Tesis Doctoral

# Estudios moleculares de la biosíntesis de carotenoides en *Cucurbita pepo*

**Ángeles Obrero Cepedello**

**Universidad de Córdoba**

TITULO: *ESTUDIOS MOLECULARES DE LA BIOSÍNTESIS DE CAROTENOIDES EN CUCURBITA PEPO.*

AUTOR: *ÁNGELES OBRERO CEPEDELLO*

---

© Edita: Servicio de Publicaciones de la Universidad de Córdoba.
Campus de Rabanales
Ctra. Nacional IV, Km. 396 A
14071 Córdoba

www.uco.es/publicaciones
publicaciones@uco.es

---

Instituto de Investigación y Formación Agraria y Pesquera
CONSEJERÍA DE AGRICULTURA, PESCA Y MEDIO AMBIENTE

UNIVERSIDAD D CÓRDOBA

**TESIS DOCTORAL**

DEPARTAMENTO DE GENÉTICA

# Estudios moleculares de la biosíntesis de carotenoides en *Cucurbita pepo*

Ángeles Obrero Cepedello

Tesis dirigida por:

Dra. Belén Román del Castillo
Dra. Clara Isabel González Verdejo

Córdoba, Marzo de 2013

Andalucía se mueve con Europa

**TÍTULO DE LA TESIS:** Estudios moleculares de la biosíntesis de carotenoides en *Cucurbita pepo*

**DOCTORANDO/A:** Ángeles Obrero Cepedello

**INFORME RAZONADO DEL/DE LOS DIRECTOR/ES DE LA TESIS**

(se hará mención a la evolución y desarrollo de la tesis, así como a trabajos y publicaciones derivados de la misma).

La Tesis Doctoral de Dña. Ángeles Obrero Cepedello se ha realizado en el Área de Mejora y Biotecnología de Cultivos del IFAPA-Alameda del Obispo de Córdoba. Durante la realización de esta Tesis Doctoral se han desarrollado y validado técnicas experimentales de gran utilidad para el grupo de investigación. Estas técnicas moleculares y biotecnológicas han permitido la publicación de dos trabajos científicos:

- *Obrero A, Die JV, Roman B, Gomez P, Nadal S, Gonzalez-Verdejo CI. 2011. Selection of reference genes for gene expression studies in zucchini (Cucurbita pepo) using qPCR. Journal of Agricultural and Food Chemistry, 59: 5402-5411*

- *Die JV, Obrero A, Gonzalez-Verdejo CI, Roman B. 2011. Characterization of the 3′:5′ ratio for reliable determination of RNA quality. Analytical Biochemistry, 49:336-338*

Igualmente se han generado otros dos trabajos que actualmente se encuentran en proceso de publicación:

- *Obrero A, Gonzalez CI, Die JV, Gómez P, Del Río-Celestino M, Román B. Carotenogenic gene expression in Cucurbita pepo and lycopene epsilon cyclase involvement in carotenoid accumulation during fruit development (enviado a Journal of Agricultural and Food Chemistry)*

- *Obrero A, Gonzalez-Verdejo CI, Román B, Gómez P, Die JV, Ampomah-Dwamena C. Characterization and expression analysis of three new Phytoene synthase genes from Cucurbita pepo*

Por todo ello, se autoriza la presentación de la tesis doctoral.

Córdoba, 18 de Marzo de 2013

Firma del/de los director/es

Fdo.: Belén Román del Castillo

Fdo.: Clara Isabel González Verdejo

**TÍTULO DE LA TESIS:** Estudios moleculares de la biosíntesis de carotenoides en *Cucurbita pepo*

**DOCTORANDO/A:** Ángeles Obrero Cepedello

## ESCRITO RAZONADO DEL RESPONSABLE DE LA LÍNEA DE INVESTIGACIÓN

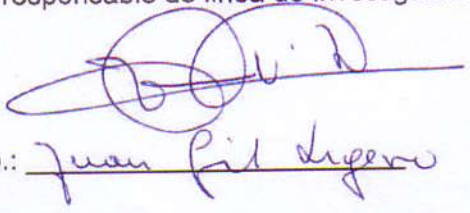
(Ratificando el informe favorable del director. Sólo cuando el director no pertenezca a la Universidad de Córdoba).

**D. Juan Gil Ligero**, Profesor Titular de la Universidad de Córdoba, Informa que el trabajo titulado: **"Estudios moleculares de la biosíntesis de carotenoides en *Cucurbita pepo*"**, realizado por Dña. Ángeles Obrero Cepedello, bajo su tutoría, en el Área de Mejora y Biotecnología del IFAPA "Alameda del Obispo" (Córdoba), se considera finalizado y puede ser presentado para su exposición y defensa como Tesis Doctoral en el Departamento de Genética de la Universidad de Córdoba.

Por todo ello, se autoriza la presentación de la tesis doctoral.

Córdoba, 19 de Marzo de 2013

Firma del responsable de línea de investigación



Fdo.: Juan Gil Ligero

El trabajo que se presenta en esta memoria de Tesis Doctoral ha sido realizado en el área de Mejora y Biotecnológia del centro IFAPA "Alameda del Obispo" de Córdoba. Ha sido financiado por los proyectos: RTA2008-00035-C02-01 "Mejora competitiva de calabacín por características de calidad de fruto y resistencia a enfermedades: desarrollo de RILs, marcadores codominantes SSR, CAPs y SNPs, y construcción de un mapa genético" y RTA2011-00044-C02-01 "Desarrollo competitivo de nuevos cultivares de calabacín y ampliación de la plataforma genómica de esta hortaliza", así como por una beca predoctoral FPI concedida por el Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA).

Parte de los resultados obtenidos en esta tesis han dado lugar a las siguientes publicaciones y comunicaciones a congresos:

**Obrero A**, Die JV, Roman B, Gomez P, Nadal S, Gonzalez-Verdejo CI. 2011 Selection of reference genes for gene expression studies in zucchini (*Cucurbita pepo*) using qPCR. *Journal of Agricultural and Food Chemistry*, 59: 5402-5411 (CAPÍTULO II)

Die JV, **Obrero A**, Gonzalez-Verdejo CI, Roman B. 2011. Characterization of the 3':5' ratio for reliable determination of RNA quality. *Analytical Biochemistry*, 49:336-338. (ANEXO)

**Obrero A**, Román B, Gómez P, Nadal S, Die JV, González-Verdejo CI. Identificación de genes de referencia para normalización en estudio de expresión génica en calabacín. XXXVII Congreso de la Sociedad Española de Genética. SEG 2009. Torremolinos, Málaga

**Obrero A**, Román B, Gómez P, Nadal S, Die JV, González-Verdejo CI. Selection of reference genes for gene expression studies in squah using qPCR. Symposium 2010. Advances in Genomics. Ghent, Belgium.

**Obrero A**, Die JV, Román B, Gómez P, Nadal S, Gónzalez-Verdejo CI. 2011. Implicación del gen PSY en la biosíntesis de carotenoides en cultivares de calabacín. Sociedad Española de Ciencias Hortícolas. Actas nº60. XIII Congreso Internacional de Ciencias Hortícolas. Almería 2012.

González-Verdejo CI**, Obrero A**, Román B, Nadal S, Gómez P. Expresión diferencial de genes dioxigenasa de ruptura de carotenoides en calabacín (*Cucurbita pepo*). Sociedad Española de Ciencias Hortícolas. Actas nº60. XIII Congreso Internacional de Ciencias Hortícolas. Almería 2012.

**AGRADECIMIENTOS**

En primer lugar, deseo expresar mi más sincero agradecimiento a mis directoras, Clara González por tantas y tan buenas ideas que me has ofrecido, por transmitirme esa calma en esos momentos que son tan necesarios. Belén Román, porque salir de tu despacho era ver las luz en muchas ocasiones, por servirme de ejemplo. Gracias a las dos por vuestro constante apoyo, guía, por mantener el humor hasta el último momento y sobre todo gracias por la libertad que me habéis brindado, me he sentido afortunada por teneros como directoras.

A Jose Die, por todo lo que me ha ayudado y más importante por todo lo que me ha enseñado. El día que llegué celebrabas el comienzo del mandato de Obama y ahora estás a unos metros de La Casa Blanca. Eres un ejemplo de esfuerzo y dedicación. Espero en estos años haberme contagiado un poco de ese espíritu.

A Pedro Gómez, no solo por su gran colaboración, sino también por su amistad y cariño. Nunca olvidaré esos viajes a Almería, gracias por abrirnos las "puertas" de tu casa.

A Salvador Nadal, por su ayuda y su constante preocupación hasta el último momento.

A Cristóbal Martínez, por su ayuda y esfuerzo cuando lo hemos necesitado, por ser el impulsor de tantas reuniones y momentos del grupo que siempre quedaran en mi recuerdo.

A Mercedes del Río, por su ayuda y colaboración, por intentar resolver todas mis dudas sobre HPLC.

A Miguel, Antonio, Cristobalina, a los de Almería, Nelly y Manuel y a todos los que han colaborado de alguna manera en este trabajo.

A Lola Madrid, Eva Córdoba y Gema González quiénes me acogieron y ayudaron desde el primer día. Gracias por vuestra amistad y cariño, gracias Lola por tu energía y tu alegría.

A Dolo, Encarni, Ana Pozo, Ana Andreu gracias por los momentos dentro y fuera del trabajo.

A mis palmeñas, a mis choperas, por ser grandes personas, por todos estos años y los que nos quedan.

Por un fututo incierto en tiempos difíciles, a mi tierra querida.

Especial agradecimiento a Charles Ampomah-Dwamena del Instituto Plant and Food Reseach de Nueva Zelanda. Porque a veces hay que seguir las señales, mereció la pena cruzar el mundo para trabajar con un gran profesional y mejor persona.

Por supuesto agradecer a mi familia y a Jesús, los que me han acompañado de cerca y por tanto "han sufrido" las consecuencias…

Quisiera remarcar el aprendizaje personal. En el transcurso de una tesis uno aprende académicamente pero también personalmente. Siempre encontramos obstáculos, pero en general, trabajar en un buen ambiente hace que todo fluya… *gracias de corazón*.

No es de ignorantes cometer errores, lo es no aprender de ellos.

*A mis padres,*
*que tanto me han dado*

## ÍNDICE GENERAL

## ÍNDICE de TABLAS

### Capítulo I



### Capítulo II



### Capítulo III



### Capítulo IV

# ÍNDICE de FIGURAS

## Capítulo I



## Capítulo II

**Capítulo III**



**Capítulo IV**

## ABREVIATURAS

**ACT** Actin

**cDNA** Complementary DNA

**CHYb** β-Ring hydroxylase

**Cq** Quantification cycle

**CAC** Clathrin adaptor complexes medium subunit family protein

**CAT1** Catalase 1

**COMAV** Instituto Universitario de Conservación y Mejora de la Agrodiversidad Valenciana

**CRTISO** Carotene isomerase

**DMAPP** Dimethylallyl diphosphate

**DXR** 1-deoxy-D-xylulose 5-phosphate reductoisomerase

**DXS** 1-deoxyxylulose-5-phosphate synthase

**EF-1A** Elongation factor-1α

**EST** Expressed Sequence Tag

**G6PDH** Glucose-6-phosphate dehydrogenase

**HDR** 1- hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase

**HELI** Helicase

**HPLC** High-performance liquid chromatography

**ICuGI** International Center for Biotechnology Information

**IFAPA** Instituto Andaluz de Investigación y Formación Agraria, Pesquera, Alimentaria y de la Producción Ecológica

**IPP** Isopentenyl diphosphate

**PP2A** Protein phosphatase 2A

**qPCR** Quantitative real-time PCR

**QTL** Quantitative Trait Loci

**LCYb** Lycopene β-cyclase

**LCYe** Lycopene ε-cyclase

**MEP** 2-C-methyl-D-erythritol 4-phosphate

**MIQE** Minimum Information for Publication of Quantitative Real-Time PCR Experiments

**NAD** NAD-dependent malic enzyme

**NADP** NADP-dependent malic enzyme

**NJ** Neighbour-Joining

**PDS** Phytoene desaturase

**PSY** Phytoene synthase

**RPL36aA** 60S Ribosomal protein L36a/L44

**TILLING** Targetin Induced Local Lesions in Genomes

**TUA** α-tubulin ubiquitin

**UBI** Ubiquitin

**UE** Unión Europea

**UFP** Ubiquitin fusion protein

**UPV** Universidad Politécnica de Valencia

**UTR** Untraslated region

**18S rRNA** 18S ribosomal RNA

**RESUMEN**

El calabacín (*Cucurbita pepo*) es un cultivo de gran importancia económica en nuestro país, centrándose su producción principalmente en el sureste andaluz. A pesar de que la importancia económica de este cultivo se ha incrementado en los últimos años, su desarrollo comercial y la obtención de nuevas variedades aún no están al nivel de otras hortícolas. La demanda del consumidor y las empresas es la de adquirir productos innovadores y con mayor calidad, dirigiéndose la tendencia futura hacia el desarrollo de nuevas variedades de calabacín con valor nutricional añadido, como es el incremento del contenido en carotenoides.

La implicación de los carotenoides en el desarrollo de las plantas, así como su importancia en la salud humana y consecuentemente una alta demanda de estos compuestos en el mercado, ha provocado un impulso en la investigación de los genes y enzimas de la ruta biosintética de carotenoides, así como de los mecanismos reguladores empleados por las plantas para determinar qué carotenoides y en qué cantidad se acumulan en ellas. Hasta el momento, existe poca información sobre estas cuestiones en *C. pepo*. En este trabajo, antes de profundizar en el estudio de los genes de la ruta y los mecanismos de regulación de ésta, se seleccionaron un conjunto de genes estables bajo diferentes condiciones para ser utilizados en el proceso de normalización de los estudios de expresión génica posteriores.

Para la realización de los siguientes trabajos, se seleccionaron tres variedades de *C. pepo* que presentaron variaciones extremas en cuanto al contenido en carotenoides. En general, las diferencias encontradas entre los distintos órganos, así como mesocarpo y exocarpo del fruto, se

correspondieron con una expresión diferencial de la mayoría de los genes entre estos tejidos. Las diferencias entre el contenido de carotenoides de los frutos de los tres cultivares, se reflejaron solo en los niveles de expresión de algunos genes carotenogénicos, siendo los patrones de expresión de los genes *LCYe* y *CpPSYA* los que más destacaron. Por otra parte, la ausencia de pigmento en el cultivar blanco es difícil de explicar observando únicamente la expresión de los genes de la ruta. Esto sugiere que, podrían existir además otros mecanismos moleculares implicados en la acumulación de carotenoides en *C. pepo* como el control a través de los genes que participan en la degradación de estos compuestos. Así, los resultados obtenidos en esta tesis abren puertas importantes para el conocimiento básico de esta ruta y la regulación de este carácter en *C. pepo*.

**SUMMARY**

Summer squash (*Cucurbita pepo*) is an economically important crop in our country, being the southeastern of Andalusia the largest producer. Although the economic importance of the crop has increased during recent years, their commercial development and breeding of new varieties are not at the level of other horticultural crops yet. Consumer and companies demand innovative products with higher quality, pointing at the future trend towards the development of new varieties of zucchini with added nutritional value, such as the increase in carotenoid content.

The involvement of carotenoids in plant development as well as their relevance in human health and consequently, a high demand in the market of these compounds, has caused an impulse in the study of the genes and enzymes of the carotenoids biosynthetic pathway and the understanding of regulatory mechanisms used by plants to determine which and how much carotenoids are accumulated in them. So far, little information is known about these issues in *C. pepo*. Before delving into the study of the metabolic pathway genes and their regulatory mechanisms, we selected a set of stable expressed genes under different conditions to be used in further process of normalization in transcriptional profiles studies.

In order to develop the following analyses, we selected three varieties of *C. pepo* that showed extreme differences in the carotenoid content. Overall, the differences among the organs as well as mesocarp and exocarp of the fruit, were consistent with the differential expression of most genes in these tissues. The differences among the carotenoid content in the fruits of the three cultivars, only were reflected in the expression levels of some carotenogenic genes, being the expression patterns of *LCYe*

and *CpPSYA* genes the most highlighted. On the other hand, the lack of pigment in the white cultivar was difficult to explain based only on the expression of the pathway genes. This suggests that in addition to expressed gene involved in the biosyhtesis pahtway, other molecular mechanisms could be acting in carotenoid accumulation in *C. pepo* such as the control through genes involved in the degradation of these compounds. Thus, the results obtained in this thesis open important doors for the basic understanding of this pathway and the regulation of this character in *C. pepo*.

# CAPÍTULO I

## 1. Importancia económica del cultivo de calabacín

El calabacín es un cultivo de enorme importancia económica del que España es uno de los principales productores y exportadores de Europa, superando incluso a Italia, un país con gran tradición de cultivo y consumo de calabacín. En las últimas décadas ha aumentado de manera significativa tanto la superficie de cultivo en España como su producción y exportaciones a otros países, incrementándose a su vez el valor económico de este cultivo. En la Figura 1 se presenta la evolución de este cultivo en los últimos años según el Anuario de estadística del Ministerio de Medio Ambiente, Medio Rural y Marino. Para entender esta tendencia en nuestro país hay que centrarse en el suroeste español donde recae el peso de la producción.

Dentro del territorio nacional, Andalucía es responsable del 82% de la producción, encabezada por la provincia almeriense a la que corresponde el 88% de la producción andaluza (Figura 2). La clave de esta productividad en Andalucía se basa en el sistema de cultivo bajo plástico que permite su producción durante todo el año.

Los datos que recoge la Junta de Andalucía y el Anuario de la Agricultura Almeriense 2012 en su balance de campaña, indican que el tomate y el calabacín fueron los productos que más aumentaron su valor, con un espectacular incremento de 61.8% en el caso del calabacín que recaudó un total de 190 millones de euros. A la hora de valorar las superficies cultivadas, el calabacín destaca con un 10% más de cultivos, ocupando ya una superficie de 5.789 hectáreas en la provincia. Con estos datos de superficie se observa una tendencia al alza en el volumen de la

producción de los cultivos almerienses, destacando el calabacín después de la berenjena, con un 18.4% más y 354.146 toneladas.

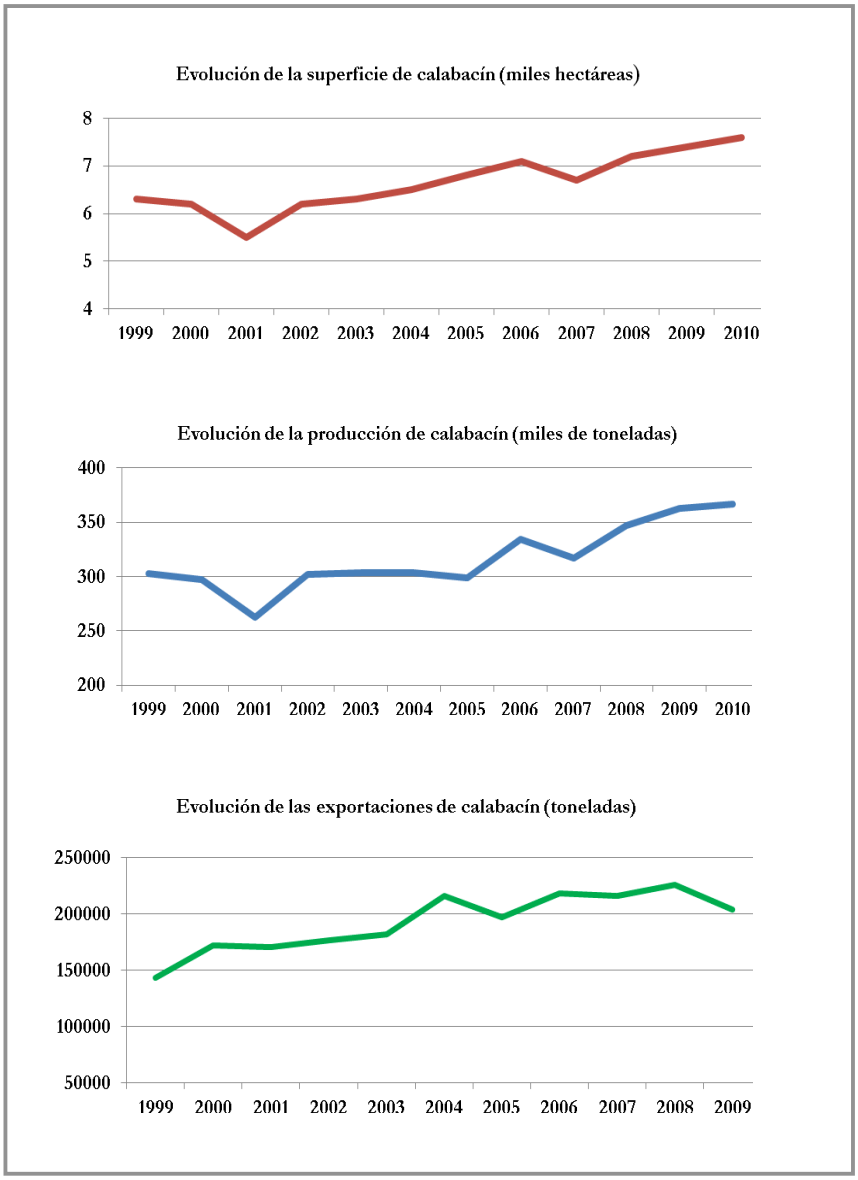



**Figura 1**. Evolución de la superficie, producción y exportación de calabacín en España. Fuente: Ministerio de Medio Ambiente, Medio Rural y Marino. Anuario de estadística 2010.

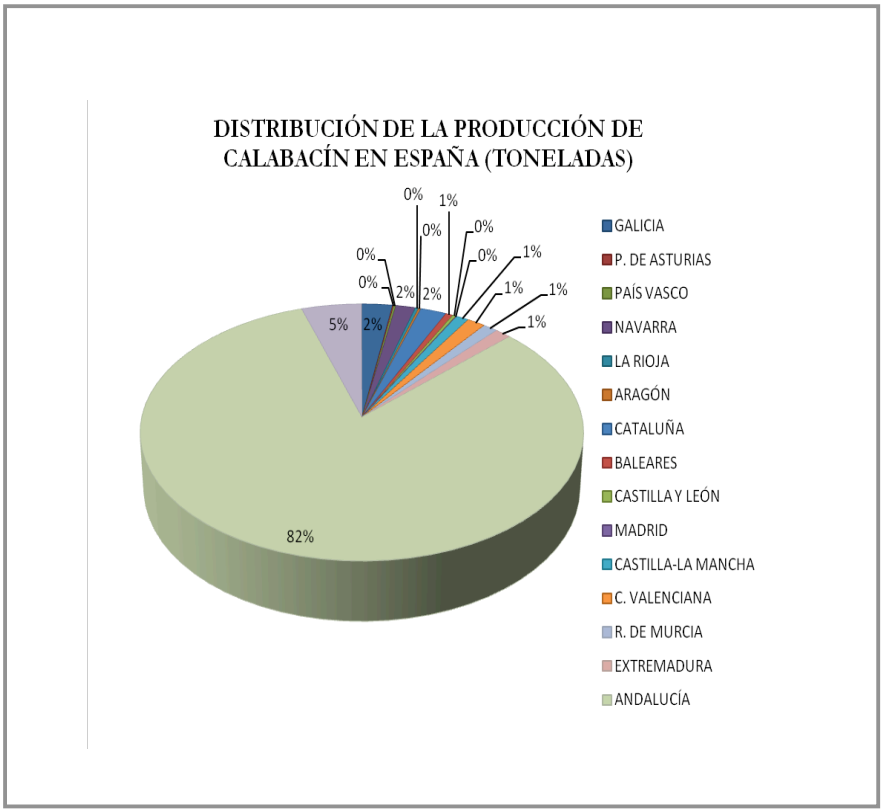



**Figura 2.** Distribución de la producción de calabacín en Andalucía. Fuente: Ministerio de Medio Ambiente, Medio Rural y Marino. Anuario de estadística 2010.

La evolución de las exportaciones ha sido paralela al de las producciones o los ingresos percibidos por el sector. Los principales destinatarios de las exportaciones siguen siendo Alemania, Francia, Holanda y Reino Unido. La Tabla 1 muestra cómo el mercado europeo está liderado por las transacciones comerciales de países de la propia Unión, pudiéndose afirmar que el calabacín español es líder indiscutible en

Europa. Sin embargo, tal y como se ha apuntado para otras hortalizas, es necesario seguir analizando la evolución de las ventas de países como Marruecos y Turquía, dado que su potencial puede convertirlos en fuertes competidores a medio plazo.

**Tabla 1**. Exportaciones de calabacín de nuestros principales competidores.

| | Miles de euros | | | Toneladas | | |
|---|---|---|---|---|---|---|
| | **10-12** | **11-12** | **% var.** | **10-11** | **11-12** | **% var.** |
| España | 175.878 | 207.646,62 | 18,06 | 222.723 | 241.713 | 8,53 |
| Holanda | 24.098 | 27.153,98 | 12,68 | 21.377 | 23.752 | 11,11 |
| Italia | 21.700 | 17.864,13 | -17,68 | 21.381 | 18.160 | -15,06 |
| Alemania | 12.486 | 13.739,56 | 10,04 | 10.383 | 16.480 | 58,72 |
| Francia | 12.122 | 14.544,14 | 19,98 | 13.510 | 15.030 | 11,25 |
| Otros intra UE-27 | 7.323 | 9.103,25 | 24,31 | 10.552 | 11.586 | 9,79 |
| **Total intra UE-27** | **253.608** | **290.051,67** | **14,37** | **299.926** | **326.720** | **8,93** |
| Marruecos | 30.140 | 26.569,49 | -11,85 | 44.464 | 33.642 | -24,34 |
| Turquía | 6.444 | 5.470,11 | -15,12 | 8.581 | 7.531 | -12,24 |
| Otros extra UE-27 | 2.318 | 3.376,33 | 45,65 | 1.797 | 2.243 | 24,81 |
| **Total extra UE-27** | **38.903** | **35.415,93** | **-8,96** | **54.841** | **43.415** | **-20,84** |

Fuente: Fundación Cajamar

## 2. Justificación del trabajo

Al sector hortofrutícola se han ido incorporando de forma paulatina, y siempre incremental, importantes adelantos tecnológicos y mejoras técnicas. La tecnología ha incidido por dos vías en el incremento de las rentabilidades. Ha contribuido a mejorar las condiciones de producción, y a mejorar las condiciones de comercialización. Pero en el modelo de agricultura actual, la acumulación e incorporaciones de elementos tecnológicos a las estructuras de producción (mejoras de cubiertas de invernaderos, de riegos…) encuentra un límite de saturación y es cada vez menos viable desde un enfoque de rentabilidad (Cajamar 2004). Según este planteamiento de **saturación tecnológica**, hay que centrarse más en tecnologías más rentables, como puede ser la biotecnología aplicada a la mejora varietal. El aumento de la producción se puede abordar desde diferentes puntos de vista como el desarrollo de nuevas variedades competitivas, vía incremento del valor añadido, más que con la propia incorporación tecnológica.

Otro punto a tener en cuenta es la **mejora genética** de este cultivo. A pesar de que la especie y el género *Cucurbita* son enormemente variables, el tipo Zucchini, el más importante a nivel económico, es muy uniforme (Paris, 2008). Por otra parte, aunque se han conseguido algunos logros de interés fundamentalmente orientados a mejorar la arquitectura de la planta, a optimizar su floración y a mejorar la resistencia a algunas enfermedades, en comparación a otras Cucurbitaceas, los avances han sido muy limitados en aspectos de interés como la mejora de la calidad del fruto. Esta carencia, se ha debido en parte al retraso de herramientas genéticas y moleculares que facilitan el aprovechamiento de esta variabilidad.

El grupo de Mejora Genética de Cucurbitáceas del COMAV- UPV junto con el Area de Mejora y Biotecnología del IFAPA en sus centros de Córdoba y Almería, han trabajado conjuntamente en la mejora competitiva del calabacín. Como resultado de este trabajo, actualmente se dispone de una colección de 53.252 unigenes, que suponen el primer transcriptoma de calabacín además de un mapa genético de alta densidad en la especie desarrollado a partir de marcadores SNP procedentes del transcriptoma y una plataforma TILLING que será de utilidad para asignar funciones a los genes de esta especie. Estas herramientas y sus diferentes aplicaciones suponen un nuevo impulso para abordar una mejora competitiva en este cultivo.

Por último, hay que destacar la clara importancia que hoy día tiene la **calidad** en los productos hortícolas. La gran preocupación actual por la salud, y el creciente interés por lo que comemos, conlleva a establecer prioridades relacionadas con la alimentación, siendo este punto importante para las empresas del sector. Sin embargo, el termino calidad puede presentar interpretaciones diferentes. Particularmente en los mercados de Europa, América del Norte y Australia, la calidad se refiere a la apariencia externa del producto. La calidad, sin embargo, no sólo debe centrarse en la apariencia, ya que incluso la aplicación más estricta de los estándares de calidad de la UE no garantiza que una fruta o verdura en particular sea sabrosa, rica en nutrientes y vitaminas, o que vaya a mantener sus propiedades aromáticas (Passan et al. 2011). Al considerar la base biológica de la calidad, hay que pensar principalmente en las características de calidad organoléptica y nutricional, en particular los componentes que son importantes para una dieta saludable, así como los procesos biológicos

implicados en su metabolismo. En este sentido, la calidad debe ser un objetivo primordial para los mejoradores, siendo importante apostar por la búsqueda de nuevos productos como alternativas que incentiven nuestros sistemas de comercialización y con ello la apertura a nuevos mercados.

Teniendo en cuenta los puntos anteriores, para ganar competitividad y rentabilidad en el sector es de vital importancia incrementar el desarrollo de nuevas variedades hortícolas con valor añadido. Desde el punto de vista de la calidad nutricional, nos hemos centrado en el estudio de los carotenoides.

## 3. Revisión bibliográfica

Los carotenoides son moléculas isoprenoides; pigmentos naturales que se encuentran en algunas bacterias, algas, plantas e incluso algunos animales aunque éstos no puedan sintetizarlos (Briton 1998). Los carotenoides que se encuentran en la dieta humana derivan principalmente de las plantas cultivadas, donde se encuentran en las raíces, hojas, tallos, semillas, frutos y flores. En general, confieren coloraciones amarilla, naranja y roja a los diferentes tejidos pero la importancia de los carotenoides en los alimentos va más allá de su rol como pigmentos naturales. Entre sus funciones biológicas principales destaca la de proteger a las plantas frente al estrés fotooxidativo (Demming-Adams y Adams 2002) y la de actuar como precursores en la síntesis de hormonas vegetales (Schwartz et al. 2003) pero además, hay que destacar su importancia nutricional. Tienen un papel fundamental en la nutrición humana como antioxidantes, precursores de la vitamina A y su consumo está cada vez más asociado a la protección de una serie de enfermedades, particularmente

ciertos tipos de cáncer y enfermedades oculares (Fenech et al. 2005; Dulinska et al. 2005). De ahí el interés de investigadores de diferentes áreas del conocimiento en estos compuestos. Dentro de la ingeniería genética, esta atención se ha reflejado en importantes avances en el estudio de los genes carotenogénicos y en la manipulación genética de plantas. La intención es incrementar los niveles de estos compuesto en la dieta, mejorando así el valor nutricional de los cultivos (Fraser y Bramdley 2004). Un ejemplo notorio es el caso del arroz dorado "golden rice", donde la manipulación de los genes apropiados de la ruta permitió obtener una variedad de arroz enriquecido en β-caroteno (Ye et al. 2000; Paine et al. 2005). La composición y acumulación de carotenoides en frutos es compleja y característica de cada especie (Gross 1987). Aunque muchas hortalizas o frutas sean fuentes de carotenoides, otras muchas presentan bajas concentraciones de estos compuestos (sobre todo en el mesocarpo). Por lo tanto, el contenido de carotenos en carne es un carácter susceptible de mejora como ocurre en calabacín.

Los genes y enzimas implicados en la biosíntesis de carotenoides han sido extensamente estudiados en el reino vegetal y se conoce que la ruta biosintética de estos compuestos está muy conservada en plantas (Bartley y Scolnik 1995; Cunningham y Gantt 1998; Hirschberg 2001). La regulación de los carotenoides se lleva a cabo durante todo el ciclo de vida de una planta. Se han descrito cambios en la transcripción génica durante el desarrollo y maduración del fruto, desarrollo de la flor o en respuesta a estímulos ambientales externos (Cazzonelli y Pogson 2010). En general, el sistema de regulación de carotenoides empleados por las plantas es complejo. Se sabe que el control de flujo de la ruta puede ser mediada, en parte, por los cambios en la expresión de los genes que codifican enzimas

de la ruta biosintética, en particular aquellos para la PSY y ciclasas, también por la disponibilidad de sustratos producidos a través de la vía del 2C-metil-D-eritrol 4-fosfato (MEP) (Cunningham 2002; Fraser et al. 2002). En plantas, esta vía MEP transcurre en los plastos y en ella se sintetizan los precursores básicos para síntesis de isoprenoides (IPP Y DMAPP), por lo que es predecible que cualquier alteración molecular o bioquímica que afecte a esta vía puede tener un efecto positivo o negativo en la producción de carotenoides (Figura 3). Así por ejemplo se ha estudiado el papel de los genes de esta ruta (Lois et al. 2000) así como el papel de la regulación transcripcional de la ruta de biosíntesis de carotenoides en muchas frutas y hortalizas como zanahoria (Clotault et al. 2008), naranja (Alquezar et al. 2008) o manzana (Ampomah-Dwamena et al. 2012). Muchos han sido los trabajos sobre la regulación de los genes de la ruta de bisíntesis en tomate, apuntando al control transcripcional de los genes de biosíntesis de carotenos como el principal mecanismo regulador del contenido y composición de carotenoides en este cultivo (Giuliano et al. 1993; Pecker el at. 1996; Isaacson et al. 2002; Giovannoni 2004). También hay descritos trabajos en otras Cucurbitaceas como sandía o calabazas (Kang et al. 2010; Nakkanong et al. 2012), sin embargo nada se sabe sobre la regulación transcripcional de los genes carotenogénicos en calabacín.

