting the existence of a catabolic mechanism responsible for presence of amygdalin in bitter almond kernels, based on the activity of β-glucosidases. β-glucosidases are catabolic enzymes belonging to the family 1 of glucoside hydrolases. They catalyze the hydrolysis of the β-glucosidic bond between rests of carbohydrates or between a carbohydrate and a rest aglucone (Morant et al., 2008). The cyanogenic glucosides are degraded by β-glucosidases. Since both compounds are chemically inert separately, only when tissue is disrupted, the cyanogenic glucosides are joined to β-glucosidases and the degradation compounds are released (Conn, 1969; Poulton, 1990; Swain et al., 1992, Poulton et al., 1994; Zhou et al., 2002; Morant et al., 2003; Zagrobelny et al., 2004; Morant et al., 2008).

Although prunasin is synthesized *de novo* in sweet and bitter cultivars, the available contents of prunasin is higher in bitter than in the sweet genotypes, especially in the fruit tissues. Prunasin is transported from the tegument to the cotyledon through the inner epidermis of tegument, which has a great PH activity, higher in the sweet kernels (Sánchez-Pérez et al., 2012). As depicted in Figure 5.2a, in bitter kernels prunasin would not come into contact with the PHs (located in the apoplast), going into the cotyledon, allowing the synthesis of amygdalin through the UGT2. On the other hand, in sweet kernels (Figure 5.2b) prunasin would pass through inner epidermis of the tegument and being degraded by PHs, located in the symplast. Therefore, the amount of prunasin available to produce amygdalin would be insignificant and this cyanogenic diglucoside would not be synthesized in the cotyledon (Sánchez-Pérez et al., 2012).

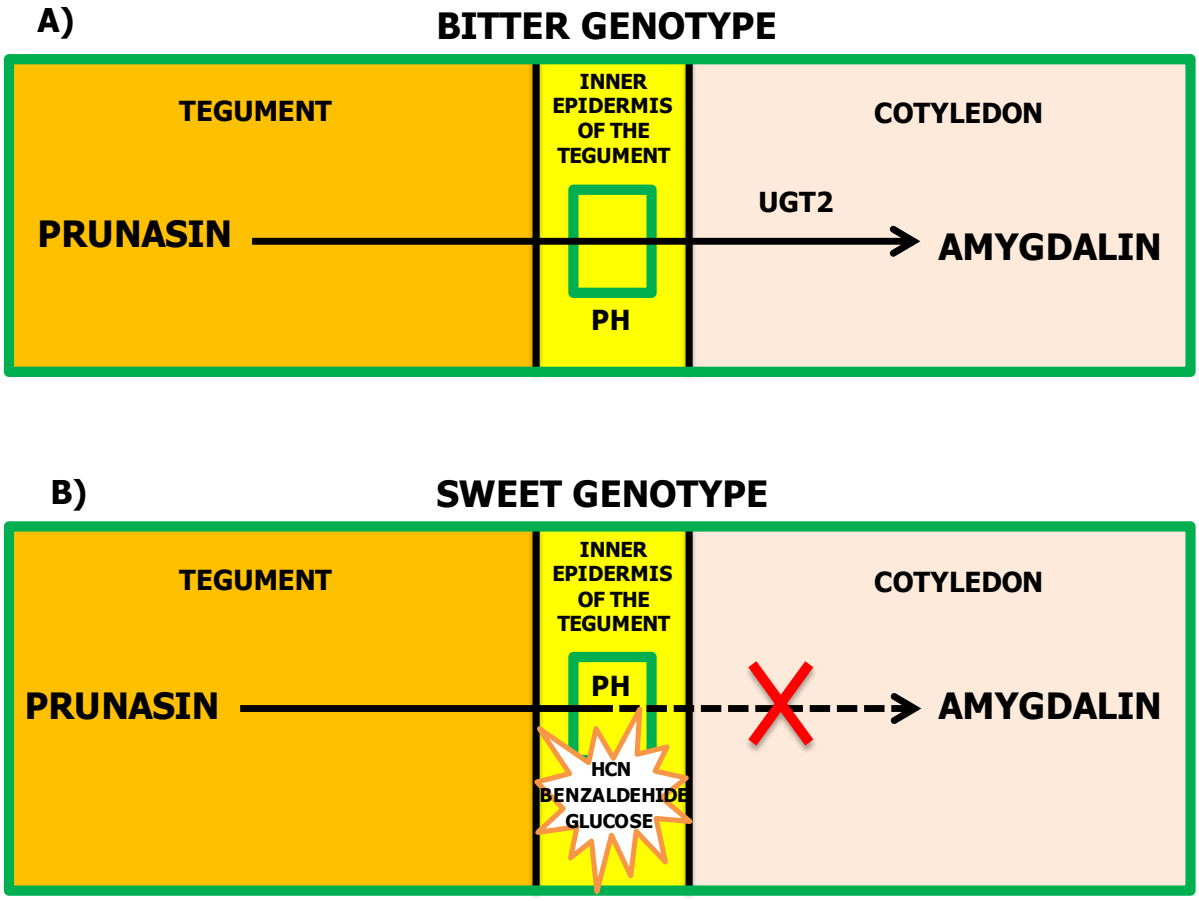


Figure 5.2. Hypothesis about accumulation of amygdalin in almonds in function of PH location in inner tegument epidermis (Sánchez-Pérez et al., 2012). a) Bitter genotype. b) In sweet genotypes prunasin is degraded by PH in the inner epidermic cells of tegument. PH: Prunasin hydrolase, UGT2: UDP-glucosyltransferase 2.

The objective of this chapter is to deep in the hypotheses about the bitterness metabolism through two complementary strategies: firstly, the isolation and characterization of the enzymes involved in the biosynthesis and degradation of cyanogenic glucosides, and secondly, through a transcriptomic analysis and identification of candidate genes that might be regulated by the transcription factors localized in the *Sk* locus, in our previous chapter.

5.2. PLANT MATERIAL

Almond fruits of the cultivars Lauranne (sweet, *SkSk*), D05-187 (bitter, *sksk*) and S3067 (bitter, *sksk*) described in the Chapter 2.1, were studied in this chapter.

5.3. METHODOLOGY

5.3.1. Characterization of catabolic enzymes

a) Plant material collection

In order to study the evolution of PHs, fruits of Lauranne, D05-187 and S3067 cultivars were harvested at the beginning (JD89), halfway (JD130) and final (JD167) stage of kernel development. "JD" stands for Julian Days, that is to say days after first of January. Tegument, nucellus+endosperm (difficult to separate) and cotyledons were isolated, grinded with liquid nitrogen, stored at -80 ºC and analysed separately. A RNA extraction from 100 mg of these tissues was carried out with the protocol Ultra Clean Plant RNA Isolation Kit (MO BIO) and by CTAB method (Annex 8.2).

b) Isolation, cloning and heterologous expression of prunasin hydrolases (*PdPh*)

As previously shown by Sánchez-Pérez et al. (2012), PH691 and PH692 were present in sweet and bitter almonds. PH691 was selected due to its high homology (73-94% of protein similarity) compared to other PHs described and characterized before in *P. serotina* (Zhou et al., 2002).

A RT-PCR was done to amplify the specific RNA by SuperScript III taq DNA Polymerase (Invitrogen) with the primers 691ATG and 691UGA (Annex 8.5). The expected band of 1,638 bp was cut from the gel and purified by the protocol Ultra Clean DNA Purification Kit (MO BIO). When a low quantity of cDNA was obtained, a PCR with the Prime STARHS (Promega, http//www.promega.com/) and an electrophoresis with marker λ DNA-Hind III was carried out (Thermo Fisher).

The β-glucosidase activity was checked in plants of *Nicotiana benthamiana* transformed with *Agrobacterium tumefaciens.*

Firstly, the nine (three cultivars by three tissues) cDNA sequences of genes were amplified by PCR Hot Master using the primers 691B1F and 692B2R. All the PCR primers contained the *attB1* and *attB2* gateway cloning sites (Annex 8.8). Gateway is a technique that provides a rapid and efficient way to move DNA sequences into multiplex vector systems.

To introduce cDNA into *A. tumefaciens* two clonations were carried out. Firstly, with the gateway BP Clonase II Enzyme Mix (Invitrogen), PCR products were introduced in pDONOR207 entry clone. Then, with the gateway LR Clonase II Enzyme Mix (Invitrogen), entry clones were cloned into destination vector pJAM1502 (Luo et al., 2007). This clone is ready to enter in *A. tumefaciens*. The destination vector (PCR product + *attB1* and *attB2* gateway cloning sites + pJAM1502) was introduced into *A. tumefaciens* (AGL1) by electroporation. Finally, transformed bacteria were plated in LB agar + rifampicin + kanamycin (Annex 8.8).

Once we had the clones ready in *A. tumefaciens*, we expressed them in *N. benthamiana*, selected for its excellent properties for rapid transient expression of genes, by agroinfiltration. *A. tumefaciens* was injected using 1 mL syringe into 4-6 weeks old *N. benthamiana* leaves, transferring the desired gene to the plant cells. After 4–5 days, leaf discs (1 cm diameter) were cut from infiltrated leaves (Annex 8.9).

c) Comparison of prunasin hydrolases from almond (PdPh) and wild cherry (PsPh)

For obtaining *PdPh* in the three cultivars in the three tissues already mentioned, the cDNA from the RT-PCR was cloned into pGEM-Teasy Vector (Promega) in *Escherichia coli* competent cells (Annex 8.6). Once positive colonies were identified by a PCR (Annex 8.5), plasmid was purified by Pure Link Quick Plasmid Miniprep Kit (Invitrogen) and with the method described in Annex 8.7. DNA

was quantified based on A280nm (Nanodrop; Thermo Scientific). Finally, positive minipreps were sent to Secugen S.L. to identify the sequence of the cDNA inserts.

Once sequences were available, they were visualized with BioEdit program and compared with other sequences from National Center for Biotechnology Information (NCBI). Nucleotide and protein sequences were aligned with the program BioEdit. Homology searches were undertaken using the BLAST network server of the National Center for Biotechnology Information (Altschul et al., 1990). For the signal peptide prediction, two web servers were used: SignalP 4.1 Server (http://www.cbs.dtu.dk/services/SignalP/) and WoLF PSORT (http:// wolfpsort.org/). Neighbor-joining phylogenetic trees were constructed using MEGA 5.5 (Tamura et al., 2011; http://www.megasoftware.net/index.html) with 1,000 bootstrap trials performed to check the homology between the sequences. For the N-glycosylation sites, NetNGlyc 1.0 Server was used (http://www.cbs.dtu.dk/services/NetNGlyc/) (Figure 5.9).

d) Detection of β-glucosidase activity

To check if the nine transformed clones of PHs had β-glucosidase activity, two methods were used.

d.1) Fast Blue BB salt. This method consists in an incubation of BNG (glucopyranoside) (Sigma Aldrich) as general substrate for β-glucosidases, in presence of Fast Blue BB salt (Sigma-Aldrich), which gives a reddish color. Protein extracts from agroinfiltrated *N. benthamiana* leaves were analysed by SDS-PAGE gel, washed in Fast Blue BB buffer and finally mixed with Fast Blue BB salt and BNG (Annex 8.11).

d.2) Umbelliferyl method. This second method uses umbelliferyl 4-methyl-β-D-glucoside (Sigma Aldrich), which is an excellent fluorogenic substrate for detecting β-glucosidases. To carry out this assay, agroinfiltrated *N. benthamiana* leaves were put in an eppendorf tube with the MES buffer and 4-methyl-umbelliferyl-β-D-

glucoside. Tubes were incubated for 15 min at 37 ºC and observed under UV light. β-glucosidase activity was detected by fluorescence (Annex 8.13).

e) Detection of PHs activity

e.1) Feigl-Anger assay. To detect the specific PH activity of the nine PdPHs, Feigl-Anger assay was carried out (Feigl and Anger, 1966). This is a qualitative probe that consists of a mixture of copper salts with a free tetra base in non-aqueous solution, which shows a distinct blue colour by oxidation of tetra base when it was exposed to the hydrogen cyanide released (Annex 8.10). A leaf disk of agroinfiltrated *N. benthamiana* with each PdPH was grinded with MES buffer and with six specific substrates: prunasin, dhurrin, linamarin, lotaustralin, amygdalin and linustatin. Feigl-Anger paper was put between the plate and the lid. Cyanogenesis was detected after two hours by the presence of blue colour.

e.2) CN assay. A second method to quantify the PH activity, the colorimetric method of Lambert et al. (1975), modified by Halkier and Møller (1989), was used. This technique detects cyanide release by spectrometry at 584 nm when the enzymes were put in contact with some substrates (prunasin, dhurrin and amygdalin) (Annex 8.12).

5.3.2 Characterization of biosynthetic enzymes

a) Isolation, cloning and heterologous expression of UGTs.

UGT1 (UGT85A19) was ordered from Genescript adding the *attB1* and *attB2* gateway cloning sites, in order to do the LR reaction and clone into the pJAM1502 by the LR Clonase II Enzyme Mix (Invitrogen).

Regarding UGT2, we created an UGT database with sequences from different plants belonging to the UGTs family's 94, 79 and 91 (Noguchi et al., 2008; Yonekura-Sakakibara and Hanada, 2011), having UGTs from 71, 73, 74 and 88 as negative controls. The reason to choose those UGT families (79, 91 and 94) was because it had been previously described in the β (1-6)-O glucosylation, as it happens in the

reaction from prunasin to amygdalin (Figure 5.1). Once the UGT database was created, we BLASTed the sweet and bitter genome assemblies versus the UGT database by using *CLC genomic workbench*, obtaining 13 putative UGTs between the two assemblies. In order to obtain full length of all of them we performed RACE with SMARTer™ RACE cDNA Amplification Kit (Clontech) as explained in Annex 8.4. Once the full length was obtained, we cloned by gateway in the pJAM1502 vector.

b) Isolation, cloning and heterologous expression of CYPs

When we BLASTed CYP79D16 and CYP71AN24 from *P. mume* versus the peach genome v.1, we found one accession with the highest similarity to CYP79D16, named ppa021236m, and two with the best hit to CYP71AN24 named ppa004152m and ppa017339m. In order to clone them, we designed primers based on peach sequences adding CACC to the starting codon, according to the kit protocol (pENTR™ Directional TOPO® Cloning Kits) from Invitrogen. Once we had the clones in the pENTR vector, we performed a LR Clonase reaction to put them into the destination vector pJAM1502.

c) Detection of UGTs activity by TnT T7 Quick for PCR DNA

With the aim of testing the activity of some of the UGT2s, we performed an *in vitro* assay by the TnT T7 Quick for PCR DNA (TnT). This is a rapid coupled transcription/translation system for optimum expression of PCR templates. PCR amplicons of putative UGTs containing a T7 promoter were added to the TnT T7 PCR Quick Master Mix and incubated for 60-90 minutes at 30 ºC (Annex 8.14).

The synthesized proteins from the TnT were used to run an assay in which the *in vitro* synthesized UGT2 was incubated with prunasin and UGP-[^{14}C]glucose to produce ^{14}C-amygdalin. The reaction was run in a TLC and the radiolabeled products were visualized using a STORM 840 PhosphorImager (Molecular Dynamics, http://www.moleculardynamics.com) (Annex 8.15).

d) Detection of CYPs and UGTs activity by LCMS

From the 13 putative UGT2 found in the almond assemblies, five were successfully cloned and heterologous expressed (UGT2 S811, UGT2 S41, UGT2 S812, UGT2 L41 and UGT2 S813). Moreover, one CY79 from *Eucalyptus yarraensis* (CYP79A34) (Neilson, 2012) and a UGT from *Lotus japonicus* UGT85K2 (Takos et al., 2011) were used in these experiments to see if we could expressed the entire biosynthetic pathway to obtain prunasin and amygdalin.

Each gene cloned in *A. tumefaciens* was grown separately and an equal amount of culture containing each gene was used in the agroinfiltration of tobacco leaves. Then 100 mg of the samples were ground in liquid nitrogen, placed in a 1.5 ml threaded tubes with 400 µL 85% methanol, boiled 5 min and put in ice. Right after, samples were centrifuged for 5 min x 20,000g and the supernatant was collected and kept at -20 ºC. 20 µL of the supernatant were mixed with 70 µL of water and 10 µL of 500 µM internal standard linamarin, and filtered through an ELISA filter plate by centrifugation (5 min x 3,000 rpm) (see conditions in Annex 16). These samples were now ready to be run by liquid chromatography-mass spectrometry (LCMS).

LC-MS is an analytical method that combines the physical separation capacity of liquid chromatography (HPLC) with the mass capacity of mass spectrometry (MS). Analytical LC-MS was carried out using an Agilent 1100 Series LC (Agilent Technologies, Germany) coupled to a Bruker HCT-Ultra ion trap mass spectrometer (Bruker Daltonics, Bremen, Germany). Samples were analyzed with the Bruker Daltonics programm Data Analysis 4.0 (see conditions in annex 8.16).

5.3.3 Identification of orthologous genes, cluster and expression analysis

The tool *OrthoMCL* was used as it provides a scalable method for constructing orthologous groups across multiple eukaryotic taxa, using a Markov Cluster algorithm to group (putative) orthologous. Firstly, the protein sequences of the *P. persica*

proteins included in CYP, UGT and BGA families were downloaded. These proteins together with the 57,982 almond proteins were given as input to *OrthoMCL*.

Secondly, almond genes previously characterized/mentioned in this chapter, as BGA genes (*PdPh691, PdPh692* and *BGA_ppa024207m,* which was a hydrolase found in the *Sk* region as it was shown in Figure 4.5), UGT2 (*L41, C02, S811, S813, UGT85A19* and *C4*) and the CYPs (*CYP79_ppa021326m, CYP71_ppa017339m and CYP71AN24_ppa004152m*) were BLASTed versus the transcripts already annotated in almond (Chapter 4). Finally, taking the best hit, the one showing highest identity and lowest e-value, we extracted the name of the gene and the expression values.

5.4. RESULTS AND DISCUSSION

5.4.1. Characterization of catabolic enzymes

5.4.1.1. Evolution of PHs during kernel development

The presence or absence of PHs in each tissue is described as follows:

STAGE 1: Beginning of development (JD 89) (Figure 5.3.a)

During this period, the immature kernel was composed by nucellus and tegument. Cotyledons were not developed yet. This development stage was similar to that observed in almond (Martínez-Gómez et al., 2008; Abarrategui, 2010) and in black cherry (Swain et al., 1992).

In this stage, DNA band characteristic of the putative *Phs* (1,638 bp) was not detected in the sweet cultivar (Figure 5.4.a). In the bitter cultivars, it was more intense in the tegument than in the nucellus.

STAGE 2: Halfway development (JD 130) (Figure 5.3.b)

In this period endosperm and cotyledons began to grow inside of the tegument, and nucellus was still visible, what was previously observed in black

cherry and almond (Swain et al., 1992; Martínez-Gómez et al., 2008; Abarrategui, 2010).

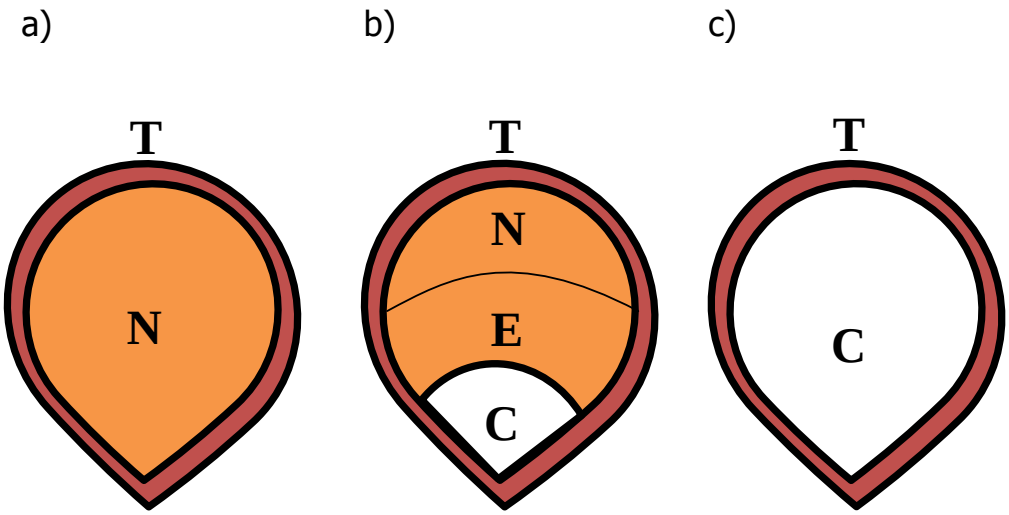


Figure 5.3. Kernel development stages: a) beginning (JD 89); b) halfway (JD 130); c) final (JD 167). T: tegument, N: nucellus, C: cotyledons, E: endosperm.

Figure 5.4.b shows that DNA bands of the amplification of *Ph691* from bitter genotypes became weaker in tegument than in Stage 1, being similar in nucellus and stronger in cotyledons. In the sweet Lauranne, *Ph691 was* amplified in the three tissues.

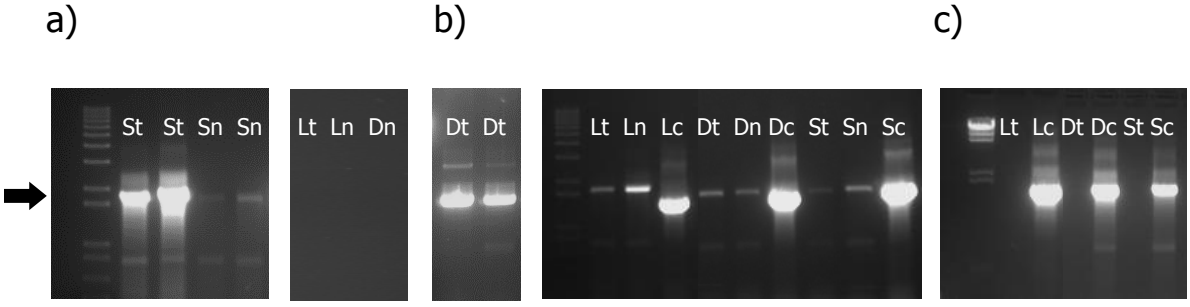


Figure 5.4. Detection of prunasin hydrolases by RT-PCR in the three development stages. a) beginning, b) halfway, c) final. Arrows indicates band size of 1638 pb. Ladder 1 kb+. L: Lauranne (sweet), D: D05-187 (bitter), S: S3067 (bitter), t: tegument, n: nucellus, c: cotyledon.

STAGE 3: Final development (JD 167) (Figure 5.3.c)

In this last stage, agreeing with Swain et al. (1992) in black cherry and Martínez-Gómez et al. (2008) and Abarrategui (2010) in almond, the kernel was only formed by cotyledons and the dried tegument.

Ph691 gene was not amplified in the tegument, only in the cotyledons (Figure 5.4.c) both in the sweet and bitter genotypes, in agreement with Abarrategui (2010).

In conclusion of this section, *Ph691* appeared in the tegument and the nucellus of the bitter genotypes at the beginning of the development, in the three tissues both bitter and sweet at the halfway of the development and only in the cotyledons at the final development.

The next step was to clone the gene in the three tissues from the two cultivars to do heterologous expression and check their activities.

5.4.1.2. Comparison of sequences of PHs clones

As previously described in almond (Sánchez-Pérez et al., 2012) and in black cherry (Zhou et al., 2002), the amplified sequences of *Ph691* gene contained 13 exons and 12 introns (Figure 5.5). The cDNA full length in our nine clones had 1,635 bp, codifying a 544 amino acid protein, which contained 5 putative N-glycosylation sites (Table 5.1).