Es importante destacar que a diferencia de otras Cucurbitaceas, el calabacín se consume en estado inmaduro y que en algunos frutos el aumento de concentración de carotenoides se produce durante la maduración o una vez cosechados (Vicente et al. 2009). Por otra parte, también se han identificado en mapas genéticos los QTLs responsables del color de fruto en diferentes especies asociados a genes bisintéticos de la ruta de carotenoides (Just et al. 2009; Ravel et al. 2013). El estudio de estos

genes y la posible colocalizacion de éstos con las regiones implicadas en el control del carácter en el mapa, sería un paso importante que permitiría desarrollar herramientas de selección para el mejorador.

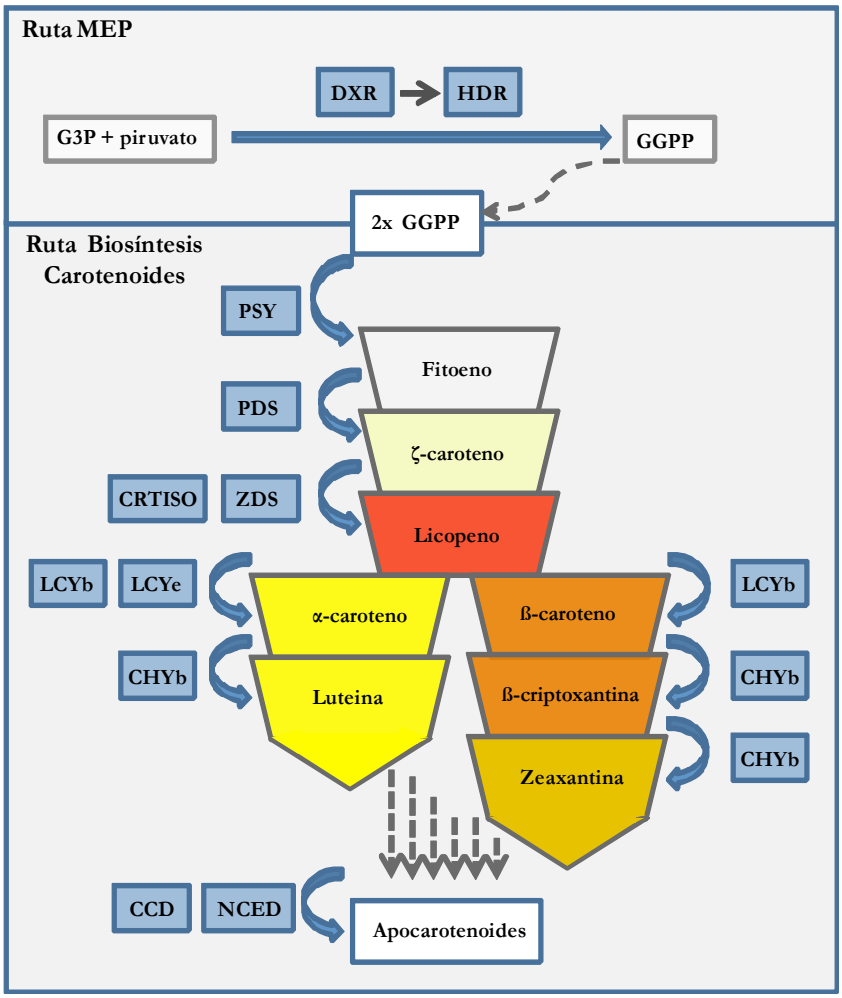



**Figura 3.** Resumen de las enzimas, carotenoides y precursores de la ruta de biosíntesis de carotenoides en planta superiores utilizados en este trabajo.

Dentro de la especie *Cucurbita pepo* (2n=2x=40), hay una gran variabilidad en cuanto a color en el fruto, lo que sugiere que diferentes mecanismos podrían regular el contenido en carotenoides dependiendo de la variedad. Para este trabajo hemos empleado variedades contrastantes en cuanto al color de la piel (Figura 4). Dentro de esta familia se encuentran especies de alto valor agronómico como el pepino, el melón o la sandía, siendo *C. pepo* la especie más polimórfica dentro de la familia y con el mayor valor económico del género (Paris 2008; Esteras et al. 2012). Tomando estas tres variedades como material de partida, en el presente estudio hemos querido analizar los mecanismos moleculares implicados en la regulación de la expresión génica de los genes carotenogénicos en *C. pepo*, además de profundizar en el conocimiento de un gen clave como es *fitoeno sintasa* (*PSY*). Para ello ha sido necesario optimizar todos los pasos previos a cualquier estudio de expresión génica que intervienen en el proceso de análisis (controles de calidad sobre el material de partida, diseño de cebadores, asi como el manejo del proceso) ya que éstos son los que darán fiabilidad y reproducibilidad a los resultados. La necesidad de aplicar una estrategia sólida de normalización basada en el empleo de múltiples genes control, se ha señalado como elemento clave en el análisis de los datos en un experimento de PCR cuantitativa en tiempo real (qPCR) (Bustin et al. 2009=guia MIQE, Gutierrez et al 2008). Por ello, es fundamental la búsqueda de genes estables para el proceso de normalización, llamados genes control o de referencia.

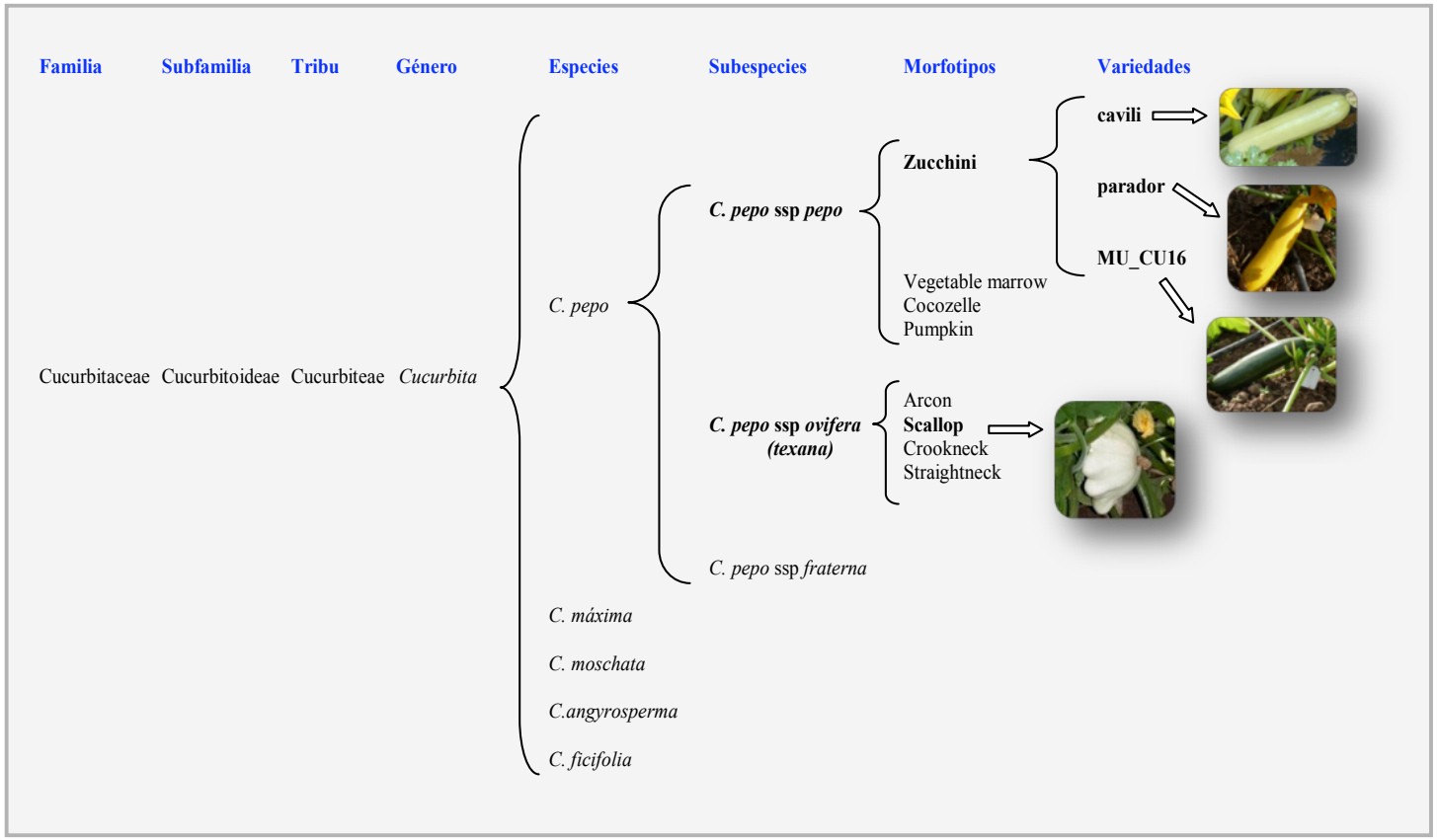



**Figura 4**. Clasificación taxonómica y variedades empleadas de *C. pepo*

## 4. Objetivos

La presente Tesis plantea como objetivo general impulsar, con herramientas biotecnológicas, el desarrollo de nuevas variedades de calabacín que mejoren la calidad del fruto, incidiendo en caracteres de interés como es el contenido en carotenoides. Este objetivo se aborda desde los siguientes objetivos específicos:

- Identificación y selección de genes de referencia en *C. pepo* para normalizar en estudios de expresión génica con PCR cuantitativa (qPCR).

- Identificación y clonación de los genes implicados en la biosíntesis de carotenoides en *C. pepo*.

- Análisis de expresión y estudio de la implicación de los genes carotenogénicos en la acumulación y contenido de carotenoides en tres variedades contrastantes de *C. pepo*.

- Clonación, caracterización y estudio del gen fitoeno sintasa (*PSY*) en *C. pepo*.

- Análisis de expresión de tres genes fitoeno sintasa (*PSY*) en tres variedades constrantastes de *C. pepo*.

## 5. Referencias

Anuario de estadística 2010. Ministerio de Medio Ambiente, Medio Rural y Marino.

Anuario de Agricultura de la Provincia de Almería. 2012. Fundación Cajamar.

Alquezar B, Rodrigo MJ, Zacarías L (2008) Regulation of carotenoid biosynthesis during fruit maturation in the red-fleshed orange mutant Cara Cara. Phytochemistry 69(10):1997–2007.

Ampomah-Dwamena C, Dejnoprat S, Lewis D, Sutherland P, Volz RK, Allan AC (2012) Metabolic and gene expression analysis of apple (*Malus* × *domestica*) carotenogenesis. Journal of Experimental Botany 63 (12):4497-4511.

Bartley GE, Scolnik PA (1995) Plant carotenoids: pigments for photoprotection, visual attraction, and human health. THE PLANT CELL ONLINE 7 (7):1027-1038.

Britton G (1998) Overview of carotenoid biosynthesis. Carotenoids 3:13-147

Bustin SA, Benes V, Garson J, Hellemans J, Huggett JF, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley G, Vandesompele J, Wittwer C (2009) The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. Clinical Chemistry 55 (4):611-622.

Cazzonelli CI, Pogson BJ (2010) Source to sink: regulation of carotenoid biosynthesis in plants. Trends in Plant Science 15 (5):266-274.

Clotault J, Peltier D, Berruyer R, Thomas M, Briard M, Geoffriau E (2008) Expression of carotenoid biosynthesis genes during carrot root development. Journal of Experimental Botany 59 (13):3563-3573

Cunningham FX (2002) Regulation of carotenoid synthesis and accumulation in plants. Pure and applied chemistry 74 (8):1409-1418

Cunningham FX, Gantt E (1998) GENES AND ENZYMES OF CAROTENOID BIOSYNTHESIS IN PLANTS. Annual Review of Plant Physiology and Plant Molecular Biology 49 (1):557-583.

Demmig-Adams B, Adams WW (2002) Antioxidants in Photosynthesis and Human Nutrition. Science 298 (5601):2149-2153.

Dulinka J, Gil D, Zagajewski J, Hartwich J, Bodzioch M, Dembinska-Kiec A, Langmann T, Schmitz G, Laidler P (2005) Different effect of beta-carotene on proliferation of prostate cancer cells. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1740 (2):189-201.

Esteras C, Nuez F, Picó B (2012) Genetic diversity studies in Cucurbits using molecular tools. . In: Genetics, Genomics and breeding of Cucurbits. Science Publishers Inc, Enfield, New Hampshire, pp 140-198

Fenech M, Baghurst P, Luderer W, Turner J, Record S, Ceppi M, Bonassi S (2005) Low intake of calcium, folate, nicotinic acid, vitamin E, retinol, β-carotene and high intake of pantothenic acid, biotin and riboflavin are

significantly associated with increased genome instability—results from a dietary intake and micronucleus index survey in South Australia. Carcinogenesis 26 (5):991-999.

Fraser PD, Bramley PM (2004) The biosynthesis and nutritional uses of carotenoids. Progress in Lipid Research 43 (3):228-265.

Fraser PD, Romer S, Shipton CA, Mills PB, Kiano JW, Misawa N, Drake RG, Schuch W, Bramley PM (2002) Evaluation of transgenic tomato plants expressing an additional phytoene synthase in a fruit-specific manner. Proceedings of the National Academy of Sciences 99 (2):1092-1097.

Giovannoni JJ (2004) Genetic Regulation of Fruit Development and Ripening. THE PLANT CELL ONLINE 16 (suppl 1):S170-S180.

Giuliano G, Bartley GE, Scolnik PA (1993) Regulation of carotenoid biosynthesis during tomato development. THE PLANT CELL ONLINE 5 (4):379-387.

Gross J (1991) Carotenoids. In: Pigments in Vegetables. Springer, pp 75-278

Gutierrez L, Mauriat M, Gunin S, Pelloux J, Lefebvre J, Louvet R, Rusterucci C, Moritz T, Guerineau F, Bellini C, Van Wuytswinkel O (2008) The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. Plant Biotechnology Journal 6 (6):609-618.

Hirschberg J (2001) Carotenoid biosynthesis in flowering plants. Current Opinion in Plant Biology 4 (3):210-218.

Isaacson T, Ronen G, Zamir D, Hirschberg J (2002) Cloning of tangerine from Tomato Reveals a Carotenoid Isomerase Essential for the Production of -β-Carotene and Xanthophylls in Plants. THE PLANT CELL ONLINE 14 (2):333-342.

Just B, Santos CF, Yandell B, Simon P (2009) Major QTL for carrot color are positionally associated with carotenoid biosynthetic genes and interact epistatically in a domesticated × wild carrot cross. Theoret Appl Genetics 119 (7):1155-1169.

Kang B, Zhao W-e, Hou Y, Tian P (2010) Expression of carotenogenic genes during the development and ripening of watermelon fruit. Scientia Horticulturae 124 (3):368-375.

Lois LM, Rodríguez-Concepción M, Gallego F, Campos N, Boronat A (2000) Carotenoid biosynthesis during tomato fruit development: regulatory role of 1-deoxy-D-xylulose 5-phosphate synthase. The Plant Journal 22 (6):503-513.

Nakkanong K, Yang JH, Zhang MF (2012) Carotenoid Accumulation and Carotenogenic Gene Expression during Fruit Development in Novel Interspecific Inbred Squash Lines and Their Parents. Journal of Agricultural and Food Chemistry 60 (23):5936-5944.

Paine JA, Shipton CA, Chaggar S, Howells RM, Kennedy MJ, Vernon G, Wright SY, Hinchliffe E, Adams JL, Silverstone AL, Drake R (2005) Improving

the nutritional value of Golden Rice through increased pro-vitamin A content. Nat Biotech 23 (4):482-487.

Paris H (2008) Summer Squash. In: Prohens J, Nuez F (eds) Vegetables I, vol 1. Handbook of Plant Breeding. Springer New York, pp 351-379.

Passam HC, Karapanos IC, Alexopoulos AA (2011) The Biological Basis of Fruit Quality. In: Breeding for Fruit Quality. John Wiley & Sons, Inc., pp 3-38. Pecker I, Gabbay R, Cunningham F, Jr., Hirschberg J (1996) Cloning and characterization of the cDNA for lycopene β-cyclase from tomato reveals decrease in its expression during fruit ripening. Plant Molecular Biology 30 (4):807-819.

Ravel C, Dardevet M, Leenhardt F, Bordes J, Joseph J, Perretant M, Exbrayat F, Poncet C, Balfourier F, Chanliaud E, Charmet G (2013) Improving the yellow pigment content of bread wheat flour by selecting the three homoeologous copies of Psy1. Molecular Breeding 31 (1):87-99.

Schwartz SH, Qin X, Zeevaart JAD (2003) Elucidation of the Indirect Pathway of Abscisic Acid Biosynthesis by Mutants, Genes, and Enzymes. Plant Physiology 131 (4):1591-1601.

Vicente AR, Manganaris GA, Sozzi GO, al. e (2009) Nutritional quality of fruits and vegetables. In: Florkowski WJ, Shewfelt RL, Brueckner B, al e (eds) Postharvest Handling: A Systems Approach (Food Science and Technology). Academic Press, London, pp 58-106

Ye X, Al-Babili S, Klöti A, Zhang J, Lucca P, Beyer P, Potrykus I (2000) Engineering the Provitamin A (β-Carotene ) Biosynthetic Pathway into (Carotenoid-Free) Rice Endosperm. Science 287 (5451):303-305.

# CAPÍTULO II

# Selection of Reference Genes for Gene Expression Studies in Zucchini (*Cucurbita pepo*) Using qPCR

Ángeles Obrero[1], Jose V. Die[2], Belén Román[1], Pedro Gómez[3], Salvador Nadal[4],

Clara I. González-Verdejo[1]

[1]IFAPA, Centro Alameda del Obispo, Área de Mejora y Biotecnología, Apdo. 14080 Córdoba, Spain.
[2]CSIC, Instituto de Agricultura Sostenible, Mejora Genética Vegetal, Apdo. 14080 Córdoba, Spain.
[3]IFAPA, Centro La Mojonera, Área de Mejora y Biotecnología, Autovía del Mediterráneo sal. 420. E-04745. La Mojonera. Almería. Spain.
[4]IFAPA, Centro Alameda del Obispo, Área de Producción agraria, Apdo. 14080 Córdoba, Spain.

## Abstract

Transcriptomic studies of the important food crop zucchini (*Cucurbita pepo*), are a fundamental tool to accelerate the development of new varieties by breeders. However, the suitability of reference genes for data normalization in zucchini has not yet been studied. The aim of this study was to assess the suitability of 13 genes for their potential use as reference genes in quantitative real-time PCR. Assays were performed on 34 cDNA samples representing plants under different stresses and at different developmental stages. The application of geNorm and NormFinder software revealed that the use of a combination of *UFP*, *EF-1A*, *RPL36aA*, *PP2A*, and *CAC* genes for the different experimental sets was the best strategy for reliable normalization. In contrast, *18S rRNA* and *TUA* were less stable and unsuitable for use as internal controls. These results provide the possibility to allow more accurate use of qPCR in this horticultural crop.

*Keywords* : *Cucurbita pepo*, qPCR, normalization, reference genes, zucchini

## 1. Introducction

Zucchini, also known as summer squash or courgette, is the edible immature fruit of *Cucurbita pepo* spp. *pepo*. It is a worldwide value crop, being an excellent dietary source of vitamins, minerals, and fiber. There is increasing interest from seed companies and zucchini growers in enhancing this vegetable's quality to address consumer desires. New agro-alimentary technologies and, in particular, the implementation of biotechnological tools in classic breeding programs, allow for the more efficient development of more productive varieties that have greater nutritional value. In this sense, transcriptomic studies are becoming increasingly important, as understanding gene expression patterns is expected to reveal new genes involved in the production of desirable characteristics.

Quantitative real-time PCR (qPCR) has become the most prominent emerging method for the quantification of mRNA levels because of its high sensitivity, good reproducibility, and wide dynamic quantification range.[1,2] To obtain accurate results it is necessary to ensure quality measures and to increase experimental transparency in qPCR analysis. In this sense, the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines show basic details that must be taken into account in the publication of qPCR results.[3] Moreover, appropriate normalization strategies are required to control for experimental error. The purpose of normalization is to minimize the impact of nonbiological variation on the results. Several normalization strategies have been proposed, each with its own advantages and disadvantages.[4] The use of one or more reference genes is currently the

preferred means of normalization in quantification, with the assumptions that expression levels of these genes should be relatively constant across the tissues and cells tested and should be unaffected by any experimental treatment.[5,6] Selection of appropriate reference genes for data normalization is essential to ensure correct interpretation of results. This is a laborious task, especially in species with limited available sequence information, such as zucchini, but it is an established procedure that is necessary for the accurate and reliable quantification of differentially regulated mRNAs. Many studies to optimize normalization genes for qPCR experiments have been published for humans, animals, yeasts, and bacteria. In recent years, this analysis has also begun in plants,[7-9] but remains unavailable for zucchini.

Previous Arabidopsis microarray analyses have shown how novel reference genes may have more stable expression than traditional reference genes.[10] Further works in plants based on these results also showed greater stability of expression in these novel reference genes.[11-14] Ten of the 13 zucchini genes evaluated in this work [18S rRNA (*18S rRNA*), 60S ribosomal protein L36a/L44 (*RPL36aA*), actin (*ACT*), α-tubulin (*TUA*), ubiquitin (*UBI*), glucose-6-phosphate dehydrogenase (*G6PDH*), ubiquitin fusion protein (*UFP*), NAD-dependent malic enzyme (*NAD*), NADP-dependent malic enzyme (*NADP*), and elongation factor-1α (*EF-1A*)] are classical reference genes that have commonly been used as internal controls for expression studies.[7,15,16] We also included three new candidate genes: helicase (*HELI*), protein phosphatase 2A (*PP2A*), and clathrin adaptor complexes medium subunit family protein (*CAC*).

The expression patterns of these genes were tested in different tissues and organs (roots, stems, leaves, flowers, and fruits), during

different fruit and flower developmental stages and under several abiotic stress conditions (salt, hormonal stress, and low temperature) using the software applications geNorm[17] and NormFinder[18] to calculate the most stably expressed reference genes and to determine the optimal number of reference genes required for reliable normalization of gene expression data.

The use of nonconstant reference genes can have a strong impact on the results of relative expression analysis.[19] Therefore, our aim was to identify internal controls that exhibit highly constant expression throughout the experimental conditions analyzed. Once the genes were identified, we chose a known expression profile of a well-studied gene in Cucurbitaceae, *CAT1*,[39] to test the impact of reference gene selection in the results using different normalization factors based on the more and least stable stability genes (determined by geNorm and NormFinder). We show that *CAT1* expression levels are significantly dependent on the choice of reference genes and that proper evaluation of reference gene stability is therefore mandatory before qPCR results in zucchini can be reported.

## 2. Material and Methods

### *2.1 Plant material*

Experiments were performed using *C. pepo* L. seeds that were routinely purchased from Nunhems Zaden BV (Haelen, The Netherlands). Seeds were germinated in Petri dishes on wet glass fiber filter paper and covered with aluminum foil, to exclude light, at 20 ºC for 7 days. Seedlings were transplanted into perlite-containing pots in trays with Hoagland nutrient solution.[20] Plants were maintained in a greenhouse at IFAPA Center Alameda del Obispo, Córdoba (Spain), from February through

March.

Material from different organs and development stages and from the stress treatments was dissected and harvested from at least five different zucchini plants to obtain a pool. The procedure was repeated with five distinct plants to obtain a second biological replicate. All samples were immediately frozen in liquid nitrogen and stored at -80 ºC until needed for RNA extraction.

*2.2 Fruit and Flower Developmental Stages.*

Two stages of flower development (before anthesis and anthesis) were collected, and three stages of fruit development (fruits of 1-5, 6-10, and 11-15 cm in length) were also included.

*2.3 Stress Treatments.*

Salt and hormone treatments were applied to 21-day-old zucchini plants by adding 100 mM NaCl or 5 μM 2, 4-dichlorophenoxy acetic acid (2,4-D) for 24 h, with control plants kept in water for the same length of time. Cold treatment was performed by exposing the zucchini plants to 10 ºC for 48 h, whereas control plants were kept at 18-25 ºC. For each stress situation, as well as for the control conditions, plant roots, stems, and leaves were sampled after treatments.

*2.4 Total RNA Isolation.*

All samples were ground to a fine powder with a mortar and pestle in liquid $N_2$. Total RNA was extracted using the TRIsure reagent (Bioline, London, U.K.) according to the manufacturer's instructions. RNA concentration and purity were determined with a biophotometer (Eppendorf, Hamburg, Germany), which calculated absorbance at 260 nm.

**Table 1.** Reference gene primer sequences and amplicon characteristics

| Gene symbol | Accesion Number | Name | Forward and Reverse Primer Sequence [ 5´-3´] | Amplicon Size (bp) | Amplicon Tm (ºC) | PCR efficiency (%) | Regresion coefficient ($R^2$) |
|---|---|---|---|---|---|---|---|
| 18S rRNA | | 18S ribosomal RNA | | | 83.27 | 95 | 0.997 |
| *RPL36aA* | HM594174 | 60S ribosomal protein L36a/L44 | [33]GATAGTCTTGCTGCACAGGGAAA<br>[98]GGTCTGACCTCCATATCCTGATTG | 66 | 79.79 | 95 | 0.998 |
| *ACT* | HM594170 | Actin | [100]CCTCTCAATCCCAAAGCTAACAG<br>[190]CGGCCTGGATAGCAACATACA | 91 | 77.10 | 97 | 0.999 |
| *TUA* | HM594172 | Alpha-tubulin | [148]TGATCTCTGCTTGGATCGTATAAGA<br>[206]CGAGGAAACCTTGAAGACCAGTA | 65 | 77.04 | 93 | 0.998 |
| *UBI* | HM594175 | Ubiquitin | [201]GCAACTGGAAGATGGTAGGACTCT<br>[265]CCAAGTGCAATGTCGTTTCCTT | 65 | 76.11 | 93 | 0.999 |
| *G6PDH* | AF260736 | Glucose-6-phosphate dehydrogenase | [399]CCCCTGCAACTCGAAGATGT<br>[462]ACGATTTCCCACCCTTGCT | 64 | 79.08 | 94 | 0.998 |
| *UFP* | CD726808 | Ubiquitin fusion protein | [168]CGGACCAGCAGAGGCTTATC<br>[223]GAGAGTTCGCCCATCCTCAA | 84 | 80.27 | 98 | 0.998 |
| *PP2A* | HM594171 | Protein phosphatase 2A regulatory subunit A | [98]TGGTAGCATCCTTTCCCAATACA<br>[164]CATGCCCGTTCAGCTTTAGC | 67 | 78.89 | 96 | 0.999 |
| *EF-1A* | HO702383 | Elongation factor-1 alpha | [82]GCTTGGGTGCTCGACAAACT<br>[148]TCCACAGAGCAATGTCAATGG | 67 | 79.03 | 97 | 0.996 |
| *CAC* | HM594173 | Clathrin adaptor complexes medium subunit | [148]GGACAAACAGAACCAACCATGA<br>[212]GGTTTCCTTTCCGTCACTGTAGA | 65 | 77.28 | 94 | 0.998 |
| *NAD* | AF260732 | NAD-dependent malic enzyme | [352]TTCCAGAGCAAATGGGCATT<br>[433]CCGTTCCCTGAACATCATCAT | 81 | 75.70 | 92 | 0.999 |
| *NADP* | AF260735 | NADP-dependent malic enzyme | [400]CCGATCACACTTTCTTGTTCCTT<br>[490]GCGTTCGTCTGTTTCGATATTTC | 91 | 79.98 | 94 | 0.998 |
| *HELI* | HM594176 | DEAD-box RNA helicase-like protein | [1]ACACTGGTCCCTCCCACACA<br>[60]GCGGGCACTTGGAGATTATC | 60 | 78.18 | 93 | 0.995 |

Only RNA samples with 260/280 ratios between 1.9 and 2.1 and 260/230 ratios of >2.0 were used for cDNA synthesis. RNA integrity was assessed by microcapillary electrophoresis with an Experion RNA StdSens Chip and an Experion bioanalyzer (Bio-Rad Laboratories, Hercules, CA), showing rRNA subunits with 18S/25S peaks on the virtual gel and electropherograms. Only RNA samples with electropherograms of high quality and RQI > 8 were used for further study (see Figure S1 of the Supporting Information)

*2.4 cDNA Synthesis.*

cDNA was synthesized from 1 μg of total RNA for each sample using the QuantiTec Reverse Transcription Kit (Qiagen, Hilden, Germany) with a blend of oligo-dT and random primers according to the manufacturer's instructions. In this kit, genomic DNA is efficiently removed in a single step; nevertheless, we included a negative control to test for contaminating genomic DNA. This control contained all of the reaction components except the reverse transcriptase. Samples without enzyme were checked by qPCR, and only when there was no amplification in minus RT controls after 40 cycles with *ACT* primers were the cDNAs used for further analysis. The cDNA samples used for qPCR assays were diluted 1:5 with nuclease-free water.

*2.5 Cloning Partial Sequences.*

*C. pepo* nucleotide sequences for *G6PDH*, *UFP*, *NAD*, and *NADP* were obtained from the GenBank database (see Table 1 for accession numbers). For *18S rRNA* gene amplification, the QuantumRNA 18S Internal Standards kit (Ambion, Austin, TX) was used. Sequences for *ACT*,

*TUA*, *UBI*, *PP2A*, *RPL36aA*, *CAC*, *EF-1A*, and *HELI* genes were not available from the NCBI public database. Specific PCR fragments for *ACT*, *TUA*, *UBI*, *PP2A*, *RPL36aA*, and *CAC* genes were cloned using degenerate primers. *ACT*, *TUA*, and *UBI* fragments were amplified using primers previously described[21] with new PCR conditions adapted to *C. pepo* material (Table 2). For *RPL36aA*, *CAC*, and *PP2A* genes, degenerate primers were designed on the basis of highly conserved regions of proteins within orthologous genes (Table 2). *EF-1A* and *HELI* fragments were cloned using *Pisum sativum* primers previously used to select appropriate reference genes in pea.[22] Amplification of gene fragments was performed using PCR with cDNA as a template. The following PCR conditions were used: an initial denaturation at 94 ºC for 35 s, 40 cycles of denaturation at 94 ºC for 35 s, annealing at 50-56 ºC (according to the gene) for 35 s, and extension at 72 ºC for 1 min, with a final extension of 7 min at 72 ºC. Taq DNA polymerase (Bioline, London, U.K.) was used in the reactions. PCR products for *ACT*, *TUA*, *UBI*, *PP2A*, *RPL36aA*, *CAC*, *EF-1A*, and *HELI* were separated by electrophoresis, purified using a Favorgen GEL/PCR purification kit (Favorgen Biotech Corp., Kaohsiung, Taiwan), cloned into the pGEM-T vector (Promega, Madison, WI), and sequenced (STAB-VIDA, Oeiras, Portugal). The sequence of the amplification product for each primer pair was compared in GenBank (NCBI) using the BLASTX algorithm and searches for homologous genes in other organisms. All sequences showed 76-100% identity, and the resulting sequence data were deposited in the GenBank database under accession no. HM594170, HM594172, HM594175, HM594171, HM594174, HM594173, HO702383, and HM594176, respectively.

**Table 2.** Degenerate primer pairs designed for PCR fragment cloning. Sequence and Tm of the primers as well as product length are given.

| Gene | Forward and reverse primers 5´ -3´ | $T_m$ (ºC) | Amplicon size (bp) |
|---|---|---|---|
| *Deg-RPL36aA* | F: ACN CAR TAY AAR AAR GGN AA<br>R: CAY TGV ARC YTC ARM ACA AT | 50º | 126 |
| *Deg-CAC* | F: TTY GGG CWA ARA TGT TTG CHC TKG G<br>R: GGM ACC TGR AAY TCC ATY TGR ATH GG | 50º | 382 |
| *Deg-PP2A* | F: TAA GTC HTG GCG YGT KCG YTA YAT G<br>R: ACC TGR TTB ACT TGR TCA AGY TTG CT | 55º | 213 |
| *Deg-ACT* | F: GTN ARY AAC TGG GAT GAY ATG G<br>R: ACA ATA CCW GTW GTR CGA CC | 56º | 231 |
| *Deg-TUA* | F: GAA GAT GCT GCH AAY AAY TTY GC<br>R: GAR AGH ACA CTG TTR TAA GGY TC | 56º | 366 |
| *Deg-UBI* | F: GAY TAC AAC ATY CAG ARG GAG<br>R: GCR AAR ATC ARC CTC TGC TG | 56º | 422 |

*2.6 Primer Design and qPCR Conditions.*

Specific primer pairs for qPCR amplification were designed to amplify products of 50-150 bp, with an optimal primer melting temperature (Tm) of 60 ºC and GC contents between 35 and 65%. For 18S rRNA gene amplification, the QuantumRNA 18S Internal Standards kit (Ambion, Austin, TX) was used (Table 1).

MFOLD software v3.2 (http://mfold.bioinfo.rpi.edu/cgi-bin/dnaform1.cgi) was subsequently used to evaluate the possible formation of secondary structures at the sites of primer binding, using the default settings with 50mM $Na^+$, 3mM $Mg^{2+}$, and an annealing temperature of 60 ºC.[23]

qPCR reactions were performed in a 96-well plate with an Mx3000P Real-Time PCR System (Stratagene, La Jolla, CA), using SYBR Green detection. Reactions were performed in a total volume of 10 µL,

containing 0.2 µL of 50× SYBR Green Solution, 5 µL of 2× SensiMix (dT) (Quantace, London,U.K.), 1 µL of cDNA(corresponding to~10 ng of total RNA), and 0.4 L (200 nM) of each gene-specific primer. Except for *RPL36aA*, the following standard thermal profile was used for all PCR reactions: polymerase activation (95 ºC for 10 min) and amplification and quantification cycles repeated 40 times (95 ºC for 1 min, 60 ºC for 1 min). For *RPL36aA*, primer annealing at 65 ºC was necessary to generate a sharp peak in melting curve analysis. Finally, a dissociation analysis of the PCR products was performed by running a gradient from 60 to 95 ºC to confirm the presence of a single product in each reaction and the specificity of the qPCR. Two biological replicates were analyzed using Mx3000P analysis software v4.1 (Stratagene). All amplification plots were analyzed with an $R_n$ threshold of 0.03 to obtain quantification cycle (Cq) values for each gene-cDNA combination. The PCR efficiency (E) was estimated using LinReg software with data obtained from the exponential phase of each individual amplification plot and the equation $(1 + E) = 10^{slope}$.[24] Table 1 shows the primer sequences, amplicon sizes, melting temperatures, and PCR efficiencies.

*2.7 Statistical Analyses.*

Gene expression levels were calculated for all individual reference genes based on the number of Cq. To analyze the stability of expression and to identify the most suitable reference genes, we used the freely available statistical algorithms geNorm v3.5n (http://medgen.ugent.be /~jvdesomp/genorm/) and NormFinder (http://www.mdl.dk /publicationsnormfinder.htm) according to the authors' recommendations. For both programs, $C_q$ values were converted

to linear scale expression quantities using the formula $Q = E^{-\Delta Cq}$ where E is the efficiency and ΔCq is the difference between the $C_q$ value of the target sample and the value of the sample with the lowest $C_q$. geNorm calculates a gene expression normalization factor for each tissue sample based on the geometric mean of a user-defined number of reference genes. The expression stability measure (*M*) is calculated as the mean of the pairwise variation of a gene compared to that of all other genes. The genes with the lowest *M* values were considered the most stable ones, whereas the highest *M* value indicated the least stable expression. The NormFinder program is another Visual Basic application tool for Microsoft Excel, used for the determination of stability of expression of reference genes. This approach is based on a variance estimation, and it focuses on finding smaller intra- and intergroup expression variation and combining both results in a stability value. To assess both intra- and intergroup variations, four subgroups were established as geNorm analysis (salt stress, hormonal stress, cold stress, and fruit and flower developmental stages). The mean values of the two biological replicates were calculated for the analysis. Normalized ratios of the target gene *CAT1* (GenBank accession no. GI: 862451) in leaves and roots under saline conditions as well as in control samples were calculated according to the geNorm manual[25]. Primers used for *CAT1* amplification are shown in Table 1.

## 3. Results

### *3.1. Permofamance of Amplification Primers*

qPCR was conducted on the 34 samples with 13 primer pairs. To optimize PCR conditions, the secondary structures of the sequences were

checked with nucleic acid-folding software (Figure S3 of the Supporting Information). All primers that generated amplicons possessing secondary structures with significant negative ΔG values (free energy in kcal/mol) at the primer binding sites were excluded from further analysis. The specificity of the amplifications was confirmed by the single-peak melting curves of the PCR products, with no visible primer dimer formation (Figure 1 and Figure S2 of the Supporting Information). PCR efficiencies varied from 92 to 98%, and correlation coefficients ($R^2$) ranged between 0.995 and 0.999 (Table 1).

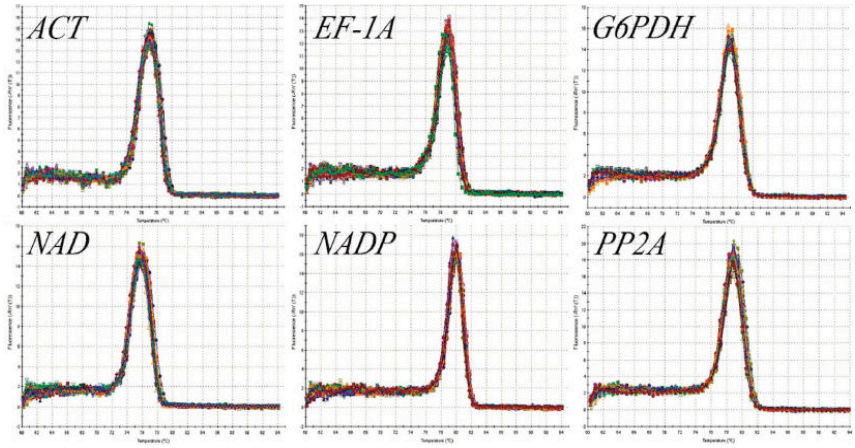



**Figure 1**. Dissociation curves of six representative reference genes, showing single peaks.

*3.2 Expression Profiles of Reference Genes.*

qPCR analysis methods were used to measure the transcription levels of 13 candidate reference genes involved in different cellular functions, from 34 different cDNA samples. The expression levels were presented as quantification cycles ($C_q$) (Figure 2), and these values were

compared. The lowest mean $C_q$ value was for *18S rRNA* (15.18), and the highest value was for *HELI* (27.40); therefore, *18S rRNA* transcript levels were about 4770-fold more abundant than those of helicase. These results indicate that the tested genes show a wide range of expression levels.

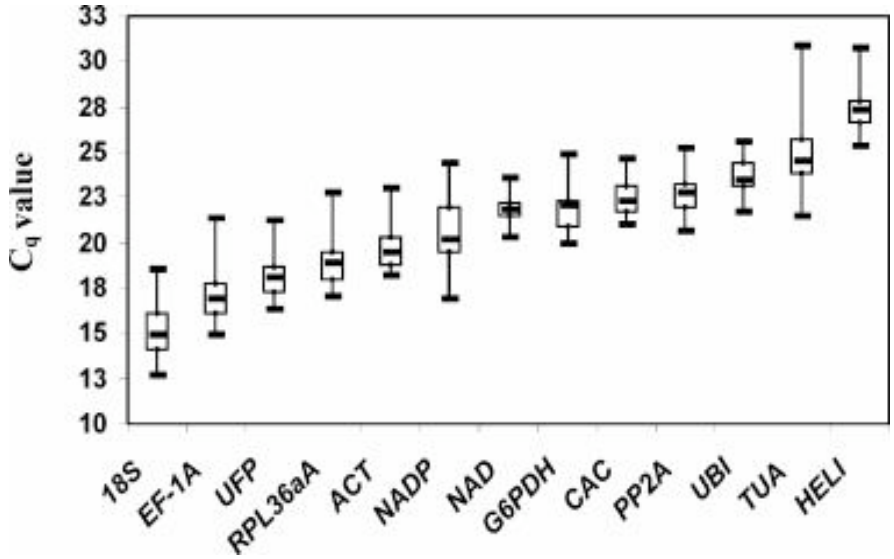



**Figure 2**. qPCR Cq values for reference genes in all samples. Boxes indicate the first and third quartiles. The vertical lines (whiskers) represent the maximum and minimum values. The central line represents the median value.

*3.3 Data Analysis.*

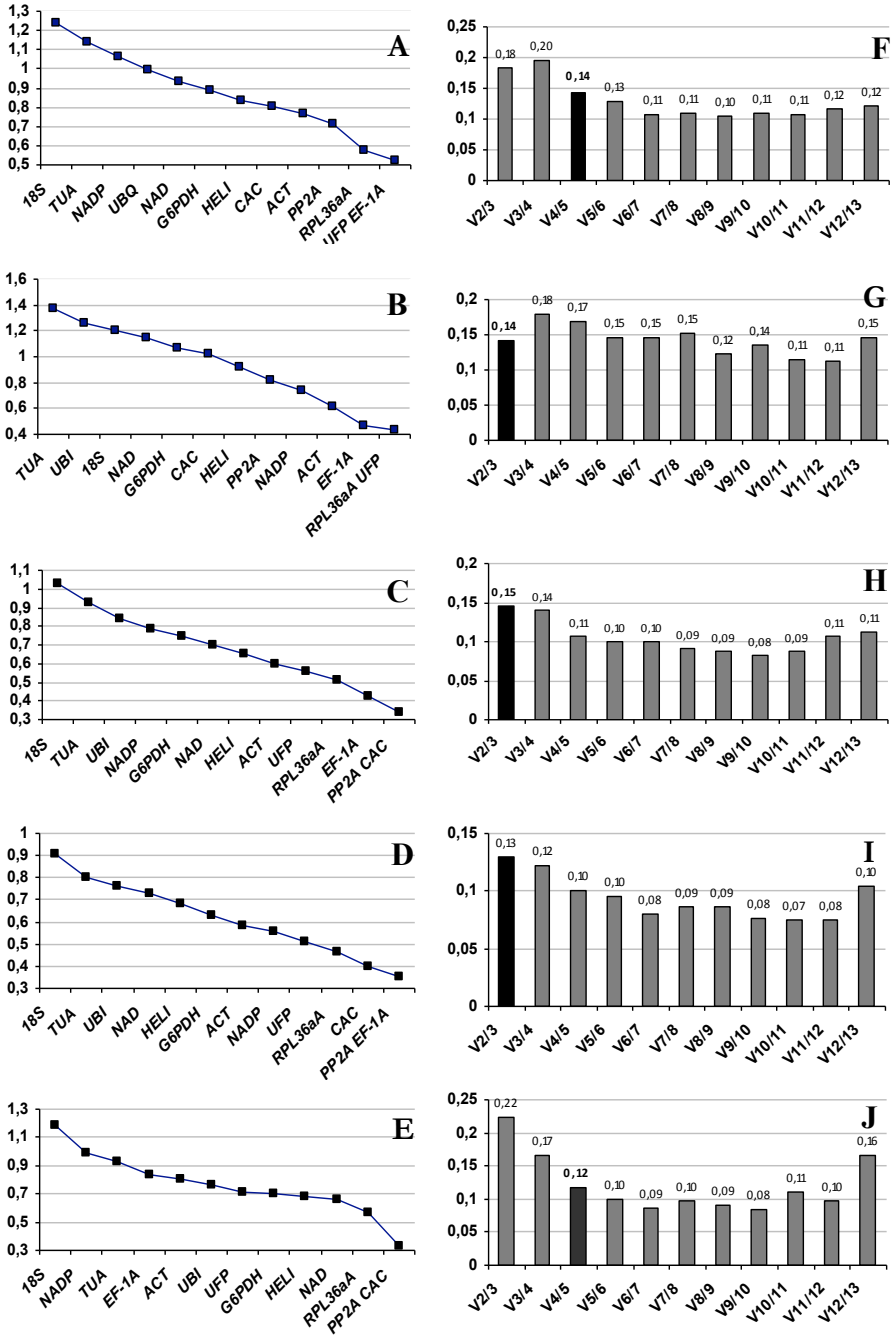
*geNorm.* Tested genes were ranked according to stability values (*M*) and were analyzed for total samples and across four subsets (Figure 3). The 13 genes' M values in the different experimental conditions were <1.4, which is below the geNorm threshold of 1.5. When all 34 tissue samples were considered together, *UFP* and *EF-1A* were selected as having the most stable expression levels, whereas *18S rRNA* was expressed most

variably (Figure 3A). The lowest values of M for fruit and flower developmental stages were for *RPL36aA* and *UFP*, whereas the highest value was for *TUA* (Figure 3B). For treatments with salt and cold stress, expression levels of *PP2A* and *CAC* were the most stable (Figure 3C, E), whereas *PP2A* and *EF-1A* were the most stable under hormonal stress (Figure 3D). For the three treatments (salt, hormonal, and cold), the highest M value was observed for *18S rRNA*.

It is crucial to determine the optimal numbers of genes that are necessary for accurate normalization in each experiment. geNorm uses the pairwise variation $V_n/V_{n+1}$ to define the optimal number of reference genes. Vandesompele et al.[17] established a threshold of 0.15, below which the addition of more genes is not necessary. Panels G, H, and I of Figure 3 show that the inclusion of a second gene did not contribute significantly to the variation of the normalization factor, as the value of $V_{2/3}$ was <0.15. Nevertheless, a third gene could be included in salt and hormonal stresses ($V_{3/4} < 0.15$) because the use of at least three reference genes is recommended by some authors.[17,26] In panels F and J of Figure 3, the values $V_{3/4} > 0.15$ and $V_{4/5} < 0.15$ indicated that the use of the four most stable genes was necessary for normalization in total samples and cold stress analyses.

*NormFinder*. The stabilities of the 13 genes were also analyzed using the NormFinder program. In Figure 4, the lowest stability value indicates the most stably expressed gene. The results obtained for NormFinder analysis were similar to those from geNorm analysis. When all samples were evaluated, four subgroups were established, taking into account the intra- and intergroup variation (fruit and flower development stages, salt stress, hormonal stress, and cold stress). In this case, *UFP* and *ACT* genes

showed the most stable expression levels, whereas *18S rRNA* was the least stable (Figure 4A). Among the subsets, *PP2A* and *CAC* were identified, as




**Average expression stability M** **Pairwise Variation V**

**Figure 3.** Expression stability and ranking of reference genes as calculated by geNorm in all samples (A), developmental stages in fruit and flower (B), salt treatment (C), hormonal treatment (D), and cold treatment (E). A lower value of average expression stability, M, indicates more stable expression. Pairwise variations calculated by geNorm to determine the minimum numbers of reference genes for accurate normalization in all samples (F), developmental stages in fruit and flower (G), salt treatment (H), hormonal treatment (I), and cold treatment (J).

was the case in geNorm, as the two most stable genes in salt, hormonal, and cold treatment. *18S rRNA* was also found by both algorithms to be the least stable gene in the three treatments (Figures 4C, D, F). Fruit and flower developmental stages showed slight differences between geNorm and NormFinder algorithms; *PP2A* and *ACT* were found to be the most stable genes, whereas both analyses ranked *TUA* as the least stable gene (Figure 4B).

*3.5 Choice of Reference Genes Affects Normalization.*

To test the impact of reference gene selection and to evaluate the suitability of the reference genes selected in this study, the differential expression of *CAT1* was calculated in leaves and roots subjected to salt stress as well as in control samples. On the basis of our data, three normalization factors (NF) were derived by calculating the geometric averages of the following gene combinations: $NF_1$, two top-ranking genes selected by geNorm and NormFinder in salt stress (*PP2A*, *CAC*); $NF_2$, three top-ranking genes identified by geNorm and NormFinder in salt stress (*EF1A*, *PP2A*, *CAC*); and $NF_3$, the least stably expressed gene according to both geNorm and NormFinder in salt stress (*18S rRNA*). Subsequently, we investigated whether these strategies resulted in

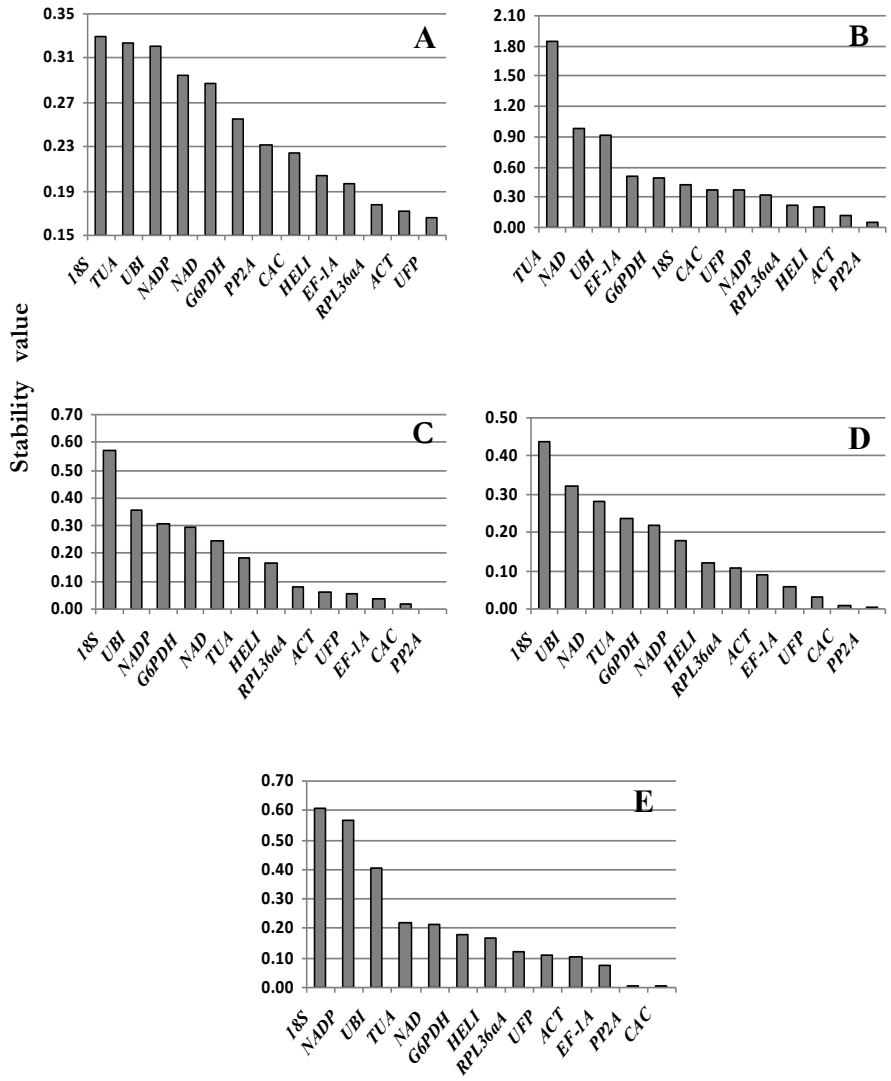



**Figure 4**. NormFinder analysis for the 13 reference genes. A lower stability value indicates more stable expression in all samples (A), developmental stages in fruit and flower (B), salt treatment (C), hormonal treatment (D), and cold treatment (E).

significantly different expression results. Figure 5A shows relative expression levels of *CAT1* after normalization with the three normalization factors. When *CAT1* transcript quantities were normalized with the least stable gene (*18S rRNA*), the standard error increased. Figure 5B shows the mean coefficients of variation (CV) for the three normalization factors; $NF_3$ exhibited the majority of the variation (CV 75%). Normalization of *CAT1* expression with $NF_1$ (CV 38%) for salt treatment resulted in <1.2-fold enrichment of transcription compared to normalization with $NF_2$ (CV 47%). On the contrary, when normalization was performed with $NF_3$ (CV 75%), nearly 4.5-fold enrichment of transcription was obtained, compared to normalization with $NF_1$

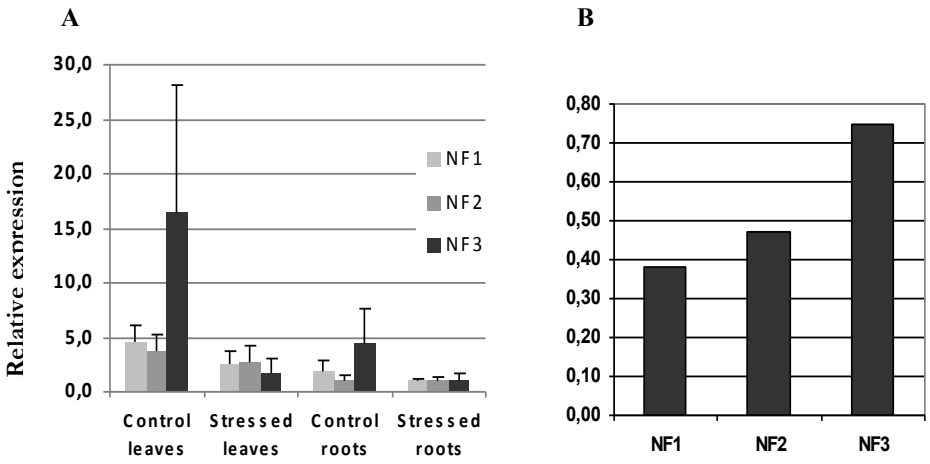



**Figure 5**. (A) Relative expression levels of *CAT1* in zucchini leaves and roots under salt treatment. *CAT1* expression was normalized with three normalization factors: $NF_1$ (*CAC/PP2A*), $NF_2$ (*CAC/PP2A/EF-1A*), and $NF_3$ (*18S*). Normalized values of *CAT1* relative expression are given as averages. Bars indicate standard errors. (B) Variation in *CAT1* after normalization. Values are given as averages of three different coefficients of variation.

## 4. Discussion

Many factors in qPCR may affect the accuracy of the results and the reliability of the conclusions, including the selection of the reference genes. Normalizing to reference gene expression is a popular method for internally controlling for error and for eliminating nonbiological variation in qPCR. To our knowledge, there is no information in the literature regarding the choice of reference genes for gene expression studies in zucchini. Compared with other species of Cucurbitaceae, zucchini is a species with limited genetic information, hence the importance and difficulty of this study. In this work, we conducted a systematic study of the expression stability of 13 reference genes in zucchini, along a series composed of 34 samples from different tissues/organs subjected to different stresses and at development stages.

Before starting qPCR quantification, we paid careful attention to preparative steps, which were performed in compliance with MIQE guidelines. Among these steps, it is considered critically important to select RNA samples with high RQI values. It is known that tissue handling in clinical applications must be carefully controlled to preserve the purity and integrity of the RNA. Nevertheless, this step receives less attention in plants. Imbeaud et al.[27] reported that working with low-quality RNA may strongly compromise the results of downstream applications, which are often labor-intensive, time-consuming, and highly expensive. For this reason, we performed careful RNA analysis by using microcapillary electrophoresis with Experion before carrying out gene expression studies. Another important factor that affects the accuracy of gene expression analysis is DNA contamination. A minus RT control demonstrated the

absence of contaminating genomic DNA, avoiding overestimation of the amount of RNA present. Primer pairs for each gene were also optimized obtaining correct efficiencies because it is known that the presence of hairpins overlapping primer annealing sites and forming secondary structures in the amplicon have negative impacts on PCR efficiency.[28] Taking into account the fact that all qPCR assays were performed with proper controls and with the checklist discussed above, we were confident that our gene expression results were accurate and reliable.

Once the quality parameters were confirmed, we studied the expression patterns of genes being evaluated for their suitability as reference genes, using two different algorithms implemented in the programs geNorm and NormFinder. These statistical models were chosen because they take PCR efficiency into account, they are freely available, and they can analyze >10 genes. In addition, geNorm supplies additional information regarding the minimal number of genes necessary for valid normalization, and NormFinder is not significantly affected by the coregulation of candidate genes.

Although the two programs are based on different statistical algorithms, they showed several common features, and only slight differences in the rankings of the best scored genes were found. For example, *UPF*, *RPL36aA*, and *EF-1A* were among the four most stable genes all samples, but *ACT* was ranked second by NormFinder and fourth by geNorm. In contrast, both programs determined that *PP2A* and *CAC* were ranked highest for different abiotic stresses. *PP2A* and *CAC* were found to be the most stably expressed genes in Arabidopsis under abiotic stresses,[10] and *CAC* also showed high stability of expression in tomato, coffee, and cucumber studies,[11,29,30] indicating that novel reference genes are

very often more stable in various experimental conditions than the commonly used internal controls. The largest differences between both algorithms were found when we analyzed flower and fruit developmental stages. geNorm ranked *RPL36aA* and *UFP* as the most stable genes, whereas NormFinder ranked *ACT* and *PP2A* highest. These discrepancies could be due to the $C_q$ values from flower samples, which were the most variable among repetitions in our analysis compared with other tissues. The differences found in the results between geNorm and NormFinder methods were expected given that the programs are based on distinct statistical algorithms. Some previous publications that identified reference genes for plant research have also found discrepancies between both algorithms,[11,13,29,31-33] and there is not a consensus regarding which is the best method. Although we have found differences between both algorithms in this results set, the two most stable genes for geNorm (*RPL36aA* and *UFP*) were within the six most stable genes for NormFinder (stability value < 0.37), and the two most stable genes for NormFinder (*PP2A* and *ACT*) were also found within the six most stable genes for geNorm (stability value M < 0.82). geNorm indicated that studies involving cold stress and total samples required at least four control genes for reliable and accurate normalization, whereas two control genes were enough for experiments within different development stages or salt and hormonal stresses. However, a third gene could be included as is recommended by Vandesompele et al.[17]

In some others plants studied, *18S rRNA* and *TUA* were identified as the most stable expressed reference genes.[7,34-36] However, in the present analysis, *18S rRNA* and *TUA* showed considerable variation in different experimental conditions by the two algorithms and were always classified

among the least reliable control genes for the four subsets and total samples. The poor stability of these genes was also found by several other authors.[30-33,37,38]

The expression of catalase (*CAT1*) was analyzed to verify the utility of the most stable reference genes as internal controls. Catalase is one of several cellular antioxidant defenses that are involved in detoxification of active oxygen species that are generated by various environmental stresses. In higher plants, several different catalase subunits are encoded by small multigene families, and differential patterns of expression have been reported in different plant tissues and developmental stages.[39,40] Esaka et al.[39] demonstrated by Northern analysis differential expression of *CAT1* in various tissues of pumpkin. In the present experiment, roots and leaves were included to determine the different expression patterns by using qPCR technology. The authors considered cold and hormonal stress in their study, so it was our aim to complete this information by considering the salt stress. *CAT1* expression was found to be influenced more by tissue type than by salt stress, with mRNA levels higher in leaves than in roots.

To check the impact of including a third gene in the NF, we performed normalization for *CAT1* gene expression data by using two ($NF_1$) or three ($NF_2$) reference genes. In this particular case, normalization of *CAT1* with three genes did not result in significant differences in *CAT1* relative expression. Because the same results were obtained with two and three reference genes, we concluded that the inclusion of a third gene was not necessary; normalization with *PP2A* and *CAC* was sufficient for accurate normalization. Nevertheless, the number of reference genes for normalization must be carefully considered for each particular experimental situation and should reflect a balance between economic adequacy,

experimental rigor, and practical considerations, with the aim of minimizing variation in NF and obtaining accurate results. Normalization with *18S rRNA* ($NF_3$) indicated significantly higher expression levels than normalization with a combination of several stably expressed genes, resulting in an overestimation of the fold changes in gene expression across control leaves and control roots. This indicates that normalizing with an improper reference gene greatly affects the apparent expression pattern of the target gene. In conclusion, this is the first detailed study on the evaluation of selected reference genes in zucchini. By a combination of two software programs for data analysis, this study showed the genes *UFP*, *EF-1A*, and *RPL36aA* to be the most stable expressed reference genes in the whole sample set and the genes *PP2A* and *CAC* to be the most stably expressed under abiotic stresses. In the context of different development stages *RPL36aA* and *UFP* were identified by geNorm and *PP2A* and *ACT* by NormFinder. Because the two programs showed high stability value for the four genes, these genes can be used for normalization in fruit and flower developmental stages.

Our identification and validation of suitable zucchini control genes will facilitate future developmental transcriptomic studies in this economically important plant. With the genome sequence of other Cucurbitaceae available, it seems possible to transfer this information to zucchini, for which transcriptomic approaches are still lacking.

## 5. Funding Sources

Financial support was provided by the Spanish Project INIARTA2008-00035-C02-01, FEDER and FSE funds (EU). AOC

acknowledges a predoctoral fellowship from INIA. JVD is researcher funded by the 'Juan de la Cierva' programme. CIGV has a contract from INIA (subprogram DOC INIA-CCAA). The authors thank the Ministerio de Ciencia e Innovación from Spain and FSE funds.

## 6. References

(1) Bustin, S. A. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J. Mol. Endocrinol.* **2000**, 25, 169–193.

(2) Pfaffl, M. W.; Hageleit, M. Validities of mRNA quantification using recombinant RNA and recombinant DNA external calibration curves in real-time RT-PCR. *Biotechnol. Lett*. **2001**, 23, 275–282.

(3) Bustin, S. A.; Benes, V.; Garson, J. A.; Hellemans, J.; Huguett, J.; Kubista, M.; Mueller, R.; Nolan, T.; Pfaffal, M. W.; Shipley, G. L.; Vandesompele, J.; Wittwer, C. T. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem*. **2009**, 55, 611–622.

(4) Huggett, J.; Dheda, K.; Bustin, S.; Zumla, A. Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun*. **2005**, 6, 279 284.

(5) Nolan, T.; Hands, R. E.; Bustin, S. A. Quantification of mRNA using real-time RT-PCR. *Nat. Protoc*. **2006**, 1, 1559–1582.

(6) VanGuilder, H. D.; Vrana, K. E.; Freeman, W. M. Twenty-five years of quantitative PCR for gene expression analysis. *BioTechniques* **2008**, 44, 619–26.

(7) Brunner, A. M.; Yakovlev, I. A.; Strauss, S. H. Validating internal controls for quantitative plant gene expression studies. *BMC Plant Biol.* **2004**, 4, 14.

(8) Gachon, C.; Mingam, A.; Charrier, B. Real-time PCR: what relevance to plant studies?. *J. Exp. Bot.* **2004**, 55, 1445–1454.

(9) Gutierrez, L.; Mauriat, M.; Guenin, S.; Pelloux, J.; Lefebvre, J. F.; Louvet, R.; Rusterucci, C.; Moritz, T.; Guerineau, F.; Bellini, C.; Van Wuytswinkel, O. The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. *Plant Biotechnol. J*. **2008**, 6, 609–618.

(10) Czechowski, T.; Stitt, M.; Altman, T.; Udvardi, M. K.; Scheible, W. R. Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiol.* **2005**, 139, 5–17.

(11) Exposito-Rodríguez, M.; Borges, A. A.; Borges-Perez, A.; Pérez, J. A. Selection of internal control genes for quantitative real-time RTPCR studies during tomato development process. *BMC Plant Biol.* **2008**, 8, 131.

(12) Remans, T.; Smeets, K.; Opdenakker, K.; Mathijsen, D.; Vangronsveld, J.; Cuypers, A. Normalisation of real-time RT-PCR gene expression measurements in Arabidopsis thaliana exposed to increased metal concentrations. *Planta* **2008**, 227, 1343–1349.

(13) Paolacci, A. R.; Tanzarella, O. A.; Porceddu, E.; Ciaffi, M. Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. *BMC Mol. Biol.* **2009**, 10, 11.

(14) Hu, R.; Fan, C.; Li, H.; Zhang, Q.; Fu, Y.-F. Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR. *BMC Mol. Biol.* **2009**, 10, 93.

(15) Radonic, A.; Thulke, S.; Mackay, I. M.; Landt, O.; Siegert, W.; Nitsche, A. Guideline to reference gene selection for quantitative real time PCR *Biochem. Biophys. Res. Commun.* **2004**, 313, 856–862.

(16) Nicot, N.; Hausman, J. F.; Hoffmann, L.; Evers, D. Housekeeping gene selection for real time RT-PCR normalization in potato during biotic and abiotic stress. *J. Exp. Bot.* **2005**, 56, 2907–2914.

(17) Vandesompele, J.; De Preter, K.; Pattyn, F.; Poppe, B.; Van Roy, N.; De Paepe, A.; Speleman, F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **2002**, 3, RESEACH0034.

(18) Andersen, C. L.; Jensen, J. L.; Orntoft, T. F. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* **2004**, 64, 5245–5250.

(19) Dheda, K.; Huggett, J. F.; Chang, J. S.; Kima, L. U.; Bustin, S. A.; Johnson, M. A.; Rook, G. A. W.; Zumla, A. The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. *Anal. Biochem.* **2005**, 344, 141–143.

(20) Hoagland, H. D.; Arnon, D. I. The water-culture method for growing plants without soil. *Calif. Agric. Exp. Stn. Circ.* **1950**, 347, 1–32.