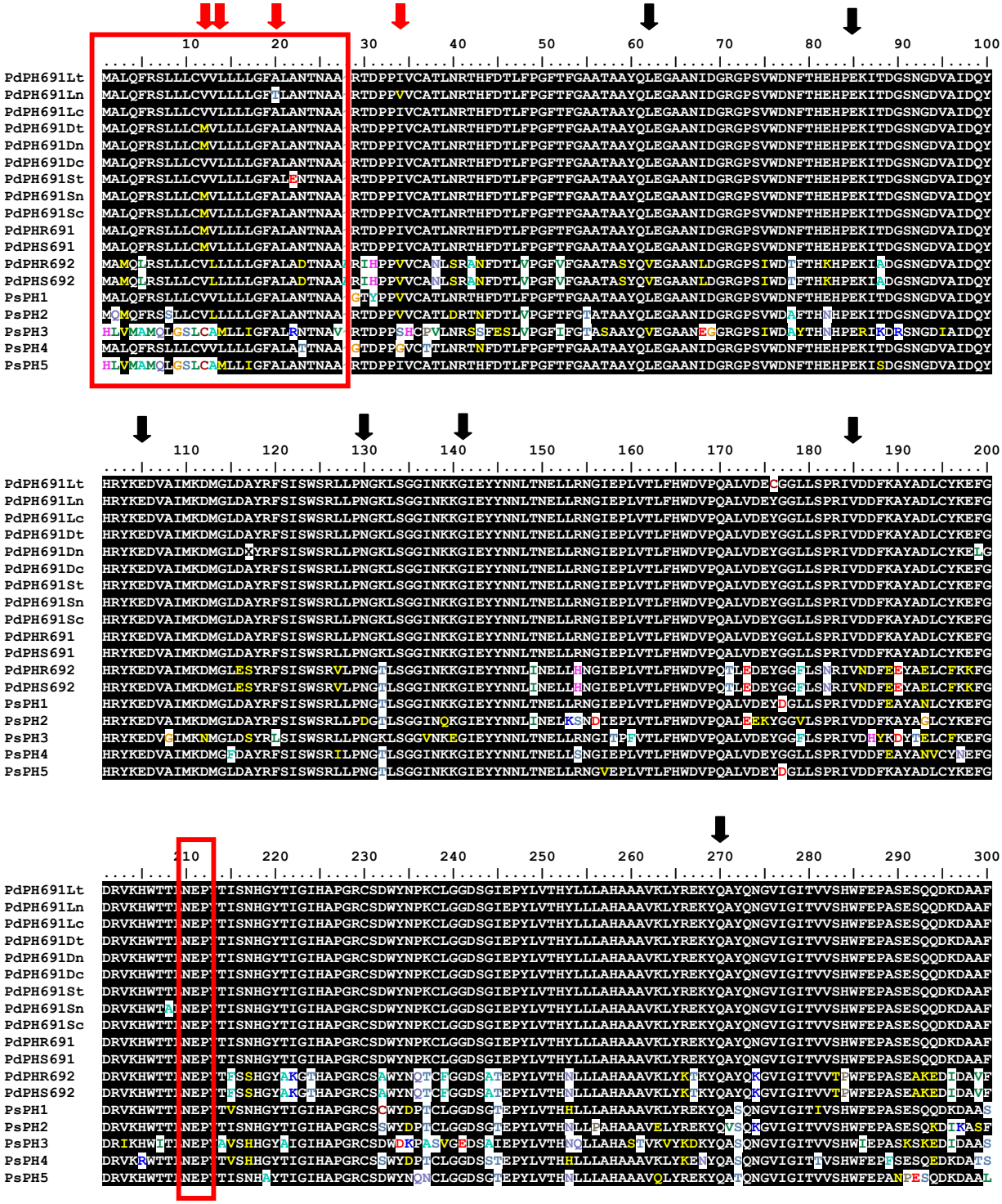
On the other hand, the five PHs identified in black cherry had between 537 and 549 amino acid, and between 2 and 8 N-glycosylation sites (Zhou et al., 2002). Another almond catabolic enzyme, mandelonitrile lyase 1, had an insert of 1,667 bp (559 amino acid) and 4 putative N-glucosylation sites (Suelves et al., 1998).

Regarding the signal peptide (Table 5.1), three aspects were analyzed: peptide size, cleavage site and the place where the protein would be leaded. The signal peptide size was 27 amino acids for most of genotypes and tissues and 23 amino acids for Lauranne tegument, Lauranne cotyledon and D05-187 cotyledon. Therefore the cleavage site was different in these three last genotypes, being

between amino acid 22 and 23 and between the 26 and 27 for the rest. According to Sánchez-Pérez et al. (2012) the predicted cleave site for PH691 was between amino acids 26-27 (TNA-AR) whereas for PH692 was between amino acids 22-23 (ALA-DT).

According to signal peptide sequences with the program SignalP 4.1 Server the protein was always leaded in vacuole, except for the tegument of S3067, which was extracellular. We could think about the hypothesis mentioned in the introduction where PHs, in the bitter genotypes, were in the apoplast. However, there is a discrepancy as the other cultivar D05-187 did not show this difference. Our results did not show a clear relation between signal peptide, predicted location versus the bitter phenotype.

Li et al. (1992) observed small differences in the N-terminal of different isoforms of PH from black cherry. Morant et al. (2008) also showed that β-glucosidases contained an N-terminal signal peptide, which placed them in the symplast or apoplast. Sánchez-Pérez et al. (2012) observed a different localization (symplastic versus apoplastic) of these catabolic enzymes in sweet and bitter cultivars. Our results disagree with these previous works since we were not able to distinguish between sweet and bitter genotypes according to their differences in the signal peptide. One reason could be that Sánchez-Pérez et al. (2012) was inmunolocalizing PHs with an antibody common to more than one PHs. It would have been nice to clone *Ph692* and see how it was behaved in the bitter teguments. Moreover, PH691s studied here had ITENG and NEP motifs characteristic of the active site of glycoside hydrolase family 1 (Table 5.1). These motifs were also described in PHs (Zhou et al., 2002) and AH (Zheng and Poulton, 1995) from black cherry. Differences in the active site between sweet and bitter genotypes were not detected.


PdPH691Lt
PdPH691Ln
PdPH691Lc
PdPH691Dt
PdPH691Dn
PdPH691Dc
PdPH691St
PdPH691Sn
PdPH691Sc
PdPHR691
PdPHS691
PdPHR692
PdPHS692
PsPH1
PsPH2
PsPH3
PsPH4
PsPH5

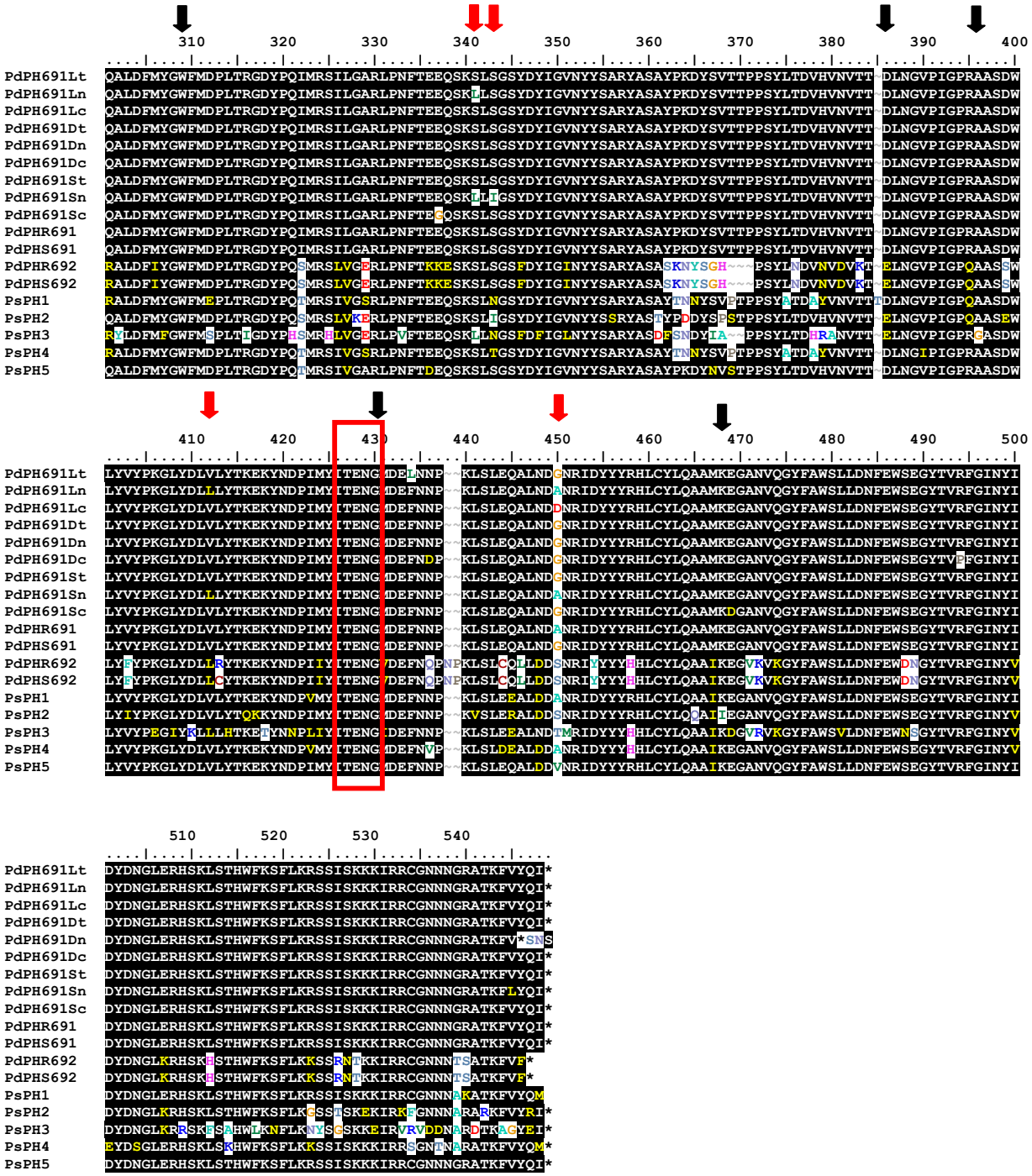


Figure 5.5. Amino acid sequences of PHs from *P. serotina* (Ps) and *P. dulcis* (Pd). Differences in amino acids are marked in colour. Red frames: signal peptide and NEP and ITENG motifs of active site. Black arrows: introns. Red arrows: SNPs. L: Lauranne (sweet), D: D05-187 (bitter), S: S3067 (bitter), t: tegument, n: nucellus, c: cotyledon. PsPH1, PsPH2, PsPH3, PsPH4 and PsPH5 (Zhou et al., 2002). PdPHR691, PdPHS691, PdPHR692 and PdPHS692 (Sánchez-Pérez et al., 2012).

Table 5.1. Comparison of major features of *P. dulcis* and *P. serotina* nucleotides and amino acid sequences. Sequence length, polypeptide length, N-glycosilation sites, ITENG/NEP motifs, signal peptide and location of the enzyme are indicated. L: Lauranne (sweet), D: D05-187 (bitter), S: S3067 (bitter), t: tegument, n: nucellus and c: cotyledon. *PsPh1, PsPh2, PsPh3, PsPh4* and *PsPh5* (Zhou et al., 2002). *PdPhR691* ,*PdPhS691, PdPhR692 and PdPhS692* (Sánchez-Pérez et al., 2012).

	Sequence length	**Polypeptide length**	**N-glycosylation sites**	**ITENG / NEP**	**Signal peptide**	**Location**
PdPh691Lt	1635	544	5	Yes	23	Vacuolar
PdPh691Ln	1635	544	5	Yes	27	Vacuolar
PdPh691Lc	1635	544	5	Yes	23	Vacuolar
PdPh691Dt	1635	544	5	Yes	27	Vacuolar
PdPh691Dn	1635	544	5	Yes	27	Vacuolar
PdPh691Dc	1635	544	5	Yes	23	Vacuolar
PdPh691St	1635	544	5	Yes	27	Extracellular
PdPh691Sn	1635	544	5	Yes	27	Vacuolar
PdPh691Sc	1635	544	5	Yes	27	Vacuolar
PsPh1	2056	549	8	Yes	23	Vacuolar
PsPh2	2059	544	2	Yes	23	Vacuolar
PsPh3	1960	537	6	Yes	27	Vacuolar
PsPh4	1911	545	7	Yes	23	Vacuolar
PsPh5	1819	542	5	Yes	23	Vacuolar
PdPhR691	1635	544	5	Yes	27	Vacuolar
PdPhS691	1635	544	5	Yes	27	Vacuolar
PdPhR692	1629	542	5	Yes	23	Vacuolar
PdPhS692	1629	542	5	Yes	23	Vacuolar

Table 5.2 shows amino acid differences between the nine PH sequences studied here. Although some nucleotide differences were detected in the exons, they were only translated in an amino acid change in eight cases (in exons 1, 9, 11 and 12). Three of these changes were found inside the signal peptide, located in the amino acids 12, 13 and 20. Differences were not consistent between the sweet and bitter cultivars or between tissues. Sánchez-Pérez et al. (2012) showed a difference

between amino acid sequences (position 414) in sweet and bitter cultivars, but we did not detect this difference in our clones.

When we compared the *Phs* studied here, the percentage of nucleotide similarity ranged between 98 and 99 in our nine sequences (Table 5.3). Differences were not found between sweet and bitter cultivars. As expected, there was a high homology (99%) with the two sequences of *Ph691* described by Sánchez-Pérez et al. (2012) in almond. Compared to black cherry sequences, *PsPh1* and *PsPh5* were the most similar (94% and 96% respectively), being *PsPh3* more different than others (Table 5.3).

The phylogenetic analysis clearly separated one group of sequences respect to the rest, based on the positions of the clones in the tree and the percentage of bootstrap support (Figure 5.6). This group contained the nine PHs characterized in this study plus, as expected, the two PdPH691s from a sweet (Ramillete, R) and a bitter (S3067, S) genotype (Sánchez-Pérez et al., 2012). According to the PHs from black cherry, PsPH5 had the amino acid sequence closer to the PHs from almond, what agrees with the nucleotide similarity described before. In this analysis, AH from *P. serotina* appears as a distinct branch, so we observed clearly the differences between all the PHs and the AH.

Table 5.2. Amino acid differences between sequences of our nine PHs. Exon, DNAg, cDNA and aa positions of the differences are indicated. They are compared with five PHs described in *P. serotina* by Zhou et al. (2002) and with two PHs described in sweet (PdPHR691 and PdPHR692) and bitter (PdPHS91 and PdPHS92) almond (Sánchez-Pérez et al., 2012) (L: Lauranne, D: D05-187, S: S3067).

		Position			Tegument			Nucellus			Cotyledon		
Difference	**EXON**	**DNAg**	**cDNA**	**aa**	**L**	**D**	**S**	**L**	**D**	**S**	**L**	**D**	**S**
M/V	**EXON 1**	34	34	12	V	M	V/I	V	M/V	M/V	M/V	M/V	M/V
G/V	**EXON 1**	38	38	13	V	V	V/G	V	V	V	V	V/G	V
A/T	**EXON 1**	58	58	20	A	A	A/T	A/T	A	A/T	A	A/K	A/T
I/V	**EXON 1**	97	97	33	I	I	I/V	I/V	I	I/V	I	I	I
L/S	**EXON 9**	2390	1019	340	S	S	L/S	L/S	S	L/S	S	S	S
I/S	**EXON 9**	2397	1025	342	S	S	I/S	S	S	I/S	S	S	S
L/V	**EXON 11**	2815	1231	411	V	V	V	L/V	V	L/V	V	V	V
A/G	**EXON 12**	3079	1340	446	G	G	A/G	A/G	G	A	D/G	G	G

		Position			***Prunus serotina***					***Prunus dulcis***			
Difference	**EXON**	**DNAg**	**cDNA**	**aa**	**PsPH1**	**PsPH2**	**PsPH3**	**PsPH4**	**PsPH5**	**PdPHR691**	**PdPRS691**	**PdPHR692**	**PdPHS692**
M/V	**EXON 1**	34	34	12	V	V	A	V	A	M	M	V	V
G/V	**EXON 1**	38	38	13	V	L	M	V	M	V	V	L	L
A/T	**EXON 1**	58	58	20	A	A	A	A	A	A	A	A	A
I/V	**EXON 1**	97	97	33	V	V	S	G	I	I	I	V	V
L/S	**EXON 9**	2390	1019	340	S	S	L	S	S	S	S	S	S
I/S	**EXON 9**	2397	1025	342	N	I	N	T	S	S	S	S	S
L/V	**EXON 11**	2815	1231	411	V	V	L	V	V	V	V	L	L
A/G	**EXON 12**	3079	1340	446	A	S	T	A	V	A	A	L	L

Table 5.3. Percentage of nucleotide similarity between our nine *Phs* and other sequences of *P. serotina* and *P. dulcis.* L: Lauranne (sweet), D: D05-187 (bitter), S: S3067 (bitter), t: tegument, n: nucellus and c: cotyledon. *PsPh1, PsPh2, PsPh3, PsPh4* and *PsPh5* (Zhou et al., 2002). *PdPhR691, PdPhS691*, *PdPhR692* and *PdPhS692* (Sánchez-Pérez et al., 2012). *Lt* stands for *PdPh691Lt*, *Ln* means *PdPh691Ln* and so on.

	SWEET			**BITTER**					
	Lt	*Ln*	*Lc*	*Dt*	*Dn*	*Dc*	*St*	*Sn*	*Sc*
PdPh691Lt	100.0	98.90	99.69	99.45	99.33	99.79	99.27	98.96	99.57
PdPh691Ln	-	100.0	98.90	98.96	98.71	98.90	98.78	99.45	98.84
PdPh691Lc	-	-	100.0	99.39	99.27	99.72	99.20	98.96	99.51
PdPh691Dt	-	-	-	100.0	99.39	99.59	99.27	99.14	99.51
PdPh691D	-	-	-	-	100.0	99.45	99.14	98.99	99.39
PdPh691D	-	-	-	-	-	100.0	99.24	99.04	99.60
PdPh691St	-	-	-	-	-	-	100.0	98.83	99.20
PdPh691S	-	-	-	-	-	-	-	100.0	99.02
PdPh691Sc	-	-	-	-	-	-	-	-	100.0
PdPhR691	99.00	99.00	99.00	99.00	99.00	99.00	99.00	99.00	99.00
PdPhS691	99.00	99.00	99.00	99.00	99.00	99.00	99.00	99.00	99.00
PdPhR692	79.00		79.00	79.00	79.00	79.00	79.00	79.00	79.00
PdPhS692			79.00	79.00	79.00	79.00	79.00	79.00	79.00
PsPh1	94.00	94.00	94.00	94.00	93.00	94.00	94.00	94.00	94.00
PsPh2	91.00	91.00	91.00	91.00	91.00	91.00	91.00	91.00	91.00
PsPh3	82.00	83.00	82.00	82.00	82.00	83.00	82.00	83.00	82.00
PsPh4	92.00	92.00	92.00	92.00	92.00	92.00	92.00	92.00	92.00
PsPh5	96.00	96.00	96.00	96.00	96.00	96.00	96.00	96.00	96.00

Noticing the similarity between amino acid sequences of our putative PHs with other sequences of different species we can deduce that almond PHs belong to BGA family of β-glucosidases for their: a) similar nucleotide and aminoacids length, b) signal peptide size, c) ITENG and NEP motifs, d) number of exons and introns and e) N-glucosylation sites. BGA (glycoside hydrolase family 1) are enzymes which catalyze glucoside hydrolysis. They have an amino acid residue (Ile/Val-Thr-Glu-Asn-Gly) motif. This family includes β-glucosidases, phosphor-β-glucosidases, thio-β-glucosidases and β-galactosidases from archaebacteria, bacteria, plants and mammals.

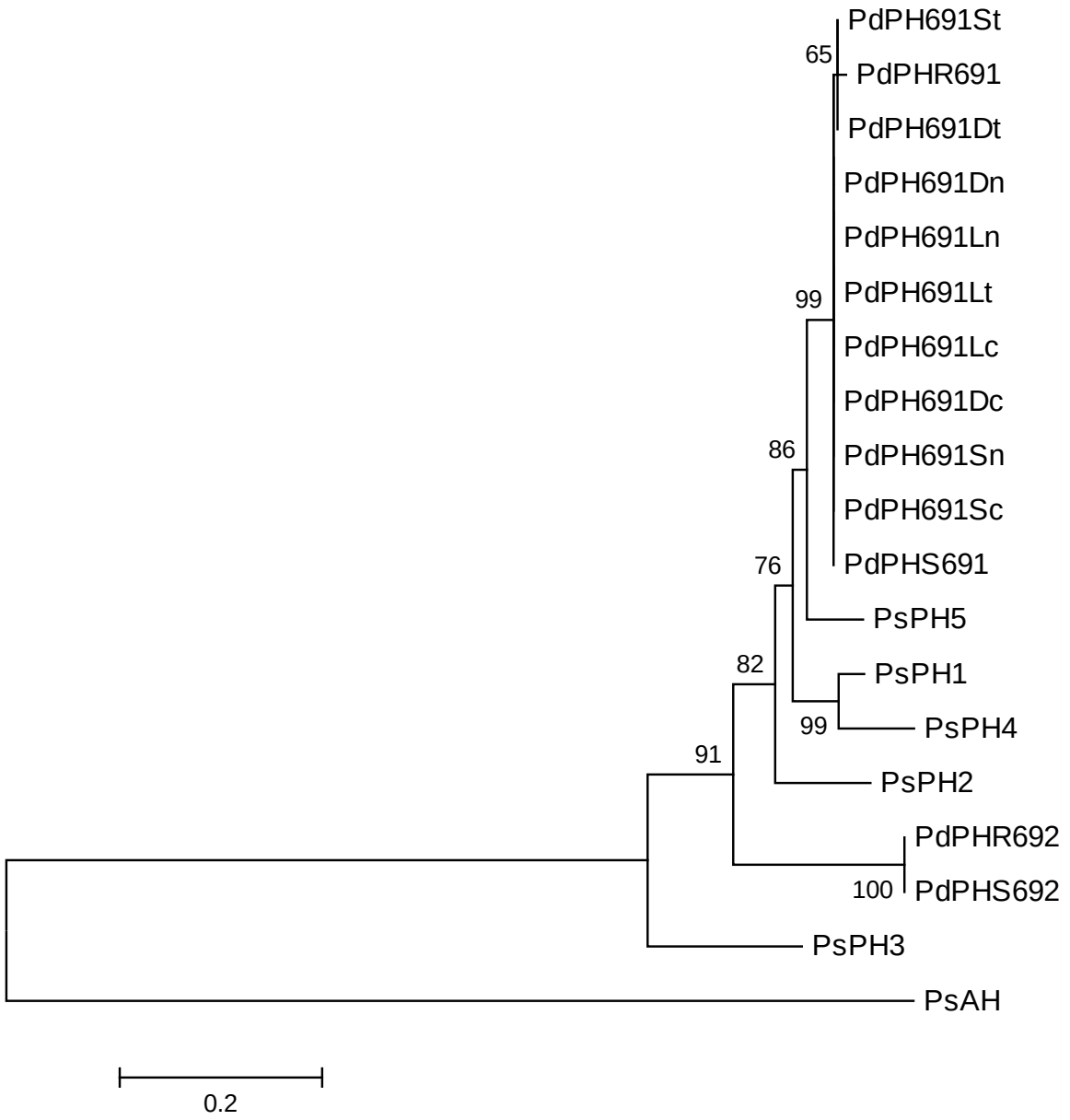


Figure 5.6. Phylogenetic tree between the amino acid sequences of our nine PHs and other PHs from *P. serotina* and *P. dulcis*. Numbers represent percentage bootstrap support (1,000 replicates). L: Lauranne (sweet), D: D05-187 (bitter), S: S3067 (bitter), t: tegument, n: nucellus, c: cotyledon. PsPH1, PsPH2, PsPH3, PsPH4 and PsPH5 (Zhou et al., 2002). PdPHR691, PdPHS691, PdPHR692 and PdPHS692 (Sánchez-Pérez et al., 2012) and PsAH (Li et al., 1992).

In conclusion of this section, we observed similarities between our PH clones compared with other PHs described before in almond and black cherry. But we could not see any difference to separate sweet and bitter cultivars at sequence level or in their location. There were small differences but no exclusively of one specific cultivar.

Therefore, the next step was to check if these clones had β-glucosidase and PH activity and if the small differences in the sequences had any influence over the prunasin hydrolase activity.

5.4.1.3. Detection of β-glucosidase activity

Fast Blue BB method.

With the first method (Fast Blue BB), the nine PHs showed β-glucosidase activity, except for D05-187 tegument and the control p19 (Figure 5.7a). Different bands of different sizes (between 50 and 75 kD) were observed. Moreover, some clones showed more than one band (Lauranne nucellus and cotyledon) what could indicate the presence of different isoforms of PHs or PH and AH migrating at the same time. These differences were not related to the sweet or bitter phenotype.

These results disagree with Abarrategui (2010), who detected a high β-glucosidase activity in cotyledons of sweet and bitter almond cultivars, very low in nucellus and endosperm, and no activity in tegument. One reason could be the time of ripening when the samples were collected, as it was shown in Sánchez-Pérez et al. (2012), the localization of PH differs from the beginning to the end of the ripening season.

Poulton (1990) suggested the presence of a common structure for most β-glucosidases with a molecular mass of 55 to 65 kD. Li et al. (1992) isolated four isoenzymes of 52 kD of AH (AH I, I´, II and II´) from black cherry. Poulton et al. (1994) described PHs of 60 kD for *P. domestica* and 68 kD for black cherry. Hu et al. (1999) characterized five isoforms of mandelonitrile lyase from black cherry with molecular masses between 57 and 59 kD. Zhou et al. (2002) described five PHs in black cherry, they had sizes between 52 and 68 kD.

Umbelliferyl method.

With the second method (umbelliferyl 4-methyl-β-D-glucoside substrate) (Figure 5.7b) β-glucosidase activity was detected in all the clones, except for D05-187 tegument which showed a low fluorescence, similar to that of p19 control (background), and similar to the Fast BB result already explained above. This agrees with Kuroki and Poulton (1987) who observed a low β-glucosidase activity from PH of black cherry with the same substrate.

We can conclude that all PHs except for D05-187 tegument had β-glucosidase activity. We could not difference between sweet and bitter cultivars in the activity and neither had a relation between the differences in the sequences and the activity of the PHs.