(21) González-Verdejo, C. I.; Die, J. V.; Nadal, S.; Jiménez-Marín, A.; Moreno, M. T.; Román, B. Selection of housekeeping genes for normalization by real-time RT-PCR: analysis of *Or-MYB1* gene expression in *Orobanche ramosa* development. *Anal. Biochem.* **2008**, 379, 38–43.

(22) Die, J. V.; Román, B.; Nadal, S.; González-Verdejo, C. I. Evaluation of candidate reference genes for expression studies in Pisum sativum under different experimental conditions. *Planta* **2010**, 232, 145–153.

(23) Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Bioinformatics* **2003**, 31, 3406–3415.

(24) Ramakers, C.; Ruijter, J. M.; Deprez, R. H.; Moorman, A. F. Moorman, Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci. Lett.* **2003**, 339, 62–66.

(25) geNorm software manual update. http://medgen.ugent.be/ ~jvdesomp/ genorm/.

(26) Derveaux, S.; Vandesompele, J.; Hellemans, J. How to do successful gene expression analysis using real-time PCR. *Methods* **2010**, 50, 227–230.

(27) Imbeaud, S.; Graudens, E.; Boulanger, V.; Barlet, X.; Zaborski, P.; Eveno, E.; Mueller, O.; Schroeder, A.; Auffray, C. Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. *Nucleic Acids Res.* **2005** 33, e56.

(28) Hoebeeck, J.; van der, L. R.; Poope, B.; De, S. E.; Yigit, N.; Claes, K.; Zewald, R.; de Jong, G. J.; De, P. A.; Speleman, F.; Vandesompele, J. Rapid detection of VHL exon deletion using real-time quantitative PCR. *Lab. Invest.* **2005**, 85, 24–33.

(29) Cruz, F.; Kalaoun, S.; Nobile, P.; Colombo, C.; Almeida, J.; Barros, L.; Romano, E.; Grossi-de-Sa, M. F.; Vaslin, M.; Alves-Ferreira, M. Evaluation of coffee reference genes for relative expression studies by quantitative real time RT-PCR. *Mol. Breed.* **2009**, 23, 607–616.

(30) Migocka, A.; Papierniak, M. Identification of suitable reference genes for studying gene expression in cucumber plants subjected to abiotic stress and growth regulators. *Mol. Breed.* **2010**, doi: 10.1007/s11032-010-9487-0.

(31) Wan, H.; Zhao, Z.; Qian, C.; Sui, Y.; Malik, A. A; Chen, J. Selection of appropriate reference genes for gene expression studies by quantitative real-time polymerase chain reaction in cucumber. *Anal. Biochem.* **2010**, 399, 257–261.

(32) Lin, Y. L.; Lai, Z. X. Reference gene selection for qPCR analysis during somatic embryogenesis in longan tree. *Plant Sci.* **2010**, 178, 359–365.

(33) Huis, R.; Hawkins, S.; Neutelings, G. Selection of reference genes for quantitative gene expression normalization in flax (*Linum usitatissimum* L). *BMC Plant Biol.* **2010**, 10, 71.

(34) Kim, B. R.; Nam, H. Y.; Kim, S. U.; Kim, S. I.; Chang, Y. J. Normalization of reverse transcription quantitative-PCR with housekeeping

genes in rice. *Biotechnol. Lett.* **2003**, 25, 1869–1872.

(35) Jarosov , J.; Kundu, J. K. Validation of reference genes as internal control for studying viral infections in cereals by quantitative real-time RT-PCR. *BMC Plant Biol.* **2010**, 10, 146.

(36) Maroufi, A.; Bockstaele, E. V.; Loose, M. D. Validation of reference genes for gene expression analysis in chicory (Cichorium intybus) using quantitative real-time PCR. *BMC Mol. Biol.* **2010**, 11, 15.

(37) Cortleven, A.; Remans, T.; Brenner, W. G.; Valcke, R. Selection of plastid- and nuclear-encoded reference genes to study the effect of altered endogenous cytokinin content on photosynthesis genes in Nicotiana tabacum. *Photosynth. Res.* **2009**, 102, 21–29.

(38) Tong, Z.; Gao, Z.; Wang, F.; Zhou, J.; Zhang, Z. Selection of reliable reference genes for gene expression studies in peach using realtime PCR. *BMC Mol. Biol.* **2009**, 10, 71.

(39) Esaka, M.; Yamada, N.; Kitabayashi, M.; Setoguchi, Y.; Tsugeki, R.; Kondo, M.; Nishimura, M. cDNA cloning and differential gene expression of three catalases in pumpkin. *Plant Mol. Biol.* **1997**, 33, 141–155.

(40) Robertson, C. Regulation of catalases in Arabidopsis. *Free Radical Biol. Med.* **1997**, 23, 489–496.

## Supporting Information

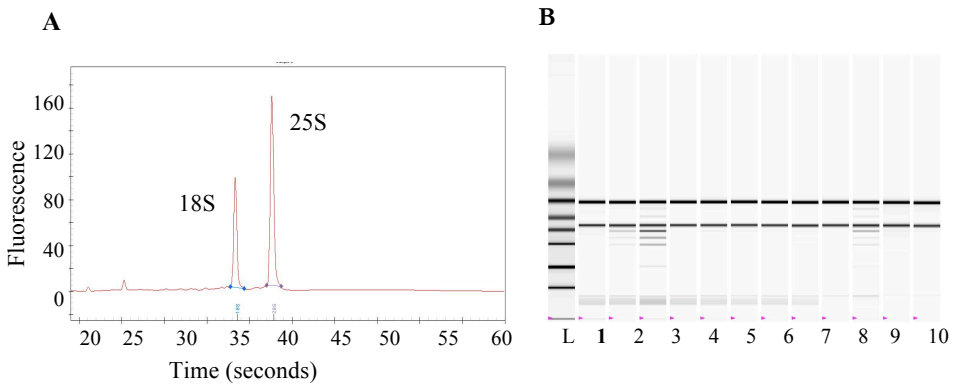



**Figure S1**. Example of good quality RNA using Experion RNA StdSens Chip. A) Electropherogram of high-quality RNA for root (sample 1, RQI =10) B) virtual gel for twelve samples (L=ladder). Samples 3 and 10 show bands corresponding to plastid.

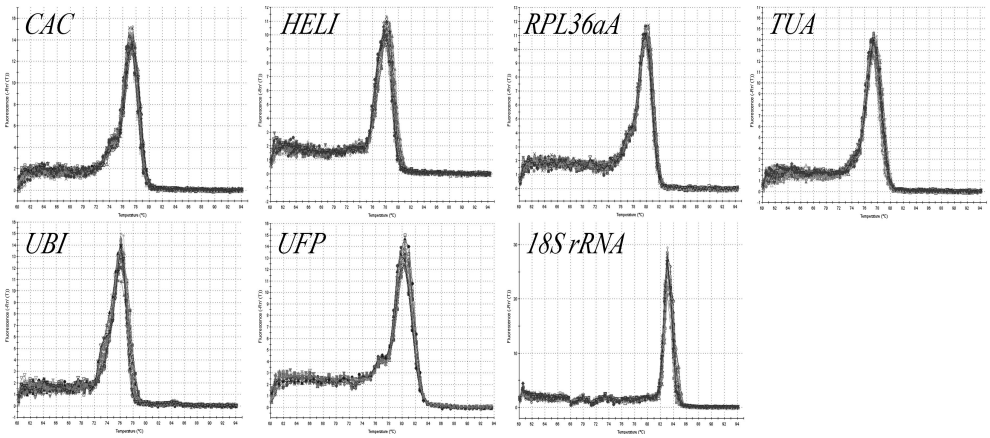



**Figure S2.** Dissociation curves of six representative reference genes showing single peaks.

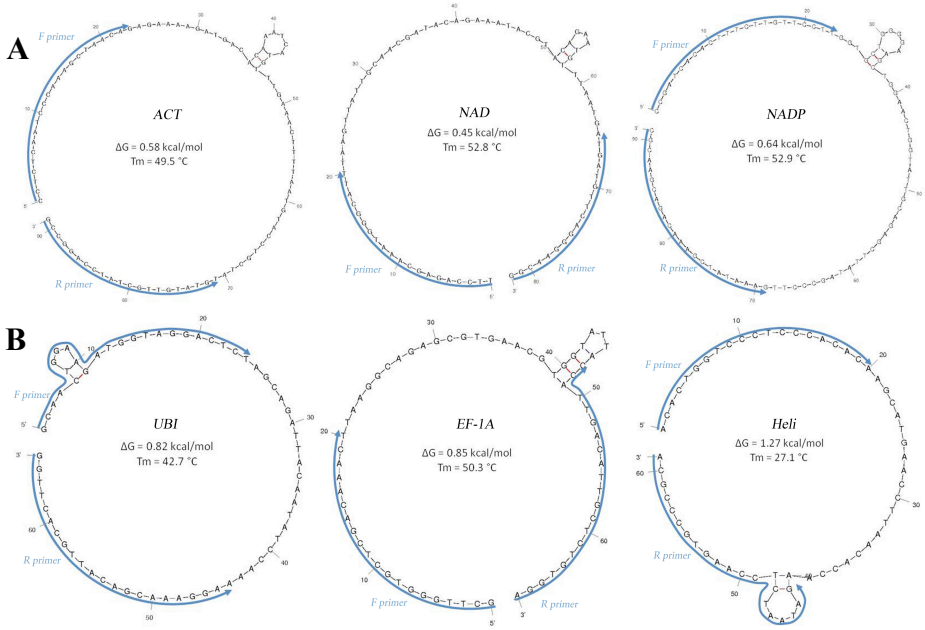



**Figure S3.** Theoretical two-dimensional secondary structures of the amplicons generated by two different primes (*NAD*, *EF-1A*). Primers are indicated by the arrows. (A) No secondary structure is present in the region where the primers anneal. (B) The secondary structure between the primer annealing sites has very small negative deltaG value, and hence does not influence the amplification efficiency.

# CAPÍTULO III

# Carotenogenic gene expression in *Cucurbita pepo* and lycopene epsilon cyclase involvement in carotenoid accumulation during fruit development

Ángeles Obrero[1], Clara I. González-Verdejo[1], Jose V. Die[2], Pedro Gómez[3], Mercedes Del Río-Celestino[3], Belén Román[1]

[1]IFAPA, Centro Alameda del Obispo, Área de Mejora y Biotecnología, Apdo. 14080 Córdoba, Spain.

[2]CSIC, Instituto de Agricultura Sostenible, Mejora Genética Vegetal, Apdo. 14080 Córdoba, Spain.

[3]IFAPA, Centro La Mojonera, Área de Mejora y Biotecnología, Autovía del Mediterráneo sal. 420. E-04745. La Mojonera. Almería. Spain.

## Abstract

The control of gene expression is a crucial regulatory mechanism in carotenoid accumulation of fruits and flowers. To date, there is still a lack of knowledge of transcriptional regulation in *Cucurbita pepo*. We investigate the role of transcriptional regulation of two isoprenoid and nine genes involved in carotenoid biosynthesis pathway in three varieties with evident differences in fruit color. Main carotenoids were measure by HPLC; exocarp of green variety showed the highest carotenoid content while skin of white fruit showed the lowest. All genes were expressed and only *PSY*, *ZDS*, *LCYe* and *CHYb* at ripe stage of white variety showed a very low expression compared with colored varieties. The expression *DXR* and *LCYe* gene exhibited a positive correlation with lutein accumulation during fruit development in flesh of yellow-orange fruit. The pronounced differences in carotenoid content in the fruits of these cultivars were not reflected in the expression levels of carotenogenic genes. This suggests that other regulatory mechanisms, in addition to gene expression, might be involved in the color absence of the white fruit variety. The results presented in this study are a first step to elucidate carotenogenesis in *C. pepo* and shows an important role of *LCYe* gene in the regulation of carotenoid accumulation during fruit development.

*Keywords*: carotenoid, *Cucurbita pepo*, lycopene ε-cyclase, lutein, transcriptional regulation

## 1. Introduction

Carotenoids are naturally occurring pigments widely present in organisms as diverse as fish, invertebrates, birds and even algae, bacteria and fungi[1]. In plants, they are responsible for the external and internal coloration of organs such as fruits, roots, flowers, and are synthesized in the plastids. In chloroplasts of green tissues, carotenoids are known to function as light-harvesting pigments and are essential for protecting tissues against photo-oxidative damage[2]. Nevertheless, in chromoplasts of fruit and flowers, they can be considered as secondary metabolites having an important role to provide different colors to tissues in order to attract pollinators[3,4]. The health benefits of carotenoids are related to its antioxidant activity, involved in removing reactive oxygen species, which may cause oxidative damage to cells[5]. They have a high potential in alleviating age-related diseases, besides some carotenoids such as β-carotene are precursors of vitamin A and are essential for normal vision[6].

Carotenoids cannot be synthesized by humans therefore they must be acquired through diet. Carotenoid engineering in crops is a potential tool to increase their content in the diet[7,8], hence it is important to understand the molecular mechanisms regulating carotenoid biosynthesis. In this regard, many advances on genetic manipulation have been achieved in several crops such as tomato, rice and potato[9-11].

Carotenoids are 40-carbon tetraterpenes built from eight 5-carbon isoprenoid units joined in such a way that the order is reversed at the centre[12]. Hydrocarbon carotenoids are collectively termed carotenes; those

containing oxygen are called xanthophylls. In plants, carotenoids are synthesized in the plastids, produced via 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway[13]. The first steps in the MEP pathway are regulated by 1-deoxyxylulose-5-phosphate synthase (DXS) and 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR), while the last committed step is regulated by 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase (HDR). This metabolic pathway produces the substrate for the carotenoid biosynthesis pathway and has been reported to influence carotenoid content[14,15]. In the carotenoid biosynthetic pathway, the first committed step is the condensation of two molecules of geranyl geranyl pyrophosphate (GGPP) to form phytoene, catalysed by the enzyme phytoene synthase (PSY). Desaturation and isomerisation of phytoene ocurring in higher plants is catalysed by phytoene desaturase (PDS), ζ-carotene desaturase (ZDS) and two isomerases, ζ-carotene isomerase (ZISO) and carotene isomerase (CRTISO), leading to the formation of lycopene. At this step the carotenoid pathway branches at the cyclization of lycopene. The lycopene β-cyclase (LCYb) catalyses the formation of β-carotene introducing two β-rings into lycopene, while lycopene ε-cyclase (LCYe) can only incorporate one β-rings and need LCYb activity to form α-carotene[16,17]. α-carotene is converted in two steps to lutein, while β-carotene after two hydroxylation steps is converted to zeaxanthin[18,19].

*Cucurbita pepo* L. is a monoecious plant divided within eight morphotypes grouped into two subspecies: subsp. *pepo* L. (Pumpkin, Vegetable Marrow, Cocozelle and Zucchini) and subsp. *ovifera* L. (Scallop, Acorn, Crookneck and Straightneck)[20]. *C. pepo* is one of the most variable genera within the *Cucurbitaceae* family and is the specie with the

greatest monetary value of the genus (21, 22). This polymorphism is particularly evident in fruit: size, shape, texture and color. As in many fruits or vegetable, carotenoids in *C. pepo* contribute to a wide range of colors and concentrations of these substances vary among species and cultivars. Although the green skin variety is the most commercial and well known, *C. pepo* shows a considerable variation in fruit color such as green, yellow and orange, ranging in intensity and shading from almost black to almost white. Moreover, accumulation of carotenoid in *C. pepo* during fruit development can lead to changes in pigmentation[23,24]. Several studies about the regulation of carotenoid during fruit development have been carried out, showing that the regulation of transcriptional level of carotenogenic genes is a principal factor controlling carotenoid acumulation[14, 25-27].

To date, although carotenoid accumulation and content have been studied in *C. pepo*[24,28] the transcriptional analysis of the biosynthetic pathway genes has been only reported in winter squashes; *C. maxima*, *C. moschata* and two lines of their interspecific inbred lines[29]. In this study, we have studied the carotenogenic gene expression profile in three *C. pepo* genotypes with different fruit peel color: white, green and yellow-orange. Two of these genotypes, Zucchini (ssp. *pepo*) and Scallop (ssp. *ovifera*) (green and white peel color respectively) are particularly of great interest since they are the parental lines of the segregating population used to construct the first SNP-based genetic map of *C. pepo*[22]. This map has already been used to infer syntenic relationships between *C. pepo* and cucumber and to successfully map QTLs that control plant, flowering and fruit traits. The third genotype 'Parador' (Zucchini) that presents a yellow-orange color from the early stages of fruit development was included in order to study the molecular basis of an early fruit coloration. Using these varieties we have studied the changes in the gene expression profiles in different tissues

and across different fruit development stages (differentiating skin and flesh). In addition, we have analyzed carotenoid content in the same samples to determine the correlations between transcript levels and the accumulation of specific carotenoids.

## 2. Material and Methods

### *2.1 Plant material*

Three cultivars of *Cucurbita pepo* L. were selected on the basis of their contrasting fruit peel colour: UPV196 white fruit, which corresponds to *C. pepo* ssp *ovifera* 'Scallop' (COMAV), 'MU_CU16' green fruit, which corresponds to *C. pepo* ssp *pepo* (COMAV) and cv 'Parador' yellow fruit, which corresponds to *C. pepo* ssp *pepo* (Gautier). Experiments were conducted from December 2010 to July 2011 under standard greenhouse conditions in "La Mojonera", Almería (Spain). Twelve samples of each cultivar were collected: leaf, female flower at two stages of development (before anthesis and in anthesis) and fruits at five stages of development (ovary, three-, five-, seven- and twenty days after pollination) (Figure 1). For each fruit sample (except the first stage) exocarp tissues was cut separately into thin slices with a vegetable peeler and mesocarp was cut into small cubes after removal of seeds. The plant materials were frozen in liquid nitrogen and then stored at -80°C for RNA isolation or freeze-dried for HPLC analysis. Both carotenoids and RNA were extracted from each sample.

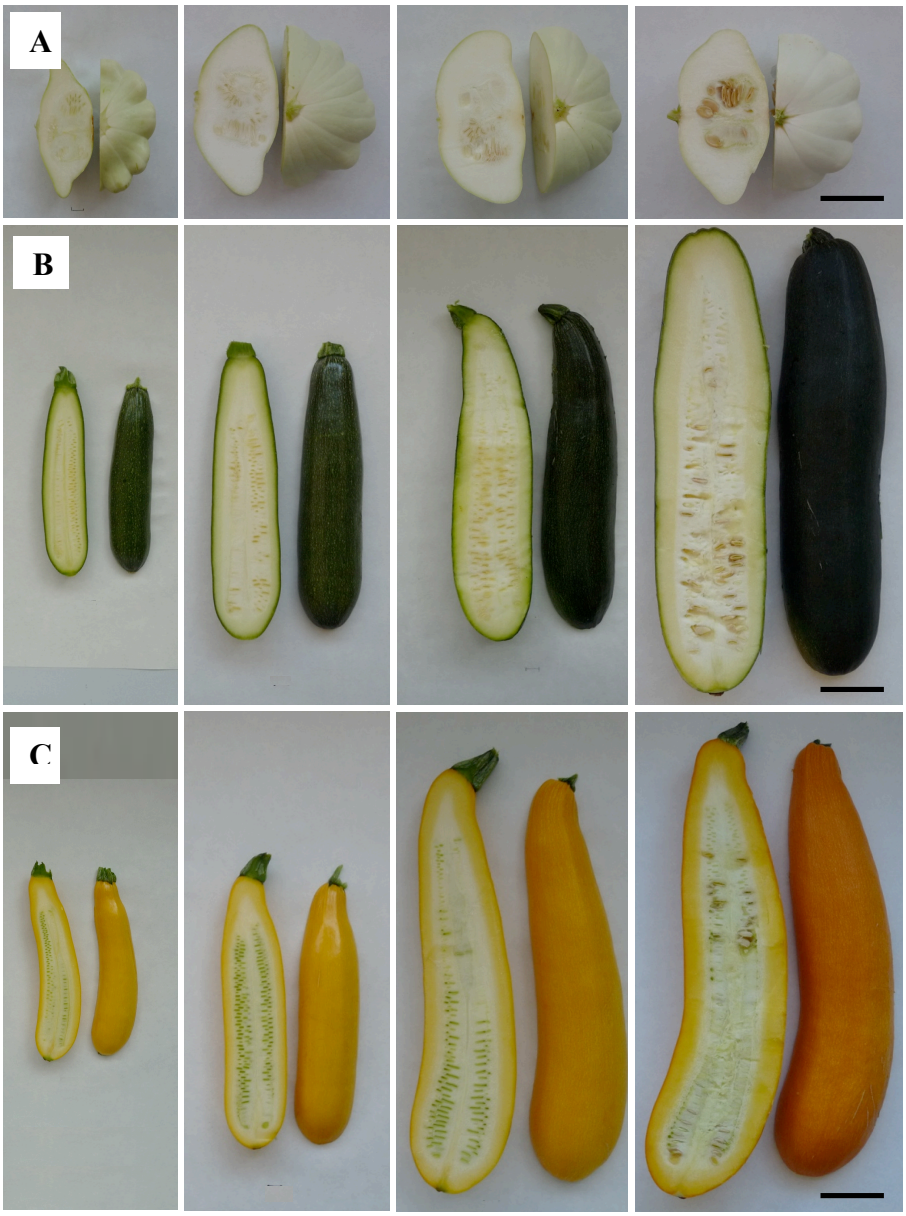



**Figure 1**. *Curcubita pepo* varieties during fruit development: (A) Scallop; (B) MU_CU16; (C) Parador. The units for scale bars represent 5 cm

*2.2. Total RNA and DNA isolation.*

All samples were ground to a fine powder in liquid $N_2$ with a pestle and mortar. Total RNA was extracted using the TRIsure reagent (Bioline, London, UK) according to manufacturer´s instructions. RNA concentration and purity were determined with NanoDrop 2000 Spectrophotometer (Thermo Scientific, Wilmington, DE) at 260nm absorbance. Only RNA samples with 260/280 ratio between 1.9 and 2.1 and 260/230 ratio greater than 2.0 were used for cDNA synthesis. RNA integrity was assessed by microcapillary electrophoresis with Experion RNA StdSens Chip and the Experion bioanalyzer (BIO-RAD Laboratories, USA) showing rRNA subunits of 18S/25S peaks on the virtual gel and electropherograms. All RNA samples showed RQI values higher than 9.
Genomic DNA from *C. pepo* leaves were isolated according to DNAzol method (Invitrogen, Carlsbad, USA) and checked by standard agarose gel.

*2.3. cDNA synthesis and genomic contamination.*

cDNA was synthesized from 1 µg of total RNA for each sample using the QuantiTec Reverse Transcription Kit (Qiagen, Hilden, Germany) with a blend of oligo-dT and random primers according to the manufacturer's instructions. In this kit, genomic DNA is efficiently removed in a single step; nevertheless, we included a negative control to test for contaminating genomic DNA. This control contained all the reaction components except the reverse transcriptase (-RT). For each of the 36 RNA samples, a quantity equivalent to the cDNA was used in –RT control and the positive control (genomic DNA) i.e. 10 ng. The absence of amplification from gDNA in cDNA samples was tested by comparison of PCR products obtained from –RT, +RT and gDNA with *EF1A* primers.

The presence of an amplification product in the "-RT" control is indicative of contaminating DNA in the sample. In 18 samples no amplification was detected in any control (-RT) after 40 cycles. In the rest of samples, genomic DNA contaminating was negligible. The sample with higher gDNA contamination had a Cq value 12 cycles higher than +RT test template, therefore the "-RT" control sample contained approximately 1000-fold less target sequence (assuming 100% efficiency, 1Cq ≈ 2-fold difference in initial template amount). Since the target template in this "-RT" control would be exclusively genomic DNA, we can conclude that 0.02% (1:4096) of the amplification in the RT sample is attributable to the genomic DNA template.

*2.4. cDNA cloning of carotenoid genes.*

*Cucurbita pepo* nucleotide sequences for *DXS*, *HDR*, *PSY*, *PDS*, *ZDS*, *CRTISO*, *LCYb*, *LCYe* and *CHYb* were not available from the public genomic database, so different strategies were applied in order to obtain these sequences. *ZDS* was cloned using degenerate primers previously described[26]. *PSY*, *PDS*, *CRTISO*, *LCYb* and *LCYe* were cloned using degenerate primers designed based on the conserved regions of the corresponding genes from other higher plants or specific primers designed based on conserved regions of the corresponding ESTs from *Cucurbitaceae*. These primer sets were designed prior to the availability of a collection of 49,610 *Cucurbita* unigenes from the two *C. pepo* subspecies: subsp. *pepo* cv. Zucchini and subsp. *ovifera* cv Scallop[30]. Considering this publication, *DXR*, *HDR* and *CHYb* were cloned using specific PCR primers designed from the sequence information generated in that study. A list of degenerate primers used is presented in Table 1.

**Table 1.** Primers used for cloning cDNA

| Gene | Forward and reverse primers 5´ -3´ | $T_m$ (ºC) | Amplicon size (bp) |
|---|---|---|---|
| *PSY* | F: GTT GGG TTG ATG AGT GTY CCT<br>R: GAA GTT TCT CCA TTT ATC AGT | 50º | 241 |
| *PDS* | F: GAG GTB TTY ATT GCH ATG TCM AAG GC<br>R: TGR TCH GCW GMR ATT TCA TCA GGA AA | 60º | 624 |
| *ZDS* | F: TAA GTC HTG GCG YGT KCG YTA YAT G<br>R: ACC TGR TTB ACT TGR TCA AGY TTG CT | 55º | 613 |
| *CRTISO* | F: AGG AGT GGA CCT TCC CAA GGA<br>R: ATG CTC AGA AAG ATG CTT CCA T | 50º | 238 |
| *LCYe* | F: TGC AGG WTT GYB AYT GTT GCW TC<br>R: ATR TAV GAC CAY TCC TCY TCR TA | 60º | 375 |
| *LCYb* | F: AGT CAA AAA TGT TGC AGA AAT<br>R: CAC CAT CCC TGC TGT TCC AC | 50º | 569 |

PCR amplification of gene fragments was performed using fruit cDNA as template. The following PCR conditions were used: an initial denaturation at 94 °C for 35s, 40 cycles of denaturation at 94 °C for 35s, annealing at 50-60 °C (depending on the primers used) for 35s and extension at 72 °C for 1 min, with a final extension of 7 min at 72 °C. Taq DNA polymerase (Bioline, London, U.K.) was used in the reactions. PCR products for *PSY*, *PDS*, *ZDS*, *CRTISO*, *LCYb* and *LCYe* were separated by electrophoresis, purified using Favorgen® GEL/PCR Purification kit (Favorgen Biotech Corp., Kaohsiung, Taiwan), cloned into the pGEM-T vector (Promega, Madison, USA) and sequenced. The sequence of the amplification product for each primer pair was compared in GenBank (NCBI) using the BLASTX algorithm and searches for homologous genes in other organisms. All sequences showed 87-100% identity and E value cutoff < 3e-40. After Cucurbita unigenes publication, same sequences were compared in Cucurbit Genomic Database (CuGenDb, htt://www.icugi.org/) and accession numbers were indicated in Table 2. Sequence data without homology after blast search in CuGenDb was

designated as *LCYe* and deposited in GeneBank (accession number JX912283).

*2.5. Primer design and qPCR analysis.*

Specific primer pairs for qPCR amplification were designed to amplify products of 60-100 bp, with an optimal primer melting temperature (Tm) of 60 ºC and GC contents between 40 and 60%. MFOLD software v3.2 (http://mfold.bioinfo.rpi.edu/cgi-bin/dnaform1.cgi) was subsequently used to evaluate the possible formation of secondary structures at the sites of primer binding, using the default settings with 50mM $Na^+$, 3mM $Mg^{2+}$, and an annealing temperature of 60 ºC[31].

The transcript levels of *DXS*, *HDR*, *PSY*, *PDS*, *ZDS*, *CRTISO*, *LCYb*, *LCYe*, *CHYb*, *PP2A* and *EF1A* were analysed using qPCR. Reactions were performed in triplicate in a 96-well plate with a Mx3000P Real-Time PCR System (Stratagene, La Jolla, USA), using SYBR Green detection chemistry. Reactions were carried out in a total volume of 15 µl that contained 7.5 µl 2x iTaq Fast SYBR Green supermix with ROX (Bio-Rad, CA, USA), 3 µl of cDNA and 0.9 µl (300 nM) of each gene-specific primer. The following standard thermal profile was used for all PCR reactions: polymerase activation (95°C for 3 min), amplification and quantification cycles repeated 40 times (95 °C for 3s, 60 °C for 30s). Finally, a dissociation analysis of the PCR products was performed by running a gradient from 60 to 95 °C to confirm the presence of a single PCR product and specificity of the qPCR (Figure S1). Table 2 shows the primer sequences, amplicon sizes, melting temperatures, and PCR efficiencies.

**Table 1.** Primers sequences used for qPCR and amplicon characteristics.

| Gene symbol | Accesion Number | Forward and Reverse Primer Sequence [ 5´-3´] | Amplicon Size (bp) | Amplicon Tm (ºC) | PCR efficiency (%) | Regresion coefficient ($R^2$) |
|---|---|---|---|---|---|---|
| *PP2A* | HM594171 | TGGTAGCATCCTTTCCCAATACA<br>CATGCCCGTTCAGCTTTAGC | 67 | 78.89 | 96 | 0.999 |
| *EF-1A* | HO702383 | GCTTGGGTGCTCGACAAACT<br>TCCACAGAGCAATGTCAATGG | 67 | 79.03 | 97 | 0.996 |
| *DXR* | PU040685 | CTTGCCCATTCTATACACCATAACA<br>AAGGCGAGGCCAAGTTATTTC | 70 | 75.41 | 100 | 0.997 |
| *HDR* | PU039913 | CGCTTAGAGGATATGGAAGTCCAA<br>ACCCTTGCCAACCACTTCAA | 75 | 76.01 | 100 | 0.998 |
| *PSY* | PU055651 | TCTCAGAGATGTTGGAGAAGATGCT<br>TCTGAAATTCCTGCCTGTGCTA | 81 | 75.96 | 100 | 0.998 |
| *PDS* | PU126694 | TAACCCCGATGAACTTTCTATGC<br>CCATCTTTGAGCCATGCTTCT | 80 | 75.65 | 100 | 0.998 |
| *ZDS* | PU046841 | CCGAGTGGTTCATGTCTAAAGGT<br>GCAACTGGACCCCACATTCT | 63 | 78.08 | 99 | 0.998 |
| *CRTISO* | PU100625 | GGTTTTACCGCCTGACACAGA<br>GGGCTCCTCTAACCTTCTCCAA | 70 | 76.11 | 100 | 0.998 |
| *LCYe* | JX912283 | TCTTGGTTTGAGTAGTCCCTGTAGTC<br>GGTTGAGGTGGAAAACAATCCTT | 75 | 74.48 | 100 | 0.999 |
| *LCYb* | PU043015 | CGGTCGAGCTACCAAAGAAGTT<br>CTCTATGCGATGCCCTTTTCA | 66 | 76.12 | 100 | 0.998 |
| *CHYb* | PU058232 | CACGACGGTCTCGTTCACAA<br>TGAGCAGCAGCGACCTTTCT | 80 | 80.51 | 98 | 0.998 |

*2.6.Carotenoids determination.*

All the solvents and chemicals were obtained from commercial sources (Sigma and Merck). The carotenoid standards were obtained from Sigma-Aldrich and Extrasynthese.

The total amount of carotenoids was determined using a spectrophotometer (Thermo Fisher Scientific, model Evolution 300). Eight millilitres of ethanol were added to 200 mg of freeze-dried sample, homogenized in a test-tube followed by a 2000 g centrifugation for 5 min. This procedure was repeated until the sample became colourless. An aliquot was taken from the upper solution for measurement of optical density at 663, 645 and 470 nm in the spectrophotometer. The total carotenoid content was calculated according to the equations of Lichtenthaler and Wellbur[32].

To determine the amount of β-carotene and lutein, two hundred milligrams of finely ground samples was extracted using a slight modification of the method described by Tadmor[33]. The carotenoids were extracted from the rehydrated sample with 8 ml ethanol containing 1 mg mL-1 butylated hydroxytoluene (BHT) using a Polytron homogenizer. All steps were carried out in darkness or under gold fluorescent light to prevent possible photodegradation of products. To hydrolyze esterified carotenoids that might complicate the chromatographic determinations, we saponified the samples[34]. One millilitre of 40% w/v KOH methanolic solution was added to each tube, and the samples were saponified for 10 min at 85ºC. The samples were cooled in an ice bath, and 2 ml of ice-cold water was added. The suspensions were extracted twice with 2 ml of hexane by vigorous vortexing followed by a 2000g centrifugation for 10 min at room temperature. The upper hexane layers were pooled and evaporated to dryness in a Savant SpeedVac apparatus and resuspended.

Immediately before injection the carotenoids were dissolved in 400 μl of an acetonitrile/methanol/dichloromethane (45:20:35 v/v/v) solution, filtered through a 0.22 μm PTFE syringe filter (Millipore) directly to amber flask for HPLC analysis, and 10 μl were injected into the chromatograph. The analyses were carried out on a HPLC apparatus equipped with binary pump, in-line vacuum degasser, autosampler injector, a Waters Symmetry C18 column (4.6 mm x 150 mm, 5 μm) and a dual λ absorbance detector (model 2487), controlled by Breeze workstation. The initial mobile phase consisted of acetonitrile/methanol (97:3, v/v/v) containing 0.05% (v/v) triethylamine. We used a linear gradient of dichloromethane from 0 to 10% in 20 min at the expense of acetonitrile, and then the dichloromethane was kept constant at 10% until the completion of the runs. The flow rate was 1.0 ml/min while the column temperature was 30º C. A dual λ absorbance detector was used to detect colored carotenoids at 450 nm.

Compounds were identified by comparison of retention times, coinjection with known standards, and comparison of their UV-visible spectra with authentic standards (β-carotene and lutein). Quantification was carried out by external standardization. Full standard curves were constructed with five different concentrations for each carotenoid in triplicate. The curves passed through or very near the origin, were linear and bracketed the concentrations expected in the samples.

*2.7.Statistical analysis.*

Expression levels of the target genes were calculated using the advanced relative quantification model with efficiency correction, multiple reference gene normalization and use of error propagation rules[35]. Based on previous results, two genes (*EF1A* and *PP2A*) were selected as

reference genes to normalize[36]. Standard error was calculated using three technical repetitions.

Transcript levels as measured by qPCR were correlated with total carotenoid content in exocarp and mesocarp of fruit development in colored varieties. Pearson correlation (r) analysis was performed and tested for statistical significance using Statistix 8.0 (Analytical Software, Tallahassee, USA).

## 3. Results

*3.1 Analysis of transcript abundance.*

Transcript levels of *DXS, HDR*, *PSY, PDS, ZDS, CRTISO, LCYb*, *LCYe,* and *CHYb* genes, measured by qPCR in the 36 samples, were detected in all cases including the white fruit cultivar. Overall, transcript levels in fruit were highest in skin compared with flesh and gene expression increased during fruit development in most of the genes.

*3.2 Analysis of transcripts in different organs.*

Gene expression of the nine genes was examined in different organs for the three cultivars ('Scallop', 'MU_CU16' and 'Parador'). The analysis in flower before anthesis (FBA), flower in anthesis (FA), leaf (L), and ovary (O) showed different expression patterns according to the gene. The results are presented in Figure 2.

The lowest expression levels were found in ovary for all the genes and, regarding varieties, in the white one 'Scallop'. Only *CRTISO* gene was up-regulated in flowers of this cultivar. *DXR*, *HDR*, *PDS* and *LCYb* transcripts exhibited similar expression patterns; the highest expression levels were obtained for flower in anthesis (FA),

principally for yellow-orange fruit cultivar 'Parador' followed by green fruit cultivar 'MU_CU16'. For *PSY, LCYe* and *CHYb* genes, the highest transcript levels were detected in leaves, in the case of *CHYb* gene it was about 25-fold higher in leaf of 'Parador' than flower before anthesis in the same cultivar. *CHYb* showed prominent difference of expression values in leaf of colored fruit varieties compared with the white fruit variety.

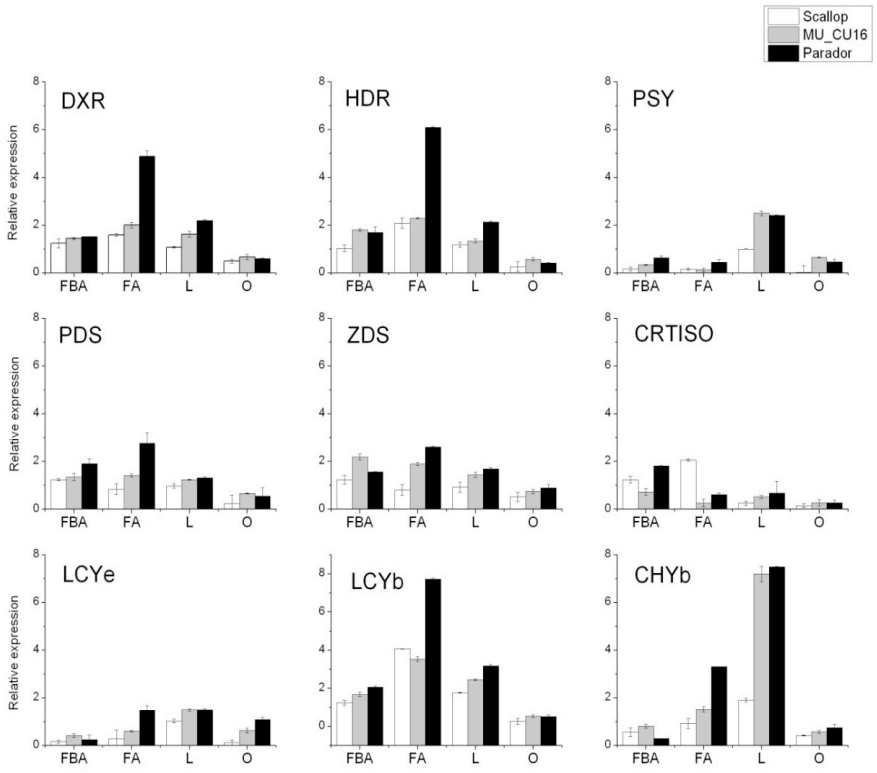



**Figure 2.** Relative expression of nine carotenoid pathway genes in different organs of the three varieties: flower before anthesis (FBA), flower in anthesis (FA) leaf (L) and ovary (O). Error bars represent the mean from three technical replicates. The expression level of *PP2A* and *EF1A* were used to normalize the mRNA levels for each sample.

*3.3 Analysis of transcripts during fruit development*

Expression levels in exocarp were higher compared with mesocarp and only *CRTISO* gene showed a similar pattern in both the skin as the flesh (Figure 3). Most of the genes showed an increasing transcript level during fruit development, being more evident in the skin than in the flesh. In the fruit exocarp (Figure 3A), *HDR*, *PDS*, *CRTISO*, and *LCYb* expression in ripe fruit (E20) was the highest for the three cultivars while *ZDS* and *LCYe* expression in ripe fruit were the highest for 'Parador' and 'MU_CU16' but not for 'Scallop' (white cultivar). At first stage of development (E3) transcripts levels were the lowest, however for *PSY* gene, expression in exocarp of 'Parador' was higher compared with 'MU_CU16' and 'Scallop', 55-fold of difference between yellow-orange and white fruit variety.

In the fruit mesocarp (Figure 3B), transcript levels of all genes increased gradually in 'Parador' variety during fruit development. The most relevant difference in transcript level was for *PSY* gene in 'Parador'. At the beginning of fruit development (3 days after pollination) transcripts were practically negligible, however at E20 stage (20 days after pollination) the difference was about 9-fold compared with fruit at E5 stage (5 days after pollination).

*3.4 Carotenoid accumulation and relationship with gene expression.*

The color changes during fruit development in *C. pepo* were different according to the variety. In 'Parador', the fruit flesh was yellow-orange in the early development stages and then turned orange at M20 stage (20 days after pollination). In 'MU_CU16' this change was opposite, fruit flesh in early development stages shows a greenish-yellow colour more intense than at M20 stage where flesh turned to a lighter color. The changes in color

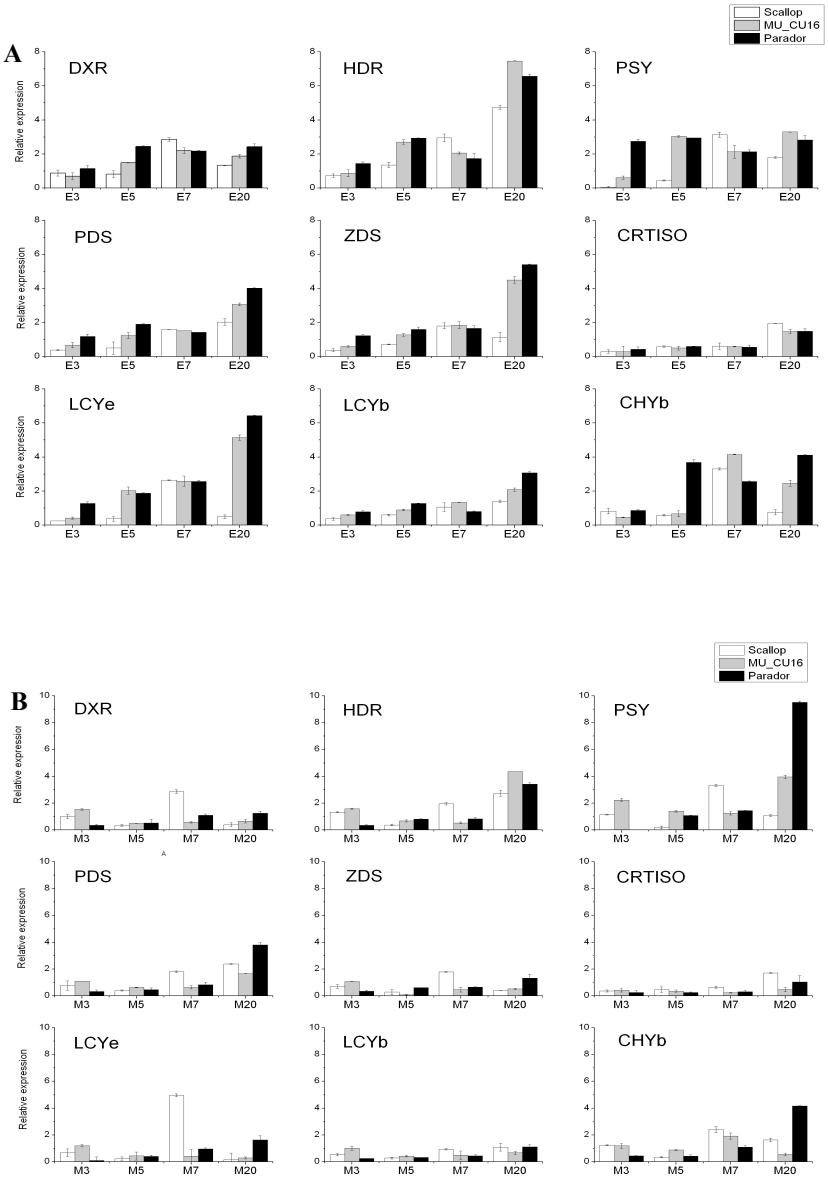



**Figure 3.** Relative expression of nine carotenoid pathway genes in exocarp (A) and mesocarp (B) of fruit development. E3, E5, E7, E20 and M3, M5, M7, M20 represent different stages (3, 4, 5 and 20 days after pollination). Error bars represent the mean from three technical replicates. The expression level of *PP2A* and *EF1A* were used to normalize the mRNA levels for each sample

flesh of white variety were more constant among development stages; in ovary, the color was greener and whiter as the fruit was developed.

The total amount of carotenoids and chlorophyll were determined using a spectrophotometer. As it is known that principal carotenoids found in *C. pepo* are lutein and β-carotene[37,28], we examined levels of these compounds in mesocarp and exocarp of all stages of fruit development and also in different organs for the three varieties by HPLC (Figure 4). The organs showed lower differences in carotenoid content among the three varieties than fruits. In flowers of the three genotypes, the principal carotenoid was lutein followed by β-carotene. In flower in anthesis (FA), lutein content increased with respect to β-carotene. In leaves (L), carotenoid content was higher for 'MU_CU16' followed by 'Parador' and 'Scallop'. (308.6 > 237.9 > 224.3 μg/g fresh weight). When comparing ovary of the three varieties, 'MU_CU16' showed the highest values (140 μg/g fresh weight).

According to the fruits, the carotenoid amount in skin of green variety 'MU_CU16' was higher than yellow-orange variety 'Parador' and carotenoid accumulation during fruit development generally increased. However, flesh of yellow-orange variety showed higher carotenoid content than the green variety, being also evident and increase of concentration along fruit development. Small amounts of carotenoids were detected in both, fruit skin and flesh of white cultivar 'Scallop'.

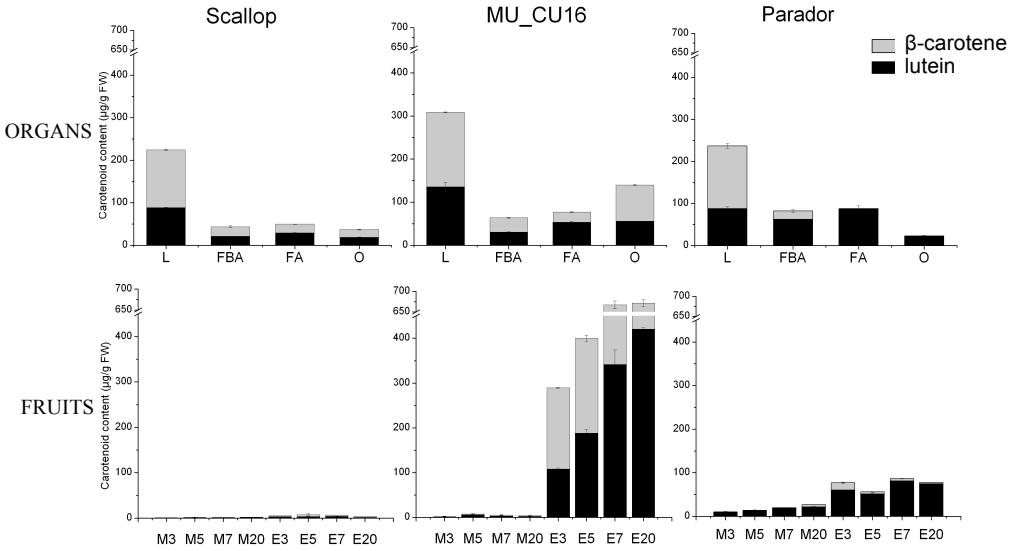



**Figure 4.** Carotenoid content in the three varieties: Scallop (white), MU_CU16 (green), Parador (yellow-orange). Error bars represent ± SD of the three replications

The relationships between carotenoid content and expression levels were studied and Pearson correlation was performed between relative gene expression for all genes during fruit development and total carotenoid concentration. The same analysis was repeated with specific carotenoid lutein and β-carotene. The correlation analysis were highly significant for *DXR* and *LCYe* expression in 'Parador' of fruit mesocarp ($r = 0.98$, $P < 0.02$; $r = 0.96$, $P < 0.04$, respectively). In exocarp of green fruit variety the highest correlation was shown for the genes *LCYb* and *LCYe* when were analyzed with lutein ($r = 0.96$, $P < 0.03$; $r = 0.93$, $P < 0.06$, respectively) (Figure 5). In white fruit variety 'Scallop', and flesh of green fruit 'MU_CU16' where total carotenoid content were low, no significant correlation were observed.

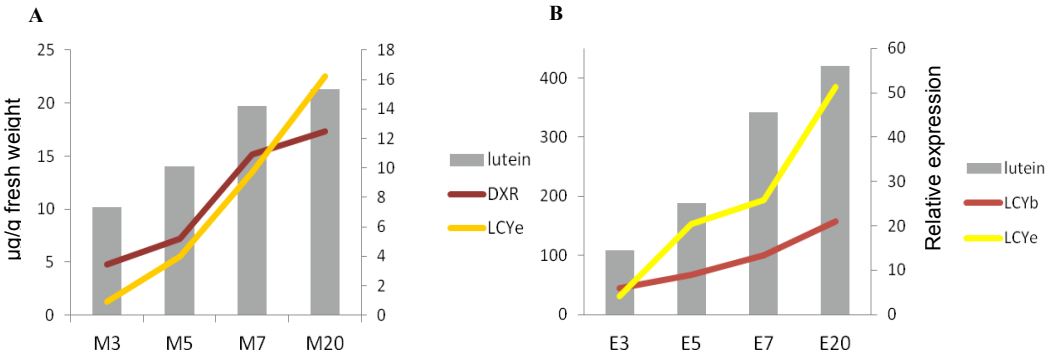



**Figure 5.** Correlation between lutein content and expression levels of *DXR, LCYb* and *LCYe* genes in fruit development. Figure (A) shows the correlation in fruit flesh of yellow-orange variety 'Parador'. Figure (B) shows the correlation in fruit skin of green variety 'MU_CU16'.

## 4. Discussion

The intensive colors of some flowers, fruits and leaves are due to a combination of various pigments including carotenoids[4]. *C. pepo* shows a significant variation in colour between different varieties, suggesting that mechanisms regulating the accumulation of carotenoids vary among genotypes. With the aim of understanding which mechanisms are regulating carotenogenesis in *C. pepo*. we have analyzed carotenoid expression of two isoprenoid and seven carotenoid genes in different organs of *C. pepo* and fruit developmental stages. The measurement of the principal carotenoids and its correlation with expression patters were also analyzed.

As expected, according to the peaks from the chromatograms, the majority of the carotenoids in all organs were lutein and β-carotene. The high levels of these two carotenoids are similar to those described in other

fruit crops and flowers such as mango, papaya, kiwifruit or japanese morning glory[38-40]. While the white color of the variety 'Scallop' is due to the absence of carotenoids, yellow-orange color of skin in 'Parador' is mainly attributed to lutein. In green fruit variety the high chlorophyll content could explain the intensive green color in the skin, masking the elevated content of lutein and β-carotene. The larger quantitative differences were detected among the fruits of the three varieties. The highest levels of these compounds were found in the skin of the green variety 'MU_CU16' following by leaves where are essential for photosynthesis. In flowers before anthesis of the three varieties, the principal carotenoid was lutein following by β-carotene. When the flowers open, lutein increases, while β-carotene content decreases. In contrast to leaves and flowers, the varieties differ dramatically in the accumulation of carotenoids in their fruits, especially in the skin. Only skin of green variety showed higher levels of carotenoid than flowers and leaves.

Regarding correlation between carotenoid content and expression levels, the higher amount of carotenoids in the exocarp compared to the mesocarp of the three varieties is correlated with the expression levels of the analysed genes. Flesh of fruit species usually showed lower total carotenoid content than the skin and expression levels concomitantly decreased in the flesh[26, 41]. In this study, for the majority of the genes analyzed, expression levels in leaves and flowers were also consistent with high carotene content in these organs. Striking results were observed for transcripts levels of *PSY* gene. Although high levels of carotenoids were obtained in flowers, very low transcripts levels were found in this organ compared with the rest of the pathway genes. Moreover, higher expression levels for this gene were obtained in ripe flesh of the orange variety. Taking into account that *PSY* is generally accepted as the most important

regulatory enzyme in the pathway and their expression can be tissue-specific[15], these results point out, as in other plant species, the existence of different *PSY* in summer squash in which the expression pattern differs according to the tissues. In tomato, high levels of expression were found in red and pink fruit for *PSY1*, however, for *PSY2* higher levels were found in petals and green tissues like leaves and sepals[42]. In other *Cucurbitaceae* such as *Cucumis melo*, *PSY1* and *PSY2* exhibit distinctive expression patterns in different tissues as well as during melon fruit development, being *PSY2* transcript higher in root[43]. Moreover, in squashes, the expression of the *PSY1* gene was correlated with the amount of carotenoid present suggesting that *PSY1* activity is related with final carotenoid concentration in ripe fruit[29].

Considering the control of carotenoid accumulation in the skin and flesh of *C. pepo* fruit, it should be noted that transcripts of all genes analyzed were detected across all samples, including both tissues of the white variety where carotenoid content was almost inexistent. In the case of the skin, *LCYe*, *ZDS* and *CHYb* presented the highest difference in the relative expression when comparing the colored and the white varieties at 20 days after pollination (E20), coinciding with a higher content of lutein. According to the flesh, although the color variation among the three cultivars is not so evident than in the skin and differences in the level of transcripts is not so marked, significant correlation values have been found for carotenoid content and expression levels along developmental stages for the genes *LCYe* and *DXR* in the yellow-orange variety 'Parador', indicating a possible implication of these genes in the trait. In skin of green fruit variety 'MU_CU16' high correlation was found for *LCYb* and *LCYe*. These results suggest that different regulatory mechanisms may be controlling carotenoid accumulation in skin and flesh. Moreover, it could

also be deduced that *LCYe* can be mainly responsible for the regulation of carotenoid accumulation in fruit of *C. pepo* since its expression seems to be related to the trait in both tissues. Considering the accumulation of lutein detected in *C. pepo* fruit, it should be mentioned that similar results have been found in fruits that also accumulate lutein. In this sense, high expression of the gene encoding *LCYe* has been found in yellow cultivar of carrot being consistent with the accumulation of lutein[44]. In citrus fruits, lutein accumulation and high expression of *LCYe* during the green stage in the flavedo has been reported[26,45]. In apple, the expression of *LCYe* was also highly correlated with carotenoid content in the skin of the fruit and was down-regulated during fruit development when lutein content was reduced[46]. In *C. pepo*, both in the skin of green fruit variety as in flesh of yellow-orange variety, accumulation of lutein increases during fruit development while *LCYe* is up-regulated. Other studies based on the genetic manipulation of *LCYe* in plants supports the hypothesis that lutein composition is largely rate-determined by *LCYe* expression[47,48]. Apart from the *LCYe* gene, *DXR* and *LCYb* expression also has shown a significant positive correlation with lutein content and subsequently can be also be involved in the regulation process.

As well as in the study of carrot[44], the differential transcriptional regulation of carotenogenic genes is not the principal cause of the absence of pigmentation of the white variety. Apart from the transcript regulation level of the carotenoid biosynthesis pathway, other regulation mechanisms at different level of the main pathway could explain the absence of carotenoids. The fact that transcripts of two representative genes of the MEP pathway *HDR* and *DXR* were detected suggests a possible accurate availability of substrates; therefore, the absence of precursors could not explain the lack of carotenoid in the the white variety. Other genes involved in

carotenoid degradation as well as other regulatory mechanisms could explain the differential accumulation. In white chrysanthemum and peach the differential expression of carotenoid cleavage dioxygenases (*CCDs*) are likely to be the major determinant in the accumulation of carotenoids[49,50]. Accumulation of carotenoid also can be affected by sink capacity. Chromoplasts are the major storage structures of carotenoid; in the white-fleshed of loquat the failure to develop normal chromoplasts seems to be the most convincing explanation for the lack of carotenoid accumulation[41]. In other plant species, genetic variation in carotenoid biosynthesis enzymes may explain the differences in carotenoid compounds[51,52]. Therefore, future studies could elucidate the differential content of carotenoids in fruit of *C. pepo*.

This work can be considered as a first step in understanding the mechanisms of regulation of carotenoids accumulation in summer squash. Further work will be needed to examine genes of interest and the possibility of allelic differences, studies of rate of degradation by carotenoid cleavage dioxygenases or post-transcriptional regulation and its relationships to carotenoid accumulation. Taking into account that two of these genotypes, 'MU_CU16' and 'Scallop', are the parental lines of the segregating population used to construct the first SNP-based genetic map of *C. pepo*[22], these genes could be located in that map in order to detect colocalization with carotenoid content QTLs in the species.

From a nutritional point of view and according to these results, three major sources of carotenoid can be found in *C.pepo*. The fleshy fruit pericarp normally comprises the major nutritive tissue. In this sense, the edible skin of 'Parador' and further 'MU_CU16' might provide high concentrations of lutein and β-carotene to the diet. On the other hand, it is

known that flesh of summer squash have low carotenoid content; however in flesh of yellow-orange fruit, lutein content is high compared with the others varieties. Considering these findings together with the fact that the skin of summer squash fruit is often removed when processed, the flesh of yellow-orange variety 'Parador' can be considered as an important product to be commercially enhanced in markets. In the same way, edible flowers of *C. pepo* that have been already used specially in the culinary arts for flavor and garnish can contribute to increase the nutritional value in foods.

## 5. Funding Sources

Financial support was provided by the Spanish Project INIA-RTA2011-00044-C02-01, FEDER and FSE funds. A. Obrero has a contract from INIA-IFAPA (Subprograma FPI-INIA), co-financed by FSE funds (Programa Operativo FSE de Andalucía 2007-2013_"Andalucía se mueve con Europa"). C.I. González Verdejo has a contract from INIA-CCAA (Subprogram DOC-INIA), which could be co-financed by FSE funds. We thank L. Madrid, M. Moya and N. Vicente-Dolera for helping with plant material.

## 6. References

(1) Britton, G.; Liaaen-Jensen, S.; Pfander, H., Carotenoids. Handbook. Birkhäuser Verlag: Basel-Boston-Berlin, 2004.

(2) Tracewell, C.A.; Vrettos, J.S.; Bautista, J.A.; Frank, H.A.; Brudvig, G.W. Carotenoid photooxidation in photosystem II. *Arch. Biochem. Biophys.* **2001**, 385, 61–69.

(3) Hirschberg, J., Carotenoid biosynthesis in flowering plants. *Curr. Opin. Plant Biol.* **2001**, 4, 210-218.

(4) Młodzińska, E. Survey of plant pigments: Molecular and environmental determinants of plants colors. *Acta Biol. Cracoviensia Ser. Botan.* **2009**, 51, 7-16

(5) Mittler, R. Oxidative stress, antioxidants and stress tolerance. *Trends Plant Sci.* **2002**, 7, 405-410.

(6) Fraser, P.D.; Bramley, P.M. The biosynthesis and nutritional uses of carotenoids. *Prog. Lipid. Res*. **2004**, 43, 228-265

(7) Giuliano, G.; Tavazza, R.; Diretto, G.; Beyer, P.; Taylor, M.A. Metabolic engineering of carotenoid biosynthesis in plants. *Trends Biotechnol.* **2008**, 26, 139-145.

(8) Farre, G.; Ramessar, K.; Twyman, R.M.; Capell, T.; Christou, P. The humanitarian impact of plant biotechnology: recent breakthroughs vs bottlenecks for adoption. *Curr. Opin. Plant Biol.* **2010**, 13, 219-225.

(9) Römer, S.; Fraser, P.D.; Kiano, J.W.; Shipton, C.A.; Misawa, N.; Schuch, W.; Bramley, P.M. Elevation of the provitamin A content of transgenic tomato plants. *Nat. Biotechnol.* **2000**, 18, 666-669.

(10) Ye, X.; Al-Babili, S.; Kloti, A.; Zhang, J.; Lucca, P.; Beyer, P.; Potrykus, I. Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science.* **2000**, 287, 303-305.

(11) Römer, S.; L.ubeck, J.; Kauder, F.; Steiger, S.; Adomat, C.; Sandmann, G. Genetic engineering of a zeaxanthin-rich potato by antisense inactivation and co-suppression of carotenoid epoxidation. *Metab. Eng.* **2002**, 4, 263-272.

(12) Rodriguez-Amaya, D.B. Carotenoids and food preparation: the retention of provitamin A carotenoids in prepared, processed, and stored foods. 1997 John Snow, Inc/OMNI Project, Brazil.

(13) Cunningham, F.X. Regulation of carotenoid synthesis and accumulation in plants. *Pure Appl. Chem*. **2002**, 74, 1409-1417.

(14) Alquezar, B.; Rodrigo, M. J.; Zacarias, L. Regulation of carotenoid biosynthesis during fruit maturation in the red-fleshed orange mutant Cara Cara. *Phytochemistry* **2008**, 69, 1997–2007.

(15) Cazzonelli, C.; Pogson, B. (2010) Source to sink: regulation of carotenoid biosynthesis in plants. *Trends Plant Sci.* **2010**, 15, 266-274.

(16) Cunningham, F.X.; Pogson, B.; Sun, Z.R.; McDonald, K.A.; Della Penna, D.; Gantt, E. Functional analysis of the β and lycopene cyclase enzymes of Arabidopsis reveals a mechanism for control of cyclic carotenoid formation. *Plant. Cell.* **1996**, 8, 1613-1626.

(17) Ronen, G.; Cohen, M.; Zamir, D.; Hirschberg, J. Regulation of carotenoid biosynthesis during tomato fruit development: expression of the gene for

lycopene epsilon-cyclase is down-regulated during ripening and is elevated in the mutant *Delta. Plant J*. **1999**, 17, 341 -51.

(18) Sandmann, G. Carotenoid biosynthesis in microorganisms and plants. *Eur. J. Biochem*. **1994** 223, 7-24.

(19) Tian, L.; DellaPenna, D. Progress in understanding the origin and functions of carotenoid hydroxylases in plants. *Arch. Biochem. Biophys*. **2004**, 430, 22 -9.

(20) Paris, H.S. A proposed subspecific classification for Cucurbita pepo. *Phytologia*. **1986**, 61, 133-138.

(21) Paris H: Summer Squash. In Handbook of Plant Breeding. Vegetables I. Part 4. Volume 1. Edited by Prohens J, Nuez, F. Springer; **2008**:351-381.

(22) Esteras, C.; Gómez, P.; Monforte, A.J.; Blanca, J.; Vicente-Dólera, N.; Roig, C.; Nuez, F.; Picó, B. High-throughput SNP genotyping in Cucurbita pepo for map construction and quantitative trait loci mapping. *BMC Genom.* **2012**, 13, 80

(23) Rodrigo, M.; Marcos, J.F.; Zacarías, L. Biochemical and molecular analysis of carotenoid biosynthesis in flavedo of orange (*Citrus sinensis* L.) during fruit development and maturation. *J. Agric. Food Chem*. **2004**, 52, 6724-6731.

(24) Tadmor, Y.; Paris, H. S.; Ayala, M.; Schaffer, A. A.; Lewinsohn, E. Dual role of the pigmentation gene B in affecting carotenoid and vitamin E content in squash (*Cucurbita pepo*) mesocarp. *J. Agric. Food Chem*. **2005**, 53, 9759 -763.

(25) Bramley, P. M. Regulation of carotenoid formation during tomato fruit ripening and development. *J. Exp. Bot*. **2002**, 53, 2107–2113.

(26) Kato, M.; Ikoma, Y.; Matsumoto, H.; Sugiura, M.; Hyodo, H.; Yano, M. Accumulation of carotenoids and expression of carotenoid biosynthetic genes during maturation in citrus fruit. *Plant Physiol*. **2004**, 134, 824 -37.

(27) Kang, B.; Zhao, W.-e.; Hou, Y.; Tian, P. Expression of carotenogenic genes during the development and ripening of watermelon fruit. *Sci. Hortic*. **2010**, 124, 368 -75.

(28) Azevedo-Meleiro, C. H.; Rodriguez-Amaya, D. B. Qualitative and quantitative differences in carotenoid composition among *Cucurbita moschata*, *Cucurbita maxima*, and *Cucurbita pepo*. *J. Agric. Food Chem*. **2007**, 55, 4027 -033.

(29) Nakkanong, K.; Yang, J.H.; Zhang, M.F. Carotenoid Accumulation and Carotenogenic Gene Expression during Fruit Development in Novel Interspecific Inbred Squash Lines and Their Parents. *J. Agric. Food Chem*. **2012**, 60, 5936–5944.

(30) Blanca, J.; Cañizares, J.; Roig, C.; Ziarsolo, P.; Nuez, F.; Picó, B. Transcriptome characterization and high throughput SSRs and SNPs discovery in Cucurbita pepo (Cucurbitaceae). *BMC Genom.* **2011**, 12, 104.

(31) Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **2003**, 31, 3406–3415.

(32) Lichtenthaler, H.K.; Wellburn, A.R. Determinations of total carotenoids and chlorophylls a and b of leaf extracts in different solvents. *Biochem. Soc. Trans.* **1983**, 603, 591-592.

(33) Tadmor, Y.; Larkov, O.; Meir, A.; Minkoff, M;, Lastochkin, E.; Edelstein, M.; Levin, S.; Wong, J.; Rocheford, T.; Lewinsohn, E. Reversed-phase high performance liquid chromatographic determination of vitamin E components in maize kernels. *Phytochem. Anal.* **2000**, 11, 370-374.

(34) Khachik, F.; Beecher, G.R. Separation of carotenol fatty acid esters by high-performance liquid chromatography. *J. Chromatogr. A.* **1988**, 449, 119-133.

(35) Hellemans, J.; Mortier, G.; De Paepe, A.; Speleman, F.; Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* **2007**, 8, R19.

(36) Obrero, A.; Die, J.V.; Roman, B.; Gomez, P.; Nadal, S.; González-Verdejo, C.I. Selection of reference genes for gene expression studies in Zucchini (*Cucurbita pepo*) using qPCR. *J. Agric. Food. Chem.* **2011**, 59:5402-5411.

(37) Murkovic, M.; Mulleder, U.; Neunteufl, H. Carotenoid content in different varieties of pumpkins. *J. Food Compos. Anal.* **2002**, 15, 633 -38.

(38) Gouado, I.; Schweigert, F.J.; Ejoh, R.A.; Tchouanguep, M.F.; Camp, J.V. Systemic levels of carotenoids from mangoes and papaya consumed in three forms (juice, fresh and dry slice). *Eur. J. Clin. Nutr.* **2007**, 61, 1180-1188.

(39) Ampomah-Dwamena, C.; McGhie, T.; Wibisono, R.; Montefiori, M.; Hellens, R. P.; Allan, A. C. The kiwifruit lycopene β-cyclase plays a significant role in carotenoid accumulation in fruit. *J. Exp. Bot.* **2009**, 60, 3765 -779.

(40) Yamamizo, C.; Kishimoto, S.; Ohmiya, A., Carotenoid composition and carotenogenic gene expression during Ipomoea petal development. *J. Exp. Bot.* **2009**, 61, 709-719.