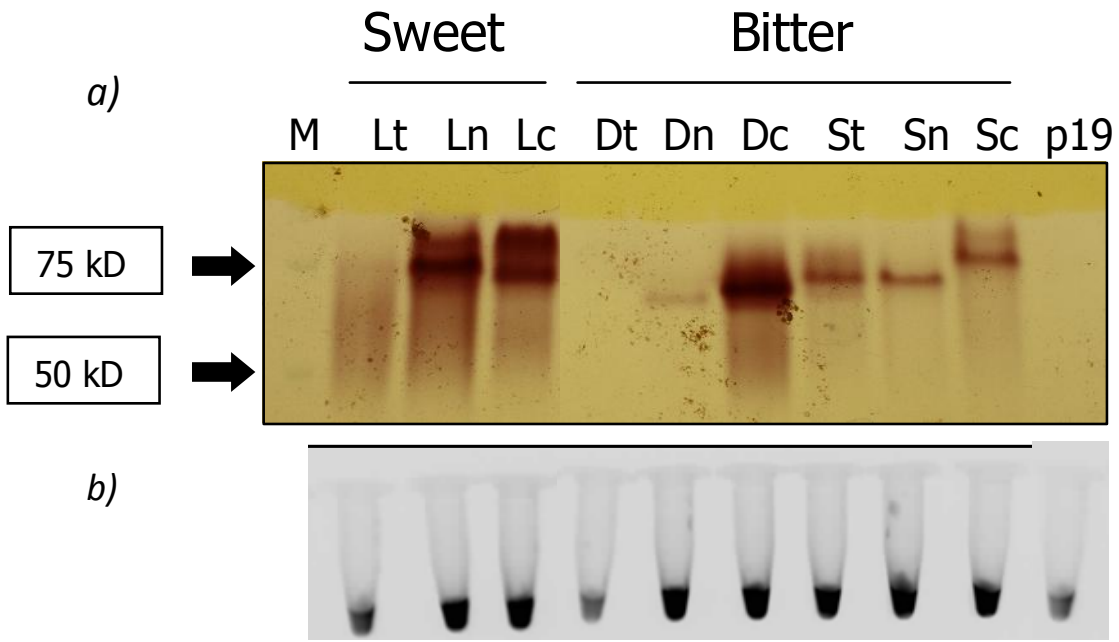


Figure 5.7. β-glucosidase activity in tissues of sweet and bitter cultivars detected by a) Fast Blue BB salt method. L: Lauranne (sweet), D: D05-187 (bitter), S: S3067 (bitter); p19: negative control; t: tegument, n: nucellus and c: cotyledon. Bands size ranged between 50 and 75 kD, and b) fluorescence under UV light using the substrate 4-methyl-umbelliferyl-β-D-glucoside.

5.4.1.4. Verification of PHs activity

Feigl-Anger assay.

The addition of prunasin to the extract from *N. benthamiana* agroinfiltrated leaves showed the PH activity of our nine PHs, both in sweet and bitter genotypes. As expected, the negative control p19 did not show any activity (Figure 5.8).

Li et al. (1992) also described specificity of PH towards prunasin in black cherry. Zhou et al. (2002) and Sánchez-Pérez et al. (2009) showed PH and AH specificity towards prunasin and amygdalin in black cherry and almond respectively. In fact, in 2009, this activity was coming from protein extracts from almond but not from heterologous expression of a prunasin hydrolase. So this is the first time a prunasin hydrolase from almond has been expressed and its activity versus prunasin has been tested.

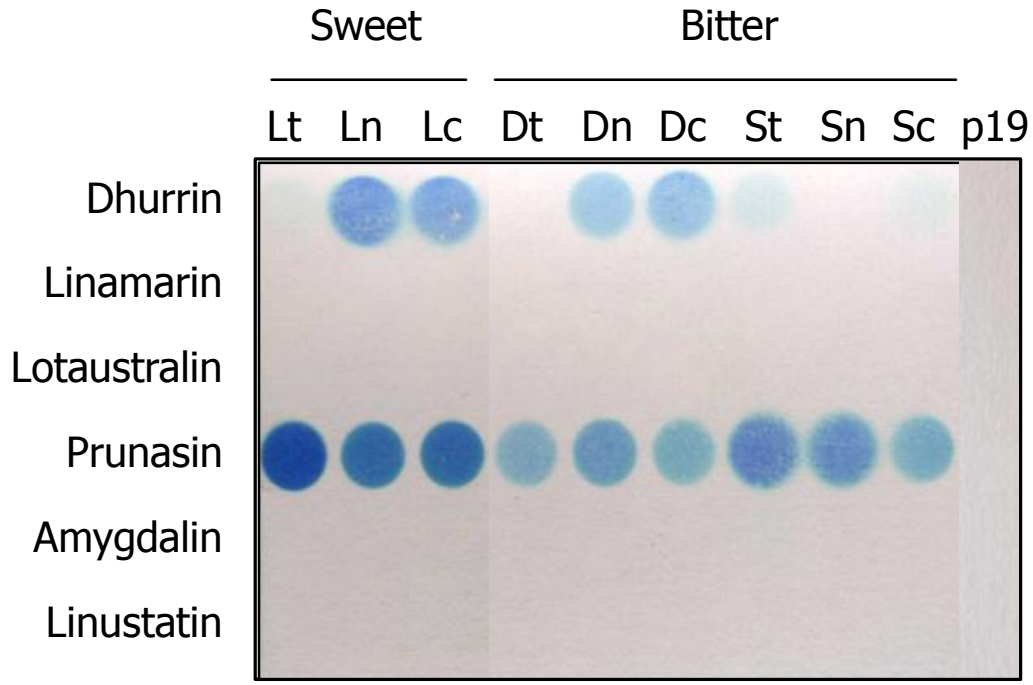


Figure 5.8. Feigl-Anger method with the nine putative PH incubated with dhurrin, linamarin, lotaustralin, prunasin, amygdalin and linustatin. p19: negative control. Blue colour indicates the positive reaction liberating CN from the PH691 activity. L: Lauranne (sweet), D: D05-187 (bitter), S: S3067 (bitter), t: tegument, n: nucellus, c: cotyledon.

Hydrolase activity was also detected in Ln, Lc, Dn and Dc with the monoglucoside dhurrin, but not with the diglucosides amygdalin and linustatin or with the other monoglucosides assayed (linamarin and lotaustralin) (Figure 5.8). According to Kuroki and Poulton (1987), PH did not show specific activity with amygdalin, dhurrin, linamarin and linustatin.

Some authors observed that only a few amino acid residues were responsible for the enzyme specificity towards the substrate through the change the active site. Zhou et al. (2002) observed that a change of two amino acids (position 220 and 394) in AH was able to confer PH activity. Lai et al. (2015) observed that only a single amino acid polymorphism in the aglycone binding region of the active site (G211 in BGD2 and V211 in BGD4) explained the difference in substrate specificity for the cyanogenic glucosides linamarin and lotaustralin. It would be nice to identify which amino acids are responsible for the aglycone specificity in those PHs, but this is something that due of time was impossible to do in this thesis.

CN assay

The assay to quantify the PH activity based on cyanide release was performed with those substrates that had a positive hydrolase activity (prunasin and dhurrin) but also to amygdalin. As expected, the results showed a high PH activity, a low hydrolase activity when incubated with dhurrin, and no activity with amygdalin, as the level detected were as the ones obtained with the negative control p19 (Figure 5.9).

With the substrate prunasin, Lauranne nucellus was the PH with the highest content of cyanide liberated, with more than 500 nmol CN^-/g agroinfiltrated leaf, followed by Lauranne cotyledon, D05-187 cotyledon and D05-187 nucellus which ranged between 300 and 400 nmol CN^-/g. The rest of PHs liberated between 100 and 200 nmol CN^-/g. Only D05-187 tegument showed a very low cyanide release, which is in agreement to what previously showed by the Fast BB and umbelliferyl assay.

With dhurrin, only Lauranne tegument had a significant cyanide release (almost 200 nmol CN^-/g). The rest of PHs showed a very low activity.

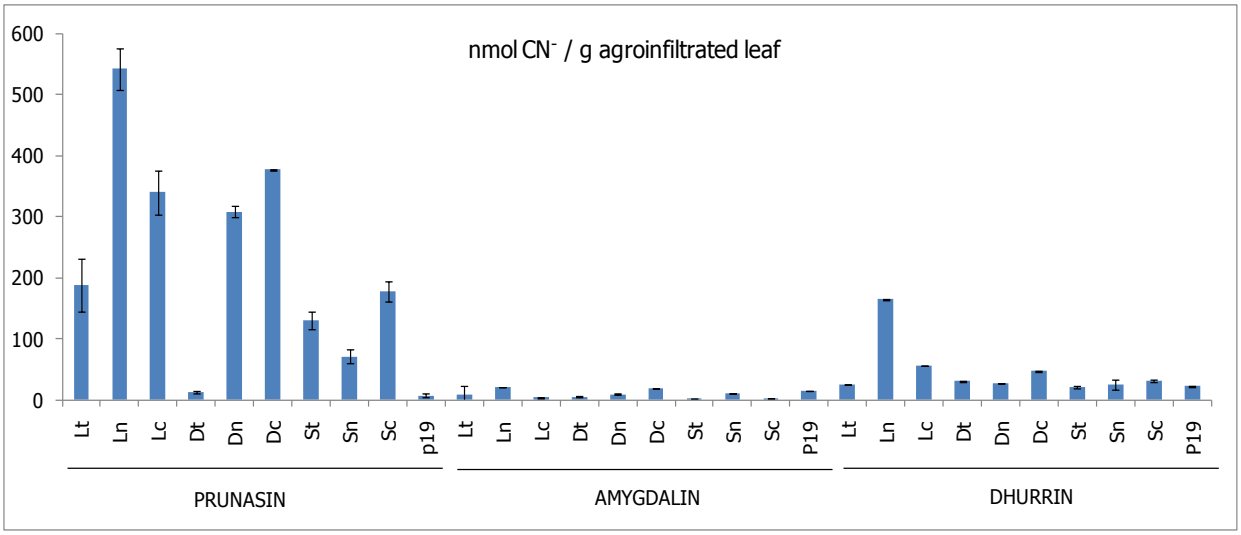


Figure 5.9. Cyanide (nmol CN/g) released from leaves of *N. benthamiana* agroinfiltrated with the nine PH691s against the substrates prunasin, amygdalin and dhurrin. L: Lauranne (sweet), D: D05-187 (bitter), S: S3067 (bitter), t: tegument, n: nucellus and c: cotyledon.

None of these techniques showed a difference related to the bitterness phenotype. However, we still need to find out why PH from D05-187 tegument did not show any β-glucosidase activity.

5.4.2. Characterization of biosynthetic enzymes

Four of the UGT-glucosyltransferases analyzed showed activity versus prunasin and/or amygdalin, which means that we found UGT1 and UGT2 activity within our candidates. First, we are going to describe the results obtained when agroinfiltrated tobacco plants were analysed by the LC-MS and second, the *in vitro* assays that were done by TnT and analysed by TLC.

Detection of biosynthetic activity by LC-MS

In order to detect UGT1 and UGT2 activity, and observing that the CYP79 analysed in almonds was not active when expressed in tobacco, we substituted the almond CYP79 by the promiscuous CYP79A34 from *E. yarraensis* (as mentioned in Methodology section). Nevertheless, the expression of CYP71AN24 was very successful, as well as UGTs *S812, L41, S813* and the already characterized UGT1-UGT85A19.

When tobacco leaves were agroinfiltrated with the three genes (CYP79A34, CYP71AN24 and UGT85K2 (from *L. japonicus*) / *S812* / *L41,* we were able to synthetize prunasin (Figure 5.10). Trace amounts of amygdalin were also observed when *S812* was also present in tobacco. *L41* gene did not show any UGT2 activity, as amygdalin was not detected. When *S813* was the third gene agroinfiltrated, then not only prunasin but also amygdalin was detected, in a higher level compared with the other UGTs previously shown (Figure 5.11). Moreover, when UGTS813 was co-infiltrated with UGT85A19 (Franks et al., 2008), no differences in metabolite profile were observed. A sum-up of these results, orthologs in peach and its localization in the peach genome is showed in Table 5.4.

Table 5.4. Five putative UGT-glucosyltransferases and their orthologs in peach and their activity detected when expressed in agroinfiltrated tobacco leaves. Four out of five showed UGT1 and/or UGT2 activity. Sc: S3067 cotyledon. Dc: D05-187 cotyledon.

Gene	Ortholog	Cultivar	Activity detected	Product
S41	*ppa005394*	Sc	-	-
L41	*ppa005106*	Dc	UGT1	Prunasin
S813	*ppa005520*	Sc	UGT1, UGT2	Prunasin, Amygdalin
S812	*ppa005532*	Sc	UGT1, UGT2	Prunasin, Amygdalin
S811	*ppa005471*	Sc	UGT2	Amygdalin

Finally, when compared to the control p19, in all the agroinfiltrations with the different constructs two more peaks were found in tobacco. The first showed a m/z of 320, which corresponds to Phe-oxime glucoside, meaning that the two CYPs were not completely coupled with the UGTs and so some intermediates were escaping of this metabolom. The second one with a m/z of 404, which corresponds to prunasin and malic acid, had also been shown in agroinfiltrated tobacco plants with CYPs from Eucalyptus and Lotus, meaning that tobacco wanted to detoxify prunasin by adding malic acid (Neilson E., personal communication).

Moreover, another UGT from sorghum named UGT85B1 was characterized in *E. coli*. It catalyzed the reaction from the CYP71E1 to obtain dhurrin (Jones et al., 1999; Bak et al., 2006; Takos et al., 2010). Four families of the family 1 of glucosyltransferases from *Barbarea vulgaris* were identified and expressed in *E. coli* (Augustin et al., 2012). They were named UGT73C10, UGT73C11, UGT73C12 and UGT73C13 and all of them had a relation with saponins in the defense of the plant (Augustin, 2012). Triterpenoids saponins are bioactive metabolites involved in the plant defense against fungi and insects.

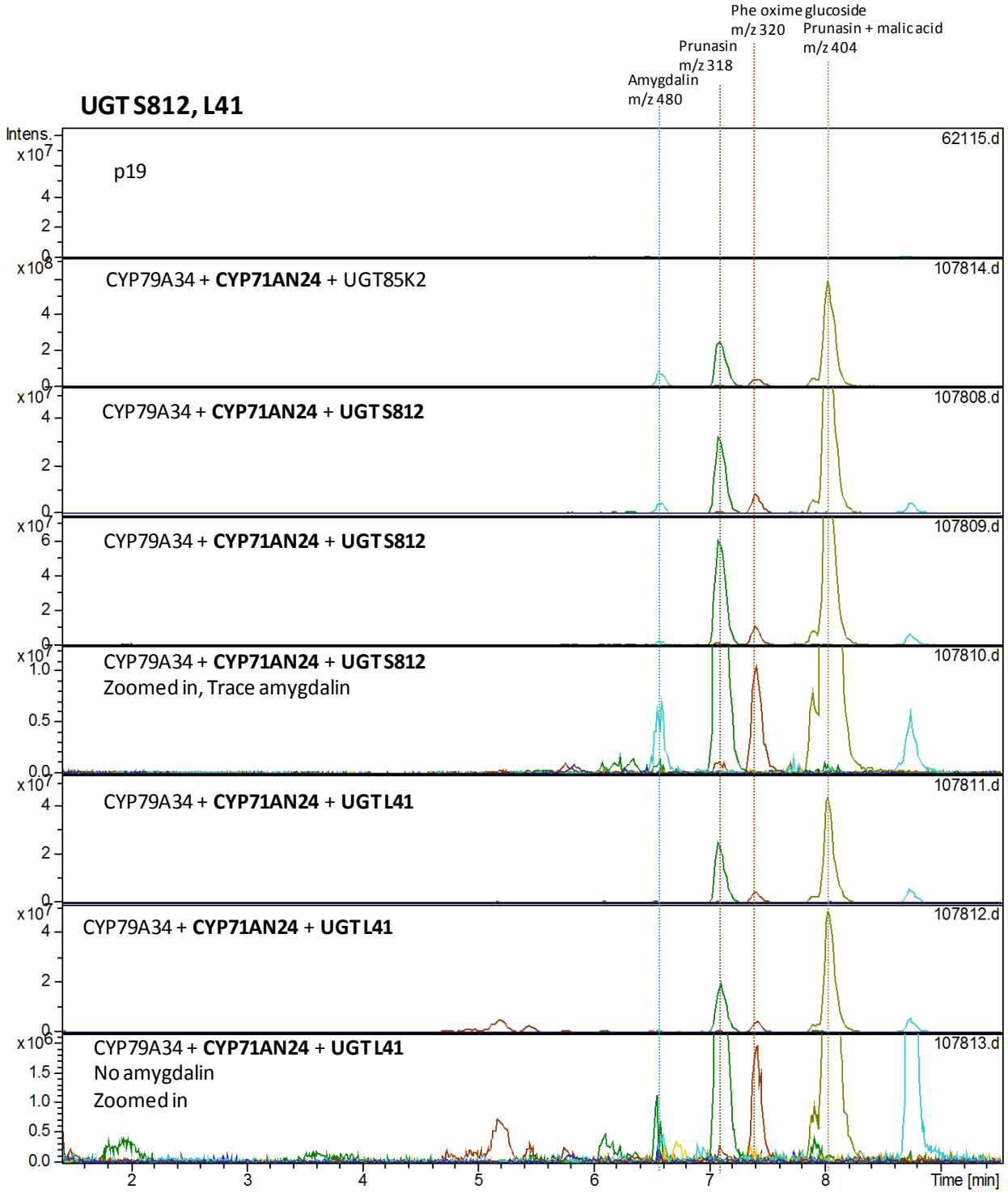


Figure 5.10. LC-MS experiment with the putative UGTs *S812* and *L41*. Constructions were done with the CYP79A34 (*E. yarraensis*) and CYP71AN24 (*P. mume*). Amygdalin, prunasin, Phe oxime glucoside and prunasin+malic acid peaks were detected.

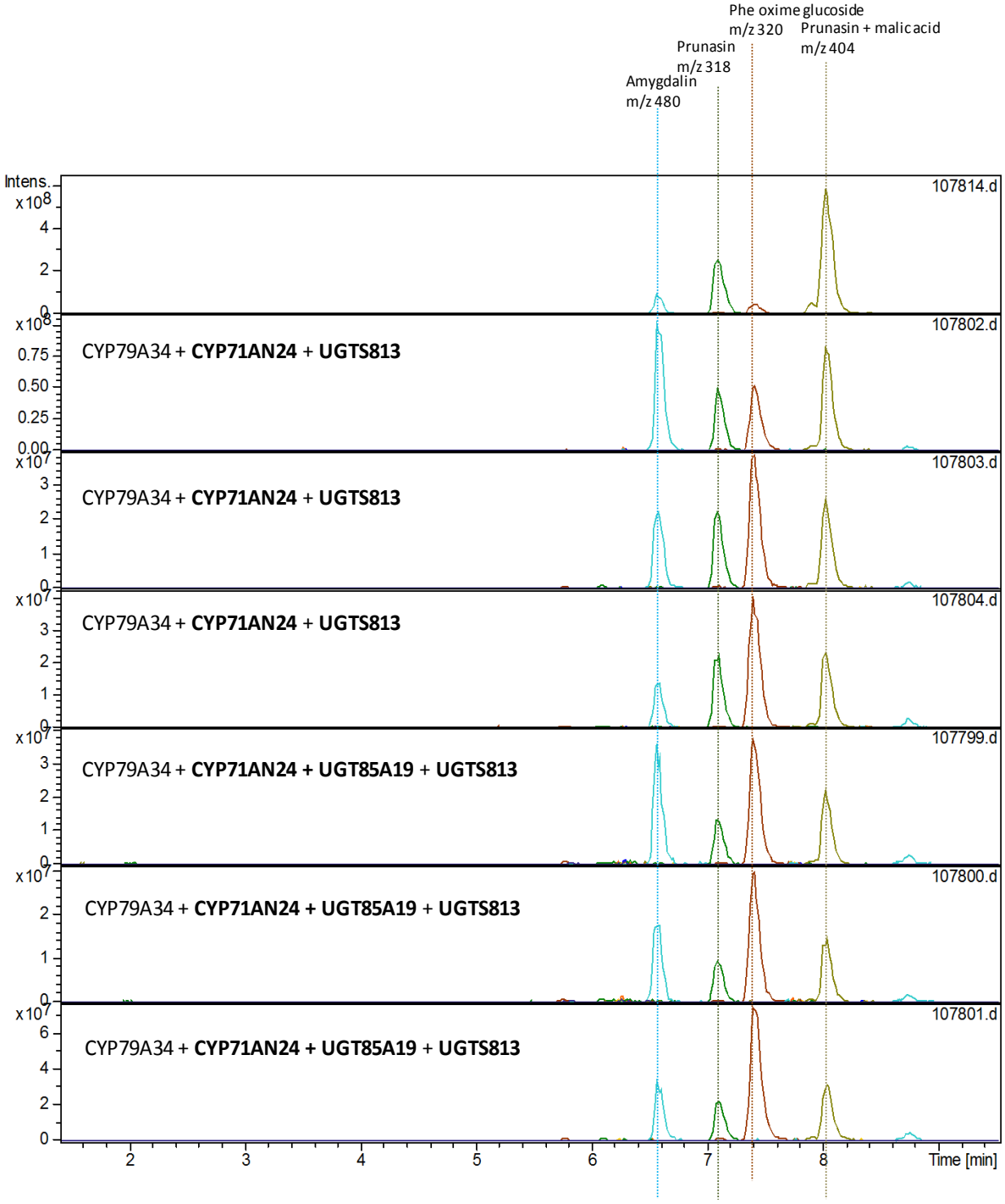


Figure 5.11. LC-MS experiment with the putative UGTS813. Constructions were done with CYP79A34 (*E. yarraensis*), CYP71AN24 (*P. mume*) and UGT85A19 (*P. dulcis*). Amygdalin, prunasin, Phe oxime glucoside and prunasin and malic acid peaks were detected.

Detection of biosynthetic activity by TLC

Only UGTS812 and UGTS811 were able to convert prunasin into amygdalin (Figure 5.12), showing their UGT2 activity. The rest of enzymes did not show UGT1 or UGT2 activity. The double bands of Figure 5.12 could be different linkage isomers of the amygdalin, as previously described by Pičmanová et al. (2015) or neoamygdalin.

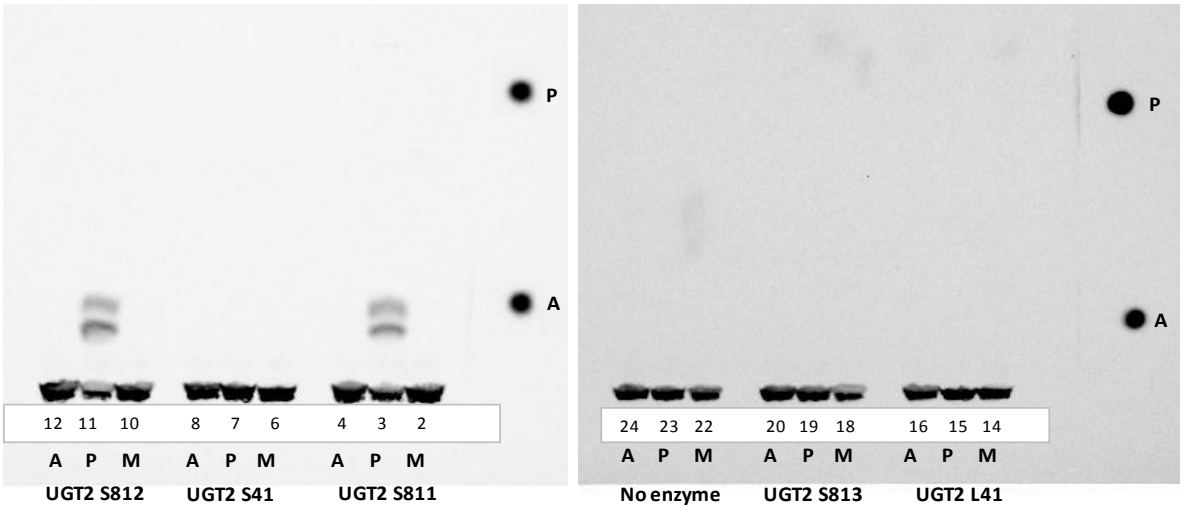


Figure 5.12. TLC experiment with the transcription translation by TNT with five UGT candidate genes *S812, S41, S811*, *S813* and *L41* (see Table 5.5). Only UGT2 S812 and S811 showed a band with the exact Rf as amygdalin as a standard. A: amygdalin, P: prunasin, M: mandelonitrile. Double bands could be different linkage isomers of the amygdalin or neoamygdalin.

In cassava, like in almond and *P. mume* (Yamaguchi et al., 2014), two CYPs belonging to the 79 and 71 families are involved in the linamarin and lotaustralin monoglucosides pathway. In the first one, starting from valine CYP79D1 (Andersen et al., 2000) and CYP71E7 (Jørgensen et al., 2011) and subsequently by UGT85K4 (Kannangara et al., 2011), linamarin is synthesized. On the other pathway, the amino acid isoleucine by the CYP79D2 (Andersen et al., 2000), CYP71E7 and UGT85K5 (Kannangara et al., 2011), lotaustralin is synthesized.

However, the second CYP71 is not present in *L. japonicus*, where there are also linamarin and lotaustralin. In the first pathway, from valine by the action of CYP79D3 and CYP736A2 and a UGT85K2, linamarin was synthesized (Forslund et al., 2004; Bjarnholt et al., 2008, Morant et al., 2008, Takos et al., 2011). In the second pathway, from isoleucine through two cytochromes CYP79D4 and CYP736A2 and a UGT85K3, lotaustralin was synthesized (Forslund et al., 2004; Bjarnholt et al., 2008; Morant et al., 2008; Takos et al., 2011). Linamarin and lotaustralin was also detected by TLC in *L. japonicus* for UGT85K2 and UGT85K3 (Takos et al., 2011).

Finally, not always two different families are involved in the synthesis or secondary metabolites. This is the case of two cytochromes CYP79E1 and CYP79E2 that were identified in the synthesis of triglochinin and taxiphyllin (both derived from Tyr) in the sea arrow grass (*Triglochin maritima*) (Nielsen and Møller, 2000).

5.4.3. Candidate gene analysis in two transcriptomes

Having the results of differential expression analysis on one hand, and a list of *P. persica* protein families related with the regulation of sweet/bitter phenotype (CYP, UGT and BGA) on the other hand, we tried to identify almond orthologous groups of these families. Then we looked if the BGA, the CYP and the UGT genes were in the same co-regulation clusters of the MYC genes in the *Sk* locus.