(41) Fu, X.; Kong, W.; Peng, G.; Zhou, J.; Azam, M.; Xu, C.; Grierson, D.; Chen, K. Plastid structure and carotenogenic gene expression in red- and white-fleshed loquat (Eriobotrya japonica) fruits. *J. Exp. Bot.* **2012**, 63, 341 - 54.

(42) Giorio, G.; Stigliani, A.L.; D'Ambrosio, C. Phytoene synthase genes in tomato (*Solanum lycopersicum* L.)—new data on the structures, the

deduced amino acid sequences and the expression patterns. *FEBS J.* **2008**, 275, 527-535.

(43) Qin, X.; Coku, A.; Inoue, K. Expression, subcellular localization, and cis-regulatory structure of duplicated phytoene synthase genes in melon (*Cucumis melo* L.). *Planta.* **2001**, 234, 737-748.

(44) Clotault, J.; Peltier, D.; Berruyer, R.; Thomas, M.; Briard, M.; Geoffriau, E. Expression of carotenoid biosynthesis genes during carrot root development. *J. Exp. Bot.* **2008**, 59, 3563 -573.

(45) Zhang, L.; Ma, G.; Shirai, Y.; Kato, M.; Yamawaki, K.; Ikoma, Y.; Matsumoto, H. Expression and functional analysis of two lycopene b-cyclases from citrus fruits. *Planta.* **2012**, 236, 1315-1325.

(46) Ampomah-Dwamena, C.; Dejnoprat, S.; Lewis, D.; Sutherland, P.; Volz, R.K.; Allan, A.C. Metabolic and gene expression analysis of apple (*Malus x domestica*) carotenogenesis. *J. Exp. Bot.* **2012**, 63, 4497-4511.

(47) Pogson, B.J.; Rissler, H.M. Genetic manipulation of carotenoid biosynthesis and photoprotection. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2000**, 355, 1395-1403.

(48) Diretto, G.; Tavazza, R.; Welsch, R.; Pizzichini, D.; Mourgues, F.; Papacchioli, V.; Beyer, P.; Giuliano, G. Metabolic engineering of potato tuber carotenoids through tuber-specific silencing of lycopene epsilon cyclase. *BMC Plant Biol.* **2006**, 6,13.

(49) Ohmiya, A.; Kishimoto, S.; Aida, R.; Yoshioka, S.; Sumitomo, K. Carotenoid cleavage dioxygenase (CmCCD4a) contributes to white color formation in chrysanthemum petals. *Plant. Physiol.* **2006**, 142, 1193-1201.

(50) Brandi, F.; Bar, E.; Mourgues, F.; Horváth, G.; Turcsi, E.; Giuliano, G.; Liverani, A.; Tartarini, S.; Lewinsohn, E.; Rosati, C. Study of "Redhaven" peach and its white-fleshed mutant suggests a key role of CCD4 carotenoid dioxygenase in carotenoid and norisoprenoid volatile metabolism. *BMC Plant Biol.* **2011**, 11, 241.

(51) Harjes, C.E.; Rocheford, T.R.; Bai, L.; Brutnell, T.P.; Kandianis, C.B.; Sowinski, S.G.; Stapleton, A.E.; Vallabhaneni, R.; Williams, M.; Wurtzel, E.T.; Yan, J.; Buckler, E.S.. Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science.* **2008**, 319, 330-333.

(52) Welsch. R.; Arango, J.; Bär, C.; Salazar, B.; Al-Babili, S.; Beltrán, J.; Chavarriaga, P.; Ceballos, H.; Tohme, J.; Beyer, P. Provitamin A accumulation in cassava (*Manihot esculenta*) roots driven by a single nucleotide polymorphism in a phytoene synthase gene. *The Plant cell.* **2010**, 22, 3348-3356.

## Supporting Information

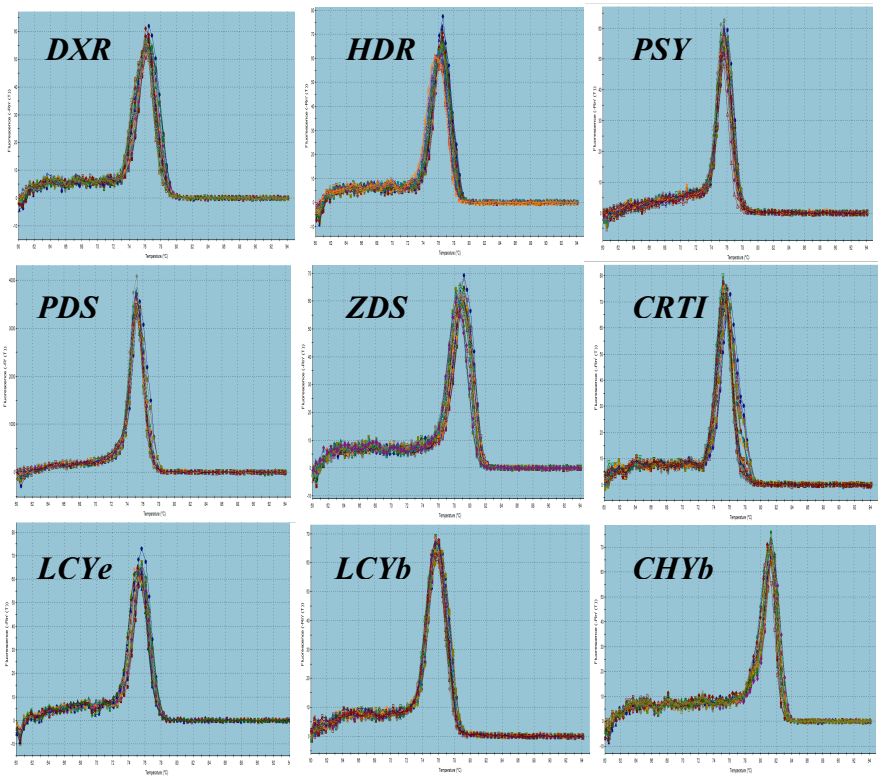



**Figure S1**. Dissociation curves of two isoprenoid and seven carotenoid pathway genes, showing single peaks.

# CAPÍTULO IV

**Characterization and expression analysis of three new phytoene synthase genes from *Cucurbita pepo***

Ángeles Obrero[1], Clara I. González-Verdejo[1], Belén Román[1], Pedro Gómez[3] , Jose V. Die[2], Charles Ampomah-Dwamena[4] ,

[1]IFAPA, Centro Alameda del Obispo, Área de Mejora y Biotecnología, Apdo. 14080 Córdoba, Spain.

[2]CSIC, Instituto de Agricultura Sostenible, Mejora Genética Vegetal, Apdo. 14080 Córdoba, Spain.

[3]IFAPA, Centro La Mojonera, Área de Mejora y Biotecnología, Autovía del Mediterráneo sal. 420. E-04745. La Mojonera. Almería. Spain.

[4]New Zealand Institute for Plant and Food Research Limited, Private Bag 92169, Auckland 1142, New Zealand

**Abstract**

The formation of phytoene through phytoene synthase enzyme (PSY) action is an essential step in the carotenoid biosynthesis pathway. Multiple *PSY* genes have been identified in many crop plants; for this reason, this phenomenon was investigated in *Cucurbita pepo*. Three new *PSYs* (*CpPSYA*, *CpPSYB*, *CpPSYC*) were cloned and characterized from three varieties of summer squash showing differences in carotenoid content. The genes sequences had high similarity with those from other plant species and their predicted proteins were significantly different from each other. Phylogenetic analysis indicated that CpPSYA and CpPSYB share homology between them and with PSYs from others cucurbits while CpPSYC is more closely related to orthologs from strawberry and carrot. Expression analysis revealed that *CpPSYA* is more highly expressed in flower compared to leaf, is regulated during fruit development and showed major differences in transcript levels between colored varieties and white variety. However, *CpPSYB* and *CpPSYC* showed higher relative expression in leaves and the expression in fruit did not show clear pattern. These results suggest that *CpPSY* genes are under different regulatory mechanisms, and they may have different roles in the various plant tissues. Future studies of *CpPSYA* could clarify the involvement of this gene in non-green tissues and the relationship with the carotenoid content of white fruit.

*Keywords*: carotenoid, *Cucurbita pepo*, phytoene synthase (PSY), transcriptional regulation

## 1. Introduction

Carotenoids, the most widespread group of pigments in nature has many structures and functions. In photosynthetic tissues the primary function is protecting the photosynthetic active pigments against photo-oxidative damage, while in nonphotosynthetic tissues like flowers and fruits the function is to attract pollinators and secure seed dispersal (Bartley and Scolnik 1995; Demming-Adams et al. 1996). Other important carotenoid functions in plants include the synthesis of regulatory molecules such as, phytohormone abscisic acid (ABA) and the strigolactones (DellaPenna and Pogson 2006). Carotenoid biosynthesis and accumulation depend on several factors like epigenetic, post-transcriptional or metabolite regulation. Recently, many studies have focused on understanding these processes with the intention of increasing levels of these compounds in the diet since they are beneficial to human health, especially in their role as biological antioxidants and vitamin A precursors (Fraser and Bramley 2004).

The plant carotenoid biosynthetic pathway is localized in the plastids and the genes and enzymes involved in this pathway have received extensive research and review (Bartley and Scolnik 1995; Cunningham and Gantt 1998; Hirschberg 2001). It is known that there are key regulatory nodes, which control the flux of metabolites through the pathway (Rodriguez-Villalón et al. 2009). In higher plants, phytoene synthase (PSY), which catalyses the first committed step in the formation of carotenoid, appears to be the most important regulatory enzyme in the pathway

(Cazzonelli and Pogson 2010). This first step is the condensation of two molecules of geranylgeranyl pyrophosphate (GGPP) to yield the first carotenoid C40 molecule, phytoene.

In Arabidopsis, a single *PSY* is present to regulate this first step in all tissues. However, in other species this regulatory step is controlled by multiple *PSYs*. In tomato, a model fruit used for the study of carotenogenesis, two *PSY* genes have been described where *PSY1* is predominant in colored tissues like mature fruit or flowers and *PSY2* in green tissues (Fraser et al. 1999; Giorio et al. 2008). In grass species such as maize, rice, and wheat there are at least two *PSYs* (Li et al. 2009). Within *Cucurbitaceae* family, such as *Cucumis melo*, two *PSYs* with different expression pattern in tissue and during fruit development have been described (Qin et al. 2011). Moreover, recent studies in wheat and *Brassica napus* confirm the high level of polymorphism of these genes (Ravel et al. 2013; Cárdenas et al. 2012)

Studies in different plant species show *PSY* gene expression levels have high influence on carotenoid accumulation. For example, during tomato fruit maturation, an increase in *PSY* transcripts is correlated with accumulation of lycopene (Bramley 2002). Similarly, an increased in *CitPSY1* transcripts of satsuma mandarin was also correlated with carotenoid accumulation during fruit maturation. Manipulation of the *PSY* activity has also resulted in increased levels of carotenoids in tobacco (Busch et al. 2002), potato (Ducreux et al. 2005) and in rice, which does not contain any carotenoids in the endosperm, but has been genetically engineered to accumulate β-carotene (Ye et al. 2000, Paine et al. 2005). It has been shown in cassava that a minor genetic alteration like an allelic polymorphism *PSY2*, can affect the catalytic efficiency of the enzyme,

resulting in an increased carotenoid accumulation in the root (Welsch et al. 2010). In addition, studies involved in association mapping and linkage analysis in peppers or wheat, have shown that *PSY* biosynthetic genes are linked to QTL for carotenoid content (Huh et al. 2001; Howitt et al. 2009). Considering these important roles, *PSY* is a preferred target for gene candidate approaches to gain insight into the molecular basis of carotenoid accumulation in fruits.

The *Cucurbitaceae* family includes numerous crops of high agronomical importance, like melons, cucumber, pumpkins, watermelons and summer squash. By comparison, research into carotenoid biosynthesis and content in summer squash is not as advanced and little is known about *PSY* genes and their potential for carotenoid improvement compared with other crop plants. Based on previous report, we selected three varieties of *Cucurbita pepo* with differences in the carotenoid content: white fruit, 'Scallop' with low content both in skin and flesh, 'MU_CU16' green fruit, with high carotenoid content in skin but low in flesh and cv 'Parador' yellow-orange fruit, with moderate carotenoid content in skin and high in flesh compared with two others. Preliminary investigation of the expression of *PSY* gene suggested the possibility of multiple enzymes in summer squash, since high levels of carotenoids were obtained in flowers, but very low transcripts levels were found in this organ compared with the rest of the pathway genes (Chapter 3).

The objectives of the present work were firstly, to identify how many *PSY* genes exist in summer squash (*Cucurbita pepo*) and secondly, to carry out a genetic expression study to understand the possible involvement of these genes in carotenoid accumulation. We have characterized three novel phytoene synthase genes in three varieties of *C.*

*pepo* grouped within two subspecies (spp *pepo* and spp *ovifera)* and analysed the differences in tissue-specific expression during fruit development.

## 2. Material and Methods

### *2.1 Plant material*

Three cultivars of *Cucurbita pepo* L. were selected on the basis of their contrasting fruit peel colour and carotenoid content: UPV196 white fruit, *C. pepo* ssp *ovifera* 'Scallop' (COMAV), 'MU_CU16' green fruit, *C. pepo* ssp *pepo* (COMAV) and cv 'Parador' yellow fruit, *C. pepo* ssp *pepo* (Gautier) (Fig. 1). Experiments were conducted from December 2010 to July 2011 under standard greenhouse conditions in "La Mojonera", Almería (Spain). Twelve samples of each cultivar were collected: leaf, flower at two stages of development (pre-anthesis and anthesis), ovary and fruits at four stages of development (three-, five-, seven- and twenty days after pollination). For each fruit sample (except the first stage) exocarp tissues was cut separately into thin slices with a vegetable peeler and mesocarp was cut into small cubes after removal of seeds. The plant materials were frozen in liquid nitrogen and then stored at -80°C for RNA isolation or freeze-dried for HPLC analysis. Both carotenoids and RNA were extracted from each sample.

### *2.2. Cloning of CpPSYA, CpPSYB, CpPSYC.*

Total RNA was extracted using the TRIsure reagent (Bioline, London, U. K.) according to manufacturer´s instructions. RNA concentration and purity were determined with NanoDrop 1000 Spectrophotometer (Thermo Scientific, Wilmington, DE) at 260nm absorvance. Total RNA from leaves, flower and fruits was treated with

Dnase I (Invitrogen) to eliminate genomic DNA present in the samples and were pooled to synthesize cDNA. cDNA was synthesized using Superscript III reverse transcriptase (Invitrogen) and two reactions were

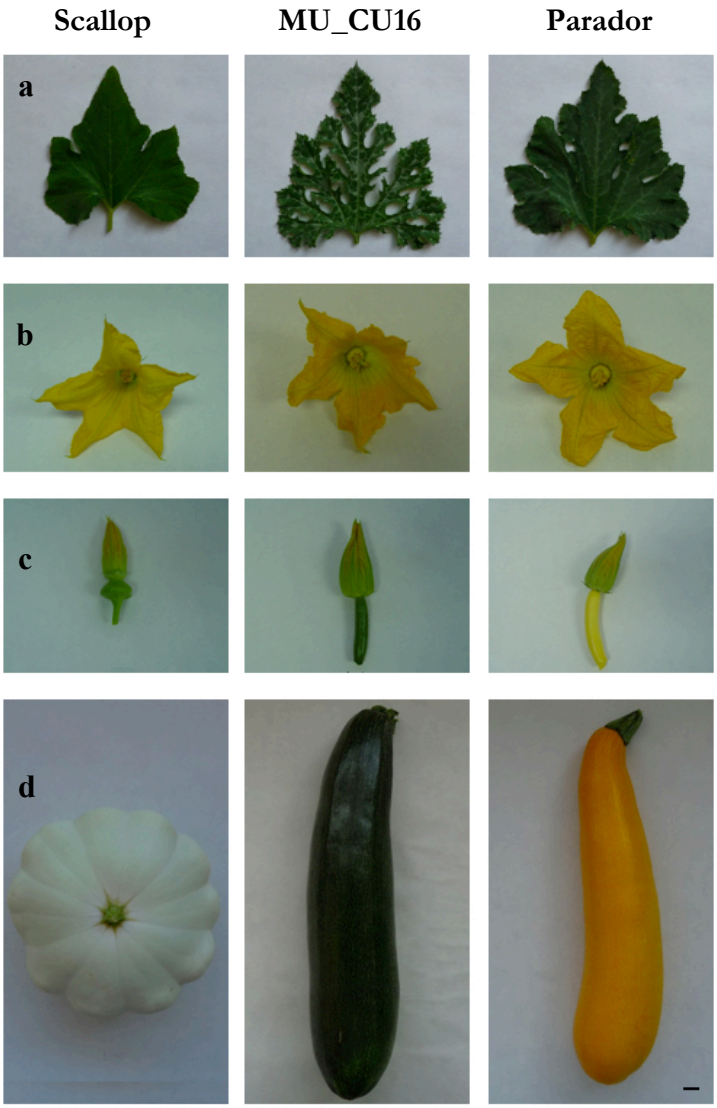



**Figure 1**. Tissues of three varieties from *Cucurbita pepo*. (**a**) leaves (**b**) flowers in anthesis (**c**) flowers before anthesis and ovaries (**d**) fruits. The units for scale bars represent 1 cm

performed separately according to the two subspecies: spp. *pepo* and spp. *ovifera*. Specific forward and reverse primer were designed in the conserved regions of *C. pepo* PSY clones from Cucurbit Genomics Database (F5OTW9201BSUAE, F5OTW9201C5BW2, F5OTW9201E4A5U, F5OTW9202HWIYN, F5OTW9202F83X0, F5OTW9202IFJ12, F5OTW9202GUGZ4, F5OTW9202GTPNY, F5OTW9201D33AL, F5OTW9202IGIWL, F5OTW9202IRKWJ, F5OTW9201CZU1X). The 3' end cDNA sequences of *CpPSY* genes were obtained using specific forward primer in combination with primers provided by GeneRacer Kit (Invitrogen). The 5' end cDNA sequences were obtained using specific reverse primer in combination with 5´-RACE System for Rapid Amplification of cDNA Ends, version 2.0 (Invitrogen). PCR 5' and 3' products were re-amplified using nested RACE PCR. Full-length cDNA copies of the *PSY* genes were generated by long distance PCR, using primers designed from the extreme 5´and 3´ ends of cDNAs. All of the cloning reactions were performed using Platinum ® Taq DNA Polymerase High Fidelity (Invitrogen). PCR products were separated by electrophoresis, purified using a DNA Clean and Concentrator kit (Zymo Research), cloned into the TOPO XL PCR Cloning Kit (Invitrogen) and sequenced. Primers used in this work are included in Table 1.

*2.3. Gene Expression Analysis.*

Total RNA was extracted using the TRIsure reagent (Bioline, London, U. K.) according to manufacturer´s instructions. RNA concentration and purity were determined with NanoDrop 2000 Spectrophotometer (Thermo Scientific, Wilmington, DE) at 260nm. Only

RNA samples with 260/280 ratio between 1.9 and 2.1 and 260/230 ratio greater than 2.0 were used for cDNA synthesis. RNA integrity was assessed by microcapillary electrophoresis with Experion RNA StdSens Chip and the Experion bioanalyzer (BIO-RAD Laboratories, USA). All RNA samples showed RQI values higher than 9.

cDNA was synthesized from 1 μg of total RNA for each sample using the QuantiTec Reverse Transcription Kit (Qiagen, Hilden, Germany) with a blend of oligo-dT and random primers according to the manufacturer's instructions. In this kit, genomic DNA is efficiently removed in a single step; nevertheless, we included a negative control to test for contaminating genomic DNA.

Specific primer pairs for qPCR amplification were designed using PRIMER3 software (Rozen and Skaletsky 2000). qPCR analysis was performed using the LightCycler system (Lightcycler 1.5, Roche). All reactions were performed using the SYBR Green master mix (Roche) following the procedure described by the manufacturer. PCR conditions were as follows: pre-incubation at 95 °C for 5 min followed by 40 cycles each consisting of 10 s at 95 °C, 10 s at 60 °C and 20 s at 72 °C. Finally, a dissociation analysis of the PCR products was performed by running a gradient from 65 to 95 °C to confirm the presence of a single PCR product and specificity of the qPCR. Each reaction sample was prepared in four technical replicates, with a negative control using water as template.

The raw data were analysed using LightCycler software version 4. Expression levels of the target genes were calculated using the advanced relative quantification model with efficiency correction, multiple reference gene normalization and use of error propagation rules (Hellemans et al. 2007). Based on previous results (Obrero et al. 2011), two genes (*EF1A* and *PP2A*) were selected as reference genes to normalize.

Table 1. Primers used for this analysis

| Primer | Sequence (5´to 3´) | Use |
|---|---|---|
| GeneRacer 3' Primer | GCTGTCAACGATACGCTACGTAACG | |
| GeneRacer 3' Nested Primer | CGCTACGTAACGGCATGACAGTG | |
| PSY_F 3' Specific Primer | GAGATGTTGGAGAAGATGCTAG | 3´RACE PCR |
| PSY_F 3' Nested Primer | GATGCTAGAAGAGGAAGA | |
| AAP 5´Primer | GGCCACGCGTCGACTAGTACGGGIIG GGIIGGGIIG | |
| AUAP 5´Primer | GGCCACGCGTCGACTAGTAC | |
| PSY_R 5´ Specific primer | AGAAATCCGAAGCTCGATAGGG | 5´RACE PCR |
| PSY_R 5´ Nested primer | CGTTAGGAAGAGACAACCAAC G | |
| PSYA_F full length | CACCCTTTTACACGATCAAACA | |
| PSYA_R full length | GTGATGTGAGTTCATTAAGGGCTA | |
| PSYB_F full length | AGT GGC CTC AAC TCA AAT GC | Full-length |
| PSYB_R full length | GCCATTGGTAGTGCCATCAT | |
| PSYC_F full length | GGTGGCTGTGAAGAGAAGATGT | |
| PSYC_R full length | ACAACGAGCGAATCGTGACA | |
| PSYA_F | GTGGGCTTCGTTGCTATTATATCG | |
| PSYA_R | GTGATGTGAGTTCATTAAGGGCTA | |
| PSYB_F | GTGGGCTTCATTGCTACTATATCGG | qPCR |
| PSYB_R | GCCATTGGTAGTGCCATCAT | |
| PSYC_F | CCTTTGGAAAAGCTGTGGTG | |
| PSYC_R | ACAACGAGCGAATCGTGACA | |

*2.4. Bioinformatics analysis.*

The nucleotide sequence and multiple sequence alignments were performed by Geneious Pro 5.5.6. Additional homology analysis was conducted using BLAST tool (http://www.ncbi.nlm.nih.gov/BLAST/) and Cucurbit Genomics Database (http://www.icugi.org/).

A phylogenetic tree was generated using the neighbor- joining method included in the CLUSTALW program and the dendogram was constructed by the MEGA5.1 program (Tamura et al. 2011). The analysis was carried out from conserved domain "Trans_IPPS_HHB" (from 138 to 415 amino acids). Bootstrap analysis of the NJ tree was performed using 1,000 replicates.

## 3. Results

*3. 1. Sequence and phylogenetic Analysis.*

Three members of *PSY* gene family (*CpPSYA*, *CpPSYB* and *CpPSYC*) were isolated from two subspecies of *C. pepo*, spp *ovifera* (accession no JX912284, JX912281 and JX912286) and ssp *pepo* (accession no JX912285, JX912282 and JX912287). Fig. 2a shows the alignment of clones generated by 3´ RACE from which three different groups of *PSYs* can be distinguished. For *CpPSYA*, the cDNA sequence obtained was 1531 pb, including 271 bp 5´untranslated region (UTR) and 90/68 bp 3´UTR (ovifera and pepo respectively). *CpPSYB*, cDNA sequence was 1236 pb, including 363 bp 5´ UTR and 63 bp 3´UTR and for *CpPSYC*, the cDNA sequence was 1173 pb, including 18 bp 5´ UTR and 166 bp 3´ UTR.

CpPSYA had a deduced protein sequence of 420 amino acids. BlastP results showed CpPSYA shared the highest homology with chloroplast phytoene synthase from *Cucurbita moschata* (97% identity). The deduced protein for CpPSYB had 412 amino acids and shared the highest homology with phytoene synthase from *Cucumis melo* (89% identity) while CpPSYC, has a deduced protein of 391 amino acids and shared the highest homology with phytoene synthase *Fragaria x ananassa* (80% identity). Alignment of the three protein sequences showed 280 identical amino acids (65.7%), CpPSYA shared higher identity with CpPSYB than CpPSYC (89% and 75% identity respectively). The major differences among three proteins were present in the N-terminus; CpPSYA and CpPSYB start with the same residues however CpPSYC is shorter than both of them and has a repeat of nine serines at the start of the protein (Fig. 2b). Because of lack genomic sequences for *C. pepo*, the three *CpPSY* genes were compared with genomic sequence from other cucurbit. The comparison of the three genes with genomic sequence of *PSY1* from *Cucumis melo* (but no *PSY2*) showed a strong similarity in the structure among all sequence (Fig. 2c). The three genes seem to conserve the six exons and five introns. The first exon of *CpPSYC* gene is shorter than *PSY1* of *C. melo*, *CpPSYA* and *CpPSYB*, however the sixth exon is longer in *CpPSYC* than in the rest of sequences. To understand the relationship between the summer squash PSY protein sequences and PSY protein sequences from a broad range of plants a phylogenetic tree was constructed (Fig. 3). The analysis showed that our proteins have a high homology to other plant phytoene synthase. CpPSYA and CpPSYB share same subclades; this group (big blue square) comprises Cucurbitaceae family (*Cucurbita pepo*, *Cucumis melo* and *Cucurbita moschata*). However CpPSYC protein clustered separately, being closely related to PSY from *Fragaria x ananassa* and PSY1 from *Daucus carota*

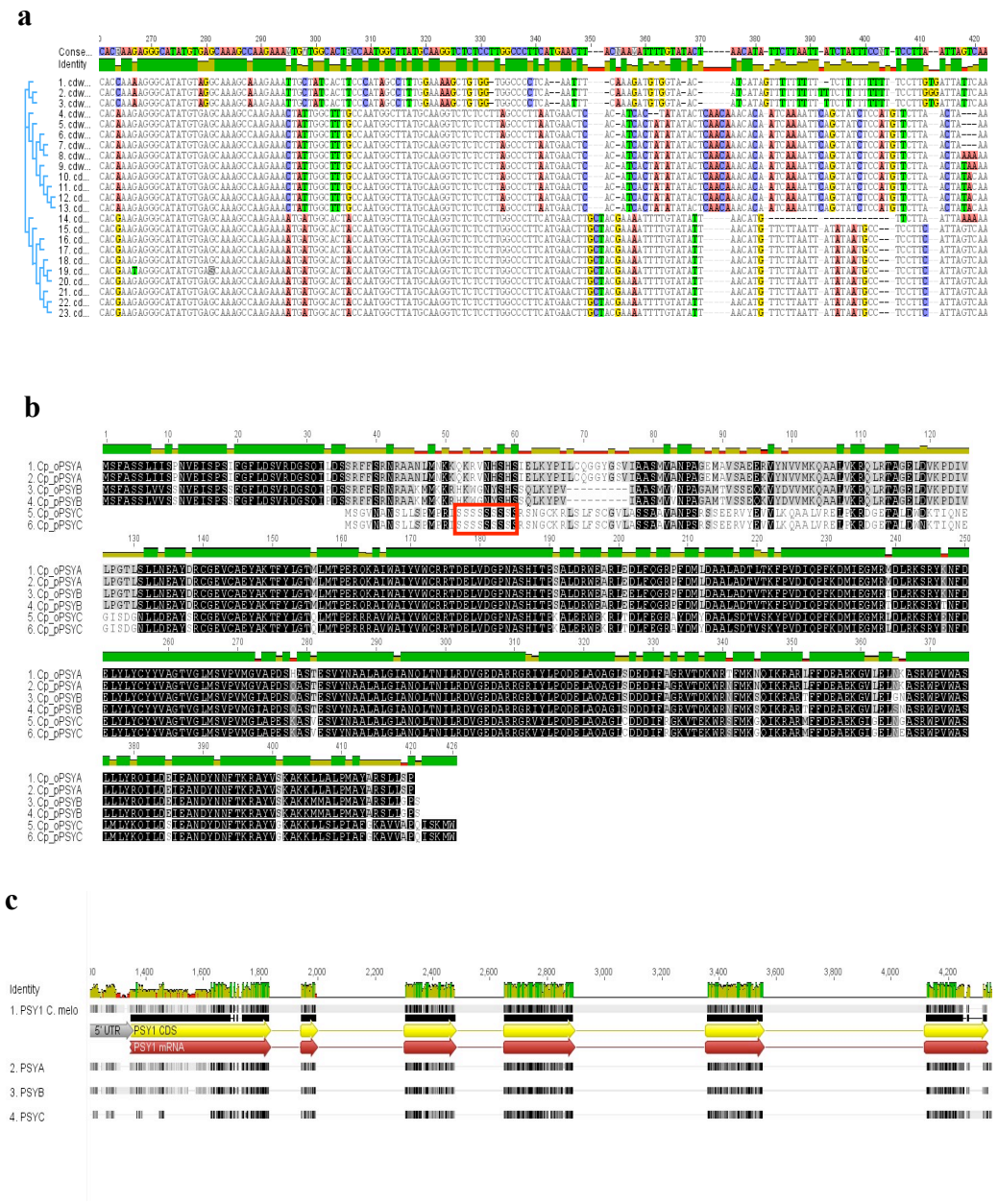



**Figure 2.** Sequence alignment of clones generated by 3´ RACE (**a**) Sequence alignment of CpPSYA, CpPSYB and CpPSYC protein. Red box indicates the repeated sequence of nine serines (b) Comparison of the structures of the *CpPSYA*, *CpPSYB* and *CpPSYC* genes with *PSY1* from *Cucumis melo* (c)

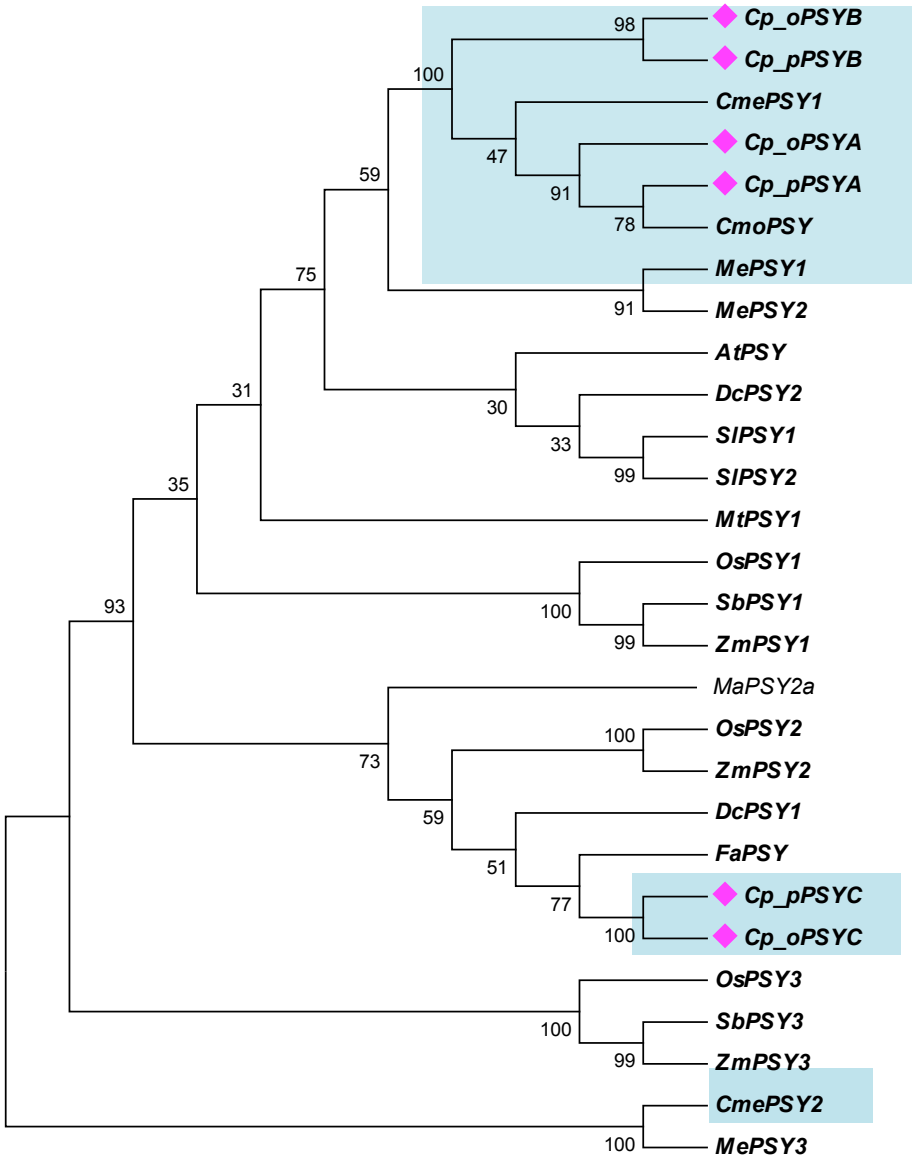



**Figure 3**. The phylogenetic tree of CpPSYs (violet diamond) and other related sequences. Numbers below the branches are the Neighbor Joining bootstrap values. Blue boxes are *Cucurbitaceae* family sequences. The abbreviation names for the PSY amino acid ssequences are as follows: rice (Os), sorghum (Sb), maize (Zm), carrot (Dc), tomato (Sl), cassava (Me), banana (Ma) melon (Cm), squash (Cmo), strawberry (Fa) and *Arabidopsis thaliana* (At)

### *3.2. qPCR analysis.*

#### *3.2.1. Transcription analysis of CpPSYA.*

In fruit skin, expression levels of 'Parador' and 'MU_CU16' increased during fruit development, showing the highest levels at 20S stage (8.9-fold higher in 'Parador' at this stage compared with same cultivar at 3S stage). Transcripts were practically absent in 'Scallop', where only were found at 7S stage (Fig. 4). In flesh, expression levels of 'Parador', as in the skin, increased during fruit development whereas 'MU_CU16' showed higher levels of transcripts at 3F and 20F stages. The relative expression of *CpPSYA* among the three cultivars differed significantly at 20F, where yellow-orange cultivar showed high expression compared with the other two individuals (9-fold and 60-fold higher in 'Parador' than 'MU_CU16' and 'Scallop' respectively). Expression pattern in 'Scallop' was very similar to skin. In the different organs, *CpPSYA* showed the highest expression levels in flowers compared with the *CpPSYB* and *CpPSYC*. 'Scallop' had lower expression levels of all three *PSYs* in all organs compared to the others cultivars (Fig. 5).