Identification of orthologous genes in the co-regulation studies

Regarding the identification of orthologous genes, most of the BGA and CYP genes were grouped within the Cluster 1, while most of the UGT genes were within the Cluster 2 (Table 5.5).

Table 5.5. Number of genes belonging to Cluster 1 and Cluster 2 of the co-regulation analysis. BGA: β-glucosidases, CYP: cytochromes, UGT: UDP-glucosyltransferase.

Cluster 1			**Cluster 2**		
BGA	**CYP**	**UGT**	**BGA**	**CYP**	**UGT**
19	30	18	7	14	24

Candidate gene analysis in two transcriptomes

The results from the transcriptomic analysis showed differential expression in diverse candidate genes belonging to the CYP family (Table 5.6), some UDP-glucosyltransferases (Table 5.7), prunasin hydrolase and amygdalin hydrolase (Table 5.8).

CYPs (Table 5.6 and Figure 5.13)

Three CYPs candidate genes were analysed from the transcriptomic analysis. For the three genes studied, the expression was always higher in the bitter genotypes at the two times analysed, especially in the genes *CYP71AN24_ppa004152m* and *CYP79_ppa021236m*, where the TMM expression values were very high.

As a conclusion, we should try to express the ortholog ppa021236m in tobacco, as it looks like it is the first gene in the amygdalin pathway.

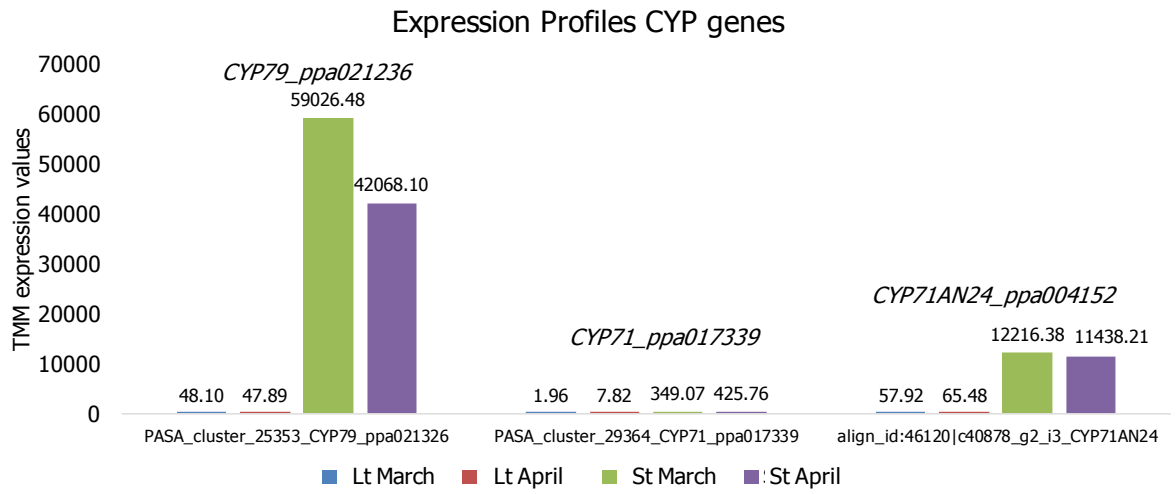


Figure 5.13. Differential expression of three CYP candidate genes in tegument of two cultivars (Lauranne sweet and S3067 bitter) in two different times (March and April).

Table 5.6. CYP candidate genes from the transcriptomic analysis.

Query	Almond Transcript	Almond Gene	% Identity	e-value
CYP79_ppa021236m	align_id:169456|asmbl_40733	PASA_cluster_25353	99.04	0
CYP71_ppa004152m	align_id:46120|c40878_g2_i3|blat.proc74062.chain_4410	align_id:46120|c40878_g2_i3	99.29	0
CYP71_ppa017339m	align_id:175712|asmbl_46989	PASA_cluster_29364	76.8	0

Table 5.7. UGT candidate genes from the transcriptomic analysis.

Query	Almond Transcript	Almond Gene	% Identity	e-value
UGT85A19|ppa004968m	align_id:200166|asmbl_71443	PASA_cluster_44980	99.74	0.00E+00
C4|ppa024453m	align_id:204090|asmbl_75367	PASA_cluster_47353	96.61	0.00E+00
L41|ppa005106m	align_id:129565|asmbl_842	PASA_cluster_503	98.95	0.00E+00
C0C2|ppa023071m	align_id:152682|asmbl_23959	PASA_cluster_14893	91.27	0.00E+00
S811|ppa005471m	align_id:159471|asmbl_30748	PASA_cluster_19119	70.67	7.00E-11
S61_9_S813|ppa005520m	align_id:166061|asmbl_37338	PASA_cluster_23166	80.77	1.00E-07

Table 5.8. Hydrolase candidate genes from the transcriptomic analysis.

Query	Almond Transcript	Almond Gene	% Identity	e-value
PdPhR691	align_id:147815|asmbl_19092	PASA_cluster_11924	71,52	0
PdPhR692	align_id:140936|asmbl_12213	PASA_cluster_7701	98,76	0
AH|ppa024207m	align_id:133837|asmbl_5114	PASA_cluster_3224	98,37	0

UGTs (Table 5.7 and Figure 5.14)

Six UGT candidate genes were analysed from the transcriptomic analysis. The expression was very low in all the evaluated candidate genes except for UGT85A19 where the expression was especially high in the bitter genotype, in April. These results agree with those of Sánchez-Pérez et al. (2008), where prunasin was only detected in the tegument of the bitter almonds. In this sense Franks et al. (2008) observed three-fold greater mandelonitrile glucosyltransferase activity in the bitter genotypes compared to the sweet ones. However, no amygdalin has ever been detected in the tegument (Sánchez-Pérez et al. 2008), which would indicate that UGT2 is not highly expressed in this tissue, what agrees with the results shown here.

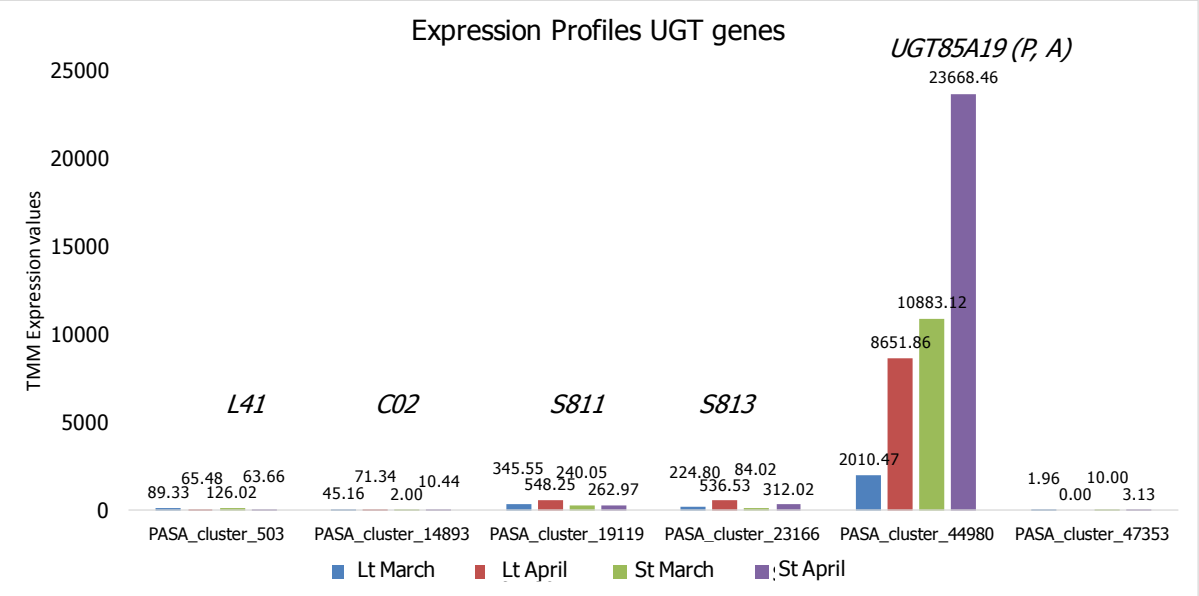


Figure 5.14. Differential expression of six UGT candidate genes in tegument of two cultivars (Lauranne (sweet) and S3067 (bitter)) in two different time points (March and April).

Hydrolases (Table 5.8 and Figure 5.15)

Three hydrolase candidate genes were found when we blasted our candidates versus the transcriptome assembly. Whereas the expression of *BGA_ppa024207m* and *PdPhR692* was very low at both time points and genotypes, the expression levels of *PdPhR691* was very high at both times and cultivars, being a little higher in March than in April. In this sense Lai et al. (2015) detected the expression of BGD (β-glucosidase) gene transcripts in *L. japonicus* tissues (flowers and apical leaves).

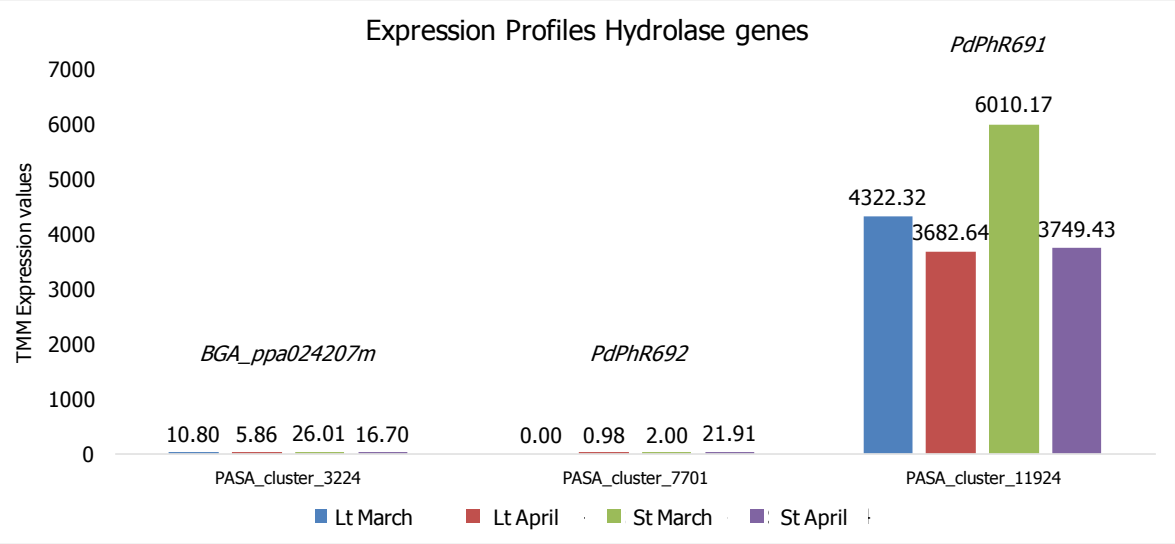


Figure 5.15. Differential expression of three hydrolase candidate genes in tegument of two cultivars (Lauranne (sweet) and S3067 (bitter)) at two different time points (March and April).

5.5. CONCLUSIONS

- Regarding the anabolic enzymes related to bitterness in almonds, five UDP-glucosyltransferases out of the original 13, and a CYP71 have been characterized in the cotyledon of bitter cultivars. When eucalyptus CYP79A34 and almond CYP71AN24 were agroinfiltrated in tobacco plants, together with three different UGTs, prunasin and /or amygdalin was synthesized. In fact, three UGTs were able to make prunasin and also three were able to make amygdalin. Surprisingly, two of them were able to make prunasin and amygdalin. To our knowledge this is the first time an UGT can make prunasin and amygdalin at the same time.

- Concerning the degradation enzymes, PdPH691 has been characterized in sweet and bitter almonds from three tissues: tegument, nucellus and cotyledon. *Ph691* was expressed in the three tissues during kernel development. Differences in the sequences of nucleotides (SNPs) and amino acids were found between the nine PHs sequences studied, but none of these differences were related to the sweet or bitter flavor. Only PH691 from D05-187 tegument showed no β-glucosidases or PH activity. In order to check if bitterness in almond was related to PH activity, other PHs would have to be studied.

- In conclusion of this chapter, further studies need to done to find out how these genes related to the amygdalin pathway are regulated. Our first attempt would be to study the transcription factors found in the small interval of 95.76 kb where the *Sk* locus is place.

6. General conclusions

6. GENERAL CONCLUSIONS.

- In our study, prunasin and amygdalin were detected for the first time in flower buds during the flower development period in almonds. Furthermore, these cyanogenic glucosides were found to play an important role in the flower bud development. Amygdalin was present in flower buds during dormancy breaking, whilst prunasin had a role in the development and opening of the flowers. The nitrogen released by the degradation of the prunasin and amygdalin most certainly promotes flower development.

- Pollen was the tissue with the highest levels of amygdalin and prunasin and their putative derivatives. Moreover, derivates prunasin acid, amide, anitrile, prunasin pentoside and prunasin anitrile pentoside were detected for the first time in the flowers of sweet and bitter cultivars. These derivative compounds may be produced from cyanogenic glucosides in an alternative turnover pathway.

- A population of 550 seedlings of R1000 x Desmayo was phenotyped, increasingly the previous analysis from 167 to 550 individuals. 10 new CAPs and 4 SSRs were located in the GL5. These markers have allowed the saturation the *Sk* locus, fine mapping the region from 3.6 Mb to 95.76 kb. Acording to the peach genome, in this region we found eleven candidate genes, of which six of them were transcription factors of the MYC family, the others were a glyoxal oxidase, a cytochrome P450, an alcohol O-acetyltransferase, a methionine lyase and a hydrolase.
- The resequencing of the almond has allowed transporting these candidate genes from the peach to the almond. The bitter assembly was proposed as the reference genome for the *Sk* gene in the GL5, annotating 10 genes inside the *Sk* locus.
- The transcriptome analysis in almond has shown candidate genes differentially expressed in sweet and bitter almonds, being able to find six candidate genes involved in the *Sk* locus: three MYC transcription factors, a P450 cytochrome, a hydrolase and a glyoxal oxidase. Several SNPs were detected in three candidate genes present in the *Sk* region of 95.76 kb, two MYC transcription factors and the

putative glyoxal oxidase. The gene (ppa022201m) encoding a MYC transcription factor had two variants, one was synonymous and the other a missense. In the gene (ppa025417m) enconding MYC transcription factor a missense mutation was observed. Finally, regarding to the gene (ppa023406m) enconding a glyoxal oxidase, out of the three SNPs observed, one was synonymous and the other two were missense.

- Regarding the anabolic enzymes related to bitterness in almonds, five UDP-glucosyltransferases out of the original 13, and a CYP71 have been characterized in the cotyledon of bitter cultivars. When eucalyptus CYP79A34 and almond CYP71AN24 were agroinfiltrated in tobacco plants, together with three different UGTs, prunasin and /or amygdalin was synthesized. In fact, three UGTs were able to make prunasin and also three were able to make amygdalin. Surprisingly, two of them were able to make prunasin and amygdalin. To our knowledge this is the first time an UGT can make prunasin and amygdalin at the same time.

- Concerning the degradation enzymes, PdPH691 has been characterized in sweet and bitter almonds from three tissues: tegument, nucellus and cotyledon. *Ph691* was expressed in the three tissues during kernel development. Differences in the sequences of nucleotides (SNPs) and amino acids were found between the nine PHs sequences studied, but none of these differences were related to the sweet or bitter flavor. Only PH691 from D05-187 tegument showed no β-glucosidases or PH activity. In order to check if bitterness in almond was related to PH activity, other PHs would have to be studied.

7. Bibliography

7. BIBLIOGRAPHY

Abarrategui L (2010) Bitterness in almonds. Master thesis. University of Copenhagen.

Akopyanz N, Bukanov N, Westblom TU, Berg DE (1992) PCR-based RFLP analysis of DNA sequence diversity in the gastric pathogen *Helicobacter pylori* . Nucleic Acid Research. 20: 6221-6225.

Alburquerque N, García-Montiel F, Carrillo A, Burgos L (2008) Chilling and heat requirements of sweet cherry cultivars and the relationship between altitude and the probability of satisfying the chill requirements. Environmental and Experimental Botany. 64: 162–170.

Alioto TS, Cruz F, Gómez Garrido J, Frías L, Ribeca P, Gut M, Alexiou K, Howad W, Morata J, Casacuberta J, Rubio Cabetas M J, Galán B, García JL, Dhingra A, Duval H, Fernández i Martí AV, Wirthensohn M, Gut IG, Arús Pere (2016) A Draft Assembly of the Almond Genome. Plant and Animal Genome XXIV Conference. The Largest Ag-Genomics Meeting in the World: San Diego, CA, 9-13.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215:403-410.

Andersen MD, Busk PK, Svendsen I, Møller BL (2000) Cytochromes P-450 from cassava (*Manihot esculenta* Crantz) catalyzing the first steps in the biosynthesis of the cyanogenic glucosides linamarin and lotaustralin. The Journal of Biochemical Chemistry. 275**:** 1966–1975.

Aranzana MJ, Carbó J, Arús P (2003a) Microsatellite variability in peach [*Prunus persica* (L.) batsch]: cultivar identification, marker mutation, pedigree inferences and population structure. Theoretical and Applied Genetics. 106: 1341-1352.

Aranzana MJ, Cosson P, Dirlewanger E, Ascasibar J, Cipriani G, Arús P, Testolin R, Abbott A, King GJ, Iezzoni AF (2003b) A set of simple-sequence repeat (SSR) markers covering the *Prunus* genome. Theoretical and Applied Genetics. 106: 819-825.

Arrázola G (2002) Análisis de glucósidos cianogénicos en variedades de almendro: Implicaciones en la mejora genética. Tesis. Universidad de Alicante. Pp. 194.

Arrázola G, Sánchez-Pérez R, Dicenta F, Grané N (2012) Content of the cyanogenic glucoside amygdalin in almond seeds related to the bitterness genotype. Agronomía Colombiana. 30: 260-265.

Arulsekar S, Parfitt DE (1986) Isoenzyme analysis procedure for stone fruits, almond, grape, walnut, pistachio and fig. HortScience 21: 928-933.

Arulsekar S, Parfitt DE, Kester DE (1986) Comparison of isozyme variability in peach and almond cultivars. Journal of Heredity. 77: 272-274.

Arús P, Messeguer R, Viruel MA, Tobutt K, Dirlewanger E, Santi F, Quarta R, Ritter E (1994a) The European *Prunus* mapping project. Euphytica. 77: 97-100.

Arús P, Olarte C, Romero M, Vargas FJ (1994b) Linkage analysis of ten isozyme genes in F_1 segregating progenies of almond. Journal of the American Society for Horticultural Science. 119: 339-344.

Augustin JM, Drok S, Shinoda T, Sanmiya K, Nielsen JK, Khakimov B, Carl Erik Olsen CE, Esben Halkjær Hansen, Kuzina V, Ekstrøm CT, Thure Hauser T, Bak S (2012) UDP-Glycosyltransferases from the UGT73C Subfamily in *Barbarea vulgaris*

Catalyze Sapogenin 3-*O*-Glucosylation in Saponin-Mediated Insect Resistance. Plant Physiology. 160: 1881–1895.

Bacarella A, Chironi G, Barbera G (1991) Aspetti tecnici, economici e di mercato del mandorlo in Sicilia. Quarderni di Ricerca di Sperimentazione (Palermo, Sicily) 40: 1–191.

Bak S, Kahn RA, Nielsen HL, Moller BL, Halkier BA (1998) Cloning of three A-type cytochromes P450, CYP71E1, CYP98, and CYP99 from Sorghum bicolor (L.) Moench by a PCR approach and identification by expression in Escherichia coli of CYP71E1 as a multifunctional cytochrome P450 in the biosynthesis of the cyanogenic glucoside dhurrin. Plant Molecular Biology. 36:393-405.

Bak S, Olsen CE, Halkier BA, Møller BL (2000) Transgenic tobacco and Arabidopsis plants expressing the two multifunctional sorghum cytochrome P450 enzymes, CYP79A1 and CYP71E1, are cyanogenic and accumulate metabolites derived from intermediates in Dhurrin biosynthesis. Plant Physiology 123: 1347-1448.

Bak S, Paquette SM, Morant M, Morant AV, Saito S, Bjarnholt N, Zagrobelny M, Jørgensen K, Osmani S, Hamann T, Simonsen HT, Sánchez-Pérez R, van Heeswijck, Jørgensen K, Møller BL (2006) Cyanogenic glycosides: a case study for evolution and application of cytochromes P450. Phytochem Reviews. 5:309-329.

Ballester J, Socías i Company R, Arús P, de Vicente MC (2001) Genetic mapping of a major gene delaying blooming time in almond. Plant Breeding. 120: 268-270.

Barros PM, Gonçalves N, Saibo NJM, Oliveira MM (2012) Cold acclimation and floral development in almond bud break: insights into the regulatory pathways. Journal of Experimental Botany 63: 4585-96.

Bertozzi T, Bennet C, Sedgley M (1998) Almond improvement in Australia. Nucis-Newsletters. 7: 10-11.

Bjarnholt N, Møller BL (2008) Hydroxynitrile glucosides. Phytochemistry. 69: 1947-1961.

Bliss FA, Arulsekar S, Foolad MR, Becerra AM, Gillen A, Warburton ML, Dandekar AM, Kocsisne GM, Mydin KK (2002) An expanded genetic linkage map of Prunus based on an interspecific cross between almond and peach. Genome 45: 520–529.

Blum MS, Woodring JP (1962) Secretion of benzaldehyde and hydrogen cyanide by the millipede Pachydesmus crassicutis. Science, N.Y. 138:512-513.

Bogatek R, Dziewanowska K, Lewak S (1991) Hydrogen cyanide and embryonal dormancy in apple seeds. Physiologia planetarium. 83: 417-421.

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of human genetics. 32: 314-31.

Boudheri K, Bendahmane A, Cardinet G, Troadec C, Moing A, Dirlewanger E (2009) Phenotypic and fine genetic characterization of the *D* locus controlling fruit acidity in peach. BMC Plant Biology. 9:59.

Browicz K, Zohary D (1996) The genus Amygdalus L. (Rosaceae): species relationships, distribution and evolution under domestication. Genetic Resources and Crop Evolution. 43:229-247.

Cantini C, Iezzoni AF, Lamboy WF, Boritzki M y Struss D (2001) DNA fingerprinting of tetraploid cherry germplasm using SSR. Journal of the American Society for Horticultural Science. 126: 205-209.

Cipriani G, Lot G, Huang HG, Marrazzo MT, Peterlunger E, Testolin R (1999) AC/GT and AG/CT microsatellite repeats in peach (*Prunus persica* (L) Batsch): Isolation, characterization and cross-species amplification in *Prunus*. Theoretical and Applied Genetics. 99: 65-72.

Claverie M, Dirlewanger E, Bosselut N, Van Ghelder C, Voisin R, Kleinhentz M, Lafargue B, Abad P, Rosso MN, Chalhoub B, Esmenjaud D (2011). The Ma gene for complete-spectrum resistance to Meloidogyne species in Prunus is a TNL with a huge repeated C-terminal post-LRR region. Plant physiology. 156:779-792.

Conn E (1969) Cyanogenic glycosides. Journal of Agricultural Food Chemistry. 17: 519-526.

Conn E (1980) Cyanogenic compounds. Annual Revision of Plant Physiology 31: 433-51.

Conn E (1993) â-Glycosidases in plants: substrate specificity. In: Esen A, editor. â-Glucosidases: Biochemistry and Molecular Biology, ACS Symposium Series 533. Washington, DC: American Chemical Society. pp. 15–26.

Crossa-Raynaud P y Grasselly C (1985) Existence de groupes d'interstérilité chez l'amandier. Options Méditerran. Serie Études 1, 43-45.

Dahler JM, McConchie CA, Turnbull CGN (1995) Quantification of cyanogenic glycosides in seedlings of three *Macadamia* (Proteaceae) species. Australian Journal of Botany. 43: 619–628.

Decroocq V, Favé MG, Hagen L, Bordenave L, Decroocq S (2003) Development and transferability of apricot and grape EST microsatellite markers across taxa. Theoretical and Applied Genetics. 106: 912-922.

Denisov VP (1988) Almond Genetic Resources in the USSR and their use in production and breeding. Acta Horticulturae. 224: 299-306.

Dicenta F (1991) Mejora genética del almendro (*Prunus dulcis* Miller) por cruzamientos intervarietales: Herencia de caracteres y selección. Tesis. Universidad de Murcia 313 pp.

Dicenta F, García JE (1993a) Inheritance of self-compatibility in almond. Heredity 70: 313-317.

Dicenta F, García JE (1993b) Inheritance of the kernel flavour in almond. Heredity 70: 308-312.

Dicenta F, Martínez-Gómez P, Ortega E, Duval H (2000) Cultivar pollinizer does not affect almond flavour. HortScience 35: 1153-1154.

Dicenta F, Martínez-Gómez P, Grané N, Martín ML, León A, Cánovas JA, Berenguer V (2002) Relationship between cyanogenic Compounds in kernels, leaves, and roots of sweet and bitter kernelled almonds. Journal of Agricultural and Food Chemistry. 50: 2149-2152.

Dicenta F, García-Gusano M, Ortega E, Martínez-Gómez P (2005) The possibilities of early selection of late flowering almonds as a function of seed germination or leafing time of seedlings. Plant Breeding. 124: 305–309.

Dicenta F, Ortega E, Martinez-Gomez P (2007) Use of recessive homozygous genotypes to assess the genetic control of kernel bitterness in almond. Euphytica. 153: 221–225.

Dicenta F, Ortega E, Martínez-Gómez P, Sánchez-Pérez R, Gambín M, Egea J (2009) Penta and Tardona: Two new extra-late flowering self-compatible almond cultivars. Acta Horticulturae. 814:189-192.

Dicenta F, Sánchez-Pérez, Rubio M, Egea J, Batlle I, Miarnau X, Palasciano M, Liparid E, Confolente C, Martínez-Gómez P, Duval H (2015) The origin of the self-compatible almond 'Guara'. Scientia Horticulturae. 197:1-4.

Dirlewanger E, Cosson P, Howad W, Capdeville G, Bosselut N, Claverie M, Voisin R, Poizat C, Lafargue B, Baron O, Laigret F, Kleinhentz M, Arús P Esmenjaud D (2004a) Microsatellite genetic linkage maps of myrobalan plum and an almond-peach hybrid - Location of root-knot nematode resistance genes. Theoretical and Applied Genetics. 109: 827-832.

Dirlewanger E, Graziano E, Joobeur T, Garriga-Caldré F, Cosson P, Howad W, Arús P (2004b) Comparative mapping and marker-assisted selection in *Rosaceae* fruit crops. Proceedings of the National Academy of Sciences of the United States of America. 101: 9891-9896.

Don RH, Cox PT, Wainwright BJ, Mattick JS (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Research. 19:4008.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry. 19: 11-15.

Egea L, García JE, Egea J, Berenguer T, Gambín M, Patiño JL (1985) Variedades de almendro de la Región de Murcia. Centro de Edafología y Biología Aplicada del Segura (CSIC). Murcia.

Egea L, García JE (1988) Lignes de recherche actuella sur l'amandier au CEBAS Murcie. Objectifs et premières observations. 2 Colloque du GREMPA. Montpellier (Nîmes). Septembre 1975. Pp. 1-11.

Egea J, Dicenta F, Berenguer T, García JE (2000) 'Antoñeta' and 'Marta' almonds. HortScience 35: 1358-1359.

Egea J, Ortega E, Martínez-Gómez Pedro, Dicenta F (2003) Chilling and heat requirements of almond cultivars for flowering. Environmental and experimental botany. 50: 79-85.

Erez A, Couvillon GA (1987) Characterization of the influence of moderate temperatures on rest completion in peach. Journal of the American Society for Horticultural Science. 112: 677-680.

Evreinoff VA (1952) Quelques observations biologiques sur l'amandier. Revue internationale de botanique appliquèe et d'agriculture tropicale. 32: 442-459.

Evreinoff VA (1958) Contribution a l'etude de l'amander. Fruits et primeurs Afrique 28:99-104.

FAO (2005) Los datos de Faostat 2005. http://faostat.fao.org.

Feigl F, Anger V (1966) Replacement of benzidine by copper ethylacetoacetate and tetra basis as spot-test reagent for hydrogen cyanide and cyanogen. Analyst (Lond.) 91: 282-284.

Felipe AJ (1977) Stadi fenologici del mandorlo. Proceedings of the 3rd GREMPA Colloquium, 3-7 October 1977, Valenzano, Bari, Italy. Edizioni Quadrifoglio, Bari. Pp: 101-103.

Felipe AJ, Socias i Company R (1985) L'amélioration génétique de l'amandier à Saragosse. Options Méditterranéennes. Serie Études 1: 9-14.

Felipe AJ, Socías i Company R (1987) 'Aylés', 'Guara' and 'Moncayo' almonds. HortScience 22: 961-962.

Felipe AJ (2000) El almendro. Volume I. El material vegetal. In: MIRA, 1st edition, pp 55-78.

Fennell A (1999) Systems and approaches to studying dormancy: introduction to the workshop. Horticultural Science 34: 1172-1173.

Fernández i Martí A, Howad W, Tao R, Alonso JM, Arús P, Socias i Company R (2011) Identification of QTL associated with self-compatibility in Prunus. Tree Genetetics and Genomes 7: 629-639.

Fernández i Martí A, Font i Forcada C, Socias i Company R (2013). Genetical analysis for physical nut traits in almond. Tree Genetenitcs and Genomes. 9 (2): 455-465.

Flematti GR, Waters MT, Scaffidi A, Merritt DJ, Ghisalberti EL, Dixon KW, Smith SM (2013). Karrikin and cyanohydrins smoke signals provide clues to new endogenous plant signaling compounds. Molecular Plant. 6:29–37.

Font i Forcada C, Fernández i Martí A, Socias i Company (2012) Mapping quantitative trait loci for kernel composition in almond. BMC genetics. 13: 47.

Fooland MR, Arulsekar S, Becerra V, Bliss FA (1995) A genetic map of *Prunus* based on an interspecific cross between peach and almond. Theoretical and Applied Genetics. 91: 262-269.

Forslund K, Morant M, Jørgensen B, Olsen CE, Asamizu E (2004) Biosynthesis of the nitrile glucosides rhodiocyanoside A and D and the cyanogenic glucosides lotaustralin and linamarin in *Lotus japonicus*. Plant Physiology. 135:71–84.

Franks TK, Yadollahi A, Wirthensohn MG, Guerin J, Kaiser BN, Sedgley M, Ford CM (2008) A seed coat cyanohydrin glucosyltransferase is associated with bitterness in almond (*Prunus dulcis*) kernels. Functional Plant Biology. 35: 236-246.

Frehner M, Scalet M, Conn EE (1990) Pattern of the cyanide-potential in developing fruits. Plant Physiol 94: 28–34.

Fukuda T, Ito H, Mukainaka T, Tokuda H, Nishino H, Yoshida T (2003) Anti-tumor Promoting Effect of Glycosides from *Prunus persica* Seeds. Biological and Pharmaceutical Bulletin. 26: 271-273.

Fulgoni VL, Abbey M, Davis P, Jenkins D, Lovejoy J, Most M, Sabate J, Spiller G (2002) Almonds lower blood cholesterol and LDL-cholesterol but not HDL-cholesterol in human subjects: results of a meta-analysis. FASEB Journal. 16:A981–A982.

García JE, Dicenta F, Berenguer T, Egea J (1996) Programa de Mejora del almendro del CEBAS (CSIC-Murcia). Fruticultura Profesional. 81: 64-70.

García-Gusano M, Martínez-García PJ, Dicenta F (2010) Seed germination time as a criterion for the early selection of late flowering almonds. Plant Breeding. 129: 578-580.

Gleadow RM, Woodrow IE (2000) Temporal and spatial variation in cyanogenic in *Eucalyptus cladocalyx*. Tree Physiology. 20: 591-598.

Gleadow RM, Møller BL (2014) Cyanogenic Glycosides: Synthesis, Physiology, and Phenotypic Plasticity. Annual Review of Plant Biology. 65: 155-185.

Godini A (2000) About the possible relationship between Amygdalus webbii Spach. And Amygdalus communis L. Nucis Newsletter. 9:17–19.

Gradziel TM, Kester DE (1994) Breeding for resistance to *Aspergillus flavus* in almond. Acta Horticulturae. 373: 111-117.

Gradziel TM, Kester DE (1996) Genetic improvements. En: Almond production manual. University of California, 70-75.

Gradziel TM, Kester DE (1998) Breeding for self-fertility in California almond cultivars. Acta Horticulturae. 470: 109-117.

Gradziel TM, Kester DE (1999) The University of California almond breeding programme: II. Breeding objectives and projects. Nucis-Newslett. 8: 10-14.

Gradziel TM, Martínez-Gómez P, Dicenta F, Kester DE (2001) The utilization of related *Prunus* species for almond variety improvement. Journal of the American Pomological Society. 55: 100-108.

Gradziel TM (2009) Almond (Prunus dulcis) Breeding. Breeding Plantation Tree Crops: Temperate Species. pp 1-31.

Grasselly C (1972) L'amandier; caractères morphologiques et physiologiques des variétés, madalité de leurs transmissions chez les hybrides de première génération. These Université de Bordeaux I, Bordeaux. Pp. 156.

Grasselly C (1976) Origine et évolution de l´amandier cultivé. Options Méditerran. 32 : 45-50.

Grasselly C, Crossa-Raynaud P (1980) L'amandier. GP Maison et Larose. París.

Grasselly C, Crossa-Raynaud (1984) El almendro. Mundiprensa. Madrid.

Guajardo V, Solís S, Sagredo B, Gainza F, Muñoz C, Gasic K, Hinrichsen P (2015) Construction of High Density Sweet Cherry (Prunus avium L.) Linkage Maps Using Microsatellite Markers and SNPs Detected by Genotyping-by-Sequencing (GBS). PLoS One. 26: 10.

Hadrys H, Balick M, Schierwater B (1992) Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. Molecular Ecology. 1: 55-63.

Halkier BA, Møller BL (1989) Biosynthesis of the cyanogenic glucoside dhurrin in seedlings of *Sorghum bicolor* (L.) Moench and partial purification of the enzyme system involved. Plant Physiology. 90:1552–1559.

Halkier BA, Nielsen HL, Koch B, Møller BL (1995) Purification and characterization of recombinant cytochrome $P450_{TYR}$ expressed at high levels in *Escherichia coli*. 322: 369-377.

Heppner J (1923) The factor for bitterness in the sweet almond. Genetics. 8: 390-392.

Heppner J (1926) Further evidence on the factor for bitterness in the sweet almond. Genetics 11: 605-606.

Howad W, Yamamoto T, Dirlewanger E, Testolin R, Cosson P, Cipriani G, Monforte AJ, Georgi L, Abbott AG y Arús P (2005) Mapping with a few plants: Using selective mapping for microsatellite saturation of the *Prunus* reference map. Genetics. 171: 1305-1309.

Hu Z, Poulton J (1999) Molecular Analysis of (R)-(+)-Mandelonitrile Lyase Microheterogenety in Black Cherry. Plant Physiology. 119: 1535-1546.

Jáuregui B (1998) Localización de marcadores moleculares ligados a caracteres agronómicos en un cruzamiento interespecífico almendro x melocotonero. PhD Thesis dissertation. Universitat de Barcelona, Barcelona, Spain.

Jenrich R, Trompetter I, Bak SS, Olsen CE, Møller BL, Piotrowski M (2007) Evolution of heteromic nitrilase complexes in Poaceae with new functions in nitrile metabolism. Proceedings of the National Academy of Sciences of the United States of America. 104: 18848-53.

Jones PR, Møller BL, Hoj PB (1999) TheUDP-glucose:p-hydroxymandelonitrile-O-glucosyltransferase that catalyzes the last step in synthesis of the cyanogenic glucoside dhurrin in Sorghum bicolor. Isolation, cloning, heterologous expression, and substrate specificity. The Journal of Biological Chemistry. 274 :35483-91.

Joobeur T, Viruel MA, de Vicente MC, Jáuregui B, Ballester J, Dettori MT, Verde I, Truco MJ, Messeguer R, Battle I, Quarta R, Dirlewanger E, Arús P (1998) Construction of a saturated linkage map for *Prunus* using an almond × peach F_2 progeny. Theoretical and Applied Genetics. 97: 1034-1041.

Jørgensen K, Morant AV, Morant M, Jensen NB, Olsen CE, Kannangara R, Motawia MS, Møller BL , Bak S (2011) Biosynthesis of the cyanogenic glucosides linamarin and lotaustralin in cassava: isolation, biochemical characterization, and expression pattern of CYP71E7, the oxime-metabolizing cytochrome P450 enzyme. Plant Phisiology.155: 282-92.

Kannangara R, Motawia MS, Hansen NK, Paquette SM, Olsen CE, Møller BL, Jørgensen K (2011) Characterization and expression profile of two UDP-glucosyltransferases, UGT85K4 and UGT85K5, catalyzing the last step in cyanogenic glucoside biosynthesis in cassava. The Plant Journal. 68: 287-301.

Kephart SR (1990) Starch gel electrophoresis of plant isozymes: A comparative analysis of techniques. American Journal of Botany 77: 693–712.

Kester DE, Asay RN (1975) Almonds. En: *Advances in fruit breeding*. Editores: Janick J y Moore JN. Purdue University Press, West Lafayette. Indiana. Pp. 387-419.

Kester DE, Raddi P, Asay R (1977) Correlations of chilling requirements for germination, blooming and leafing within and among seedling populations of almond. Journal of the American Society for Horticultural Science. 102: 145-148.

Kester DE, Asay RN, Micke WC (1984) Solano, Sonora and Padre almonds. HortScience 19: 138-139.

Kester DE, Micke W, Gradziel T, Weinbaum SA (1990) Almond cultivar research at the University of California. 8 Colloque du GREMPA. Nîmes, Francia.

Kester DE, Gradziel TM, Grasselly C (1991) Almonds. En: *Genetic resources of temperate fruit and nut crops*. Editores: Moore JN y Ballington JR. Wageningen. Holanda. Pp. 701-758.

Kester DE, Gradziel TM (1996) Almonds. En: Fruit breeding, Volume III: Nuts. Editores: Janick J y Moore JN. Wiley: New York Pp: 1-97.

Koepke T, Schaeffer S, Harper A, Dicenta F, Edwards M, Henry RJ, Møller BL, Meisel L, Oraguzie N, Silva H, Sánchez-Pérez R, Dhingra A (2013) Comparative genomics analysis in Prunoideae to identify biologically relevant polymorphisms. Plant Biotechnology Journal. 11: 883–893.

Konieczny A and Ausubel FM (1993) A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. The Plant Journal. 4: 403-410.

Kovalev NV, Kostina KF (1935) A contribution of the study of the genus *Prunus*. Leningrad.

Kuroki GW, Poulton JE (1987) Isolation and characterization of multiple forms of prunasin hydrolase from black cherry (*Prunus serotina* Ehrh.) seeds. Biochemistry and Biophysis. 255: 19-26.

Lai D, Pièmanová M, Abou Hachem M, Motawia MS, Olsen CE, Møller BL, Rook F, Takos AM (2015) Lotus japonicus flowers are defended by a cyanogenic â-glucosidase with highly restricted expression to essential reproductive organs. Plant Molecular Biology. 89: 21-34.

Ladizinsky G (1999) On the origin of almond. Genetic Resources and Crop Evolution 46:143–147.

Lalli DA, Decroocq V, Blenda AV, Schurdi-Levraud V, Garay L, Le Gall O, Damsteegt V, Reighard GL, Abbott AG (2005) Identification and mapping of resistance gene analogs (RGAs) in Prunus: a resistance map for Prunus. Theoretical and applied genetics. 111:1504-13.

Lambert JL, Ramasamy J, Paukstelis JV (1975) Stable reagents for the colorimetric determination of cyanide by modified König reactions. Analytical Chemistry. 47: 916-918.

Lamont BB (1994) Injury-induced cyanogenesis in vegetative and reproductive parts of two *Grevillea* species and their F1 hybrid. Annals of Botany. 71:537-542.

Lecouls AC, MJ Rubio-Cabetas, JC Minot, R Voisin, A Bonnet, G Salesses, E Dirlewanger, D Esmenjaud (1999) RAPD and SCAR markers linked to the Ma1 root-knot nematode resistance gene in 'Myrobalan' plum (Prunus cerasifera Ehr.). Theoretical and Applied Genetics. 99:328–335.

Li CP, Swain E, Poulton JE (1992) *Prunus serotina* amygdalin hidrolase and prunasin hydrolase. Plant Physiology. 100: 282-290.

Lieberei R, Selmar D, Biehl B (1985) Metabolization of cyanogenic glucosides in *Hevea brasiliensis.* Plant Systematics and Evolution. 150: 49-63.

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. The American Journal of Human Genetics. 44: 397-401.

London-Shafir I, Shafir S, Eisikowitch D (2003) Amygdalin in almond nectar and pollen – facts and possible roles. Plant Systematics and Evolution. 238: 87-95.

López J, Prezioso P (2001) A better way to optimize: two-step gradient PCR. Eppendorf Bionews Application Notes, 16:3-4.

Lovejoy JC, Most MM, Lefevre M, Greenway FL, Rood JC (2002) Effect of diets enriched in almonds on insulin action and serum lipids in adults with normal glucose tolerance or type 2 diabetes. The American Journal of Clinical Nutrition. 76:1000-1006.

Loyd RC, Gray E (1970) Amount and distribution of hydrocyanic acid potential during the life cycle of plants of three sorghum cultivars. Agronomy Journal. 45: 159**-**162.

Luo J , Nishiyama Y, Fuell C, Taguchi G, Elliott K, Hill L, Tanaka Y, **Kitayama M, Yamazaki M, Bailey P, Parr A, Michael AJ, Saito K, Martin C (2007)** Convergent evolution in the BAHD family of acyltransferases: identification and characterization of anthocyanin acyltransferases from Arabidopsis thaliana. The Plant Journal. 50: 678–695.

Martin GB, Williams JGK, Tanksley SD (1991) Rapid identification of markers linked to Pseudomonas resistance pene in tomato by using random primers and near isogenic lines. Proceedings of the National Academy of Sciences of the United States of America. 88: 2336-2340.

Martins M, Tenreiro R, Oliveira MM (2003) Genetic reladness of Portuguese almond cultivars assessed by RAPD and ISSR markers. Plant Cell Reports. 22: 71-78.

Martínez-Gómez P, Arulsekar S, Potter D, Gradziel TM (2003a) An extended interspecific gene pool available to peach and almond breeding as characterized using simple sequence repeat (SSR) markers. Euphytica. 131: 313-322.

Martínez-Gómez P, Arulsekar S, Potter D, Gradziel TM (2003b) Relationships among peach and almond and related species as detected by SSR markers. Journal of the American Society for Horticultural Science. 128: 667-671.

Martínez Gómez P, Sánchez Pérez R, Dicenta F (2008) Fruit development in almond for fresh consumption. Journal of the American Pomological Society. 62: 82-86.

Matsuoka N, Ikeda T, El-Aasr M, Manabe H, Murakami Y, Deguchi H, Nohara T (2011) Study of the chemical constituents of Pruni Cortex and its related parts. Journal of Natural Medicines. 65: 166–171.

McCarty CD, Leslie JW, Frost HB (1952) Bitterness of kernels of almond 3 peach hybrids and their parents. Proceedings of the American Society for Horticultural Science. 59: 254-258.

Mentzer C, Favre-Bonvin J (1961) Sur la biogenèse du glucoside cyanogénétique des feuilles de laurier-cerise (Prunus lauro-cerasus). C R Acad Sci Ser III Sci Vie 253: 1072–1074.

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant análisis: A rapid meted to detect markers in specific genomic regions by using segregating populations. Proceedings of the National Academy of Sciences of the United States of America. 88: 9828-9832.

Møller BL, Seigler DS (1991) Biosynthesis of cyanogenic glycosides, cyanolipids and related compounds. In BK Singh, ed, Plant Amino Acids, Biochemistry and Biotechnology. Marcel Dekker, New York, pp563–609.

MorantM, Bak S,Møller BL, Werck-Reichhart D (2003) Plant cytochromes P450: tools for pharmacology, plant protection and phytoremediation. Current Opinion in Biotechnology. 14: 151–162.

Morant AV, Jørgensen K, Jørgensen C, Paquette SM, Sánchez-Pérez R, Møoller BL, Bak S (2008) â-Glucosidases as detonators of plant chemical defense. Phytochemistry 69: 1795-1813.

Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H (1986) Specific enzymatic amplification of DNA *in vitro*: The polymerase chain reaction. Cold Spring Harbor Symposia on Quantitative Biology 51: 263-273.

Muncharaz M (2004) El almendro: Manual técnico. Mundi Prensa. Pp: 200.

Nahrstedt A (1972) Zur Cyanogenese von Prunus Avium. Phytochemistry 11: 3121-3126.

Neale DB, Williams CG (1991) Restriction fragment length polymorphism mapping in conifers and applications to forest genetics and tree improvement. Canadian Journal of Forest Research. 21: 545-554.

Neff MM, Neff JD, Chory, Pepper AE (1998) dCAPS, a simple technique for the

genetic analysis of single nucleotide polymorphisms: experimental applications in

Arabidopsis thaliana genetics. Plant Journal. 14:387–392.

Neilson EH, Goodger JQD, Woodrow IE (2006) Novel aspects of cyanogenesis in *Eucalyptus camphora* subsp. *Humeana.* Functional Plant Biology. 33:487-496.

Neilson EH, Goodger JQD, Motawia MS, Bjarnholt N, Frisch T, Olsen CE, Møller BL, Woodrow IE (2011) Phenylalanine derived cyanogenic diglucosides from *Eucalyptus camphora* and their abundances in relation to ontogeny and tissue type. Phytochemistry. 72: 2325–34.

Neilson EH (2012) Characterisation of Cyanogenic Glucoside Synthesis in Eucalyptus. University of Melbourne.

Neilson EH, Goodger JQD, Woodrow IE, Møller BL (2013) Plant chemical defense: at what cost? Trends in Plant Science. 18: 250-258.

Niedzwiedz-Siegieñ I (1998) Cyanogenic glucosides in *Linum usitatissimum*. Phytochemistry 49:59-63.

Nielsen JS, Møller BL (2000) Cloning and expression of cytochrome P450 enzymes catalyzing the conversion of tyrosine to p-hydroxyphenylacetaldoxime in the biosynthesis of cyanogenic glucosides in Triglochin maritima. Plant Physiology. 122: 1311-1321.

Noguchi A, Fukui Y, Iuchi-Okada A, Kakutani S, Satake H, IwashitaT, Nakao M, Umezawa T, Ono E (2008) Sequential glucosylation of a furofuran lignan, (+)-sesaminol, by Sesamum indicum UGT71A9 and UGT94D1 glucosyltransferases. Plant Journal. 54: 415-427.

Okie WR, Hancock JF (2008) Plums. In Hancock JF (ed.) Temperate fruit crop breeding: germplasm to genomics. Springer, New York, p. 337-357.

Okolie PN, Obasi BN (1993) Diurnal variation of cyanogenic glucosides, thiocyanate and rhodanese in cassava. Phytochemistry. 33: 775–778.

Oracz K, El-Maarouf-Bouteau H, Kranner I, Bogatek R, Corbineau F, Bailly C (2009). The mechanisms involved in seed dormancy alleviation by hydrogen cyanide unravel the role of reactive oxygen species as key factors of cellular signaling during germination. Plant Physiology. 150: 494-505.

Paran I, Michelmore RW (1993) Development of reliable PCR based markers linked to downy mildew resistance genes in lettuce. Theoretical and Applied Genetics 85: 985–993.

Peace C, Bassil N, Main D, Ficklin S, Rosyara UR, Stegmeir T, Sebolt A, Gilmore B, Lawley C, Mockler TC, Bryant DW, Wilhelm L, Iezzoni A (2012) Development and evaluation of a genome-wide 6K SNP array for diploid sweet cherry and tetraploid sour cherry. PLoS One. 7: 12.

Pičmanová M, Neilson E, Motawie M, Erik Olsen C, Sánchez-Pérez R, Gray C, Flitsch S, Agerbirg N, Silvestro D, Lindberg Moller B, Jorgensen K and Bjarnholt N (2015) A recycling pathway for cyanogenic glycosides evidenced by the comparative metabolic profiling in three cyanogenic plant species. The Biochemical Journal. 469: 375-389.

Pirona R, Eduardo I, Pacheco I, Da Silva Linge C, Miculan M, Verde I, Tartarini S, Dondini L, Pea G, Bassi D, Rossini L (2013) Fine mapping and identification of a candidate gene for a major locus controlling maturity date in peach. BMC Plant Biology. 13:166.

Poulton JE (1990) Cyanogenesis in plants. Plant Physiology. 94: 401-405.

Poulton JE (1993) Enzymology of Cyanogenesis in Rosaceous Stone Fruits. In A Esen, ed, Symposium Series 533. American Chemical Society, Washington, DC, pp 170-190.

Poulton JE, Ping Li C (1994) Tissue level compartmentation of (R)-amygdalin and amygdalin hydrolase prevents large-scale cyanogenesis in undamaged *Prunus* seeds. Plant Physiology. 104: 29-35.

Prudencio A, Dicenta F, Martínez-Gómez P (2016) Chilling and heat requirements of Tardona, the latest flowering almond cultivar released in the world. (In progress)

Rho JR, Jun CS, Ha YA, Yoo MJ, Cui MX, Baek HS, Lim JA, Lee YH, Chai KY (2007) Isolation and Characterization of a New Alkaloid from the Seed of *Prunus persica* L. and its Anti-inflammatory Activity. Bulletin of the Korean Chemical Society. 28: 1289-1293.

Roberts EH (1973) Oxidative processes and the control of seed germination. In: W Heydecker, ed, Seed Ecology. Butterworths, London, pp 189–218.

Roberts EH, and Smith RD (1977) Dormancy and the pentose phosphate pathway. In AA Khan, ed, The Physiology and Biochemistry of Seed Dormancy and Germination. Elsevier North-Holland Biomedical Press, Amsterdam, pp 385–411.

Robinson ME (1929) Cyanogenic in plants. Biochemical Reviews of the Cambridge Philosophical Society 5: 126-141.

Rovira M, Clavé J, Romero M, Santos J, Vargas FJ (1997) Self-compatibility in almond progenies. Acta Horticulturae. 470: 66-71.

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods in Molecular Biology. 132: 365-386.

Ruiz D, Campoy JA, Egea J (2007) Chilling and heat requirements of apricot cultivars for flowering. Environmental and Experimental Botany. 61:254–263.

Sabate J, Haddad E (2001) Almond-rich diets simultaneously improve plasma lipoproteins and alpha-tocopherol levels in men and women. Annals of Nutrition and Metabolism. 45:596.