#### *3.2.2. Transcription analysis of CpPSYB*

The expression patterns for *CpPSYB* in fruit skin were different for the three cultivars examined. In 'Parador', the stages with higher transcript level were the first one and the last one (3S and 20S), in 'MU_CU16' were

middle stages (5S and 7S) and in 'Scallop' the later stages (7S and 20S). Overall in fruit flesh, expression pattern in the three cultivars were different with respect to skin but very similar to expression pattern of the flesh *CpPSYA*, the principal difference was that relative expression in *CpPSYB* was increased compared with *CpPSYA* in all stages (Fig. 4). In terms of tissue-specific expression, the transcript levels were higher in leaves than in flowers and ovary, showing flowers and ovary similar relative expression values (Fig. 5).

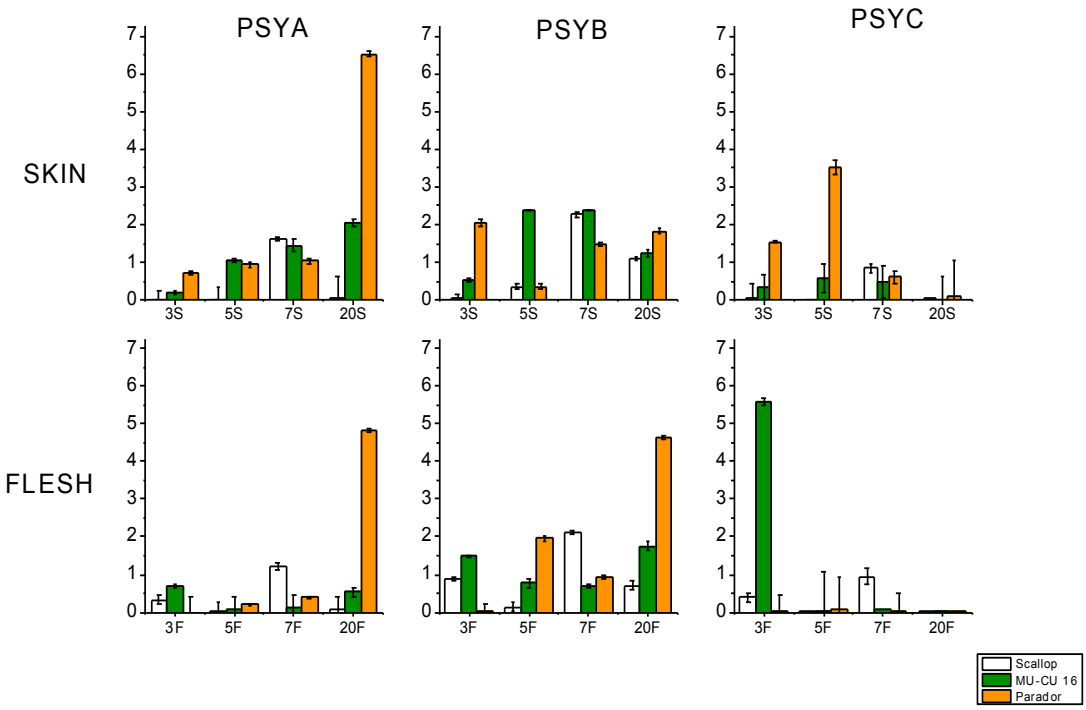



**Figure 4**. Relative expression of *CpPSYA*, *CpPSYB* and *CpPSYC* genes in skin and flesh of fruit development (3, 4, 5 and 20 days after pollination). Error bars represent the meanfrom four technical replicates. The expression level of *PP2A* and *EF1A* were used to normalize the mRNA levels for each sample

*3.2.3. Transcription analysis of CpPSYC*

The *CpPSYC* transcription pattern was quite unpredictable. In fruit skin of 'Parador' and 'MU_CU16', *CpPSYC* expression increased at the first stages but at 20S stage, there was a reduction in expression level when the fruit was ripe. In fruit flesh, there was a decrease in transcript levels in all cultivars, with the highest expression of *CpPSYC* at 3F stage in green fruit cultivar, 15.5-fold and 622-fold more than 'Scallop' and 'Parador' respectively (Fig. 4). The highest transcript level was observed in leaves from 'Parador' and 'MU_CU16' but at a reduced level in 'Scallop'. In ovary, *CpPSYC* expression was also increased in colored fruit varieties compared with the white fruit one. As with other genes, 'Scallop' showed the lowest expression levels in all organs (Fig. 5).

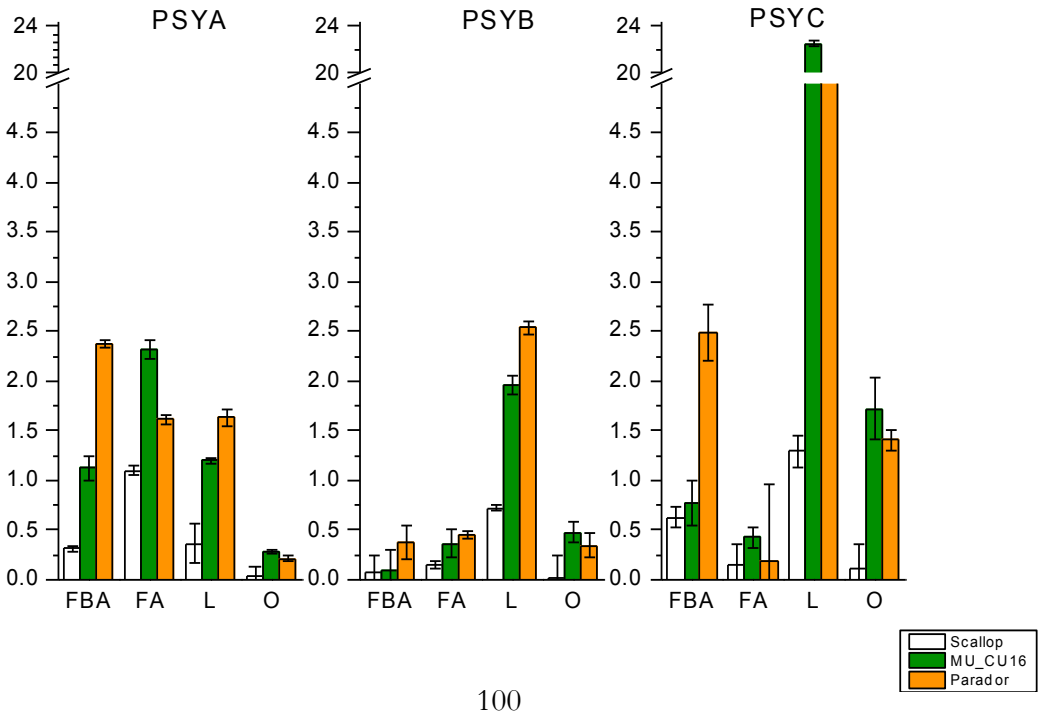


**Figure 5**. Relative expression of *CpPSYA*, *CpPSYB* and *CpPSYC* genes in different organs of the three varieties: flower before anthesis (FBA), flower in anthesis (FA) leaf (L) and ovary (O). Error bars represent the mean from four technical replicates. The expression level of *PP2A* and *EF1A* were used to normalize the mRNA levels for each sample.

## 4. Discussion

Gene identification and characterization can be utilized to define diverse functions and their relationship to fruit skin and flesh color. *PSY* is a critical regulatory enzyme within the carotenoid biosynthesis pathway and therefore plays an important role in the formation of color in flowers and fruits (Hirschberg 2001; Cazzonelli and Pogson 2010). Because of the existence of *PSY* gene families in several crop species one of the objectives of this study was to determine if different *PSY* genes are present in *C. pepo* (Bartley and Scolnik 1993; Busch et al. 2002; Li et al 2008; Arango et al. 2010; Qin et al 2011). *CpPSY* was isolated by RACE technique and we found that there are at least three *PSY* genes. Both homology and phylogenetic analysis indicated that the genes have a high degree of conservation. Comparison of the three genes with *PSY1* gene structure from other Cucurbitaceae (*Cucumis melo*) indicated that these *CpPSY* genes may have similar exon/intron structure, containing six exons and five introns, as previously described in other species (Wang et al. 2009; Rodríguez-Suárez et al. 2011; Zhao et al. 2011). Exons 1 and 6 of *CpPSYC* were significantly shorter and longer respectively than in *CpPSYA* and *CpPSYB*, resulting in a smaller predicted protein. The phylogenetic analysis

showed CpPSYA and CpPSYB are closely related, while CpPSYC is more similar to PSY from other species.

The expression analysis of *CpPSYA* gene revealed: i) Transcript levels of *CpPSYA* in flesh and skin were different between colored and white fruit varieties, which were accentuated in the mature stages (20S and 20F). ii) Transcript levels in *CpPSYA* increased during fruit development in skin of 'Parador' and 'MU_CU16' and in flesh of 'Parador'. This pattern seems to correlate with carotenoid accumulation previously measured in these tissues (Chapter 2). iii) In contrast to *CpPSYB* and *CpPSYC*, transcript levels of *CpPSYA* in flowers are higher than in leaves. *PSY* gene expression is known to be highly regulated and show tissue-specific expression pattern (Dobrowolska 2006). Several studies have shown that transcription of *PSY* genes is regulated developmentally, with elevated expression in chromoplast-containing tissue (Fraser et al 2002; Rodrigo et al. 2004). At the early stages of fruit development, carotenoids are located with chlorophyll within the chloroplasts and assist in photosynthesis. However, during ripening, chloroplasts are transformed into chromoplasts, where carotenoids accumulate to high concentrations. In tomato, *PSY1*, which is expressed in ripening fruit and petals, contributes to phytoene formation in chromoplasts while *PSY2* is expressed in all plant organs, with high levels in leaves (Bartley and Scolnik 1993; Giorio et al. 2008). Considering these points *CpPSYA* could be preferentially acting in chromoplast-rich tissues.

The expression pattern for *CpPSYB* gene was similar to *CpPSYA* in flesh fruit. However in skin, the correlation with carotenoid content was not so evident such as *CpPSYA*. For *CpPSYC,* transcription pattern was difficult to explain. The highest expression was found in leaves of

'MU_CU16' and 'Parador' and lowest in mature fruit of the three varieties. A recent study in banana cultivars, *PSY1* and *PSY2a* mRNA was detected in leaf and green fruit tissue of two cultivars with low and high carotenoid content, but not in ripe fruit suggesting an increasing degradation of these transcripts at this stage of fruit development (Mlalazi et al. 2012). Interestingly the phylogenetic tree showed that *CpPSYC* is more closely relate to *MaPSY2a* (*Musa acuminata*) than *CpPSYA* and *CpPSYB*. On the other hand, in some species where several PSYs genes have been found, there appears to be a class of PSYs induced by abiotic stress condition (Li et al. 2008; Welsch et al. 2008). *CpPSYC* does not seem to be involved in carotenoid content; future studies could elucidate whether this gene is induced under stress conditions and has a differential expression with respect to *CpPSYA* and *CpPSYB* in those conditions.

Biosynthetic genes involved in carotenogenesis have been linked to QTL for carotenoid content (Wong et al. 2004; Harjes et al. 2008). Howitt et al. (2009) identified QTLs for endosperm lutein content in wheat that colocalized with the genes encoding *LCYe* and *PSY*. In the case of the *PSY* gene, the allelic variation resulted in a change in enzymatic activity as alternative splicing of the PSY-A1 allele gave four transcripts, of which only one was functional. A study in barley also identified alternative splicing in *PSY1* (Rodríguez-Suárez et al. 2011). In addition, an amino acid change in *LCYe* that correlated with an increase in lutein, has been identified. A previous report showed that differences in fruit color among the three *C. pepo* varieties were explained by qualitative and quantitative differences in pigments accumulation. Yellow-orange flesh of 'Parador' showed the highest levels of carotenoid content compared with the flesh of

the other two varieties, lutein being the determinant factor for the color of this cultivar (Chapter 3). In that study, the *LCYe* gene expression appeared to be directly involved in lutein accumulation in this species while the expression pattern of *CpPSYA,* in this study, also suggest its involvement in the carotenoid accumulation. Therefore, both these genes may have important roles in controlling carotenoid accumulation in summer squash. It will be interesting to study their possible co-localization with the color fruit QTLs already identified in the first SNP-based genetic map of *C. pepo*.

This study was aimed at characterizing *PSY* genes in summer squash. Overall the results suggest, as in other species, that the differences observed among the three genes could be due to a functional diversification of PSY homologues to allow the accumulation of carotenoids in non- photosynthetic tissues like fruits and flowers. In addition, a better understanding about the role of these genes under stress environmental conditions could be carried out considering their implication pointed out in previous reports (Li et al. 2008; Welsch et al. 2008). Further research efforts should be directed toward analysis of the predicted proteins which could help to know what transcripts lead to functional proteins.

## 5. Acknowledgments

Financial support was provided by the Spanish Project INIA-RTA2011-00044-C02-01, FEDER and FSE funds. A. Obrero has a contract from INIA-IFAPA (Subprograma FPI-INIA), co-financed by FSE funds (Programa Operativo FSE de Andalucía 2007-2013_"Andalucía se mueve con Europa"). C.I. González Verdejo has a contract from INIA-

CCAA (Subprogram DOC-INIA), which could be co-financed by FSE funds.

## 6. References

Arango J, Wust F, Beyer P, Welsch R (2010) Characterization of phytoene synthases from cassava and their involvement in abiotic stress-mediated responses. Planta 232:1251–1262

Bartley GE, Scolnik PA (1993) cDNA cloning, expression during development, and genome mapping of PSY2, a second tomato gene encoding phytoene synthase. J Biol Chem 268:25718– 25721

Bartley GE, Scolnik PA (1995) Plant carotenoids: pigments for photoprotection, visual attraction, and human health. Plant Cell 7:1027–1038

Bramley PM (2002) Regulation of carotenoid formation during tomato fruit ripening and development. J Exp Bot 53:2107–2113

Busch M, Seuter A, Hain R (2002) Functional analysis of the early steps of carotenoid biosynthesis in tobacco. Plant Physiol 128:439–453

Cárdenas PD, Gajardo HA P, Huebert T, Parkin IA, Iniguez-Luy FL, Federico ML (2012) Retention of triplicated phytoene synthase (*PSY*) genes in *Brassica napus* L. and its diploid progenitors during the evolution of the Brassiceae. Theor Appl Genet 124:1215–28

Cunningham FX, Gantt E (1998) Genes and enzymes of carotenoid biosynthesis in plants. Annu Rev Plant Phys 49:557–583

Cazzonelli CI, Pogson BJ (2010) Source to sink: regulation of carotenoid biosynthesis in plants. Trends Plant Sci 15:266–274

DellaPenna D, Pogson BJ (2006) Vitamin synthesis in plants: tocopherols and carotenoids. Annu Rev Plant Biol 57:711–738

Demmig-Adams B, Gilmore AM, Adams WW (1996) Carotenoids 3: in vivo function of carotenoids in higher plants. FASEB J. 10:403-412.

Dobrowolska A (2006) Molecular view on the carotenogenesis in plants – a mini review. Pol J Food Nutr Sci 15:385–389.

Ducreux LJ, Morris WL, Hedley PE, Shepherd T, Davies HV, Millam S, Taylor MA (2005) Metabolic engineering of high carotenoid potato tubers containing enhanced levels of beta-carotene and lutein. J Exp Bot 56:81–89

Fraser PD, Bramley PM (2004) The biosynthesis and nutritional uses of carotenoids. Prog Lipid Res 43:228–265

Fraser PD, Romer S, Shipton CA, Mills PB, Kiano JW, Misawa N, Drake RG, Schuch W, Bramley PM (2002) Evaluation of transgenic tomato plants expressing an additional phytoene synthase in a fruit-specific manner. Pro Natl Acad Sci U.S.A. 99:1092-1097

Fraser PD, Kiano JW, Truesdale MR, Schuch W, Bramley PM (1999) Phytoene synthase-2 enzyme activity in tomato does not contribute to carotenoid synthesis in ripening fruit. Plant Mol Biol 40: 687–698.

Giorio G, Stigliani AL, D'Ambrosio C (2008) Phytoene synthase genes in tomato (*Solanum lycopersicum* L.)—new data on the structures, the deduced amino acid sequences and the expression patterns. FEBS J 275:527–535

Harjes CE, Rocheford TR, Bai L, Brutnell TP, Kandianis CB, Sowinski SG, Stapleton AE, Vallabhaneni R, Williams M, Wurtzel ET, Yan JB, Buckler ES (2008) Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. Science 319:330–33

Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol 8: R19

Hirschberg J (2001) Carotenoid biosynthesis in Xowering plants. Curr Opin Plant Biol 4:210–218

Howitt CA, Cavanagh CR, Bowerman AF, Cazzonelli C, Rampling L, Mimica JL, Pogson BJ (2009) Alternative splicing, activation of cryptic exons and amino acid substitutions in carotenoid biosynthetic genes are associated with lutein accumulation in wheat endosperm. Funct Integr Genomic 9:363–376

Huh JH, Kang BC, Nahm SH, Kim S, Ha KS, Lee MH, KimBD (2001) A candidate gene approach identified phytoene synthase as the locus for mature fruit color in red pepper (*Capsicum spp.*). Theor Appl Genet 102:524–530

Li F, Tsfadia O, Wurtzel E (2009) The phytoene synthase gene family in the Grasses: subfunctionalization provides tissue-specific control of carotenogenesis. Plant Signal Behav 4:208–211

Li F, Vallabhaneni R, Wurtzel ET (2008) PSY3, a new member of the phytoene synthase gene family conserved in the Poaceae and regulator of abiotic stress-induced root carotenogenesis. Plant Physiol 146:1333–1345

Mlalazi B, Welsch R, Namanya P, et al. (2012) Isolation and functional characterisation of banana phytoene synthase genes as potential cisgenes. Planta 236:1585-1598

Obrero A, Die JV, Roman B, Gomez P, Nadal S (2011) Selection of reference genes for gene expression studies in Zucchini (*Cucurbita pepo*) using qPCR. J Agric Food Chem 59:5402-5411

Paine JA, Shipton CA, Chaggar S, Howells RM, Kennedy MJ, Vernon G, Wright SY, Hinchliffe E, Adams JL, Silverstone AL, Drake R (2005) Improving the nutritional value of Golden Rice through increased pro-vitamin A content. Nat Biotechnol 23:482–487

Qin X, Coku A, Inoue K, Tian L (2011) Expression, subcellular localization, and cis-regulatory structure of duplicated phytoene synthase genes in melon (*Cucumis melo* L.). Planta 234:737–48

Ravel C, Dardevet M, Leenhardt F, Bordes J, Joseph JL, Perretant MR, Exbrayat F, Poncet C, Balfourier F, Chanliaud E, Charmet G (2013) Improving the yellow pigment content of bread wheat flour by selecting the three homoeologous copies of *Psy1*. Mol Breed 31:87-89

Rodrigo MJ, Marcos JF, Zacarias L (2004) Biochemical andmolecular analysis of carotenoid biosynthesis in flavedo of orange (*Citrus sinensis* L.) during fruit development and maturation. J Agric Food Chem 52:6724–6731

Rodriguez-Suarez C, Atienza SG, Piston F (2011) Allelic variation, alternative splicing and expression analysis of *Psy1* gene in *Hordeum chilense* Roem. et Schult. PLoS One 6:e19885

Rodriguez-Villalon A, Gas E, Rodriguez-Concepcion M (2009) Phytoene synthase activity controls the biosynthesis of carotenoids and the supply of their metabolic precursors in dark-grown Arabidopsis seedlings. The Plant Jour 60:424–435

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

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. Mol Biol Evol 28:2731–2739

Wang J, He X, He Z, Wang H, Xia X (2009) Cloning and phylogenetic analysis of phytoene synthase 1 (*Psy1*) genes in common wheat and related species. Hereditas 146: 208–256.

Welsch R, Wust F, Bar C, Al-Babili S, Beyer P (2008) A third phytoene synthase is devoted to abiotic stress-induced abscisic acid formation in rice and defines functional diversification of phytoene synthase genes. Plant Physiol 147:367–380

Welsch R, Arango J, Bär C, Salazar B, Al-Babili S, Beltrán J, Chavarriaga P, Ceballos H, Tohme J, Beyer P (2010) Provitamin A accumulation in cassava (*Manihot esculenta*) roots driven by a single nucleotide polymorphism in a phytoene synthase gene. Plant Cell 22:3348– 3356

Wong JC, Lambert RJ, Wurtzel ET, Rocheford TR (2004) QTL and candidate genes phytoene synthase and zeta-carotene desaturase associated with the accumulation of carotenoids in maize. Theor Appl Genet 108:349–359

Ye X, Al-Babili S, Kloti A, Zhang J, Lucca P, Beyer P, Potrykus I (2000) Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science 287:303–305

Zhao D, Zhou C, Sheng Y, Liang G, Tao J (2011) Molecular Cloning and Expression of Phytoene Synthase, Lycopene Beta-cyclase, and Beta-carotene Hydroxylase Genes in Persimmon (*Diospyros kaki* L.) Fruits. Plant Mol Biol Rep 29:345–351

# CAPÍTULO V

## 1. Discusión general

Desde que empezara el interés por conocer los mecanismos moleculares implicados en la acumulación de carotenoides, un primer paso y foco de muchos trabajos ha sido el estudio de la transcripción de los genes de la ruta biosintética. Se sabe que variaciones en la expresión de estos genes coinciden con cambios en el contenido de carotenoides durante el desarrollo de órganos que acumulan estos pigmentos, como frutos y flores (Kato et al. 2004; Yamamizo et al. 2010). Por este motivo, el punto de partida de este trabajo ha sido comprobar si los genes carotenogénicos son responsables, al igual que ocurre en otros cultivos, de la variación en el contenido de carotenoides en *Cucurbita pepo*. Aunque la regulación transcripcional de los genes de la ruta de biosíntesis es un factor fundamental que determina el contenido y composición final de carotenoides, existen evidencias de la regulación por parte de otros mecanismos muy diversos, que incluyen el control de la regulación postranscripcional, control epigenético, control del flujo metabólico a través de mecanismos de retroalimentación, así como el control en la disponibilidad de sustrato, acumulación, almacenamiento y degradación de estos compuestos (Cazzonelli y Pogson 2010). Para comenzar el estudio los genes de la ruta biosintética de carotenoides en *C. pepo* es imprescindible disponer de material vegetal adecuado en el que se incluyan variedades contrastantes en cuanto a contenido en carotenoides, así como de información genética de los genes de interés.

En los tejidos verdes de las plantas, la composición de carotenoides es la típica de los cloroplastos donde son esenciales para realizar la fotosíntesis, siendo generalmente luteína y β-caroteno los mayoritarios. La

mayoría de los frutos adquieren un color característico dependiendo del carotenoide dominante que lo compone. Así, la coloración roja del tomate y la sandía se debe fundamentalmente al licopeno (Lois et al. 2000; Lewinsohn et al 2005) mientras que el color naranja de la zanahoria se debe a la abundancia de β-caroteno (Nicolle et al. 2004). Los frutos que presentan coloración verde como muchas variedades de calabacín, el kiwi o el pimiento presentan una composición muy similar a la observadas en las hojas verdes (McGhie y Ainge 2002; Montean 2006). A pesar de la amplia gama de variedades de calabacín, la información es escasa en cuanto al contenido y composición de carotenoides en esta especie. Algunos trabajos previos en variedades de calabacín de color verde han mostrado como principales compuestos β-caroteno y luteína (Del Río-Celestino et al. 2010). Dos variedades contrastantes para color de piel, 'MU_CU16' (ssp. *pepo*) y 'Scallop' (ssp. *ovífera*) (color de piel verde y blanco respectivamente), son particularmente interesantes por la reciente publicación de las secuencias de sus transcriptomas, además de ser las líneas parentales de la población segregante utilizada para el primer mapa genético basado en SNPs de *C. pepo*. Una tercera variedad 'Parador' (ssp. *pepo*) fue elegida por presentar coloración amarillo-naranja en la carne desde el comienzo de su desarrollo. Así pues, en 'MU_CU16' encontramos abundancia de β-caroteno y luteína, destacando grandes diferencias en cuanto a concentración de estos compuestos en el exocarpo comparado con las otras variedades. Esto nos llevó a pensar que al igual que en las hojas, la acumulación de carotenoides no es apreciable a simple vista ya que las clorofilas podrían estar enmascarando el color característico de estos compuestos otorgando un verde intenso. Sin embargo, 'Parador' donde la concentración en clorofilas es mucho menor, presenta un color amarillo-naranja principalmente atribuido a la luteína, siendo la variedad que mostró mayor concentración

de este compuesto en la carne. La determinación del contenido de carotenoides en piel en estas variedades fue de suma importancia ya que el alto contenido de estos compuestos (por tanto de nutrientes) en el cultivar verde no es apreciable a simple vista. En el cultivar blanco 'Scallop' solo se encontraron trazas de estos compuestos.

En el momento de iniciarse esta tesis doctoral las secuencias de *C. pepo* disponibles en bases de datos tales como NCBI o ICuGI eran pocas, por lo que muchas secuencias parciales de genes fueron clonadas para los distintos trabajos realizados. Una vez obtenidas estas secuencias, y tras el estudio de expresión de los genes carotenogénicos, pudimos obtener una pista sobre cuáles de ellos podrían tener un papel relevante en la regulación de carotenoides en *C. pepo* para poder enfocar estudios futuros.

En general, los resultados mostraron concordancia entre contenido en carotenoides y expresión para todos los genes. En el caso de hojas, flores y piel de frutos, unos mayores niveles de expresión se correspondieron con una mayor acumulación de pigmentos. Sin embargo en carne, los niveles relativos de expresión fueron mucho menores, coincidiendo con un notable descenso del contenido en carotenoides. Estos resultados sugieren que existen mecanismos diferentes regulando la acumulación de pigmento en ambos tejidos del fruto. Por otra parte los resultados obtenidos por Esteras et al. 2012 ponen de manifiesto que los genes responsables del color en la piel y carne se encuentran localizados en distintos grupos de ligamiento. También se observó, que los mayores cambios de expresión y acumulación de pigmentos se produjeron en el estadio maduro (20 días después de polinización). Es importante destacar que en la fase de maduración de muchos frutos (y flores) es dónde se producen los mayores cambios en cuanto a acumulación y composición de

carotenoides. Además en esta fase, la carotenogénesis está controlada por mecanismos reguladores distintos a los utilizados en los tejidos fotosintéticos (Thelander et al. 1986; Bramley 2002). Otra generalidad encontrada fue que todos los genes analizados se expresaron, incluso en el cultivar blanco. Esto nos llevó a pensar que la ausencia de carotenoides en el fruto de este cultivar podría deberse a la regulación a otro nivel distinto de la ruta biosintética principal. Un estudio paralelo de genes que codifican enzimas implicadas en la degradación de carotenoides mostró que el transcrito del gen dioxigenasa de ruptura de carotenoides *CpCCD4* presentó niveles más altos en el cultivar blanco que en los coloreados, y pudiendo, por lo tanto, estar implicado en la ausencia de color en 'Scallop'.

El análisis de los resultados de expresión de los genes de la ruta, nos hizo centrar nuestra atención principalmente en dos genes: *LCYe* y *PSY*. Cuando comparamos la piel de las tres variedades donde existen las mayores diferencias en cuanto al contenido de carotenoides, el gen *LCYe* mostró la mayor diferencia de transcritos entre las variedades de color y la blanca, sobre todo en estadio maduro. Por otra parte, observando la evolución del contenido de carotenoides conforme el desarrollo del fruto, en la piel de 'MU_CU16', *LCYe* mostró una correlación alta con el contenido en luteína, de igual modo una alta correlación entre niveles de expresión y luteína se encontró en la carne de 'Parador' donde los niveles de este pigmento son mayores con respecto a 'Scallop' y 'MU_CU16'. La variación de expresión en este gen y su relación con la acumulación de luteína ha sido descrita en varias especies como zanahoria o manzana (Clotault et al. 2008; Ampomah-Dwamena et al. 2012). Teniendo en cuenta los resultados obtenidos para este gen y la acumulación de luteína en estas variedades, podríamos deducir una posible implicación de este gen en el contenido de carotenoides en fruto de calabacín. En cuanto al gen *PSY*, los

niveles de expresión obtenidos en el desarrollo de la carne de 'Parador' fueron diferentes al resto de los genes, mostrando este cultivar los mayores niveles de transcritos en el estadio maduro.

Por otra parte, los genes que codifican fitoeno sintasa (PSY) han sido clonados en diversas plantas. En *Arabidopsis thaliana* parece estar codificada por un único gen, mientras que en otras especies puede existir más de un gen que codifique PSYs funcionales (Fraser y Bramley, 2004). La baja expresión de este gen en las flores con respecto al resto de los genes de la ruta hizo sospechar la posible presencia de genes duplicados para esta enzima. Con el fin de determinar si en *C. pepo* ha ocurrido un proceso de duplicación como ocurre en otras especies, en el tercer trabajo de la tesis se aislaron las secuencias codificantes para este gen y se obtuvieron tres enzimas diferentes, que presentaron alto grado de conservación cuando se compararon con PSY de otras especies. El análisis filogenético mostró como CpPSYA y CpPSYB se agrupaban dentro de la familia de las Cucurbitaceas, sin embargo CpPSYC se agrupaba con PSY de otras especies como fresa (*Fragaria x ananassa*) y zanahoria (*Daucus carota*). El patrón de expresión para los tres genes fue diferente, sugiriendo que distintos mecanismos intervienen en la regulación de estos genes en los diferentes tejidos. El análisis de los resultados nos permitió esclarecer cual de los tres genes podría tener una mayor implicación en el contenido de carotenoides en fruto, siendo *CpPSYA* un buen candidato para futuros estudios y a tener en cuenta en los programas de mejora de la especie. En este sentido, de las tres fitoeno sintasas descritas en gramíneas, *PSY1*, *PSY2* y *PSY3*, solo *PSY1* está relacionada con el contenido en carotenoides en el endospermo (Li et al. 2009). Así pues, muchos trabajos han estado enfocados en la estructura de este gen y en su implicación en el contenido

de carotenoides (Howitt et al. 2009; Rodríguez-Suárez et al. 2011; Ravel et al 2012).

La técnica nexo común utilizada en los distintos trabajos que hemos planteado en esta tesis ha sido el uso de la PCR en tiempo real para medir la expresión génica. La precisión de los resultados obtenidos por este método depende en gran medida de una normalización adecuada, para lo que es fundamental el uso de genes expresados de manera estable, conocidos como genes de referencia (Gutierrez et al 2008, Guénin et al 2009). Sorprendentemente, este enfoque no ha sido muy utilizado en estudios de plantas, donde la tendencia ha sido utilizar "housekeeping genes" como genes de referencia sin una validación apropiada. Además, actualmente se sabe que existen nuevos genes de referencia más estables que los utilizados tradicionalmente (Czechowski et al. 2005). En el primer trabajo de la presente tesis se han identificado un conjunto de genes estables en diferentes tejidos, estadios del desarrollo del fruto y bajo diferentes estreses. Estos genes serán de gran utilidad a la hora de plantear cualquier trabajo de expresión génica en calabacín ya que hasta el momento se han utilizado genes de referencia sin validar (p.e. 18SrRNA) demostrándose que hay "novel genes" como *CAC* o *PP2A* considerados más estables y que darán mayor fiabilidad a los resultados. Estos genes ya han sido de utilidad en varios estudios, entre los que se encuentran el tercer y cuarto capítulo donde han sido aplicados como base para el análisis de datos.

## 2. Perspectivas futuras

El estudio de los genes de la ruta biosintética de carotenoides y su regulación es un avance hacia el conocimiento de la carotenógenesis en *C. pepo*. A corto plazo, se podría continuar avanzando en el estudio de las

diferentes *PSYs* y profundizar en los mecanismos de regulación para este gen, así como el estudio de mutaciones alélicas o splicing alternativo. La expresión de este gen está regulada por la luz, por señales del desarrollo y por estreses abióticos (Bramley 2002; Lu y Li 2008; Welsch et al 2008). Además han sido identificados motivos en la región promotora que puede mediar la regulación transcripcional de este gen (Welsh et al. 2003). Por otra parte, un estudio en yuca (*Manihot esculenta*), identificó un polimorfismo de un solo nucleótido (SNP) en *MePSY2* que aumentaba significativamente la formación y acumulación de carotenoides en las raíces de yuca (Welsch et al. 2010). Además, estudios recientes sobre splicing alternativo han demostrado como los diferentes transcritos originados influyen en el contenido de carotenoides (Howitt et al. 2009; Rodriguez-Suarez 2012). Por otra parte, también se podría profundizar en otros genes de la ruta (un gen candidato podría ser la ciclasa *LCYe*) o en genes relacionados con la degradación de carotenoides.

En los últimos años, la rápida evolución de distintas herramientas genéticas para este cultivo, como la publicación del primer transcriptoma de la especie (Blanca et al. 2011), o la publicación del primer mapa genético basado en SNPs (Esteras et al. 2012) han abierto la posibilidad de explorar numerosos procesos biológicos y moleculares que a medio plazo pueden ser de utilidad para asistir a la mejora de calabacín. En muchas especies han sido caracterizados los genes involucrados en la biosíntesis de carotenoides y se ha visto como algunos de ellos están asociados a QTLs para el contenido en carotenoides (Wong et al. 2004; Chander et al. 2008). Sin embargo, hay que mencionar que existen varios ejemplos en los que los QTLs que afectan la variación de este carácter no están relacionados con

genes de la ruta (Li et al. 2001; Liu et al. 2003). La colección de marcadores codominantes específicos, la posibilidad de identificar QTLs en el mapa cuyos parentales son las líneas utilizadas en este trabajo, así como la caracterización de genes y su posible colocalización en el mapa con las regiones que controlan la variación del carácter, puede ser un avance importante para la selección asistida por marcadores en la mejora de esta especie.

## 3. Referencias

Ampomah-Dwamena C, Dejnoprat S, Lewis D, Sutherland P, Volz RK, Allan AC (2012) Metabolic and gene expression analysis of apple (*Malus* × *domestica*) carotenogenesis. Journal of Experimental Botany 63 (12):4497-4511.

Blanca J, Canizares J, Roig C, Ziarsolo P, Nuez F, Pico B (2011) Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (Cucurbitaceae). BMC Genomics 12 (1):104

Bramley PM (2002) Regulation of carotenoid formation during tomato fruit ripening and development. Journal of Experimental Botany 53 (377):2107-2113.

Cazzonelli CI, Pogson BJ (2010) Source to sink: regulation of carotenoid biosynthesis in plants. Trends in Plant Science 15 (5):266-274.

Chander S, Guo YQ, Yang XH, Zhang J, Lu XQ, Yan JB, Song TM, Rocheford TR, Li JS (2008) Using molecular markers to identify two major loci controlling carotenoid contents in maize grain. Theoret Appl Genetics 116 (2):223-233.

Clotault J, Peltier D, Berruyer R, Thomas M, Briard M, Geoffriau E (2008) Expression of carotenoid biosynthesis genes during carrot root development. Journal of Experimental Botany 59 (13):3563-3573

Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible W-R (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. Plant Physiol 139 (1):5-17.

Del Río-Celestino M, Gómez P, Villatoro-Pulido M, Moya M, Domínguez-Pérez I, Martínez-Valdivieso D, Font R, Muñoz Serrano A, Alonso-Moraga A QUANTIFICATION OF CAROTENOIDS IN ZUCCHINI (CUCURBITA PEPO L. SUBSP. PEPO) CULTIVARS CULTIVATED IN ALMERIA BY LIQUID CHROMATOGRAPHY. In: XXVIII International Horticultural Congress on Science and Horticulture for People, 2012. pp 183-187

Esteras C, Gomez P, Monforte A, Blanca J, Vicente-Dolera N, Roig C, Nuez F, Pico B (2012) High-throughput SNP genotyping in *Cucurbita pepo* for map construction and quantitative trait loci mapping. BMC Genomics 13 (1):80

Fraser PD, Bramley PM (2004) The biosynthesis and nutritional uses of carotenoids. Progress in Lipid Research 43 (3):228-265.

Guenin S, Mauriat M, Pelloux J, Van Wuytswinkel O, Bellini C, Gutierrez L (2009) Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. Journal of Experimental Botany 60 (2):487-493.

Gutierrez L, Mauriat M, Gunin S, Pelloux J, Lefebvre J, Louvet R, Rusterucci C, Moritz T, Guerineau F, Bellini C, Van Wuytswinkel O (2008) The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. Plant Biotechnology Journal 6 (6):609-618.

Howitt C, Cavanagh C, Bowerman A, Cazzonelli C, Rampling L, Mimica J, Pogson B (2009) Alternative splicing, activation of cryptic exons and amino acid substitutions in carotenoid biosynthetic genes are associated with lutein accumulation in wheat endosperm. Funct Integr Genomics 9 (3):363-376.

Kato M, Ikoma Y, Matsumoto H, Sugiura M, Hyodo H, Yano M (2004) Accumulation of Carotenoids and Expression of Carotenoid Biosynthetic Genes during Maturation in Citrus Fruit. Plant Physiology 134 (2):824-837.

Lewinsohn E, Sitrit Y, Bar E, Azulay Y, Meir A, Zamir D, Tadmor Y (2005) Carotenoid Pigmentation Affects the Volatile Composition of Tomato and Watermelon Fruits, As Revealed by Comparative Genetic Analyses. Journal of Agricultural and Food Chemistry 53 (8):3142-3148.

Li F, Tzfadia O, Wurtzel ET (2009) The phytoene synthase gene family in the Grasses: Subfunctionalization provides tissue-specific control of carotenogenesis. Plant Signaling & Behavior 4 (3):208-211

Li L, Paolillo DJ, Parthasarathy MV, DiMuzio EM, Garvin DF (2001) A novel gene mutation that confers abnormal patterns ofβ-carotene accumulation in cauliflower (*Brassica oleracea* var. botrytis). The Plant Journal 26 (1):59- 67.

Liu Y-S, Gur A, Ronen G, Causse M, Damidaux R, Buret M, Hirschberg J, Zamir D (2003) There is more to tomato fruit colour than candidate carotenoid genes. Plant Biotechnology Journal 1 (3):195-207.

Lois LM, Rodríguez-Concepción M, Gallego F, Campos N, Boronat A (2000) Carotenoid biosynthesis during tomato fruit development: regulatory role of 1-deoxy-D-xylulose 5-phosphate synthase. The Plant Journal 22 (6):503-513.

Lu S, Li L (2008) Carotenoid Metabolism: Biosynthesis, Regulation, and Beyond. Journal of Integrative Plant Biology 50 (7):778-785.

McGhie TK, Ainge GD (2002) Color in fruit of the genus *Actinidia*: carotenoid and chlorophyll compositions. Journal of Agricultural and Food Chemistry 50 (1):117–121

Muntean E, Lazar V, Muntean N (2006) HPLC-PDA analysis of carotenoids and chlorophylls from *Cucurbita pepo* L.convar Giromontiina fruits. Buletinul USAMV-CN 62:94-99.

Nicolle C, Simon G, Rock E, Amouroux P, Rémésy C (2004) Genetic Variability Influences Carotenoid, Vitamin, Phenolic, and Mineral Content in White, Yellow, Purple, Orange, and Dark-orange Carrot Cultivars. Journal of the American Society for Horticultural Science 129 (4):523-529.

Ravel C, Dardevet M, Leenhardt F, Bordes J, Joseph J, Perretant M, Exbrayat F, Poncet C, Balfourier F, Chanliaud E, Charmet G (2013) Improving the yellow pigment content of bread wheat flour by selecting the three homoeologous copies of *Psy1*. Molecular Breeding 31 (1):87-99.

Rodríguez-Suárez C, Atienza SG, Pistón F (2011) Allelic Variation, Alternative Splicing and Expression Analysis of *Psy1* Gene in *Hordeum chilense* Roem. et Schult. PLoS ONE 6 (5):e19885.

Thelander M, Narita J, Gruissem W (1986) Plastid differentiation and pigment biosynthesis during tomato fruit ripening. Current Topics in Plant Biochemistry and Physiology 5:128-141.

Welsch R, Arango J, Bär C, Salazar B, Al-Babili S, Beltrán J, Chavarriaga P, Ceballos H, Tohme J, Beyer P (2010) Provitamin A Accumulation in Cassava (*Manihot esculenta*) Roots Driven by a Single Nucleotide Polymorphism in a Phytoene Synthase Gene. THE PLANT CELL ONLINE 22 (10):3348-3356.

Welsch R, Medina J, Giuliano G, Beyer P, von Lintig J (2003) Structural and functional characterization of the phytoene synthase promoter from *Arabidopsis thaliana*. Planta 216 (3):523-534.

Welsch R, Wüst F, Bär C, Al-Babili S, Beyer P (2008) A Third Phytoene Synthase Is Devoted to Abiotic Stress-Induced Abscisic Acid Formation in Rice and Defines Functional Diversification of Phytoene Synthase Genes. Plant Physiology 147 (1):367-380.

Wong JC, Lambert RJ, Wurtzel ET, Rocheford TR (2004) QTL and candidate genes phytoene synthase and ζ-carotene desaturase associated with the accumulation of carotenoids in maize. Theoret Appl Genetics 108 (2):349-359.

Yamamizo C, Kishimoto S, Ohmiya A (2010) Carotenoid composition and carotenogenic gene expression during *Ipomoea* petal development. Journal of Experimental Botany 61 (19933319):709-719.

# CAPÍTULO VI

## Conclusiones

1. Se ha identificado un conjunto de genes estables para su aplicación en el proceso de normalización en estudios de expresión génica con qPCR y se ha demostrado que el uso de genes no adecuados puede influir en la exactitud de los resultados. Este estudio proporciona una guía de nuevos genes de referencia para ser utilizados en futuros estudios transcriptómicos en *C. pepo*.

2. El análisis de carotenoides en los frutos de los tres cultivares presentaron grandes diferencias en cuanto al contenido de estos compuestos. La mayor acumulación de carotenoides en carne, principalmente luteína, se observó para el cultivar 'Parador'. Aunque aparentemente por su color amarillo-naranja, este cultivar sugiere también tener el mayor contenido de carotenoides en piel, es el cultivar verde 'MU_CU16' el que mostró mayor concentración, siendo los carotenoides mayoritarios β-caroteno y luteína.

3. Los niveles de transcritos de los principales genes que participan en la biosíntesis de carotenoides en *C. pepo* son mayores en flores, hojas y piel del fruto que en carne, coincidiendo con una mayor concentración de pigmentos en estos tejidos. Esto demuestra que, en general, la regulación de los genes de la ruta es un factor fundamental que determina la acumulación de estos compuestos.

4. La expresión diferencial entre los cultivares de color y el blanco de algunos genes como *LCYe*, junto con otros mecanismos regulatorios podrían explicar las grandes diferencias encontradas en cuanto a

contenido en carotenoides entre las tres variedades 'Scallop', 'MU_CU16' y 'Parador'.

5. En *C. pepo* existen al menos tres genes *fitoeno sintasa*: A, B y C. CpPSYA y CpPSYB mostraron mayor similtud entre ellas y se agruparon con especies de Cucurbitaceas, mientras que CpPSYC mostró una divergencia mayor. Los patrones de expresión para los tres genes fueron diferentes, siendo mayores para *CpPCYC* en hoja y para *CpPSYA* en flores. Además el aumento en la expresión de *CpPSYA* se corresponde con el incremento de carotenoides en fruto por lo que este gen podría tener mayor implicación en la acumulación de pigmentos en los cromoplastos.

**Conclusions**

1. A set of stable genes for further normalization application in qPCR expression studies has been identified. It has been shown how the use of unsuitable genes may influence the accuracy of results. This study provides a guide of new reference genes for future use in transcriptomic studies of *C.pepo*

2. The carotenoids analysis in fruits of three cultivars has shown significant differences in the content of these compounds. The highest accumulation of carotenoids in flesh, mainly lutein, was observed for the cultivar 'Parador'. Although its yellow-orange color might suggest a higher carotenoid content in skin, the green cultivar 'MU_CU16 showed the highest concentration, being β-carotene and lutein the major carotenoids.

3. The transcriptional levels of the key genes involved in the carotenoid biosynthesis are higher in flower-, leaf- and fruit skin tissues than flesh tissues. This correlates with higher concentration of these pigments in these tissues. This demonstrates that in general, regulation of the pathway genes is a critical factor that determines the accumulation of these compounds.

4. The differential expression among the colored and white cultivars detected for some genes, such as *LCYe*, in combination with other regulatory mechanisms could explain the large differences found in terms of carotenoid content among the varieties 'Scallop', 'MU_CU16' and 'Parador'.

5. There are at least, three *phytoene synthase* genes in *C. pepo*: A, B and C. CpPSYB and CpPSYA showed more similarity between them and were grouped within Cucurbit species, however CpPSYC showed major divergence. The expression patterns for the three genes were different, being higher for leaf in *CpPCYC* and flower in *CpPSYA*. Moreover, the increase of the expression in CpPSYA was concordant with the increase of carotenoid in the fruit. Thus, this gene might be playing a major role in pigments accumulation in chromoplasts.

# ANEXO

Analytical Biochemistry 419 (2011) 336–338

Contents lists available at SciVerse ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

Notes & Tips

# Characterization of the 3′:5′ ratio for reliable determination of RNA quality

Jose V. Die [a,*], Ángeles Obrero [b], Clara I. González-Verdejo [b], Belén Román [b]

[a] Mejora Genética Vegetal, Instituto de Agricultura Sostenible (IAS-CSIC), 14080 Córdoba, Spain
[b] Biotecnología y Mejora, IFAPA Alameda del Obispo, 14080 Córdoba, Spain

ARTICLE INFO

*Article history:*
Received 12 April 2011
Received in revised form 6 August 2011
Accepted 8 August 2011
Available online 16 August 2011

*Keywords:*
3′:5′ Ratio
qPCR
RNA quality

ABSTRACT

Determination of RNA quality is a critical first step in obtaining meaningful gene expression data. The PCR-based 3′:5′ assay is an RNA quality assessment tool. This assay is a simple, fast, and low-cost method of selecting samples for further analysis. However, its practical applications are unexploited primarily because of the absence of an experimental threshold. We show that, by anchoring the 5′ assay a specific distance from the 3′ end of the sequence and by spacing the 3′ at a distance of a number of nucleotides, a cutoff determines whether a sample is suitable for downstream quantification studies.



Nowadays the reverse transcription quantitative PCR (RT-qPCR)[1] is the most rapid, sensitive, accurate, and precise method available for gene expression analysis [1]. The technique requires a combination of various steps with several conditions that have a direct impact on the conclusions. Therefore, minimizing the method's variability and maximizing its reproducibility by quality assessing every component of the qPCR workflow and adhering to common guidelines for the analysis and accurate interpretation of the data are essential [2–4]. Prior studies have reported that the quality of the template may be the most important determinant of the reproducibility and biological relevance of qPCR results [5]. This notion is supported by the well-known impact that the use of degraded RNA has on the interpretation of data [6,7]. Despite these obvious implications, RNA quality check is deficient or absent from a substantial number of published papers [8]. Therefore this is an issue that requires more attention than it has received to date.

Methods for assessing the quality of total RNA have evolved from highly sample-consuming, low-throughput, qualitative gel-based techniques to an easily automated, high-throughput, quantitative technology. Questions have recently been raised regarding the ability of traditional methods based on the 28S/18S ratio or optical density measurements to measure RNA integrity [9,10]. Currently, lab-on-chip technology is the most efficient system for ensuring RNA quality. This technique offers accuracy, reproducibility, and high sensitivity [11]. Although its applications are expected to increase in the future, a major drawback of this technology is that it mainly assesses the ribosomal RNA profile, providing total RNA quality, but is not focused on the quality of the messenger RNA molecules. Moreover, the cost of automated electrophoresis stations renders the method unfeasible for laboratories with resource constraints.

In the absence of an alternative, the 3′:5′ assay has been proposed in order to assess mRNA integrity [4]. It is based on the evidence that cDNA yield from sequences near the 5′ end of partially degraded mRNAs is significantly lower than that from sequences near the 3′ end, provided that reverse transcription proceeds from the poly(A)-tail [12]; i.e., poor RNA quality adversely affects the synthesis of first-strand cDNA resulting in the underrepresentation of the 5′ moiety of the transcript. The assay has been adopted by microarray users for many years [13], although its practical utility remains to be determined as several different cutoff criteria for the 3′:5′ ratio have been proposed. Furthermore, the term 3′:5′ assay is not used unambiguously by researchers. These inconsistencies have hampered the systematic application as a quality assessment tool.

Since the availability of a practical and reliable cutoff prior to RT and qPCR assays has the potential to save substantial costs in wasted reagents and technical time, we investigated which 3′:5′ ratio can determine the suitability of RNA samples for downstream applications in qPCR-based studies. In this work, we considered this value to depend on the nature of the samples (i.e., fresh frozen), the target's abundance, location between two primer sets and distance from the 3′ end. Our hypothesis was that employing a specific target design enables the identification of samples with sufficient RNA quality to be suitable for downstream applications.

We designed seven sets of specific primer pairs at various positions on the ubiquitin cDNA sequence (UBQ; DFCI Medicago Gene Index, TC112803) to amplify products of 60–110 bp, with an optimal primer melting temperature ($T_m$) of 60 °C and GC contents between 35% and 65%. To maximize PCR amplification efficiencies, control for secondary structure of single-stranded DNA and/or at

* Corresponding author. Fax: +34 957499252.
*E-mail address:* jdieramon@ias.csic.es (J.V. Die).

[1] *Abbreviations used:* NRQ, normalized relative quantity; RQ, relative quantity; RT-qPCR, reverse transcription quantitative polymerase chain reaction.

0003-2697/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2011.08.012

the sites of primer binding was conducted using MFOLD v3.2 software [14]. Using Trizol reagent (Invitrogen, CA, USA), total RNA was extracted from roots and leaves ($n = 3$) obtained from the pools of five plants of the model legume *Medicago truncatula* to minimize sample-to-sample variation. RNA was quantified using a ND-1000 Nanodrop spectrophotometer (Thermo Scientific, MA, USA). Prior to cDNA synthesis, DNase treatment was performed using Turbo DNAseI (Ambion, TX, USA) according to the manufacturer's protocol. The absence of contaminating genomic DNA was verified by qPCR. RNA was artificially degraded by heat exposure. Subsequently, samples were subjected to microfluidic electrophoresis analyses performed on the Experion system (Bio-Rad, CA, USA). Each sample ($n = 9$) was classified as (i) high quality (RQI $\geqslant$9; 0 min at 70 °C), (ii) optimal quality (RQI: 7–9; 10 min at 70 °C) or, (iii) partly degraded (RQI: 6–7; 30 min at 70 °C). Total RNA electropherograms were in accordance with the degradation exposure, indicating a progressive reduction in size of the 18S and 25S peaks and an elevation of the baseline, resulting in a decrease of the RQI, as depicted in Fig. 1. RT was performed on 2 μg total RNA using oligo($dT_{15}$) and Superscript III (Invitrogen) following the manufacturer's instructions. Gene expression analysis was performed on 1.5 μl of a 1:4 dilution of cDNA, 7.5 μl of 2× SYBR Green master mix (Promega, WI, USA), and 200 nM of each primer pair in a total volume of 15 μl. PCRs were performed on a Mx3000P real-time PCR system (Stratagene, CA, USA). Reactions were performed in duplicate on each of the 9 samples across the seven targets. Variations in Cq among technical replicates were assessed (max variation <0.30) and the mean Cq of replicates was employed in subsequent calculations. The experimental data, primer sequences, and target assays are outlined in the Supplementary material. Formation of specific predicted PCR products was confirmed by melting curve analysis. Relative quantity (RQ) of template was calculated using the comparative $C_T$ method [15]. Next, expression data were normalized using a normalization factor consisting of the geometric mean of the expression levels determined by the two assays located closer to the 3′ end. This provides the normalized RQ (NRQ) of the specific assay for each biological replicate (Fig. 1D–F). A log transformation of NRQ data was subsequently applied and analysis of variance was employed to compare quality groups using the transformed NRQ. Results exhibiting $P < 0.05$ were considered significant, and the least significant difference (LSD) was calculated at the 5% level to compare expression levels obtained from the various groups.

When designing a primer pair for use in qPCR, screening the 3′ region of a cDNA may be a useful approach, as it reduces the risk of comparing samples with differing degrees of RNA degradation, which predominantly starts at the 5′ end. Accordingly, higher RNA quality is associated with an elevated signal from the 5′ primer set. Conversely, lower RNA quality is associated with downregulated levels of 5′ transcripts and relatively upregulated levels of 3′ transcripts. This gene-specific primer design strategy, however, is not possible in many cases in which limited genetic information is available. We did not observe statistically significant differences in expression levels between intact and degraded RNA samples from two target templates designed within the ~800-nt sequence closest to the 3′ end of the transcript. Nevertheless, significant differences above this threshold were evident between maximum- and minimum-quality groups (Supplementary material). To determine a practical cutoff 3′:5′ ratio value to identify suitable cDNAs for further analysis, we first defined the 3′:5′ ratio as the fold change difference in expression values between the two assays. Each amplicon may represent a 3′ or 5′ assay, as that designation is merely a spatial attribute. The 3′ target assay is designed to amplify an amplicon near the

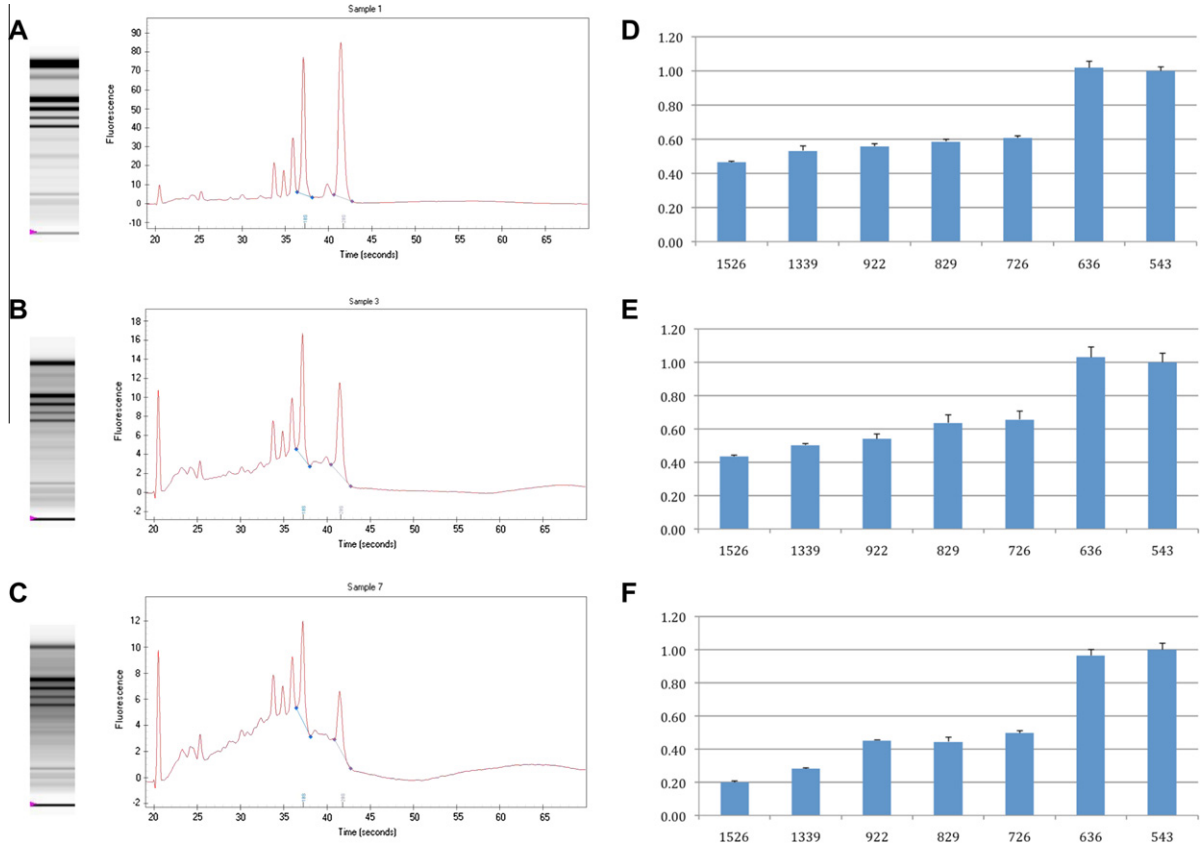
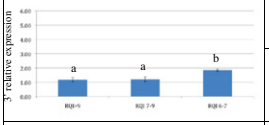
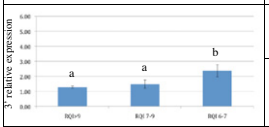
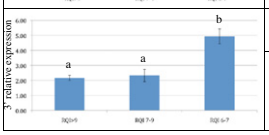



**Fig.1.** Virtual gel image and electropherograms from three representative RNA leaf samples showing different degrees of degradation according to the RQI (A) 9.1, (B) 7.2, or (C) 6.0. (D–F) Expression profiles of different amplicons throughout the UBQ gene. Normalized values are rescaled to 543 assay. Mean expression level (+SEM) from three samples is shown for each quality group. The *X* axis represents amplicon distance from the 3′ end.

**Table 1**
Recommended 3′:5′ ratios and expression patterns of 3′ amplicons designed at different positions of the cDNA sequence from three different quality groups.

| | *P* value | Mean 3’: 5’ratio | Max. 3’:5’ fold difference | Uncertain 3’: 5’ ratio |
|---|---|---|---|---|
|  | 0.0015 | 1 : 0.83 | 1.37 | 1.37-1.77 |
|  | <0.001 | 1 : 0.67 | 1.76 | 1.76-1.97 |
|  | <0.001 | 1 : 0.43 | 2.73 | 2.73-4.43 |



Expression level of 5′ amplicon was set as 1. Error bars indicate 95.45% confidence intervals. Statistical differences are denoted by different letters above bars. The numbers in the UBQ sequence schema represent positions of 5′ and 3′ amplicons with respect to the end of the sequence.

3′ end, and the 5′ target assay targets the 5′ end of the mRNA sequence. Next, we decided to anchor a 5′ assay at 1526 nt from the end of the sequence, and investigated whether the 3′ targets expression levels were significantly different from the 5′ assay, employing the mean expression level of various exons across the UBQ gene that were located 700, 900, and 1100 bp distance from the 5′ assay. Statistical analysis indicated that the 3′:5′ ratio of the lowest quality group differed when the 3′ assay was designed targeting sequences 700 bp ($P < 0.01$), 900 bp ($P < 0.001$), and 1070 bp downstream ($P < 0.001$). Finally, on the basis of the observed 5′ mean expression level (setting at 1), we employed the mean 3′ relative expression value in optimal-quality samples and the 95.45% confidence interval (+95.45% CI, upper interval) as the upper limit for scoring a suitable cDNA derived from intact RNA, and mean value of the partly degraded samples and the 95.45% confidence interval (−95.45% CI, lower interval) as lower limit for discarding a cDNA (Table 1).

Some conclusions from the aforementioned: (1) to characterize successfully the extent of RNA degradation in the starting material, 5′ amplicon of the 3′:5′ ratio assay must be designed at a sufficient distance to the 3′ end of the sequence. Otherwise, measured gene expression from starting materials of equal quality, a critical condition for obtaining meaningful gene expression data, will not be ensured. This has a second meaning suggesting a “security zone” close to the 3′ end that may be utilized, through careful primer design, to enable the quantification of mRNA levels. As noted above, this method’s relevance is limited by the availability of information in public databases regarding gene structures; (2) as determined from the fixed distance from the 5′ amplicon to the 3′ end sequence, there are various ratio thresholds, depending on the targets’ spatial separation, that can discriminate between various integrity levels of the RNA samples. Results indicate that 3′:5′ ratio values to 2.73 provide full assurance that the input material is of high quality. Moreover, values over the limit of 4.43 for the same distance between amplicons may characterize cDNAs as unreliable for gene expression studies. The inclusion of samples within this safety margin does not demonstrate that the material is unsuitable for further qPCR assays, but does not guarantee that the sample is in a maximum-quality criteria compliant way.

Note that the usefulness of the 3′:5′ ratio relies on the oligo(dT) priming method for cDNA synthesis because, with this method, the progress of the reverse transcriptase is dependent on the intactness of the mRNA. Moreover, experiments may be conducted in order to determine whether 3′:5′ ratios are RT enzyme dependent or use more cell tissue types or targeting the 5′ and 3′ end of other reference genes with different transcript abundance. Furthermore, it would be important to confirm these results in formalin-fixed paraffin-embedded tissues. However, data within the ratios shown in Table 1 can ensure that observed biological differences are not due to poor RNA integrity.

To summarize, the quality criteria proposed in this study should provide a useful guideline for future studies using qPCR to assess mRNA levels. The suggested cutoffs derived from specific distances between amplicons may help determine whether a certain cDNA is suitable for downstream quantification assays. Although we have conducted the experiment in plant tissues, it should be applicable for the assessment of gene expression data obtained from other cell populations.

## Acknowledgments

Jose V Die is a researcher funded by the ’Juan de la Cierva’ programme of the Spanish *Ministerio de Ciencia e Innovación*.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ab.2011.08.012.

## References

[1] H.D. VanGuilder, K.E. Vrana, W.M. Freeman, Twenty-five years of quantitative PCR for gene expression analysis, Biotechnology 44 (2008) 619–626.
[2] S. Derveaux, J. Vandesompele, J. Hellemans, How to do successful gene expression analysis using real-time PCR, Methods 50 (2010) 227–230.
[3] S. Bustin, V. Benes, J.A. Garson, J. Hellemans, J. Huggett, M. Kubista, et al., The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments, Clin. Chem. 55 (2009) 611–622.
[4] T. Nolan, R.E. Hands, S.A. Bustin, Quantification of mRNA using real-time RT-PCR, Nat. Protoc. 1 (2006) 1559–1582.
[5] S.A. Bustin, T. Nolan, Analysis of mRNA expression by real-time PCR, in: J. Logan, K. Edwards, N. Saunders (Eds.), Real-Time PCR: Current Technology and Applications, 2009, pp. 111–135.
[6] J. Vermeulen, K. De Preter, S. Lefever, J. Nuytens, S. Derveaux, J. Hellemans, et al., Measurable impact of RNA quality on gene expression results from quantitative PCR, Nucleic Acids Res. (2011), doi:10.1093/nar/gkr065.
[7] S. Fleige, M.W. Pfaffl, RNA integrity and the effect on the real-time qRT-PCR performance, Mol. Aspects Med. 27 (2006) 126–139.
[8] J. Murphy, S.A. Bustin, Reliability of real-time reverse-transcription PCR in clinical diagnostics: gold standard or substandard?, Expert Rev Mol. Diagn. 9 (2009) 187–197.
[9] C. Strand, J. Enell, I. Hedenfalk, M. Fernö, RNA quality in frozen breast cancer samples and the influence on gene expression analysis—a comparison of three evaluation methods using microcapillary electrophoresis traces, BMC Mol. Biol. 8 (2007) 38.
[10] O. Mueller, K. Hahnenberger, M. Dittmann, H. Yee, R. Dubrow, R. Nagle, D. Ilsley, A microfluidic system for high-speed reproducible DNA sizing and quantitation, Electrophoresis 21 (2000) 128–134.
[11] I. Riedmaier, M. Bergmaier, M.W. Pfaffl, Comparison of two available platforms for determination of RNA quality, Biotechnol. Biotechnol. Equip. 24 (2011) 2154–2159.
[12] G.H. Swift, M.J. Peyton, R.J. MacDonald, Assessment of RNA quality by semi-quantitative RT-PCR of multiple regions of a long ubiquitous mRNA, Biotechniques 28 (2000) 524–531.
[13] H. Auer, S. Lyianarachchi, D. Newsom, M.I. Klisovic, G. Marcucci, K. Kornacker, Chipping away at the chip bias: RNA degradation in microarray analysis, Nat. Genet. 35 (2003) 292–293.
[14] M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction, Nucleic Acids Res. 31 (2003) 3406–3415.
[15] T.D. Schmittgen, K.J. Livak, Analyzing real-time PCR data by the comparative CT method, Nat. Protoc. 3 (2008) 1101–1108.

## Internet resources

The Gene Index Project: http://compbio.dfci.harvard.edu/tgi/