Salazar JA, Ruiz D, Egea J, Martínez-Gómez P (2013) Transmission of Fruit Quality Traits in Apricot (*Prunus Armeniaca* L.) and Analysis of Linked Quantitative Trait Loci (QTLs) using Simple Sequence Repeat (SSR) Markers. *Plant Molecular Biology Reporter.* 31: 1506-1517.

Sánchez-Pérez R, Dicenta F, Gradziel, TM, Arús P, Martínez-Gómez P (2004) Application of molecular markers in almond breeding programmes. Nucis-Newsletter. 12: 9-12.

Sánchez-Pérez R, Martínez-Gómez P, Dicenta F, Egea J, Ruiz D (2005a) Level and Transmission of genetic heterozygosity in apricot, explored by simple sequence repeat markers. Genetic Resources and Crop Evolution. 53: 763-770.

Sánchez-Pérez R, Ruíz D, Dicenta F, Egea J y Martínez-Gómez P (2005b) Application of simple sequence repeat (SSR) markers in apricot breeding: molecular characterization, protection, and genetic relationships. Scientia Horticulturae. 103: 305-315.

Sánchez-Pérez R, Howad W, Dicenta F, Arús P. and Martínez-Gómez (2007) Mapping major genes and quantitative trait loci controlling agronomic traits in almond. Plant Breeding. 126: 310-318.

Sánchez-Pérez R, Ortega E, Duval H, Martínez-Gómez P, Dicenta F (2007) Inheritance and relationships of important agronomic traits in almond. Eupytica. 155: 381-391.

Sánchez-Pérez R, Jørgensen K, Erik Olsen C, Dicenta F, Møller BL (2008) Bitterness in almonds. Pant Physiology. 146: 1040-1052 .

Sánchez-Pérez R, Jøorgensen K, Motawia MS, Dicenta F, Møller LB (2009) Tissue and cellular localization of individual â-glycosidases using a substrate-specific sugar reducing assay. The Plant Journal. 60: 894-906.

Sánchez-Pérez R,Howad W,Garcia-Mas J,Arus P,Martinez-Gomez P, Dicenta F (2010) Molecular markers for kernel bitterness in almond. Tree Genetics & Genomes. 6:237–245.

Sánchez-Pérez R, Belmonte FS, Borch J, Dicenta F, Møleer BL, Jørgensen K (2012) Prunasin Hydrolases during Fruit Developmentin Sweet and Bitter Almonds. Plant Physiology. 158: 1916-1932.

Saura-Calixto FS, Bauza M, de Toda FM, Argamenteria A (1981) Amino acids, sugars and inorganic elements in the sweet almond (*Prunus amygdalus*). Journal of agricultural and food chemistry. 29: 509–511.

Schappert PJ, Shore JS (2000) Cyanogenesis in *Turnera ulmifolia* L. (Turneraceae): II. Developmental expression, heritability and cost of cyanogenesis. Evolutionary Ecology Research. 2: 337-352.

Schirra M (1997) Postharvest technology and utilization of almonds. Horticultural Reviews. 20:267–292.

Schuelke M (2000) An economic method for the fluorescent labelling of PCR fragments. Nature Biotechnology. 18:233-234.

Scorza R, Okie WR (1990) Peaches (Prunus) In: Moore JN, Ballington HJ (Eds.) Genetic Resources of Temperate fruit and nut crops, vol I. ISHS Wageningen, The Netherlands, pp 177-231.

Sedgley M, Collins G (2002) Almond improvement in Australia. Fruits 57: 129-134.

Selmar D (1988) Mobilization and utilization of cyanogenic glycosides. Plant Physiology. 86: 711-716.

Selmar D (1990) Cyanogenic lipids. Plant Physiology. 93: 631-636.

Selmar D, Irandoost Z, Wray V (1996) Dhurrin-6´-glucoside, a cyanogenic diglucoside from *Sorghum bicolor*. Phytochemistry 43: 569–572.

Silva C, García-Mas J, Sánchez AM, Arús P, Oliveira MM (2005) Looking into flowering time in almond (*Prunus dulcis* (Mill) D. A. Webb): the candidate gene approach. Theoretical and Applied Genetics. 110: 959-968.

Socias i Company R, Felipe AJ (1987) Pollen tube growth and fruit set in a self-compatible almond selection. Horticultural Science 22, 113-116.

Socias i Company R, Felipe AJ (1992) Self-compatibility and autogamy in 'Guara' almond. Journal of Horticultural Science 67: 313-317.

Socias i Company R (1998) Fruit tree genetics at a turning point: the almond example. Theoretical and Applied Genetics. 96: 588–601.

Socias i Company R, Felipe AJ, Gomez Aparisi J (1999) A major gene for flowering time in almond. Plant Breeding. 118: 443-448.

Socias i Company R (2002) The relationship of Prunus webbii and almond revisited. Nucis-Newsletter. 11: 17–19.

Socias i Company R, Felipe AJ (2006) "Belona" and "Soleta", two new autogamous almonds. Nucis 13: 12-15.

Socias i Company R, Kodad O, Alonso JM, Felipe AJ (2008) 'Mardía' Almond. Horticultural Science 43: 2240-2242.

Sonneveld T, Robbins TP, Boškoviæ R, Tobutt KR (2001) Cloning of six cherry self-incompatibility alleles and development of allele-specific PCR detection. Theoretical and Applied Genetics. 102: 1046-1055.

Sosinski B, Gannavarapu M, Hager LE, Beck LE, King GJ, Ryder CD, Rajapakse S, Baird WV, Ballard RE y Abbott AG (2000) Characterization of microsatellite markers in peach (*Prunus persica* (L) Batsch). Theoretical and Applied Genetics. 101: 421-428.

Suelves M, Puigdoménech P (1998) Molecular cloning of the cDNA coding for the (R)-(+)-mandelonitrile lyase of Prunus amygdalus: temporal and spatial expression patterns in flowers and mature seeds. Planta. 206: 388-393.

Sunyaev SR, Eisenhaber F, Rodchenkov IV, Eisenhaber B, Tumanyan VG, Kuznetsov EN (1999) PSIC: profile extraction from sequence alignments with position-specific counts of independent observations. Protein Engineering. 12: 387-394.

Swain E, Li CP, Poulton JE (1992) Development of the potential for cyanogenesis in maturing black cherry (*Prunus serotina* Ehrh.) fruits. Plant Physiology. 98: 1423-1428.

Swain E, Li CP, Poulton JE (1992) Tissue and subcellular localization of enzymes catabolizing (R)-amygdalin in mature *Prunus serotina* seeds. Plant Physiology. 100: 291-300.

Swain E, Poulton JE (1994) Immunocytochemical localization of prunasin hydrolase and mandelonitrile lyase in stems and leaves of *Prunus serotina*. Plant Physiol 106: 1285-1291.

Swamy, B.G.L. and K.V. Krishnamurthy (1980) From flower to fruit. Embryology of flowering plants. Tata MacGraw-Hill, New Delhi.

Takos A, Lai D, Mikkelsen L, Abou Hachem M, Shelton D, Motawia MS, Olsen CE, Wang TL, Martin C, Rook F (2010) Genetic screening identifies cyanogenesis-deficient mutants of Lotus japonicus and reveals enzymatic specificity in hydroxynitrile glucoside metabolism. The plant cell. 22: 1605-1619.

Takos AM, Knudsen C, Lai D, Kannangara R, Mikkelsen L, Motawia MS, Olsen CE, Sato S, Tabata S, Jørgensen K, Møller BL, Rook F (2011) Genomic clustering of cyanogenic glucoside biosynthetic genes aids their identification in Lotus japonicus and suggests the repeated evolution of this chemical defence pathway. The plant journal: for cell and molecular biology. 68: 273-286.

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. 28:2731–273.

Tanaka O, Cleland CF, Ben-Tal Yosef (1983) Effect of Ferricyanide, Ferrocyanide and KCN on Growth and Flowering in the Short-Day Plant Lemna paucicostata 6746. Plant Cell Physiology. 24: 705-711.

Tankskey, S.D. and T.J. Orton (1983) Isozymes in plant genetics and breeding, part A. Elsevier, Amsterdam.

Tavassolian I, Rabiei G, Gregory D, Mnejja M, Wintherson MG, Hunt PW, Gibson JP, Ford CM, Sedgley M, Wu SB (2010) Construction of an almond linkage map in an Australian population Nonpareil × Lauranne. BMC Genomics. 11: 551.

Taylorson RB, Hendricks SB (1973) Promotion of seed germination by cyanide. Plant Physiology. 52: 23-27.

Testolin R, Marrazo T, Cipriani G, Quarta R, Verde I, Dettori T, Pancaldi M, and Sansavini S (2000) Microsatellite DNA in peach (*Prunus persica* (L.) Batsch) and it use in fingerprinting and testing the genetic origin of cultivars. Genome, 43:512–520.

Testolin R, Messina R, Lain O, Marrazo T, Huang G, Cipriani G (2004) Microsatellites isolated in almond from an AC-repeat enriched library. Molecular Ecology Notes. 4: 459-461.

Van Ooijen JW (2011) JoinMap 4.1, Software for the Calculation of Genetic Linkage Maps, Plant Research International. Wageningen, the Netherlands.

Vargas FJ, Romero MA, Vila J (1980) Information sur le programme d'amélioration de l'amandier par croisements de variétés de la Diputación de Tarragona (Espagne). Options Méditerran. Serie Études 81/I: 43-45.

Vargas FJ, Romero MA, Vila J (1982) Selección y mejora de variedades de almendro en la Diputación de Tarragona. I Convención Nacional de la Almendra. Murcia.

Vargas FJ, Romero MA, Rovira M, Girona J (1984) Amélioration de l'amandier par croisements de variétés. Résultats préliminaires à Tarragone (Espagne). Options Méditerran. Serie Études 84/II: 101-122.

Vargas FJ, Romero MA (1988) Comparación entre descendencias de cruzamientos intervarietales de almendro en relación con la época de floración y la calidad del fruto. 7 Colloque du GREMPA. Reus, Junio 1987. Pp: 59-72.

Vargas FJ, Romero MA (1992) Ensayo de producción de selecciones de almendro obtenidas por el IRTA en Tarragona (España) y de otras variedades de referencia. Primeros resultados. 8 Colloque du GREMPA. Nîmes, 26-27 Juin 1990. Rapport EUR 14081, pp. 115-119.

Vargas FJ, Clavé J, Romero MA, Batlle I, Rovira M (1997) Autogamy studies on almond progenies. Acta Horticulturae 470: 74-81.

Vargas FJ, Romero MA, Batlle I (2001) Kernel taste inheritance in almond. In: Ak B.E. (ed.) 11 GREMPA Seminar on pistachios and almonds. Cahiers Options Méditerranéennes 56: 129-134.

Vargas FJ, Romero MA, Clavé J, Vergés J, Santos J, Batlle I (2006) Four New Almond Varieties Released by IRTA: Vayro, Marinada, Constantí and Tarraco. Nucis 13: 9-12.

Vavilov NI (1930) Wild progenitors of the fruit trees of Turkistan and Caucasus and the problem of the origin of fruit trees. IX International Horticultural Congress report and proceedings. London August 7-15, 1930. Pp 271-286.

Verde I, Lauria M, Dettori MT, Vendramin E, Balconi C, Micali S, Wang Y, Marrazzo MT, Cipriani G, Hartings H, Testolin R, Abbott AG, Motto M, Quarta R (2005) Microsatellite and AFLP markers in the *Prunus persica* [L. (Batsch)] x P. Ferganensis BC1 linkage map: saturation and coverage improvement. Theoretical and applied genetics. 111: 1013-1021.

Vezvaei A, Hancock TW, Giles LC, Clarke GR, Jackson JF (1995) Inheritance and linkage of isozyme loci in almond. Theoretical and Applied Genetics. 91: 432-438.

Viruel MA, Messeguer R, de Vicente MC, García-Mas J, Puigdomènech P, Vargas F, Arús P (1995) A linkage map with RFLP and isozyme markers for almond. Theoretical and Applied Genetics. 91: 964-971.

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. Plant Journal. 33: 949–956.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research. 23: 4407-4417.

Watkins R (1979) Cherry, plum, peach, apricot and almond. Prunus spp. In: N. W. Simmonds (ed.), Evolution of crop plants. p. 242–247. Longman, London.

Webber BL, Woodrow IE (2008) Intra-plant variation in cyanogenesis and the continuum of foliar plant defense traits in the rainforest tree *Ryparosa kurrangii* (Achariaceae). Tree Physiology. 28:977–84.

Weinberger JH (1950) Chilling requirements of peach varieties. Procedings of the American Society for Horticultural Science. 56: 122-128.

Werner DJ y Creller MA (1997) Genetic studies in peach: inheritance of sweet kernel and male sterility. Journal of the American Society for Horticultural Science. 122: 215-217.

Wheeler JL, Mulcahy C, Walcott JJ, Rapp GG (1990) Factors affecting the hydrogen cyanide potential of forage sorghum. Australian Journal of Agricultural Research. 41: 1093–1100.

Wilsie C (1966) Cultivos: aclimatación y distribución. Editorial Acribia. Zaragoza.

Williams JG , Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research. 18: 6531-6535.

Wirthensohn MG, Chin WL, Franks TK, Baldock G, Ford CM, Sedgley M (2008). Characterizing the flavor phenotypes of almond (*Prunus dulcis* Mill.) kernels. The Journal of Horticultural Science and Biotechnology.83: 462–468.

Wünsch A, Hormaza JI (2002a) Cultivar identification and genetic fingerprinting of temperate fruit tree species using DNA markers. Euphytica 125: 56-67.

Wünsch A, Hormaza JI (2002b) Molecular characterisation of sweet cherry (*Prunus avium* L.) cultivars using peach (*Prunus persica* L.) SSR sequences. Heredity. 89: 56-63.

Xu X, Yu T, Xu R, Shi Y, Lin X, Xu Q, Qi X, Weng Y, Chen X (2016) Fine mapping of a dominantly inherited powdery mildew resistance major-effect QTL, *Pm1.1*, in cucumber identifies a 41.1 kb region containing two tandemly arrayed cysteine-rich receptor-like protein kinase genes. Theoretical and Applied Genetics. 129:507–516.

Yamaguchi T, Yamamoto K, Asano Y (2014) Identification and characterization of CYP79D16 and CYP71AN24 catalyzing the first and second steps in L-phenylalanine-derived cyanogenic glycoside biosynthesis in the Japanese apricot, Prunus mume Sieb. et Zucc. Plant Molecular Biology. 86:215–223.

Yonekura-Sakakibara K, Hanada K (2011) An evolutionary view of functional diversity in family 1 glycosyltransferases. Plant Journal. 66:182-193.

Zagrobelny M, Bak S, Rasmussen AV, Jørgensen B, Naumann CM, Møller BL (2004) Cyanogenic glucosides and plant-insect interactions. Phytochemistry 65: 293–306.

Zagrobelny M, Bak S, Møller BL (2008) Cyanogenesis in plants arthropods. Phytochemistry 69 (7): 1457–1468.

Zheng L, Poulton JE (1995) Temporal and spatial expression of amygdalin hydrolase and (R)-(+)- mandelonitrile lyase in black cherry seeds. Plant Physiology 109: 31-39.

Zhou J, Hartmann S, Shepherd BK, Poulton JE (2002) Investigation of the microheterogeneity and aglycone specificity-conferring residues of black cherry prunasin hydrolases. Plant Physiology 129: 1252-1264.

8. Protocol annexes

8. PROTOCOL ANNEXES

ANNEX 8.1. DNA EXTRACTION

(Method "CTAB" (Doyle and Doyle, 1987), with the modifications of Sonneveld et al. (2001).

MATERIAL

- Steel balls of 3 mm diameter.
- Eppendorf tubes of 1.5 and 2 ml.
- Pliers.

REACTIVES

- Bi-distilled water.
- CTAB (Hexadecyltrimethylammonium bromide) (Sigma).
- Sodium chloride (Sigma).
- EDTA (Sigma).
- PVP 40 (Sigma).
- Tris (Trizma Base) (Sigma).
- Liquid nitrogen.
- Chloroform: isoamyl alcohol (24:1) (Sigma).
- Isopropanol (Sigma).
- Ethanol 100% (Panreac).
- RNase A (Roche).

INSTRUMENTATION

- Bench-top unit Retsch MixerMill MM 200 with vials for Eppendorf tubes.
- Thermostatic bath.
- Extraction hood.
- Mini centrifuge 1-14 (Sigma-Aldrich).

METHOD

Before starting, prepare the extraction buffer (work in the fumehood): 2% CTAB, 1.4 M NaCl, 20 mM EDTA, 100 mM Tris HCl pH 8 and 2% PVP 40. Autoclave and keep at room temperature (rt).

1. Put 2-4 leaf discs or 2 young and small leaves in a 2ml eppendorf tube and keep at -80 ºC. When we are going to do the extraction put in liquid nitrogen. The plant material is very important; with young leaves the results are much better.
2. Crush the samples in a mortar, or in a special machine for grinding, with 2 metal balls always with liquid nitrogen. We have to get powder. Keep always frozen the samples, if the defrost the extraction is not valid. This step is very important, powder is necessary to break the tissues.
3. Prepare the buffer that we are going to use, add 2% of β-mercaptoethanol and preheat at 65 ºC in a bath.

4. Add 750 µL of the preheated buffer to each sample and incubate 30 min at 65 ºC in a thermoblock and mix, by vortex, every 5 min. This step is to break the cell wall and the cellular membrane.
5. Add 750 µL of chloroform isoamyl alcohol (24:1) and mix well by inversion with the hand for 2 min.
6. Centrifuge the samples 14,000 rpm 10 min at room temperature and take the supernatant to a new eppendorf tube with the pipette. Separate the nucleic acid of the rest.
7. Repeat the step 5, if the supernatant is not clear.
8. Add 450 µL of cold isopropanol (-20 ºC) to precipitate the DNA and mix by inversion with the hand. When DNA precipitation occurs, you may see the DNA strands (cottony mass).
9. Centrifuge the samples 14,000 rpm 10 min at room temperature and remove the supernatant, being careful not to lose the pellet.
10. Wash the pellet with 500 µL cold ethanol 70% (-20 ºC), vortex.
11. Let the pellet with the ethanol overnight at -20 ºC or for 2 hours.
12. The next day centrifuge 14,000 rpm 10 min, remove the supernatant and repeat the step 9, centrifuge again.
13. Remove the ethanol and dry the tubes in horizontal position for 2 hours in the fumehood. It is very important to remove all the ethanol. However, the longer you let the samples drying the harder will be to resuspend the pellet later.
14. Resuspend the pellet in 40-100 µL TE buffer pH8.
15. RNase treatment: Add 1 µL RNase 1 mg/ml in 100 µL elution and let overnight at 4 ºC.
16. Incubate the next day at 37 ºC for 1 hour to digest the RNA. This step is to remove all the RNA.
17. Keep the samples at -20 ºC.
18. To quantify the DNA: Nanodrop. Good quality: Ratio 260/280: between 1.8-2.0. PCR with specific primers and electrophoresis.

ANNEX 8.2. RNA EXTRACTION BY THE METHOD "CTAB"

MATERIAL

- Steel balls of 3 mm diameter.
- Eppendorf tubes of 1.5 and 2 ml.
- Pliers.

REACTIVES

- Bidistilled water.
- CTAB (Hexadecyltrimethylammonium bromide) (Sigma).
- Sodium chloride (Sigma).
- EDTA (Sigma).
- PVP 40 (Sigma).
- Tris (Trizma Base) (Sigma).
- Liquid nitrogen.
- Chloroform: isoamyl alcohol (24:1) (Sigma).
- Isopropanol (Sigma).
- Ethanol 100% (Panreac).
- RNase A (Roche).

INSTRUMENTATION

- Bench-top unit Retsch MixerMill MM 200 with vials for Eppendorf tubes.
- Thermostatic bath.
- Fume hood.
- Mini centrifuge (Sigma 1-13).

METHOD

Before starting, prepare the extraction buffer (work in the fume hood): 2.5% CTAB, 2M NaCl, 25mM EDTA,100 mM Tris HCl pH8, 2% PVP, 20 mM hydrous $CaCl_2$. Autoclave and keep at room temperature (rt).

1. Plant material: 200-400 mg young leaves in a 2ml eppendorf tube and keep it at -80 ºC. When you are going to do the extraction put in liquid nitrogen.
2. Crush the samples in a mortar with liquid nitrogen and transfer to a 2ml eppendorf tube. We have to get powder. Avoid to defrost the samples.
3. Prepare the buffer that we are going to use, add 5% β-mercaptoethanol and preheat at 65 ºC in a bath.
4. Add 900 µL of the preheated buffer to each sample and incubate 30 min at 65 ºC in a thermoblock and mix by vortex every 5 min.
5. Add 400 µL of chloroform isoamyl alcohol (24:1) and mix well by inversion with the hand for 2 min.
6. Mix by vortex.
7. Centrifuge 16,000 x g 10 min 4 ºC.
8. Take the supernatant to a new 1.5 ml eppendorf tube.

9. Add chloroform isoamyl alcohol (24:1) in the same volume as supernatant.
10. Repeat steps 6, 7 and 8.
11. Add LiCl 10M to the final concentration of 3 M.
12. Mix by inversion with the hand.
13. Precipitate at -20 ºC overnight.
14. Centrifuge 21,000 x g 10 min 4 ºC.
15. Remove the supernatant and dissolve the pellet in 500 µL TE + SDS 1% + NaCl 1M buffer preheated at 65 ºC.
16. Add 500 µL of chloroform isoamyl alcohol (24:1).
17. Mix by inversion.
18. Centrifuge 16,000 x g 10 min 4 ºC.
19. Transfer the supernatant to a new 1.5 ml eppendorf tube and add 0.7 volumes of isopropanol.
20. Mix by inversion.
21. Centrifuge 21,000 g 15 min 4 ºC.
22. Remove the supernatant and wash with 500 µL ethanol 70%.
23. Vortex
24. Centrifuge 16,000 x g 10 min 4 ºC.
25. Remove the supernatant and dry the tubes in horizontal position in the fume hood until all the ethanol is removed.
26. Resuspend the pellet in DEPC-water in ice.
27. DNase treatment: turbo DNA-free Kit (Ambion, Life Technologies).
28. Keep the samples at -80 ºC.
29. To quantify the DNA: Nanodrop. Good quality: Ratio 260/280: 2

ANNEX 8.3. cDNA SYNTHESIS

MATERIAL

- Pipettes and tips.
- Plates and sterile PCR tubes (0.2 ml).
- Sterile tubes (1.5 ml).

REACTIVES

- Sterile distilled water.
- 5x iScript reaction mix (Bio Rad).
- Reverse transcriptase (Bio Rad).

INSTRUMENTATION

- Thermocycler Eppendorf Mastercycler.

METHOD

The kit used for the synthesis of cDNA was the iScript cDNA synthesis kit (Bio Rad). In this protocol, the following components were used in a total volume of 20 µL:

- 4 µL 5x iScript reaction mix
- 1 µL Reverse transcriptase
- Water
- RNA template (up to 1 µg RNA)

And these compounds were incubated in a thermocycler with the next conditions:

- 5 min at 25 ºC
- 30 min at 42 ºC
- 5 min at 85 ºC
- 4 ºC

cDNA product was used for a direct PCR or the RACE experiment.

ANNEX 8.4. RACE

MATERIAL

- Pipettes and tips.
- Plates and sterile PCR tubes (0.2 ml).
- Sterile tubes (1.5 ml).

REACTIVES

- Buffer HF
- dNTP 10mM
- GSP: Gene Specific Primers
- UPM: Universal Primer Mix
- Phusion Polymerase

INSTRUMENTATION

- Thermocycler Eppendorf Mastercycler.

METHOD

1. 5´-RACE and 3´-RACE PCR reactions were done to generate the 5´and 3´cDNA fragments following manufacturer instructions (ClonTech). The full length reaction was done with the next conditions:

- 5'-3'-RACE-Ready cDNA: 2.5 µl
- Buffer HF: 10 µL
- dNTP 10mM: 1 µL
- UPM (10X): 5 µL
- GSP 5´or 3´ (10 µM): 1 µL
- Water: 30 µL
- Phusion Polymerase: 0,5 µL
- Total: 50 µL

- 30 s at 98 ºC
- 10 s at 98 ºC
- 30 s at 68 ºC
- 1 min at 72ºC

30 cycles

- 5 min at 72 ºC
- 12 ºC

Primers used to perform RACE by SMARTer™ RACE cDNA Amplification Kit (Clontech).

		Position	Nt	Tm	Total size	Expected size
C0_3RACEF2	CGAAGTTGTAGAGAGATTGAGGGTGTGT	421	28	65	2000	1579
C0_5RACER2	ACCTTGGACCGTCAACTTCACACGA	402	25	70	2000	1598
C1_3RACEF2	CGAACAAAGGGTCGTGGACTGGTTT	631	25	70	2000	1369
C1_5RACER2	CCCTTCCTCATCCACCACAACCA	934	23	65	2000	1066
C2_3RACEF2	CAAATCGAGCCACTTCCCCTCAGTC	697	25	70	2000	1303
C2_5RACER2	CGTCCTCTTCTCATGCTGCCTTGCT	276	25	70	2000	1724
C3_3RACEF2	GACAGCATTACTGGCGACGAAAGCA	538	25	70	2000	1462
C3_5RACER2	GTTTTCGGTCCAACCACTGGCTCAT	769	25	70	2000	1231
C4_3RACEF2	TGGTGCCAGAGTCCCCAACTACAGA	158	25	70	2000	1842
C4_5RACER2	CACCACCGGCCCTGCAAGAATCACT	162	25	75	2000	1838
L41_3RACEF2	CGGTGGGTGGATTCTTGACTCACTG	830	25	70	2000	1170
L41_5RACER2	GAATAGGATCAAAACCCGCCCGAAC	909	25	70	2000	1091
L43_3RACEF2	AAGGCATGTTGGGTGCGGATGCACT	402	25	75	2000	1598
L43_5RACER2	GTGCATCCGCACCCAACATGCCTTC	424	25	75	2000	1576
S31_3RACEF2	ACGGTGTTGCTTGTGCTGCGATAAC	140	25	70	2000	1860
S31_5RACER2	CTTGTATTTGCTGACGGGTGCGGTCCA	2	25	75	2000	1998
S41_3RACEF2	AGAGGATAGGAGGGCGTGGACTGGT	740	25	70	2000	1260
S41_5RACER2	TTCTTCTTGGCTCGGCATGGCCTCA	645	25	75	2000	1355
S61_3RACEF2	CCCTCCCCATCTCATGTCCACTCTC	267	25	70	2000	1733
S61_5RACER2	TCTGGCTTTAGGGCTGTGAGGATGG	353	25	70	2000	1647
S811_3RACEF2	GGGTTGGGCTCCGCAGGCAAACATA	685	25	75	2000	1315
S811_5RACER2	AATTCCCTTCTCCCCCACCCTCTCT	675	25	69	2000	1325
S812_3RACEF2	GTCCACTCTCAAAACCGCCTTCGAC	279	25	70	2000	1721
S812_5RACER2	CCTGTGGAGCCCAACCATCCACTAT	997	25	70	2000	1003
S813_3RACEF2	ACTGCGGATGGAACTCAGTGTTGGA	770	25	70	2000	1230
S813_5RACER2	TGATCGGCTGGTCAAGATGCATAGG	853	25	70	2000	1147

ANNEX 8.5. RT-PCR

MATERIAL

- Pipettes and tips.
- Plates and sterile PCR tubes (0.2 ml).
- Sterile tubes (1.5 ml).

REACTIVES

- Sterile distilled water.
- 2X Reaction Mix (a buffer containing 0.4 mM of each dNTP, 3.2 mM $MgSO_4$) (Invitrogen)
- SuperScript III RT / Platinum Taq Mix Reverse transcriptase (Invitrogen).
- Primers:
  691ATG: ATGGCATTGCAATTCCGCTCTTTGCTCTTGTG
  691UGA: TCAAATTTGATACACAAATTTGGTAGCCCTA

INSTRUMENTATION

- Thermocycler Eppendorf Mastercycler.

METHOD

A RT-PCR was done to obtain cDNA by SuperScript III Taq DNA Polymerase

1. 1µg of total RNA was used with the primers 691ATG and 691UGA. The elements and conditions of amplification were:
   - 2x reaction mix: 15µl
   - Template RNA (0.01-1µg): 4µl
   - Forward: 0.6 µl
   - Reverse: 0.6 µl
   - SuperScript III RT / Platinum *Taq* Mix: 1.2 µl
   - Water: 8.6 µl
   - Total: 30 µl

   - 30 min at 55 ºC
   - 2 min at 94 ºC
   - 15s at 94ºC
   - 30s at 59 ºC

   30 cycles

   - 2:30 min at 68 ºC
   - 5 min at 68 ºC
   - 12 ºC

2. PCR products were run in an agarose gel 1% in TAE 1X with 1 Kb Plus Ladder (Invitrogen) and then it was observed in GeneSnap Program (Syngene) with a size band expected of 1,650 bp.
3. The bands were cut and purified with QIAquick Gel Extraction Kit Protocol (QIAGEN).

ANNEX 8.6. CLONATION IN *Escherichia coli*

MATERIAL

- Pipettes and tips.
- Plates and sterile PCR tubes (0.2 ml).
- Sterile tubes (1.5 ml).
- Petri dishes.

REACTIVES

- Ice.
- 2X rapid ligation buffer (Promega).
- pGEM®-T Easy Vector (Promega).
- T4 DNA ligase (Promega).
- *E. coli* Top10 (Invitrogen).
- SOC medium.
- LB Broth.
- Agar bacteriological (Scharlau).
- Kanamycin (50 µg/mL) (Sigma-Aldrich).
- 5x buffer (Promega).
- Sterile water.
- $MgCl_2$ 25Mm (Biotools).
- dNTP 10mM (Biotools).
- Polimerase *Go Taq* (Promega).
- 1 Kb plus ladder (Invitrogen).
- TAE 1X
- Agarose D-1 LOW (Conda).
- Gel red (Invitrogen)

INSTRUMENTATION

- Fridge (4 ºC).
- Thermoblock.
- Incubator (37 ºC).
- Thermocycler.
- Precision scale.
- Microwave
- Horizontal Electrphoresis System (Sub-Cell GT, Biorad).
- Power supply PowerPac 3000 V (Biorad).
- GeneTools Image Analyzer (Syngene) with ultraviolet light.
- Fume hood

Primers for cloning PHs in almond	Tm
27F: AATTATGAAGGATATGGGGTTGG	57
27R: CACCTCTGTTACCTCCAAGCA	57
Primers for cloning CYPs in almond	
21326F: CACCATGAAGCGTTATTGGCACC	66
21326R: TGTCTGGTACACGTGAGCTGGC	62
4152Fcacc: CACCATGGCTCTTCTAACACTT	56.4
4152F: ATGGCTCTTCTAACACTTTTCAA	55
4152R: AGGGGAGTATGGTGTTGGAAC	57
17339F: CACCATGGCGGTACTTTCACTT	61
17339R: AGGGGAGTATGGTGTTGGAATAAG	60

METHOD

1. The inserts of cDNA from the RT-PCR were cloned into pGEM®-T easy Vector (Promega).
2. The ligation was done with the next elements at 4 ºC temperature overnight:
    - 2X buffer: 5µL
    - pGEM®-T easy: 0.5 µL
    - RT-PCR product: 3.5 µL
    - T4 ligase: 1µL
3. After ligation at 4 ºC, chemical transformation took place by adding 2 µL of ligation to *E. coli* TOP10 competent cells, in a total volume of 50 µL.
4. Mix was put 20 min in ice, heat shock for 45 s at 42 ºC, 2 min in ice, and then 950 µL SOC medium was added, incubated for 1,5 h at 37 ºC with shaking.
5. Finally 100 µL of each transformation culture was spread on LB agar+ kanamycin plate and left overnight at 37 ºC.
6. To identify positives colonies of *E. coli* containing the insert, PCR Go Taq (Promega) was made, with the nested primers (27F and 27R). The elements and conditions of amplification were:
    - 5x buffer: 4 µL
    - Water: 13.9 µL
    - MgCl2 (25mM): 0.8 µL
    - dNTP (10mM): 0.4 µL
    - Forward: 0.4 µL

- Reverse: 0.4 µL
- Polymerase *Go Taq (Promega)*: 0.1 µL
- Colony taken with a tip
- Total: 20 µl

- 2 min at 95 ºC,
- 30s at 95 ºC,
- 30 s at 57 ºC, } 30 cycles
- 1 min at 72 ºC,
- 5 min at 72 ºC
- 12 ºC.

7. An electrophoresis gel was made in 1% agarose gel with 1kb Plus Ladder (Invitrogen), the size of the band expected was 1,050 bp.

ANNEX 8.7. MINIPREPS DNA PLASMID ISOLATION

MATERIAL

- Sterile tubes (1.5 ml).
- Pipettes and tips.

REACTIVES

- LB Broth.
- Ice.
- Ethanol 100%.
- Sterile water
- Tris-HCl 1M (Sigma-Aldrich).
- EDTA 0.5 M (Sigma-Aldrich).
- NaOH 5M (Panraec).
- AcK 3 M (Scharlau).
- Glycerol 100%.
- ECORI (BioLabs).

INSTRUMENTATION

- Incubator (37 ºC).
- Centrifuge SIGMA 1-14 Microfuge.
- Freezer -80 ºC.

METHOD

It is a method which isolates the DNA plus the vector from the bacteria.

Steps:

1. Pick positive *E. coli* colonies up with a sterile tip and put in 4 ml LB + kanamycin, conditions 37 ºC shaking 180-220 rpm with O_2 overnight.
2. Put 1.5 ml of the LB culture + antibiotic in a 2ml Eppendorf (twice, 3 ml).
3. Centrifuge 8,000 rpm 3 min 4 ºC.
4. Remove supernatant and repeat the step before. Centrifuge 8,000 pm 3 min 4 ºC.
5. Remove supernatant.
6. Add 250 µL of solution A. Vortex 10 seconds to resuspend the cell debris.
7. Add 250 µL of solution B. Incubate room temperature 2-3 min.
8. Add 250 µL of solution C. Incubate 5 min on ice.
9. Centrifuge 14,000 rpm 15 min 4 ºC.
10. Take the supernatant and put in another 1.5 ml eppendorf tube on ice and add 1ml ethanol 100%. Centrifuge 13,000 rpm for 15 min 4 ºC.
11. Remove the supernatant and the rests of ethanol.

12. Elution in 60-70 µL of sterile water and resuspend the pellet.
13. Keep at -20 ºC.

Solution A:

- -Tris-HCl 1M 2,5 ml
- EDTA 0.5 M 2 ml
- Distilled water up to 250 ml
- Adjust pH 8 and autoclave

Solution B:

- NaOH 5 M 8 ml
- SDS 10%
- Distilled water up to 200 ml

Solution C:

- AcK 3 M
- Distilled water up to 250 ml
- Adjust pH 6 and autoclave

14. To be sure that the cDNA really was inserted in the vector, the minipreps were digested with ECO RI enzyme restriction in a total volume of 15 µL. 4 µL of minipreps was added, incubated at least 3 hours and check with an electrophoresis in 1% agarose gel.
15. Glycerols stocks were done for positive colonies, for this 800 µL *E. coli* LB culture were put plus 200 µL of glycerol 80% and stored at -80 ºC.

ANNEX 8.8. GATEWAY CLONIING

MATERIAL

- Sterile tubes (1.5 ml).
- Plates and sterile PCR tubes (0.2 ml).
- Pipettes and tips.
- Petri dishes

REACTIVES

- Ice.
- Sterile Water.
- 10x buffer (Thermo Fisher).
- 10 mM dNTP (Biotools).
- HotMaster Taq (Thermo Fisher).
- BP Clonase II Enzyme Mix (Invitrogen).
- pDONR207 entry clone (Invitrogen).
- *E. coli* DH5α (Invitrogen).
- SOC medium
- LB Broth.
- Agar bacteriological (Scharlau).
- Minipreps: genelute HP Plasmid Miniprep (Sigma-Aldrich).
- Restriction enzyme *Bgl* II (BioLabs).
- Glycerol 100%.
- LR Clonase II Enzyme Mix (Invitrogen)
- Vector pJAM1502 (Luo et al., 2007).
- Kanamycin (Sigma-Aldrich).
- Rifampicin (Sigma-Aldrich).
- Gentamicin (Sigma-Aldrich).
- *Agrobacterium tumefaciens AGL1.*
- Primers:

  Att B1:GGGGACAAGTTTGTACAAAAAAGCAGGCT
  Att B2:GGGGACCACTTTGTACAAGAAAGCTGGGT
  691B1F:GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGCATTGCAATTCC
  GCTCTTTGCTCTTGTG
  691B2R:TCAAATTTGATACACAAATTTGGTAGCCCTATCAAATTTGATACAC
  AAATTTGGTAGCCCTA

INSTRUMENTATION

- Thermoblock.
- Incubator (37 ºC).
- Thermocycler.
- Fume hood.
- Water bath.
- MicroPulse Electroporator (Bio-Rad).

METHOD

cDNA sequences of genes were amplified by PCR Hot Master using the primers 691B1F and 692B2R. All the PCR primers contained the attB1 and attB2 Gateway cloning sites, which is a technique that provides a rapid and efficient way to move DNA sequences into multiplex vector system.

PCR att B1 */ att* B2 *sites*
The cDNA sequences of genes were amplified by PCR with HotMaster with the next elements and conditions:

- Water: 20.1 µl
- 10x buffer: 3 µl
- 10 mM dNTP: 0.6 µl
- Forward (691B1F): 1.5 µl
- Reverse (691B2R): 1.5 µl
- DNA: 3 µl
- HotMaster Taq: 0.3 µl
- Total: 30 µl

- 2 min at 94 ºC
- 20 s at 94 ºC
- 35 cycles of 20 s at 57 ºC
- 2 min at 65 ºC
- 5 min at 65 ºC
- 12 ºC

Primers used were 691B1F and 692B2R. All the PCR primers contained the *att* B1 and *att* B2 Gateway cloning sites. So the flanking sites *att* were added to the insert.

Gateway BP Clonase II Enzyme Mix
In order to introduce cDNA in *Agrobacterium tumefaciens*, PCR products were cloned by gateway recombination. Two clonation steps were carried out. Firstly, with the gateway BP Clonase II Enzyme Mix (Invitrogen), PCR products were introduced in pDONR207 entry clone. Following a normal chemical transformation in *E. coli* DH5α (Invitrogen).

The BP reaction consisted in a ligation with the pDONR vector to obtain the entry clone with the flanking sites attL.

BP reaction:
- *att* P1-*ccd* B-*att* P2: entry clone
- pDONR 207 vector (Invitrogen)
- *E. coli* DH5α (Invitrogen)

- Antibiotic gentamicin

Ligation:

- *att* B-PCR product (150ng)
- 1 µL donor vector (150ng)
- To 8 µL water
- +2 µL BP clonase
- Overnight room temperature

1. Add 1 µL proteinase K+10 min 37 ºC to stop the reaction.
2. Transform 5 µL of the ligation+50 µL of *E. coli.*
3. Incubate 30 min on ice.
4. Heat-shock cells by incubating at 42 ºC for 45 sec in a bath.
5. Incubate 2 min on ice.
6. Add 900 µL SOC medium.
7. Incubate 1:30 hours 37 ºC shaking 225 rpm.
8. Spread on plates LB agar + gentamicin.
9. Make PCR colonies to check the postives.
10. Pick colonies in LB + gentamicin overnight 37 ºC 225 x rpm.
11. Use Plasmid minipreps kit (Sigma) and measure in Nanodrop.
12. Use restriction enzymes *ECO RV* and *Xho* I.
13. Make Glycerols and keep them at -80 ºC.

Gateway LR Clonase II Enzyme Mix

Then, with the gateway LR Clonase II Enzyme Mix (Invitrogen), entry clones were cloned into destination vector pJAM1502 (Luo et al., 2007), that is ready to enter in *Agrobacterium tumefaciens.* The bacteria used in transformation were *E. coli* DH5α (Invitrogen).

LR reaction, which consists in a new ligation with the destination vector pJAM to obtain the expression clone with the flanking sites *aattB*, this clone is ready to enter in *Agrobacterium tumefaciens.*

LR reaction:

- *att* R1-*ccd* B-*att* R2
- destination vector
- pJAM1502 (vector) (Luo et al., 2007)
- *E. coli* DH5 α
- Antibiotic kanamycin

1. Ligation:
    - 1-7 µL entry clone (150ng)
    - 1 µL destination vector (150ng)
    - To 8 µL water
    - +2 µL LR clonase
    - Overnight room temperature
    - Add 1 µL proteinase K+10 min 37C

2. Transform 5 µL of the ligation+50 µL of *E. coli.*

3. Incubate 30 min on ice.

4. Heat-shock cells by incubating at 42 ºC for 45 sec in a bath.

5. Incubate 2 min on ice.

6. Add 900 µL SOC medium.

7. Incubate 1:30 hours 37 ºC shaking 225 rpm.

8. Spread on plates LB agar+ kanamycin.

9. Make PCR colonies to check the postives.

10. Pick colonies in LB + kanamycin overnight 37 ºC 225 rpm.

11. Use Minipreps: genelute HP plasmid minipreps kit (sigma) and measure in Nanodrop.

12. Use restriction enzyme *Bgl* II.

13. Make Glycerols and keep them at -80 ºC.

Agrotransformation

The destination vector (PCR product + *att* B1 and *att* B2 Gateway cloning sites + pJAM1502) was introduced into *Agrobacterium tumefaciens* (AGL1) by electroporation with the next conditions: 2,2 V, 25 µF, 400Ω in a MicroPulser Electroporator (Bio-Rad)**.** Finally bacteria were plate in LB agar + rifampicin + kanamycin.

ANNEX 8.9. AGROINFILTRATION

MATERIAL

- *Nicotiana benthamiana* plants
- Syringe 1 ml (Sigma-Aldrich)
- Pipettes and tips

REACTIVES

- *Agrobacterium tumefaciens* (AGL1).
- Lb Broth.
- Kanamycin (Sigma-Aldrich).
- Rifampicin (Sigma-Aldrich).
- Sterile water.
- MES Buffer 10 mM (Sigma-Aldrich).
- MgCl2 10mM (Biotools).
- Acetosyringone 100µM (Sigma-Aldrich).

INSTRUMENTATION

- Greenhouse.
- Centrifuge.

METHOD

1. Plants of *Nicotiana benthamina* (autor) were prepared in a greenhouse, 4-6 weeks old plants were used in the agroinfiltration.
2. Overnight cultures of *A. tumefaciens* (AGL1) containing expression constructs of cDNAs in pJAM1502 and the gene-silencing inhibitor protein p19 (Voinnet et al., 2003) were grown in LB, containing suitable antibiotics kanamycin and rifampicin.
3. The next day agro cells were harvested by centrifugation at 3,000 rpm for 10 min and resuspended to an OD600 of 2.0 in 1 mL of acetosyringone solution:
    - 50 mL water
    - 0.5 mL MES 10 mM
    - 0.5 mL MgCl2 10mM
    - 5 µL acetosyringone 100µM
4. After 2 hours of incubation at room temperature, a combined solution was made in a total volume of 2 mL:
    - 0.5 mL of p19 OD600 2.0
    - 0.5 mL of Agro solution OD600 2.0
    - 1 mL of Acetosyringone solution.
5. *N. benthamiana* leaves were agroinfiltrated using 1 mL syringe. After 4–5 days, leaf discs (1 cm diameter) were cut from infiltrated leaves. Four agroinfiltrations were carried out.

ANNEX 8.10. FEIGL-ANGER (CYANIDE RELEASE QUALITATIVE METHOD)

MATERIAL

- *N. benthamiana* leaves.
- Pipettes and tips.

REACTIVES

- copper ethylacetoacetate (Alfa Aesar).
- 4.4´-tetramethyldiaminodiphenylmethane (Sigma-Aldrich).
- Whatman 3MM paper.
- MES buffer (Sigma-Aldrich).
- Cyanogenic monoglucosides: prunasin, dhurrin, linamarin, lotaustralin.
- Cyanogenic diglucosides: amygdalin and linustatin.

INSTRUMENTATION

- 96-well microtiter plate (Thermo Scientific).

METHOD

According to the methods described by Feigl / Anger (1996):

Feigl-Anger paper was prepared by separately dissolving 5 g of copper ethylacetoacetate and 4.4´-tetramethyldiaminodiphenylmethane in 0.5 liters each, and combining both solutions.

1. Whatman 3MM paper cut to 8 x 12 cm size was wetted with the solution and dried. The paper was stored until use.
2. It was used for 96-well microtiter plate, each one was added 200µL of 20 mM MES buffer pH 6.5 plus 15 µL substrate 4 mM.
3. The substrates used were the cyanogenic monoglucosides: prunasin, dhurrin, linamarin, lotaustralin and the cyanogenic diglucosides amygdalin and linustatin.
4. A leaf disk of *N. benthamiana* plant agroinfiltrated was added and grinded in the well, three discs per construction were made. Feigl-Anger paper was put over the plate and the lid.
5. Cyanogenesis was detected after two hours by exposure, seeing a blue spot when there was a positive reaction. When the cyanide is released, oxidation of a tetrabase in presence of a cupper salt takes place.

ANNEX 8.11. FAST BLUE BB SALT ASSAY (β-GLUCOSIDASE ACTIVITY DETECTION)

MATERIAL

- *N. benthamiana.*
- Pipettes and tips.
- Sterile tubes (1.5 ml).
- Mortar.
- Foil paper.

REACTIVES

- Liquid nitrogen.
- MES Buffer (Sigma-Aldrich).
- SDS PAGE gel 12% (Bio-Rad).
- Bromophenol Blue (Sigma-Aldrich).
- Glycerol.
- SDS distainer.
- Coommasie Brilliant Blue R250 (Sigma-Aldrich).
- Acetic Acid (Sigma-Aldrich).
- Ethanol.
- Sterile Water.
- Sodium citrate 50 mM (Sigma-Aldrich).
- Phosphate 100 Mm (Sigma-Aldrich).
- Fast Blue BB salt (Sigma-Aldrich).
- 6 bromo-2-naphthyl β-D-glucopyranoside (Sigma-Aldrich).
- Dimethylformamide (Sigma-Aldrich).

INSTRUMENTATION

- Centrifuge
- Nanodrop (Thermofisher).
- Plastic box.
- Scale precision.
- Criterion Vertical Electrophoresis Cell (Bio-Rad).
- PowerPac 3,000 V (Biorad)
- Fume hood.

METHOD

1. All in cold. Firstly, agroinfiltrated *N. benthamiana* leaves samples were grinded in liquid nitrogen.
2. 20 mM MES buffer pH 6 was added in a proportion 1:3.
3. Protein extracts were centrifuged at 4 ºC 30 min 20,000 x g.
4. The supernatant was collected and quantified based on A280 (Nanodrop; Thermo Scientific)

5. Protein extracts were analyzed by SDS–PAGE (12% gel) following application of protein (20–80 µg) combined with 10% bromophenol blue (2 µl), 100% glycerol (10 µl) and 0.5% SDS (10 µl) in a final volume of 40 µl.
6. At the end of electrophoresis (2 h, 175 V, 4 ºC) gels were washed (2 x 10 min) in Fast Blue BB buffer (50 mM sodium citrate, 100 mM phosphate (pH 5.8).
7. Solution A (15 mg Fast Blue BB salt + 20 ml Fast BB Buffer) and solution B (20 mg 6 bromo-2-naphthyl β-D-glucopyranoside + 200 µL DMF (dimethylformamide)) were mixed and put with the gel by incubation 2 hours at 37 ºC shaking in foil paper.

ANNEX 8.12. CYANIDE RELEASE (QUANTITATIVE METHOD)

MATERIAL

- Sterile tubes (1.5 ml).
- Pipettes and tips.
- Microtiter plate (Sigma-Aldrich)

REACTIVES

- MES Buffer (Sigma-Aldrich).
- Liquid nitrogen.
- NaOH (Sigma-Aldrich).
- Glacial acetic acid (Sigma-Aldrich).
- N-chlorosuccinimide (Sigma-Aldrich).
- Succinimide (Sigma-Aldrich).
- Pyridine (Sigma-Aldrich).
- Barbituric acid (Fluka).
- KCN (Sigma-Aldrich).

INSTRUMENTATION

- Incubator
- Scanner

METHOD

According to Lambert 1975:

1. The reaction mix (total volume: 100 µL) contained 10 µL of substrate 1 mM, 10 µL of protein and 80 µL of MES 20 mM buffer pH 6.
2. Samples without substrates prunasin, amygdalin and dhurrin were used as control.
3. Following incubation (10 min, 30 ºC, 300 rpm shaking) and freeze closed Eppendorf tubes in liquid nitrogen to avoid loss of volatile HCN formed. While samples thawed at room temperature, 40 mL of 6 M NaOH was added.
4. A 60 mL aliquot from each sample was transferred to a 96-well microtiter plate.
5. Sequentially the following reagents were added to each well: 12.5 µL of 100% glacial acetic acid, 50 µL of reagent A (dissolve 50mg of N-chlorosuccinimide in 50 mL of water and add and dissolve 125 mg of succinimide) and 50 µL of reagent B (mix 15 mL of pyridine with water to make 50 mL, while starring, add 3 g of barbituric acid from Fluka).
6. After 5 min incubation at room temperature, wells were scanned between 450 and 700 nm with a peak reading made at 584 nm.
7. Released cyanide was calculated against standards of KCN made in 1 M NaOH.

ANNEX 8.13. UMBELLYFERIL SUBSTRATE (β-GLUCOSIDASE ACTIVITY DETECTION)

MATERIAL

- Pipettes and tips.
- Sterile tubes (1.5 ml).

REACTIVES

- MES Buffer (Sigma-Aldrich).
- 4-methyl-umbelliferyl-β-D-glucoside (Sigma-Aldrich).

INSTRUMENTATION

- GeneTools image analyzer (Syngene) with ultraviolet light.

METHOD

1. Leaf discs were put in a tube with 150 µL MES buffer 20 mM pH 6 and 6 µL of general substrate 4-methyl-umbelliferyl-β-D-glucoside 25 mM, another experiment was made without the leaf disc.
2. All tubes were put 15 min at 37 ºC to evaporate ethanol. Both proves were observed under UV.
3. β-glucosidase activity was detected by fluorescence.

ANNEX 8.14. TnT

MATERIAL

- Pipettes and tips.
- Sterile tubes (1.5 ml).

REACTIVES

- Substrates: TCP, prunasin, amygdalin, nothing.
- TnT T7 PCR Quick Master Mix (Promega).
- Methionine 1 mM.
- (35S) Methionine.

INSTRUMENTATION

- Thermoblock.

METHOD

1. Mix in an eppendorf tube the next components:

- TnT T7 PCR Quick Master Mix: 40µl
- Methionine 1 mM: 3 µl
- (35S) Methionine: (1,000 Ci/nmol at 10 mCi/ml): 0.25 µl
- PCR template: 5-7 µl
- Total: 50 µl

2. Incubate 30 ºC 90 min.

3. Take 10 µl for the TLC.

Primers for the TNT experiment to express in vitro putative UGT2 from almond:

NAME of gene	Sequence
TNTS811F	GGATCCTAATACGACTCACTATAGGGAACAGCCACCATGGTTTACTCTGAGCAC
TNTS811R	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCACTTAAAAGTAGATATTTGTTTCC
TNTS61_14F	GGATCCTAATACGACTCACTATAGGGAACAGCCACCATGGATTCTTCTCAGCA
TNTS61_14R	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCATTCCTTTCTCATACAAAGTTG
TNTS41F	GGATCCTAATACGACTCACTATAGGGAACAGCCACCCACCATGGATTCCAGTGATC
TNTS41R	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCTAGACCTTCTTGGAAGCATTTC
TNTS619F	GGATCCTAATACGACTCACTATAGGGAACAGCCACCCACCATGGTTTTCTCTGACCAAAGAAG
TNTS619R	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAGATTCCCCTTCCCGTACA
TNT L41_ppa5106F	GGATCCTAATACGACTCACTATAGGGAACAGCCACCATGGAGAAGGAGAAGAAG
TNT L41_R	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCTAAATTGTCAACCTTGGAGGG

ANNEX 8.15. TLC

MATERIAL

- Tubes 10 ml.
- Pipettes and tips.
- Sterile tubes (1.5 ml).

REACTIVES

- UDP-glucose C14 (50 nM).
- Prunasin.
- Amygdalin.
- Tricine 20 mM (Sigma-Aldrich).
- Methanol.

INSTRUMENTATION

- TLC silica.
- Glass box.
- Phospho Image Cassette.
- Fume hood.

METHOD

1. First of all these components were mixed in a tube:

   - TnT protein total: 10 µl
   - UDP glucose C14 (50 nM): 2µl
   - Substrate: TCP/Prunasin /Amygdalin/Mandelonitrile: 2 µl 25 mM
   - $MgCl_2$ (250 mM): 1 µl
   - $CaCl_2$ (50mM): 1µl
   - TAPS-HCl buffer (add 1mM DTT to the total concentration of 0.1 mM): 34 µl.
   -Total amount: 50 µl

2. The mix was incubated 1 hour 30 ºC.
3. Stop the assay by adding 3 volumes (150 µl) of ice-cold 100% MeOH and vortex.
4. Heat 45 ºC for 15 min and put on ice for 10 min.
5. Centrifuge 12,000 x g for 10 min.
6. Filter the supernatant (96 well filter polyvinylidene difluoride; 0.45 mm).
7. Centrifuge for 2 min (3,000 x g, room temperature).
8. Transfer to eppendorf tubes
9. Concentrate in Scan Vac 40 ºC 2 hours.
10. Resuspend the pellet in 20 µl 100% MeOH.

11. Load on TLC silica gel. Radiolabed products were separated by thin layer chromatography (TLC) (silica gel 60 F254 plates; Merck).
12. The samples were loaded in a line 1.5 cm separated of the bottom of the TLC. The volume loaded was 10+10 µl.
13. TLC silica was put in a glass box with 1 cm of methanol 100%. As a reference, 10 µl of 40 mg/ml of prunasin and amygdalin were added next to the samples.
14. Remove the methanol and add 200 ml of the mix buffer: ethyactate, acetic acid and methanol.
15. Run the TLC for 2-3 hours until samples reached the top and leave it dry.
16. Put in the phospo image cassette and develop it in one or 2 days.
17. Products were visualized using a STORM 840 PhosphorImager (Molecular Dynamics, http://www.moleculardynamics.com).

ANNEX 8.16. METHANOL EXTRACTION AND LC-MS

MATERIALS

- Sterile tubes
- Elisa plate.
- Morter.
- HPLC vials.

REACTIVES

- Methanol 85%.
- Linamarin.
- Liquid nitrogen.
- Prunasin.
- Amygdalin.

INSTRUMENTATION

- Bath.
- Centrifuge.
- Agilent 1100 Series LC (Agilent Technologies).

METHOD

1. To extract cyanogenic glucosides and their derivates, once the samples were collected and keep at -80 ºC, they were grinded with a morter and pestle and liquid nitrogen and weighed frozen.
2. Then between 50 and 100 mg of the samples were added to 400 µL methanol 85% in a threaded tube of 1,5 ml.
3. The samples with methanol were boiled 5 min in a bath and they were put in ice.
4. Right after, they were centrifuged 5 min 2,000 x g and the supernatant was collected and taken out to a HPLC tube and keep at -20 ºC.
5. 20 µL of this supernatant was filtered with 70 µL of water and 10 µL Linamarin (internal standard) 500 µM (final concentration 50 µM) in an ELISA filter (5x dilution). The order was lid + ELISA plate + ELISA filter.
6. The mix from the filtering was centrifuged 5 min 3000 x rpm and 60 µL were transferred to a HPLC vial in a HPLC tube and sent to analyse by LC-MS.
7. LC–MS/MS was carried out using an Agilent 1100 Series LC (Agilent Technologies) coupled to a Bruker HCT-tra ion trap mass spectrometer (Bruker Daltonics). A Zorbax SB-C18 column (Agilent; 1.8 μm, 2.1×50 mm), maintained at 35 ºC, was used for separation. The mobile phases were: A, water with 0.1% (v/v) HCOOH and 50 mM NaCl; B, acetonitrile with 0.1%(v/v) HCOOH. The gradient programme was: 0–0.5 min, isocratic 2% B; 0.5–7.5 min, linear gradient 2%–40% B; 7.5–8.5 min, linear gradient 40%–90% B; 8.5–11.5 min isocratic 90% B; 11.6–17 min,

isocratic 2% B. The flow rate was 0.2 ml · min−1 but increased to 0.3 ml · min−1 in the interval 11.2–13.5 min. ESI–MS^2 was run in positive mode.

8. The data was analyzed by the Bruker Daltonics programme Data Analysis 4.0. Extracted ion chromatograms for specific $[M + Na]^+$ adduct ions and their MS^2 profiles were used to identify the compounds. Standard series of prunasin, amygdalin, prunasin acid, prunasin amide and prunasin anitrile spanning a range of concentration from 7.5 µM to 125 µM were used for absolute quantification (Pičmanová et al. 2015).

9. Standards, prunasin and amygdalin, were also sent in concentrations 7.5, 15, 31, 62.5 and 125 µM.

ANNEX 8.17. SNP$_S$ ANALYSIS

MATERIAL

- Pipettes and tips.
- Plates and sterile PCR tubes (0.2 ml).
- Sterile tubes (1.5 ml).

REACTIVES

- Ice.
- Phusion Reaction Buffer (HF) (Phusion).
- $MgCl_2$ (50 mM) (Biotools).
- dNTP (10mM) (Biotools).
- Phusion DNA polymerase (Phusion)
- Bi-distilled sterile water
- Polimerase *Go Taq* (Promega).
- 1 Kb plus ladder (Invitrogen).
- TAE 1X
- Agarose D-1 LOW (Conda).
- Gel red (Invitrogen).
- "QIAquick Gel Extraction kit" (Qiagen).
- "Nucleo Spin Extract II" kit (Macherey Nagel).

INSTRUMENTATION

- Fridge (4 ºC).
- Thermocycler.
- Horizontal Electrphoresis System (Sub-Cell GT, Biorad).
- Power supply PowerPac 3000 V (Biorad).
- GeneTools Image Analyzer (Syngene) with ultraviolet light.

METHOD

1. Primers used for the SNP analysis based on Koepke et al. (2013) were the following:

gene from peach	Forward	Reverse
From the SNPs found in Koepke et al. 2013		
ppa001291m	AAATGCGGAAGGAGGAAAAT	GAGCCTGCTTATTGAATGT
ppa001981m	TTCTCGCTATTAGATCCTCC	ATAGTAGTCCAGTAAAGGGG
ppa003453m	CTCCTTAGCACTGTCATC	GGGAGTCCATTTCAACTG
ppa003514m	GAGGATATGGTGCTTGCTGA	GGCTCAAGTGATGAAGAT
ppa004278m	GCACTTGCTTGGTTATTCGAT	TCCGATCTCTCACAATGTCCA
ppa008772m	CTTGCTTATGATGCGGCA	CTCATAATTACCGCCTCG
ppa006138m	GTGCAGTAAGCAAGAATTGA	GCTTCTTTTAGCCTCAGTAT
ppa022964m	CACTGTCTATGGGAAACTAC	CAGTAGAGTTATGTTGTTCCC
ppa022964m	TGTTTTGTTTTGCAGCATGG	TTTGGAATGGTAATGGAAGG
ppa023181m	AAATCAGTAAGCCCCATCAT	AATAAACTTCTCCCACCATC
ppa024141m	ACAAAAGACGACGATGAGAA	ATGTCAGCCCTTATACTTCA
ppa024207m	TCGTGTTTGTGTCGTCATTT	CGATGTGGGATAAAGTGGTG

2. Once the primers were designed, a screening on R1000 and Desmayo parental and on a few individuals of the F_1 population was carried on by Polymerase Chain Reaction (PCR) technique. In particular, the proofreading Phusion® High-Fidelity DNA Polymerase (New England Biolabs) was used as described in the following program:

- Phusion Reaction Buffer (HF): 5µL
- MgCl2 (50 mM): 0.75 µL
- dNTP (10mM): 0.5 µL
- Forward primer (10 µM): 1.25 µL
- Reverse primer (10 µM): 1.25 µL
- Phusion DNA polymerase: 0.25 µL
- DNA template: 1 µL
- Bi-distilled sterile water: 13.3 µL
- Total: 20 µL

- 30 s at 98 ºC
- 10 s at 98 ºC
- 30 s
- 30 s at 72 ºC

30 cycles

- 7 min at 72 ºC
- 12 ºC

3. Each PCR product was then checked by electrophoresis, on 2-3% Agarose gel in 1X TBE.
4. The PCR products were purified, by using the "Nucleo Spin Extract II" kit (Macherey Nagel, Germany) or "QIAquick Gel Extraction kit" (Qiagen), depending on the size and the amount amplified, and once purified they were sent to sequence to "Macrogen Europe" (Netherlands) or "Eurofins Genomics" (Germany).
5. The electropherograms obtained were analyzed by using the software "CLC Sequence Viewer 7" (http://clcbio.com), aiming to identify Single Nucleotide Polimorphisms (SNPs) between the two parents and the offspring of the F_1 population.

ANNEX 8.18. DEVELOPMENT OF CAP$_S$

MATERIAL

- Pipettes and tips.
- Plates and sterile PCR tubes (0.2 ml).
- Sterile tubes (1.5 ml).

REACTIVES

- Ice.
- Bi-distilled sterile water
- 1 Kb plus ladder (Invitrogen).
- TAE 1X.
- Agarose D-1 LOW (Conda).
- Gel red (Invitrogen)
- Restriction enzymes (New England Biolabs)
- Buffer for the restriction enzymes (New England Biolabs)

INSTRUMENTATION

- Fridge (4 ºC).
- Incubator
- Horizontal Electrphoresis System (Sub-Cell GT, Biorad).
- Power supply PowerPac 3000 V (Biorad).
- GeneTools Image Analyzer (Syngene) with ultraviolet light.

METHOD

1. Primers used for CAPs development were the following:

NAME of gene from peach (GDR code)	Forward	Reverse
Primers for the CAPS development		
ppa003882m	CATAACGTCGCCAAGGAGAT	CATCCTTGCCAAAATCCACT
ppa018792m	ACGTTGTCTCGTTCGTGGTT	AGGTGCTGCAAAGACACTGA
ppa006282m	GTTTCGCTCGATTGGGTCTC	ATCATTTCCCGCCTGAATGC
ppa005388m	GCTTCAAGGCAAGATTGGAG	ATTCCACAATTCGGTGGTTC
ppa001838m	GGTTGTTCTGGGAGATGGAA	ACTTGACCGCAACCAAAATC

2. In some SNPs detected, a restriction enzyme analysis was developed to convert them into CAPS by using the software "CAPS Designer" (genomics.net/tools/caps_designer/caps_input.pl).

3. Digestion reactions with the endonucleases were performed according with the protocol of Neff et al. (1998).

4. The digestion mix used for every restriction enzyme, performed for each sample, was the following:

- Restriction enzyme: 0.1µL
- Buffer: 1 µL
- Water: 5.4 µL
- PCR product: 3.5 µL
- Final volume: 10 µL

5. All samples were incubated in a water bath at specific temperature and time of incubation for every enzyme (*New England Biolabs*):

Enzyme	Reaction temperature (°C)
Alu I	37
Msp I/*Hpa* II	37
Hpy 188I	37
Hpy 188III	37
Hpy CH4V	37
Bsa WI	60
Taq I	65

6. The digested fragments were visualized by Agarose or MetaPhor Agarose gel electrophoresis.

7. All CAPS markers were used to genotype a large RxD segregating F_1 population (550 individuals). Marker segregation data were analysed through the JoinMap 4.1 software, in order to develop a genetic map.

ANNEX 8.19. MICROSATELLITES ASSAY

MATERIAL

- Pipettes and tips.
- Plates and sterile PCR tubes (0.2 ml).
- Sterile tubes (1.5 ml).

REACTIVES

- Ice.
- Reaction Buffer (Thermo Scientific).
- dNTP (10mM) (Biotools).
- "DreamTaq™ DNA Polymerase" (Thermo Scientific™)
- Bi-distilled sterile water
- 1 Kb plus ladder (Invitrogen).
- TBE 1X
- Agarose D-1 LOW (Conda).
- Gel red (Invitrogen)
- "QIAquick Gel Extraction kit" (Qiagen)
- HiDi formamide (Applied Biosystems

INSTRUMENTATION

- Electrophoretic capillary (ABI-3500 Genetic Analyzer, Applied Biosystems
- Fridge (4 ºC).
- Thermocycler.
- Horizontal Electrphoresis System (Sub-Cell GT, Biorad).
- Power supply PowerPac 3000 V (Biorad).
- GeneTools Image Analyzer (Syngene) with ultraviolet light.

METHOD

SSR marker	Forward Primer Sequence	PCR product	Expected fragment size
locus	5'- 3	(bp)	(bp)
UDA-045	F: CCATCCCAGGCCTTAGTACA R: GGAGGATGCTATTGGGGTCT	168	"R": 168/168 "D": 168/152
EPDCU2584	F: TTCAGCTCATCTAGTTTCATCACC R: CACGGTTCGAACAACATCTG	132	"R": 132/130 "D": 132/130
CPDCT028	F: TGAACGTTGCACTCCTTCAC R: ACCACCACCATAACCACCAT	190	"R": 190/190 "D": 186/166
BPPCT037	F: CATGGAAGAGGATCAAGTGC R: CTTGAAGGTAGTGCCAAAGC	148	"R": 131/132 "D": 119/148

1. SSR markers UDA045 and CPDCT028 were tested by "DreamTaq™ DNA Polymerase" (Thermo Scientific™) with the following PCR conditions:

- Reaction Buffer: 2 µL
- dNTP (10mM): 0.4 µL
- Forward primer (10 µM): 1 µL
- Reverse primer (10 µM): 1 µL
- Taq DNA polymerase: 0.08 µL
- DNA template: 1 µL
- Bi-distilled sterile water: 13.9 µL
- Total: 19.38 µL

- 3 min at 94 ºC
- 30 s at 94 ºC
- 1 min at 56 ºC
- 1 min at 72 ºC

40 cycles

- 5 min at 72 ºC
- 12 ºC

2. The PCR products were visualized on 2% Agarose gel (Lonza) in 1X TBE.

3. On the other hand, SSR markers EPDCU2584 and BPPCT037 were analyzed by Capillary Electrophoresis Technique, designing with "Primer3plus: a Forward primer modified by adding a M13* tail (5'TGTAAAACGACGGCCAGT3') of 18 bp to the 5'end (Schuelke, 2000), and a Reverse primer not labeled. Moreover, a universal M13* primer that was labeled with Fam (blue) or Hex (green) fluorescent dyes (Sigma Genosys) was added.

4. Amplification reactions were carried out using a "DreamTaq™ DNA Polymerase" (Thermo Scientific™), using the following conditions:

- Reaction Buffer: 1.25 µL
- dNTP (10mM): 0.25 µL
- Forward M13 primer (10 µM): 0.4 µL
- Reverse primer (10 µM): 2 µL
- Universal primer (10 µM): 1 µL
- Taq DNA polymerase: 0.05 µL
- DNA template: 1 µL
- Bi-distilled sterile water: 5.9 µL
- Total: 11.85 µL

- 3 min at 94 ºC
- 30 s at 94 ºC
- 1 min at 56 ºC
- 1 min at 72 ºC

40 cycles

- 5 min at 72 ºC
- 12 ºC

5. Amplification products (1.2 µl) were added to 15 µl HiDi formamide (Applied Biosystems, Foster City, CA) and 0.3 µl GeneScan^{TM}-500 ROX^{TM} Size Standard and, after a denaturation at 94 °C for 5', they were run on a 36 cm electrophoretic capillary (ABI-3500 Genetic Analyzer, Applied Biosystems, HITACHI, Foster City, CA, USA).

6. The electropherograms transferred to the *Workstation*, were analyzed with the software "GeneMapper v. 5.0".

ANNEX 8.20. Genome resequencing and transcriptome analysis.

Step 1: Almond reference genome

As it can be observed, bitter assembly was superior respect to the sweet one in all the measured parameters (Table 4.10). In addition to the general statistics obtained from the entire genomes, we tried to find the best assembly for the *Sk* locus. The main objective was to evaluate which one of the two draft genomes had the most complete *Sk* region. For assessing this, the provided *Sk* locus molecular markers (see Results 4.4.1) and the *Prunus persica* complete reference genome (v.139) were used. First, the *Sk* locus molecular markers were aligned (BLAT software (v.35)) against both almond genomes to identify the scaffolds carrying the molecular markers. The molecular markers were associated to more than one scaffold both in the sweet and in the bitter assemblies. With the aim to identify all the possible scaffolds associated with the *Sk* locus the following strategy was followed. First, the molecular markers were aligned against the *P. persica* genome in order to identify the *Sk* locus region on *Prunus*. Once the *Sk* region had been targeted, both almond assemblies were aligned against *P. persica* (v.139) using *Satsuma* (v.3.1.0). *Satsuma* is a whole-genome synteny alignment program that takes two genomes, computes alignments, and then keeps only the parts that are orthologous, i.e. following the conserved order and orientation of the features, such as protein coding genes.

Step 2. Genome annotation

Once we have obtained the most complete genome sequence of almond, we annotated the genes and other important genome-encoded features. Genome annotation involved three main steps: 1. Transcript clustering; 2. Structural annotation: Identification of genomics elements and 3. Functional annotation: Attach biological information to genomic elements.

Step 2.1 Transcript clustering: First, *cd-hit-est* (v4.6) tool was used in order to cluster similar transcripts of the transcriptome assembly into clusters that met a similarity threshold (95% of identity). In Table 4.11, we can see some statistics of the transcriptome assembly before and after *cd-hit-est* analysis. The purpose of this step is to remove potential redundancy in the transcriptome assembly.

Step 2.2 Structural Annotation (PASA): The 138585 clustered transcripts together with the *ab initio* identified transcripts were used to annotate almond genome using *PASA* (v.2.0.2). *PASA*, acronym for Program to Assemble Spliced Alignments, is an eukaryotic genome annotation tool that exploits spliced alignments of expressed transcript sequences to automatically model gene structures. *PASA* also identifies and classifies all splicing variants supported by the transcript alignments. 94869 transcripts were annotated by *PASA*, clustered within 66332 genes. Furthermore, all the identified genes falling within *Sk* locus, were manually checked using Integrative Genomics Viewer (IGV v.2.3.60). IGV is a high-performance visualization tool for interactive exploration of genomic datasets. Each annotated gene was manually processed and adjusted if necessary.

Step 2.3 Functional analysis: InterProScan and blast-P: The annotated transcripts were further analyzed using *TransDecoder* (v2.0.1) tool. *TransDecoder* identifies candidate coding regions within transcript sequences. Using the 94869 transcript sequences as an input, *TransDecoder* was able to extract the sequences of 57982 proteins. Of those, 40994 were complete (starting with a methionine and with a final stop codon). The coding sequences obtained from *TransDecoder* were functionally annotated with the InterPro database. *InterProScan* is the software package that allows sequences to be scanned against InterPro's signatures. InterPro provides functional analysis of proteins and classify them into families. This step also let the production of Gene Ontology (GO) and KEGG annotations. In addition to InterPro annotations, using protein sequences of our reconstructed genome, a BlastP analysis was performed against *Arabidopsis thaliana* (TAIR10) and *Prunus mume*,

since this last one is very well annotated and it is more similar to almond than Arabidopsis. However, Arabidopsis has a very accurate functional annotation of the proteins, so in this way, our reconstructed proteins will be annotated by similarity to *Arabidopsis*. Only those hits with a value equal or lower than 0.01 were kept.

Step 3. Differential expression analysis, RNAseq:

Using RNASeq data, a precise quantification of gene expression levels on almond individual samples was carried out. All changes in gene expression levels among sweet and bitter samples were captured, compared and analyzed. RNAseq analysis involved different steps: 1. Data pre-processing: Trimming and mapping; 2. statistical analysis of the data; 3. Differential analysis.

Step 3.1 Data pre-processing: RNA sequencing experiment, as previously mentioned, was performed on 4 samples. A paired-end tag sequencing strategy was chosen, in which short tags are extracted from the ends of long DNA fragments for ultra-high-throughput sequencing. The high quality reads (obtained after trimming) were aligned against the *almond* reference genome (Step 1). For the alignment, the latest version of Spliced Transcripts Alignment to Reference (*STAR)* software (*v2.4.0*) was chosen. *STAR* outperforms other aligners by more than a factor of 50 in mapping speed of RNA-Seq reads, while at the same time improving alignment sensitivity and precision. To check the mapping quality of the alignment, *samstat* software was used.

Step 3.2 Statistical analysis: All the statistical analyses were performed with *R*. First, the clusterization of the samples was evaluated with a PCA analysis on the counts normalized with the Trimmed Mean Normalization (TMM) method. Without biological replicates it is not possible to evaluate the quality of the experiment, however it is interesting to notice that the distance on the PC1 axis is lower for the

samples coming from the same day, suggesting that "time" is the variable explaining most of the difference between the samples.

Step 3.3 Differential expression (DE) analysis: The next step has been the differential analysis to identify the genes that are differentially expressed between bitter and sweet almond. A first approach has been to join the samples from the different days considering them as replicates, however the variability was so high that it was not possible to identify any differentially expressed genes. As a consequence, an analysis without replicates was performed with the package *edgeR*. Without biological replicates the variance (dispersion) of the samples cannot be calculated so a specific pipeline has been carried out to estimate the variance from house-keeping genes. Anyway it must be stressed out that with this approach the dispersion is greatly under-estimated and this can lead to a high number of false positives. The two comparisons that were made were:

- Lt1180314 vs St1180314 (transcriptome data from samples taken in March)
- Lt1160414 vs St3160414 (transcriptome data from samples taken in April)

The DE genes were filtered for FDR <=0.05 and fold-change > |1.5|.

Step 3.4. Network analysis: For the network analysis, the *WGCNA* algorithm has been used that creates clusters of genes that are significantly co-regulated based on the Pearson correlation coefficient.

Step 4. Variant calling

Variant calling pipeline involves the identification of genomic variants present in the studied sample in comparison to the reference genome. This process is divided into further steps: 1) Mapping and data pre-processing; 2) Variant Calling; 3) Variant Filtering; 4) Variant annotation.

The first step includes the alignment of the reads against the reference genome (already done for Step 4) and some pre-processing steps to make the data suitable for variant calling analysis.

Once the data has been pre-processed, variant discovery process can be carried out, i.e. identify the sites where the data displays variation relative to the reference genome, and calculate genotypes for each sample at that site. For the variant calling analysis, a pipeline called *SUPER* (v4.0) was used. Simply Unified Pair-End Read (SUPER) workflow is a dynamic and fast tool to identify sequence variation such as SNPs, DIPs and Structural variations (SVs) developed by *Sequentia Biotech* team. The fourth step includes the attachment of functional information to DNA variants, a crucial step in linking sequence variants with changes in phenotype.

Step 4.1 Mapping and pre-processing: The resulting alignment files obtained after mapping (Step 4. DE analysis) were pre-processed to make them adequate for variant calling analysis. The following steps were performed to the alignment files: a) Keep high confidence reads: using *Samtools* (0.1.19) the reads mapping with a mapping quality less than 30 were removed and b) Remove duplicated reads: mitigate the effects of PCR amplification bias introduced during library construction. The software *Picard-tools* (1.127), specifically *MarkDuplicates* tool, was used for this purpose.

Step 4.2 Variant Calling: Using *SUPER workflow* (v4.0) the sites where the samples display variation relative to the reference genome were identified. The sequence variations include SNPs and INDELs (small insertins/deletions).

Step 4.3 Variant Filtering: A critical step of variant calling analysis is to refine the call set to reduce the amount of false positives. Several filters were applied in order to obtain the most accurate and reliable variants: a) Variant Quality: Those variants with a quality less than 30 were removed from downstream analysis and b)

Genotype Depth: SNP/indels with less than 5 reads of coverage were removed from downstream analysis.

Step 4.4 Variant Annotation: SnpEff tool was chosen as the best software to perform this analysis. *SnpEff* (v4.1b) is a genetic variant annotation and effect prediction toolbox. It annotates and predicts the effects of variants on genes (such as amino acid changes). It classifies the variants as intergenic, intronic, non-synonymous SNP, frameshift deletion, large-scale duplication, etc.